

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 QTL mapping in experimental crosses is an important component of forward genetics. In a companion article, we describe the mean-variance approach to QTL mapping, which addresses both mean and variance heterogeneity. Here, we present *vqtl*, an R package that implements mean-variance QTL mapping. The goal of *vqtl* is to provide a straightforward tool for geneticists to perform mean-variance genome scans, assess the statistical significance of their findings, and communicate their results with informative plots. Because this package is interoperable with the popular R/*qt1* 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, 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 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 many 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.

Recent variance QTL (vQTL) analyses have challenged that assumption by seeking to identify genetic factors that influence the extent of residual variation (Paré *et al.* 2010; Rönnegård and Valdar 2011, 2012; Cao *et al.* 2014). In the companion pieces, we describe an elaboration of vQTL analysis, which we term “mean-variance QTL mapping”, and use it to discover new QTL from existing data resources [companion B, C].

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 *qt1* package and is available on CRAN, so it is easy to get started. Here, we demonstrate typical use of the *vqtl* package to analyze four simulated traits and discover three QTL.

SIMULATION OF AN ILLUSTRATIVE DATASET

We used the popular R/*qt1* 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/*qt1*’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 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.

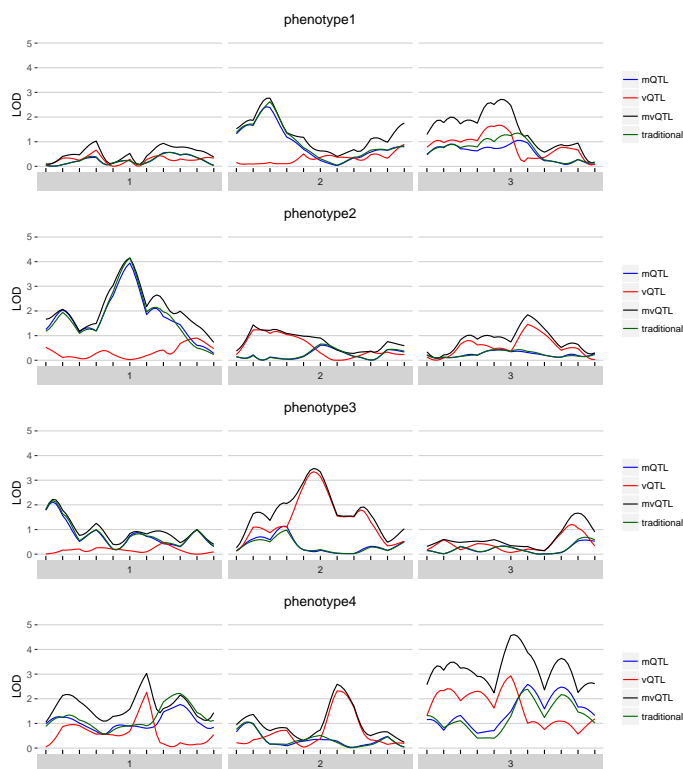


Figure 1 For each of the four simulated phenotypes, the genome scan shows the LOD score of each test – mean, variance, and joint – in blue, red, and black, respectively. The traditional test is in green and globally similar to the mean test.

GENOME SCANS

The central function for genetic mapping in package `qt1` is `scanone` (Broman *et al.* 2003). Analogously, the central function for genetic mapping in package `mvqt1` is `scanonevar`. It takes three required inputs:

1. `cross` contains the genetic and phenotypic information from an experimental cross. This object can be the same `cross` object used in package `qt1`.
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`. This option can be useful in assessing the trends in the various genetic components of the phenotype across the genome or in assessing the trends in covariate effects across the genome.

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, a 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 [CompanionC].

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 signals 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. Statistical methods for empirical FWER control in mean-variance QTL mapping are described in [CompanionC].

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

ASSESS THE SIGNIFICANCE OF RESULTS

Given the many tests involved in a genome scan, some multiple hypothesis testing correction must be made. We advocate using

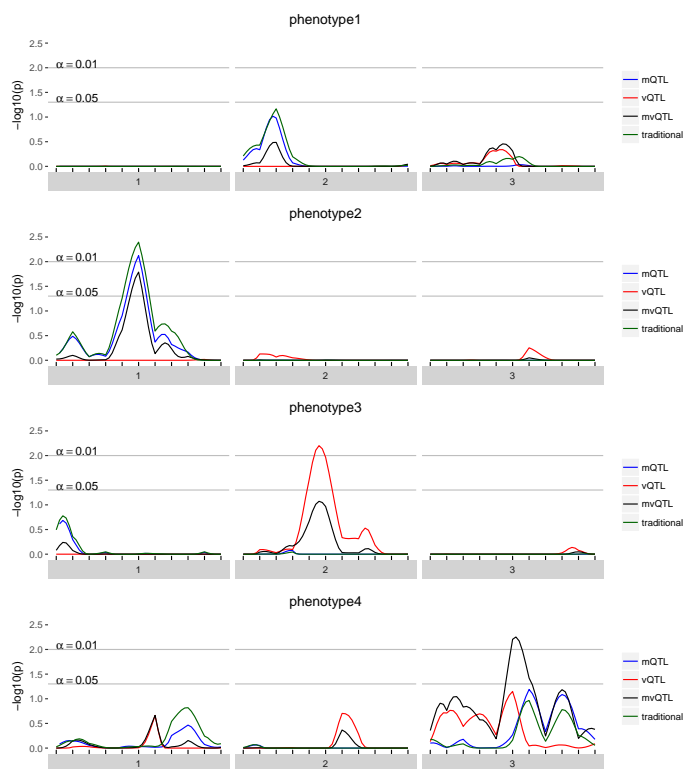


Figure 2 For each of the four simulated phenotypes, the genome scan shows the $-\log_{10}$ of the FWER-corrected p -value of each test – mean, variance, and joint – in blue, red, and black, respectively. Thus, a value of 3 implies that the quantity of evidence against the null is such that we expect to see this much or more evidence once per thousand genome scans when there is no true effect.

a permutation-based approach, detailed in [CompanionC] to control the family wise error rate (FWER). 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.

This approach involves conducting many permuted genomes scans, each executing a model comparison that maintains all mean and variance effects of covariates and any non-focal genetic factors. 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 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, 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-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.

COMMUNICATE SIGNIFICANT FINDINGS

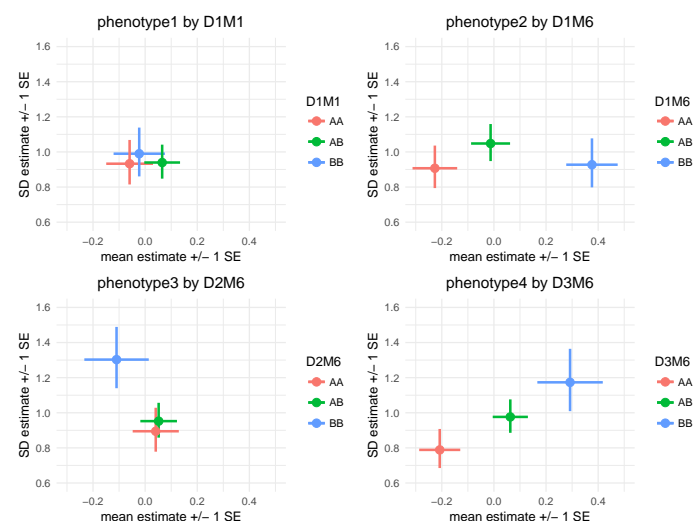


Figure 3 `mean_var_plots` show the estimated genotype effects at a locus, with mean effects on the horizontal axis and variance effects on the vertical axis. Horizontal lines indicate standard errors for mean effects and vertical lines indicate standard errors for variance effects. For phenotype1, the 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 without either total horizontal or vertical separation illustrates an mvQTL with neither mean nor variance effect strong enough to define an mQTL or vQTL.

Having identified some QTL, we want to visualize the estimated genetic and covariate effects at the QTL. Because the `vqt1` package models both mean and variance effects, existing plotting utilities aren't able to display the entirety of the modeling results. To understand and communicate the results of a `vqt1` 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. For brevity, we show only the more useful “model-based” version.

In each `mean_var_plot` in Figure A3, 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 A3; 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.

The identification of an mQTL is consistent with horizontal distances between genotype groups that exceed the horizontal standard error bars. The identification of a vQTL is consistent with vertical distances between genotype groups that exceed the vertical standard error bars. And the identification of an mvQTL is consistent with euclidean distance between genotype groups that exceed the standard error density implied by both the horizontal and vertical standard errors.

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

Finally, 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 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.

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 `vqt1` for mean-variance QTL mapping in experimental crosses. 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. In the case of covariates, they can be accommodated, which can increase power and control false positive rate, as described in a companion article [CompanionC].

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](https://doi.org/10.5281/zenodo.1173799). The

package `vqt1` and its documentation are freely available on CRAN at <https://CRAN.R-project.org/package=vqt1>.

ACKNOWLEDGEMENTS

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

- Broman, K. W., H. Wu, S. 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.
- Lander, E. S. and S. Botstein, 1989 Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185.
- 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.
- Visscher, P. M., R. Thompson, and C. S. Haley, 1996 Confidence intervals in QTL mapping by bootstrapping.
- 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, et al., 2014 Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* **5**: 5087.

APPENDIX

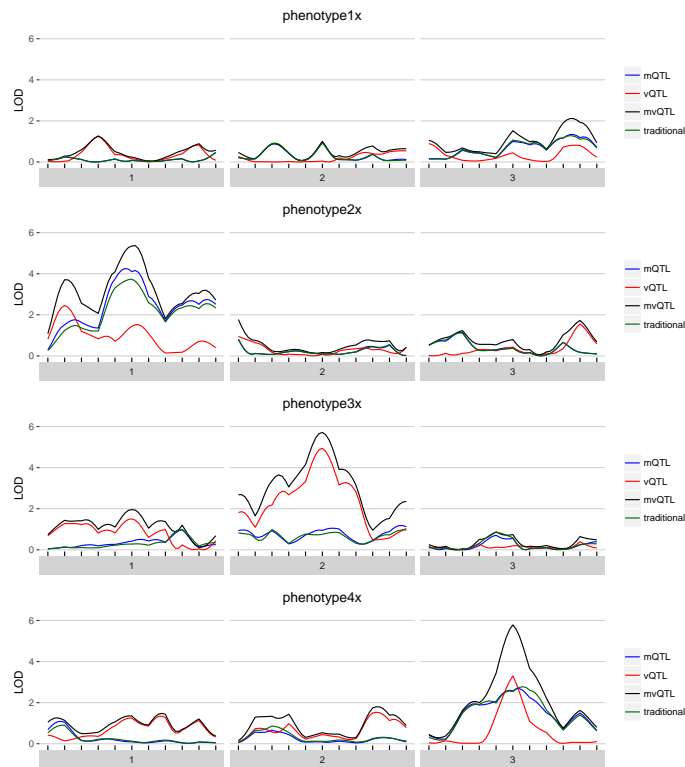


Figure A1 For each of the four simulated phenotypes with background variance heterogeneity, the genome scan shows the LOD score of each test – mean, variance, and joint – in blue, red, and black, respectively. The traditional test is in green and globally similar to the mean test.

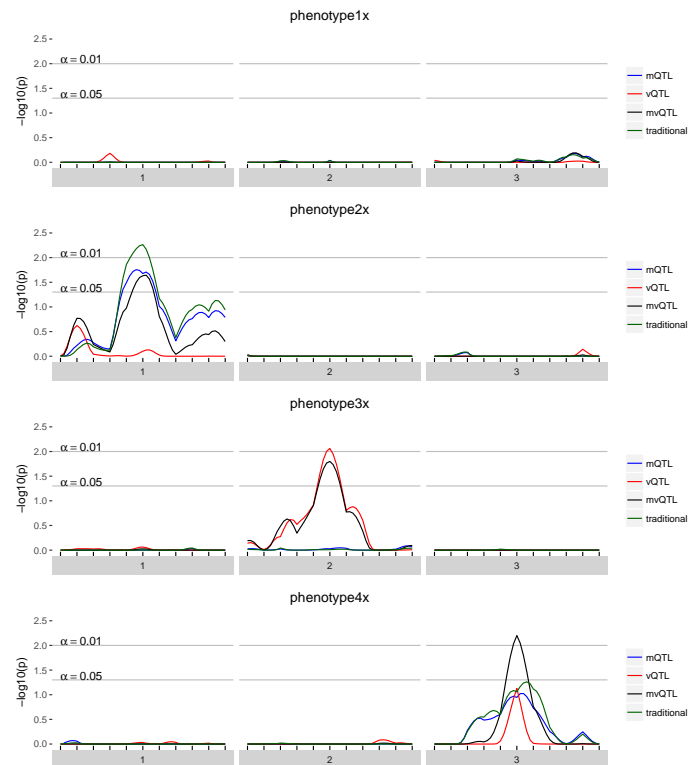


Figure A2 For each of the four simulated phenotypes with background variance heterogeneity, the genome scan shows the $-\log_{10}$ of the FWER-corrected p -value of each test – mean, variance, and joint – in blue, red, and black, respectively. Thus, a value of 3 implies that the quantity of evidence 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.

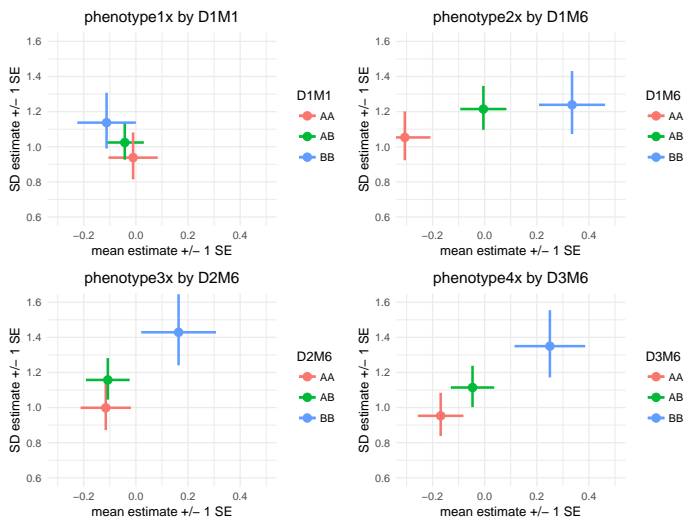


Figure A3 mean_var_plots show the estimated genotype effects at a locus, with mean effects on the horizontal axis and variance effects on the vertical axis. Horizontal lines indicate standard errors for mean effects and vertical lines indicate standard errors for variance effects. For phenotype1x, the 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.