vqtl: An R package for Efficient QTL Mapping on Phenotypes with Heterogeneous Variance

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

ABSTRACT Existing methods for QTL mapping in experimental crosses assume that the extent of residual variation is constant across all individuals. When this assumption is not met, the simultaneous mean-variance modeling approach described in the companion article can yield higher power, better protection against false positives, and detection of the rarely-considered class of QTL that influence phenotype variance, vQTL. The R package vqtl implements this statistical model, a significance testing procedure, and facilitates visualization and interpretation of 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 many geneticists to analyze the results of their experimental crosses with vqtl, possibly discovering new QTL. Here, we demonstrate a typical usage case.

KEYWORDS
QTL mapping, variance heterogeneity

Introductory material, motivation, rationale, etc. Introductory material, motivation, rationale, etc. Introductory material, motivation, rationale, etc. Introduce mQTL, vQTL, and mvQTL nomenclature.

SIMULATION

We will illustrate the usage and interpretation of the vqtl package through the analysis of an experimental cross simulated with R/qtl. The simulated population consists of 100 male and 100 female F2 offspring, with 5 chromosomes of length 100 cM tagged by 30 equally-spaced markers each. We simulate four phenotypes

1. phenotype1 consists only of random noise
2. phenotype2 has one large-effect mQTL at the 15th marker on chromosome one
3. phenotype3 has one large-effect vQTL at the 15th marker on chromosome two
4. phenotype2 has one large-effect mvQTL at the 15th marker on chromosome three

CONDUCTING A GENOME SCAN

Analogous to the central scanone function in package qtl, package vqtl provides function scanonevar. Like scanone, scanonevar takes as input an object of class cross, and the specification of a a single phenotype contained in the cross and returns as output a statistic for each genetic locus in the cross that summarizes the association between that locus and the phenotype. However, scanone only tests for association between each locus and the phenotype mean, while scanonevar tests for associations with mean and variance. Thus scanonevar returns not one, but three, statistics for each genetic locus: one that summarizes the association of the locus with phenotype mean, one that summarizes the association of the locus with phenotype variance, and one that jointly summarizes the association between the locus, phenotype mean, and phenotype variance.

1. ASSESSING THE STATISTICAL SIGNIFICANCE OF FINDINGS

Figure 1 shows the log of the odds at each locus for each of the four tests as a measure of association. This measure is traditional in QTL mapping, but in this scenario where multiple different tests are conducted at each locus, it suggesets a misleading interpretation, that points of different colors with the same LOD score represent an equal quantity of evidence against the null (no genetic effect). In fact, as is visible in the top panel of figure 1, the LOD score that results from differnet tests is different under the null. For example, the LOD score of the DGLM-joint test is always higher than the other three tests because all the other tests are "nested" inside the DGLM joint test.

Manuscript compiled: Wednesday 22nd June, 2016
[1]Correspondence e-mail: william.valdar@unc.edu

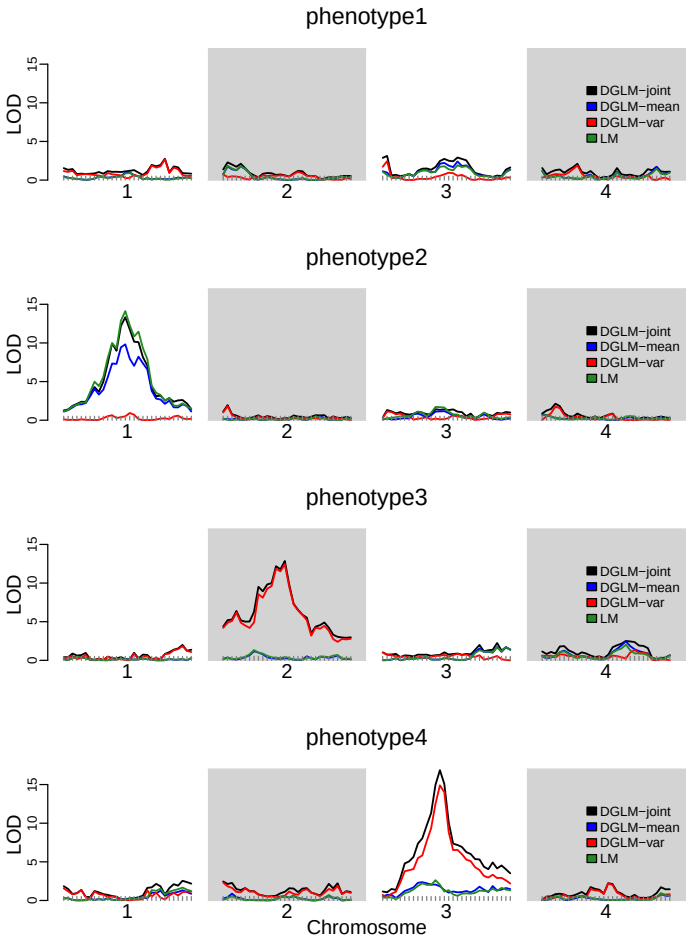

Figure 1 For each of the four simulated phenotypes, we have the three new tests in black, blue, and red. The traditional test is in green and clearly similar to the blue test in the vast manority of loci.