INVESTIGATIONS

QTL Mapping on a Background of Variance Heterogeneity – Supplementary Materials

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

[*]Department of Genetics, [†]Bioinformatics and Computational Biology Curriculum, [‡]and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC

ABSTRACT Standard QTL mapping procedures seek to identify genetic loci affecting the phenotypic mean while assuming that all individuals have the same residual variance. But when the residual variance differs systematically between groups, perhaps due to a genetic or environmental factor, such standard procedures can falter: in testing for QTL associations, they attribute too much weight to observations that are noisy and too little to those that are precise, resulting in reduced power and and increased susceptibility to false positives. The negative effects of such "background variance heterogeneity" (BVH) on standard QTL mapping have received little attention until now, although the subject is closely related to work on the detection of variance-controlling genes. Here we use simulation to examine how BVH affects power and false positive rate for detecting QTL affecting the mean (mQTL), the variance (vQTL), or both (mvQTL). We compare linear regression for mQTL and Levene's test for vQTL, with tests more recently developed, including tests based on the double generalized linear model (DGLM), which can model BVH explicitly. We show that, when used in conjunction with a suitable permutation procedure, the DGLM-based tests accurately control false positive rate and are more powerful than the other tests. We also find that some adverse effects of BVH can be mitigated by applying a rank inverse normal transform. We apply our novel approach, which we term "mean-variance QTL mapping", to publicly available data on a mouse backcross and, after accommodating BVH driven by sire, detect a new mQTL for bodyweight.

KEYWORDS
Cao's tests, reweighting, vQTL, variable transformation, background variance heterogeneity, heteroskedastic, heteroscedastic

Manuscript compiled: Thursday 1st November, 2018
[1]Corresponding author: William Valdar, 120 Mason Farm Road, Genetic Medicine Building, Suite 5113, Campus Box 7264, Chapel Hill, NC, 27599, USA, william.valdar@unc.edu

SUPPLEMENTARY MATERIALS

Simulation Details:

In simulation with BVH present, the group-wise effects on the log standard deviation were $\gamma = [-0.4, -0.2, 0, 0.2, 0.4]$. Though $\overline{\gamma} = 0$, the exponential transform connecting these effects to the standard deviation results in a simulated phenotype with slightly more total variance than one without BVH. Therefore, the additive effect of the locus on phenotype mean was adjusted when BVH was introduced, in order to maintain a constant percent variance explained by the mean effect. The following values were used in the simulation.

	no BVH	yes BVH
null	$\alpha = 0, \theta = 0$	$\alpha = 0, \theta = 0$
mQTL	$\alpha = 0.22, \theta = 0$	$\alpha = 0.25, \theta = 0$
vQTL	$\alpha = 0, \theta = 0.17$	$\alpha = 0, \theta = 0.17$
mvQTL	$\alpha = 0.18, \theta = 0.14$	$\alpha = 0.2, \theta = 0.136$

null locus and mQTL in the absence of BVH: All observations have standard deviation 1.

vQTL in the absence of BVH: The genotype-wise standard deviations implied by the additive effect of 0.17 on the log standard deviation are approximately: [0.84, 1.00, 1.19].

mvQTL in the absence of BVH: The genotype-wise standard deviations implied by the additive effect of 0.14 on the log standard deviation are approximately: [0.87, 1.00, 1.15].

null locus and mQTL in the presence of BVH: The covariate-wise standard deviations implied by the effects of [-0.4, -0.2, 0, 0.2, 0.4] on the log standard deviation are approximately: [0.67, 0.82, 1.00, 1.22, 1.49].

vQTL in the presence of BVH: Locus and covariate effects on the residual variance combine additively on the log standard deviation scale, yielding 15 distinct standard deviations:

	genotype		
covar	-1	0	1
1	0.57	0.67	0.79
2	0.69	0.82	0.97
3	0.84	1.00	1.19
4	1.03	1.22	1.45
5	1.26	1.49	1.77

mvQTL in the presence of BVH: Locus and covariate effects on the residual variance combine additively on the log standard deviation scale, yielding 15 distinct standard deviations:

	genotype		
covar	-1	0	1
1	0.59	0.67	0.77
2	0.71	0.82	0.94
3	0.87	1.00	1.15
4	1.07	1.22	1.40
5	1.30	1.49	1.71

AUC Table

		BVH absent			BVH present		
test	procedure	mQTL	vQTL	mvQTL	mQTL	vQTL	mvQTL
SLM	standard	0.931	0.502	0.860	0.926	0.495	0.858
	RINT	0.930	0.499	0.855	0.935	0.493	0.866
	residperm	0.930	0.503	0.857	0.928	0.495	0.857
	locusperm	0.931	0.503	0.858	0.927	0.494	0.858
Cao_M	standard	0.930	0.504	0.862	0.930	0.499	0.861
	RINT	0.929	0.497	0.858	0.933	0.495	0.865
	residperm	0.930	0.502	0.861	0.926	0.494	0.860
	locusperm	0.929	0.501	0.859	0.926	0.495	0.859
$DGLM_M$	standard	0.929	0.514	0.863	0.964	0.503	0.914
	RINT	0.929	0.510	0.858	0.963	0.506	0.911
	residperm	0.925	0.503	0.856	0.962	0.496	0.908
	locusperm	0.925	0.500	0.854	0.961	0.493	0.906
Levene's test	standard	0.500	0.912	0.836	0.489	0.887	0.804
	RINT	0.486	0.908	0.818	0.485	0.877	0.782
	residperm	0.502	0.916	0.842	0.497	0.890	0.809
	locusperm	0.502	0.916	0.841	0.496	0.890	0.808
Cao_V	standard	0.501	0.937	0.870	0.593	0.930	0.864
	RINT	0.485	0.925	0.844	0.495	0.882	0.795
	residperm	0.500	0.933	0.864	0.503	0.887	0.802
	locusperm	0.497	0.932	0.863	0.497	0.883	0.802
$DGLM_V$	standard	0.503	0.933	0.864	0.503	0.937	0.864
	RINT	0.491	0.919	0.838	0.440	0.891	0.799
	residperm	0.501	0.928	0.859	0.417	0.887	0.797
	locusperm	0.501	0.928	0.856	0.497	0.931	0.858
Cao_{MV}	standard	0.900	0.908	0.941	0.910	0.901	0.940
	RINT	0.897	0.890	0.930	0.896	0.838	0.918
	residperm	0.896	0.902	0.938	0.871	0.858	0.911
	locusperm	0.897	0.904	0.938	0.873	0.859	0.911
$DGLM_{MV}$	standard	0.900	0.901	0.938	0.941	0.905	0.959
	RINT	0.893	0.887	0.929	0.933	0.849	0.941
	residperm	0.892	0.893	0.933	0.912	0.856	0.933
	locusperm	0.894	0.894	0.933	0.938	0.897	0.955

Table S1 Area under the curve for the six non-null scenarios and all 32 test-procedures.

Standard Error Table

		BVH absent				BVH present			
test	procedure	null	mQTL	vQTL	mvQTL	null	mQTL	vQTL	mvQTL
SLM	standard	0.0044	0.0089	0.0045	0.0098	0.0044	0.0090	0.0044	0.0098
	RINT	0.0043	0.0089	0.0044	0.0098	0.0044	0.0089	0.0043	0.0098
	residperm	0.0043	0.0089	0.0044	0.0098	0.0043	0.0090	0.0044	0.0098
	locusperm	0.0043	0.0090	0.0044	0.0098	0.0044	0.0090	0.0044	0.0098
Cao_M	standard	0.0044	0.0089	0.0044	0.0098	0.0045	0.0090	0.0043	0.0098
	RINT	0.0044	0.0089	0.0042	0.0098	0.0045	0.0089	0.0043	0.0098
	residperm	0.0044	0.0090	0.0043	0.0098	0.0044	0.0091	0.0042	0.0098
	locusperm	0.0043	0.0090	0.0042	0.0098	0.0043	0.0091	0.0042	0.0098
$DGLM_M$	standard	0.0046	0.0089	0.0045	0.0098	0.0047	0.0073	0.0046	0.0094
	RINT	0.0045	0.0089	0.0045	0.0098	0.0046	0.0073	0.0045	0.0094
	residperm	0.0043	0.0091	0.0042	0.0098	0.0044	0.0075	0.0043	0.0095
	locusperm	0.0043	0.0091	0.0042	0.0098	0.0043	0.0076	0.0043	0.0095
Levene's test	standard	0.0041	0.0042	0.0094	0.0098	0.0043	0.0041	0.0098	0.0096
	RINT	0.0041	0.0039	0.0095	0.0097	0.0042	0.0039	0.0098	0.0093
	residperm	0.0043	0.0044	0.0093	0.0098	0.0044	0.0043	0.0097	0.0096
	locusperm	0.0042	0.0044	0.0093	0.0098	0.0044	0.0043	0.0097	0.0096
Cao_V	standard	0.0045	0.0044	0.0087	0.0098	0.0067	0.0067	0.0087	0.0098
	RINT	0.0041	0.0040	0.0091	0.0098	0.0042	0.0042	0.0098	0.0095
	residperm	0.0043	0.0043	0.0088	0.0098	0.0044	0.0043	0.0098	0.0096
	locusperm	0.0043	0.0042	0.0089	0.0098	0.0044	0.0042	0.0098	0.0096
$DGLM_V$	standard	0.0045	0.0044	0.0088	0.0098	0.0046	0.0043	0.0087	0.0098
	RINT	0.0041	0.0039	0.0092	0.0098	0.0030	0.0029	0.0098	0.0094
	residperm	0.0043	0.0043	0.0090	0.0098	0.0028	0.0027	0.0098	0.0093
	locusperm	0.0043	0.0043	0.0090	0.0098	0.0044	0.0042	0.0090	0.0098
Cao_{MV}	standard	0.0044	0.0096	0.0095	0.0086	0.0062	0.0094	0.0094	0.0085
	RINT	0.0042	0.0097	0.0098	0.0090	0.0043	0.0096	0.0098	0.0094
	residperm	0.0043	0.0097	0.0096	0.0088	0.0045	0.0098	0.0098	0.0095
	locusperm	0.0043	0.0097	0.0096	0.0088	0.0044	0.0098	0.0098	0.0095
$DGLM_{MV}$	standard	0.0046	0.0096	0.0096	0.0087	0.0047	0.0086	0.0095	0.0077
	RINT	0.0043	0.0097	0.0098	0.0091	0.0038	0.0088	0.0098	0.0087
	residperm	0.0043	0.0097	0.0097	0.0089	0.0031	0.0095	0.0098	0.0091
	locusperm	0.0043	0.0097	0.0097	0.0089	0.0045	0.0087	0.0097	0.0080

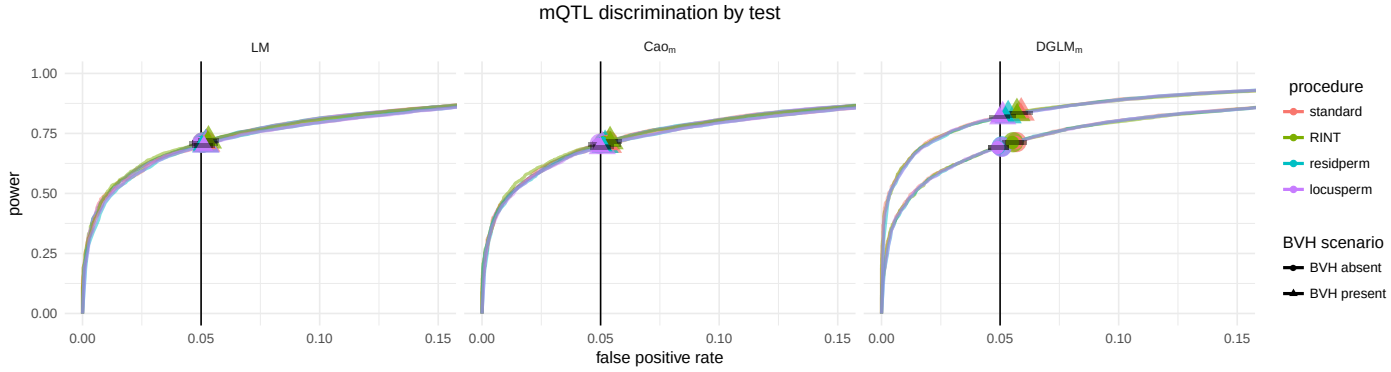
Table S2 Standard errors for the values in Table S2.

ROC Curves

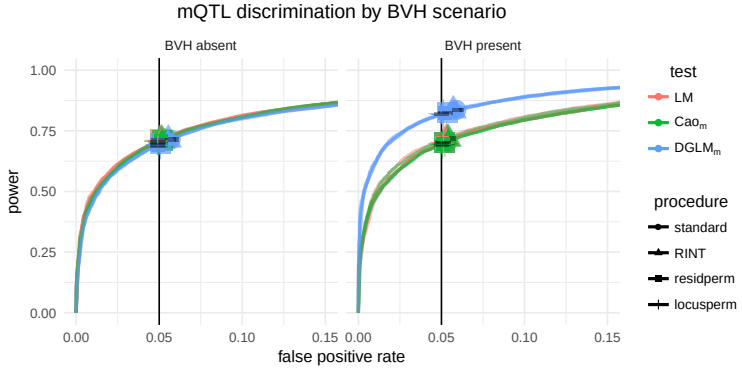
The receiver operating characteristics (ROC) curve plots the empirical power as a function of the empirical FPR, reflects the ability of a test-procedure combination to discriminate between QTL and null loci. Specifically, for a given test-procedure and cutoff c, the FPR is defined as the fraction null simulations in which the nominal p-value p was less than c. This quantity is best thought of as the "empirical FPR" and is generally superior to the nominal FPR, since the nominal FPR depends on the assumption that the test is appropriately calibrated while the empirical does not. The power is the fraction of non-null experiments in which the nominal p-value is less than c.

The ROC curve cannot immediately distinguish between tests that accurately control FPR and those that do not. We added a symbol to each ROC curve at the point where $c = 0.05$. In cases where the point falls on the vertical line at FPR = 0.05, it reflects accurate FPR control. In cases where the point falls to the left or right of the vertical line it reflects a conservative or anti-conservative test, respectively.

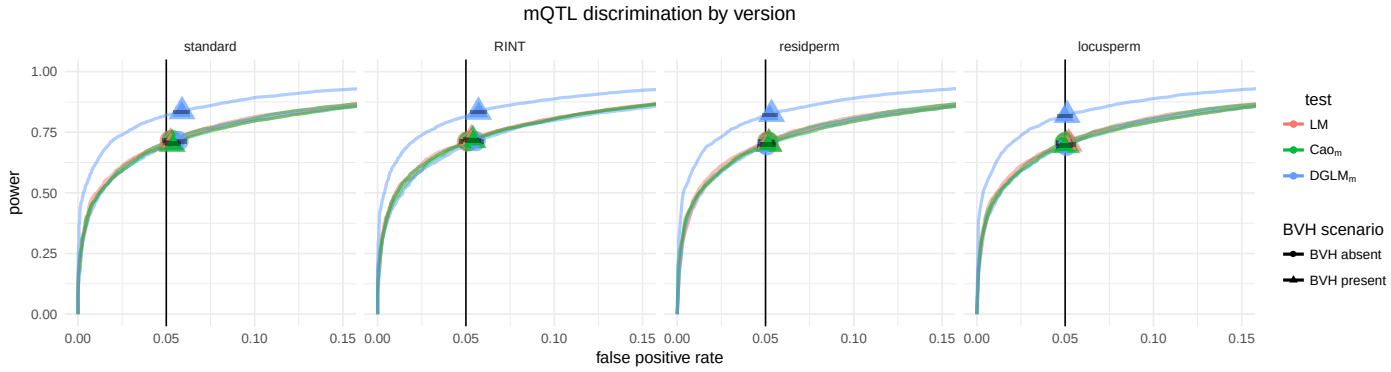
Figure S1 ROC Curves for mQTL tests in the detection of mQTL. The same 28 ROC curves are plotted three times, organized by (a) test, (b) BVH scenario, and (c) significance assessment procedure to allow for comparisons across all dimensions.



(a) All test-evaluations accurately control FPR. DGLM$_M$ with BVH of known source is the most powerful test.

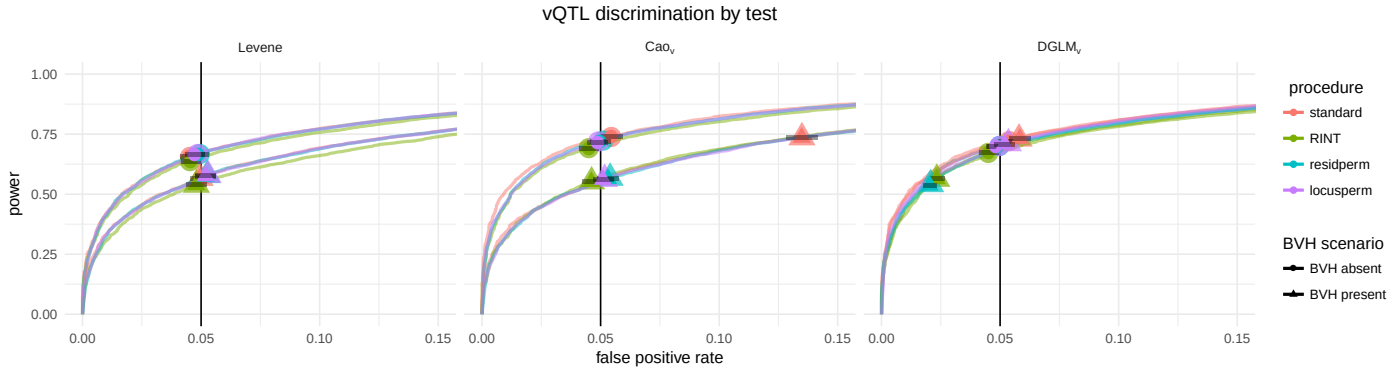


(b) Within BVH scenarios, all mQTL tests perform equivalently.

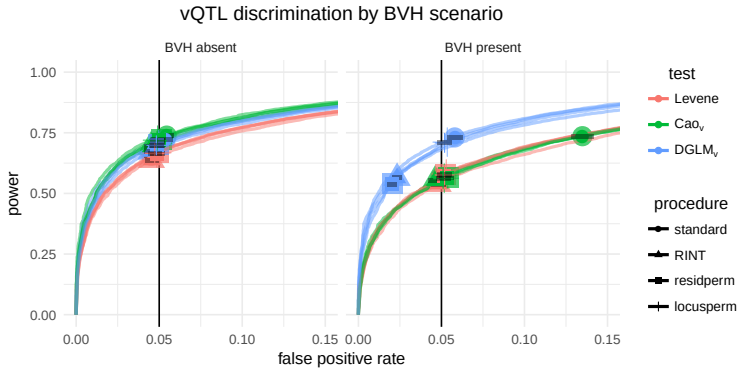


(c) DGLM$_M$ outperforms all other tests across all evaluation methods.

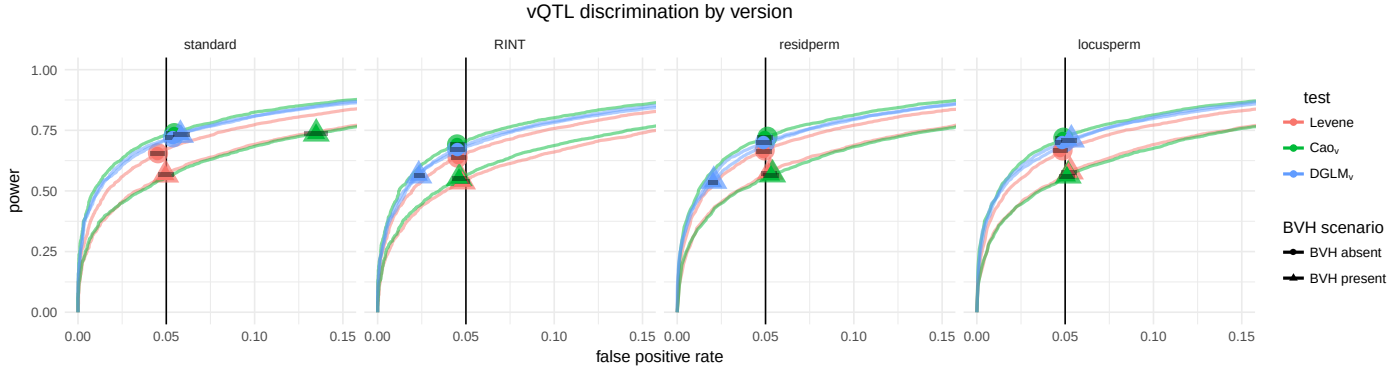
Figure S2 ROC Curves for vQTL tests in the detection of vQTL. The same 28 ROC curves are plotted three times, organized by (a) test, (b) BVH scenario, and (c) significance assessment procedure to allow for comparisons across all dimensions.



(a) Levene's test accurately controls FPR in all scenarios.Cao$_V$ and DGLM$_V$ have inflated FPR in the presence of BVH of unknown source. DGLM$_V$'s RINT and residperm procedures are anti-conservative in the presence of BVH of known source.

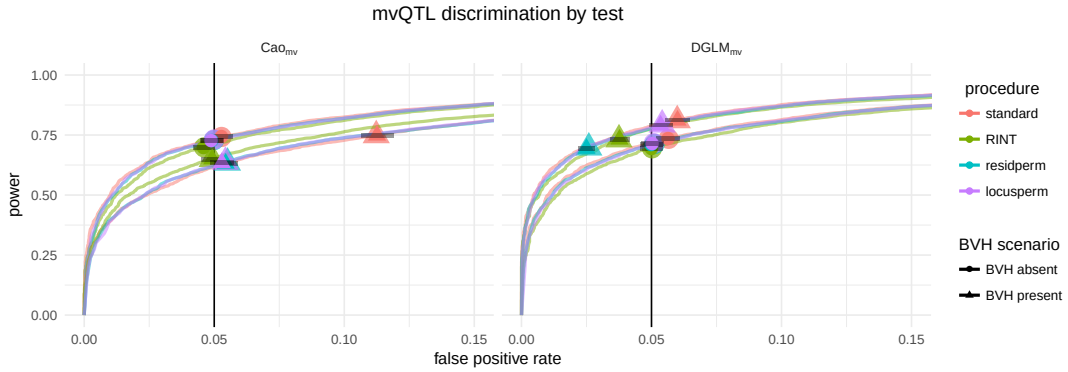


(b) In the absence of BVH, Levene's test is less powerful than Cao$_V$ and DGLM$_V$. In the face of BVH of unknown source, all tests suffer decreased power, except DGLM$_V$'s standard procedure, which fails to accurately control FPR. In the scenario with BVH of known source, DGLM$_V$ recovers most of the power lost with introduction of BVH, but its RINT and residperm procedures are anti-conservative.

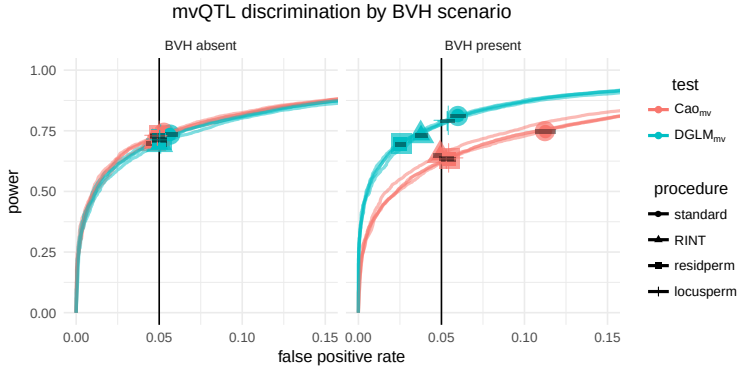


(c) The only procedure of DGLM$_V$ that accurately controls FPR across all BVH scenarios is locusperm. Its standard procedure is anti-conservative in the presence of BVH of unknown source and its RINT and residperm procedures are conservative in the presence of BVH of known source.

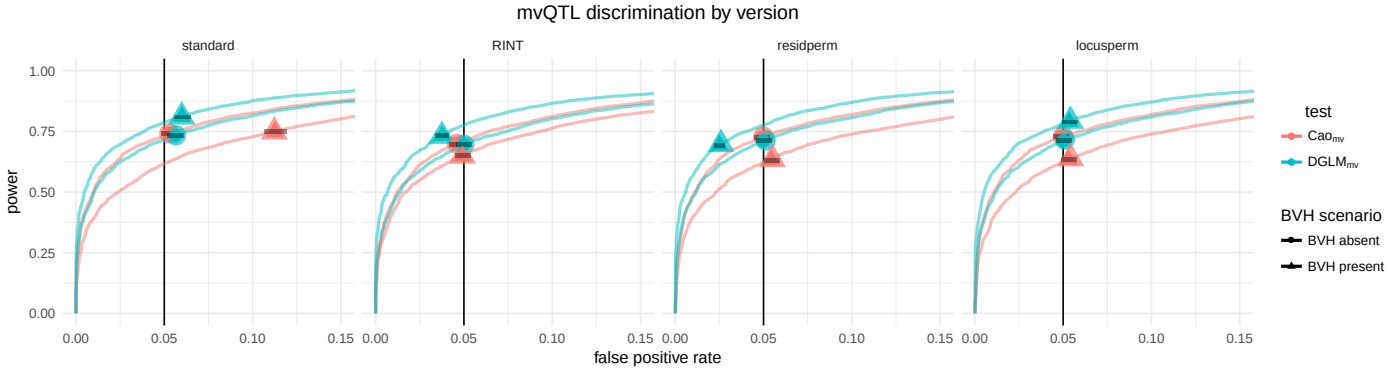
Figure S3 ROC Curves for mvQTL tests in the detection of mvQTL. The same 28 ROC curves are plotted three times, organized by (a) test, (b) BVH scenario, and (c) significance assessment procedure to allow for comparisons across all dimensions.



(a) Cao_{MV} and DGLM_{MV} both suffer a decrease in discrimination (down and right shift of ROC curve) in the presence of BVH of unknown (or unmodeled) source. Only DGLM_{MV} can accommodate the source when it is known and therefore can achieve superior discrimination in that case. The standard and locusperm procedures of DGLM_{MV} accurately control FPR.



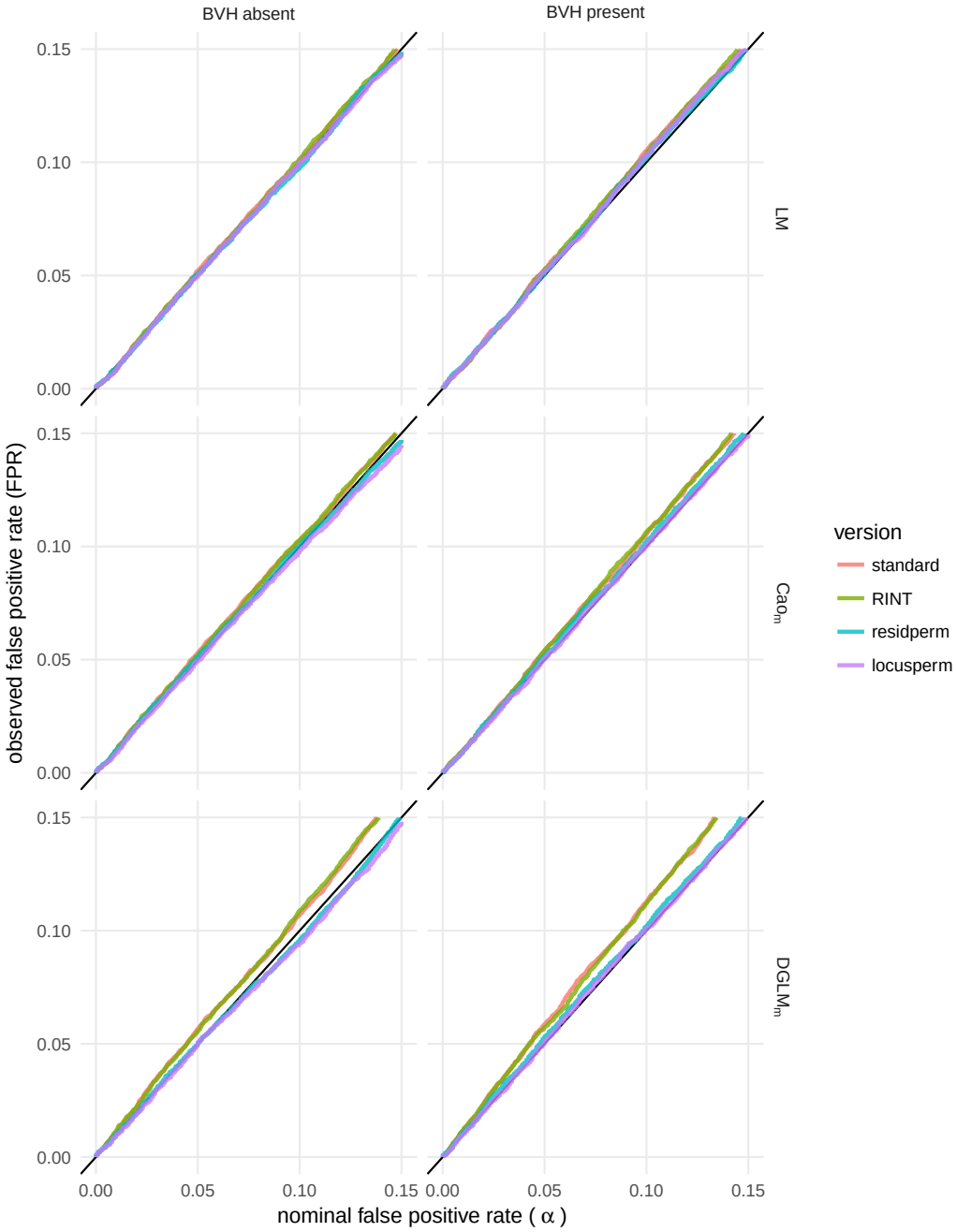
(b) In the absence of BVH, both mvQTL tests accurately control FPR and have similar power. In the presence of BVH of unknown source, the standard procedure of both mvQTL tests is anti-conservative and the other three procedures maintain FPR control but suffer a decrease in power compared with the no-BVH scenario. Only DGLM_{MV} can incorporate information on the BVH-driving covariate. It achieves increased power and accurately controls FPR in its standard and locusperm procedures and is conservative in its RINT and locusperm procedures.



(c) Only the locusperm procedure accurately controls FPR in all scenarios. The standard procedure of both tests are anti-conservative in the presence of BVH of known source and the RINT and residperm procedures are conservative in DGLM_{MV} the presence of BVH of known source.

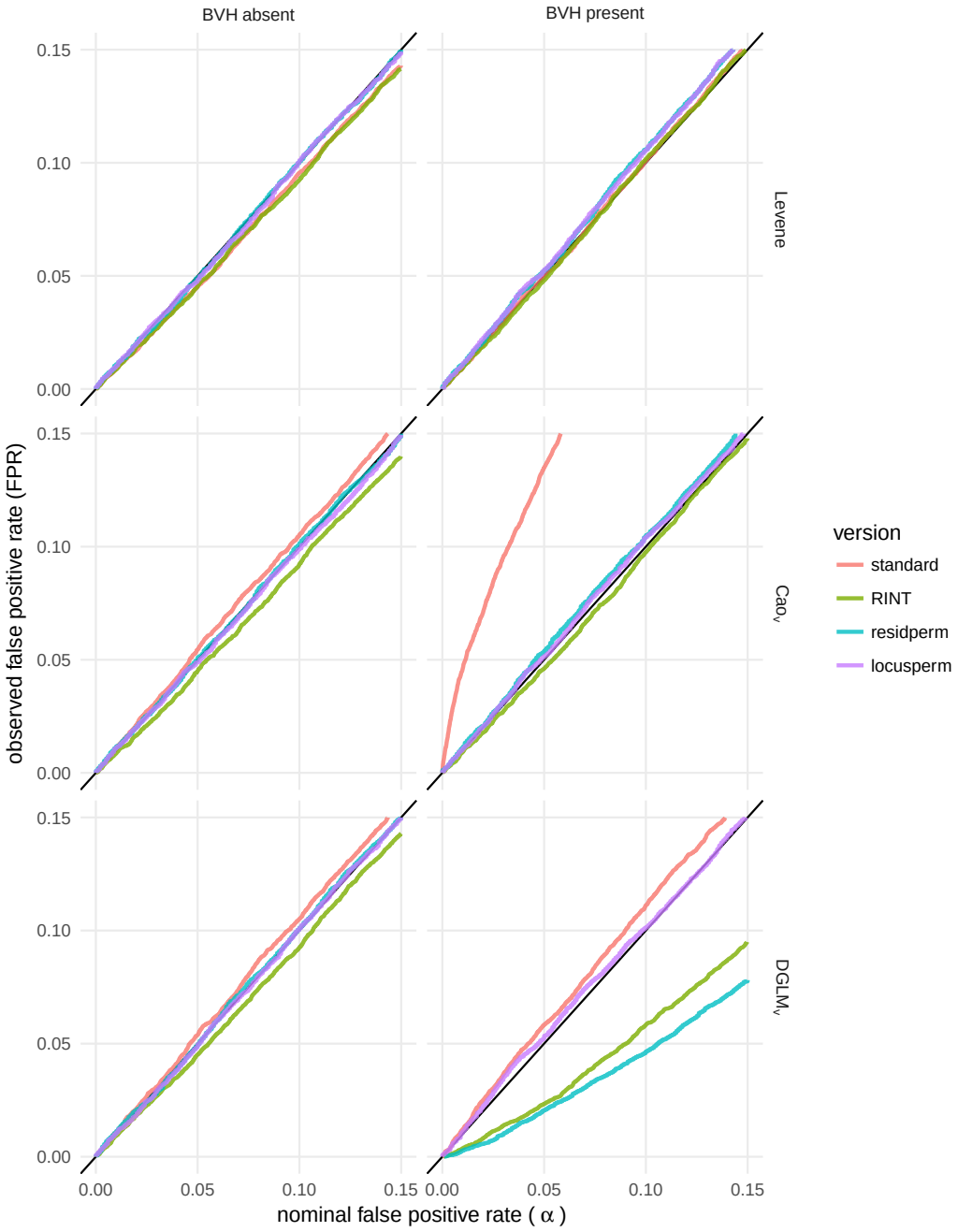
QQ Plots

Figure S4 The empirical false positive rate of each test with each procedure for each nominal false positive rate, α, in [0, 0.1].



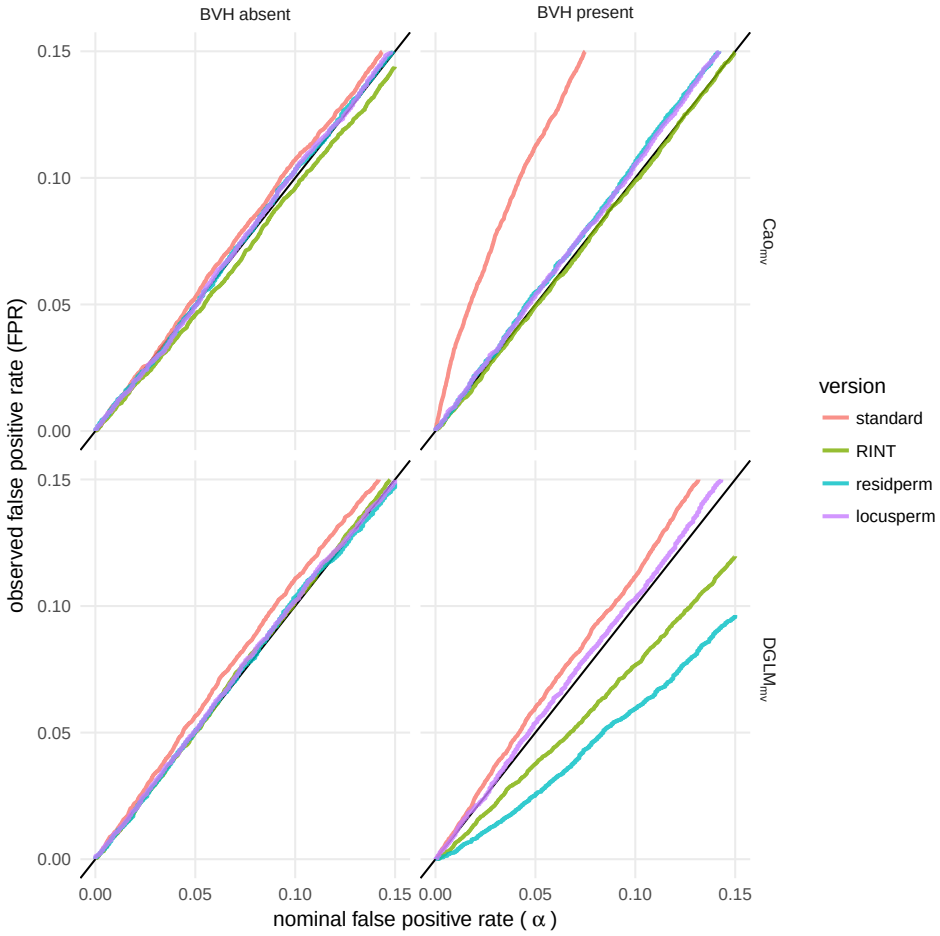
A test that accurately controls FPR will have empirical FPR $= \alpha$ for all value of α. All mQTL tests accurately control FPR.

Figure S5 The empirical false positive rate of each test with each procedure for each nominal false positive rate, α, in [0, 0.1].



A test that accurately controls FPR will have empirical FPR $= \alpha$ for all value of α. Amongst vQTL tests, Cao$_V$ has conservative behavior in the presence of BVH when the standard procedure is used, and DGLM$_V$ has anti-conservative behavior when the RINT and residperm procedures are used.

Figure S6 The empirical false positive rate of each test with each procedure for each nominal false positive rate, α, in $[0, 0.1]$.



A test that accurately controls FPR will have empirical FPR $= \alpha$ for all value of α. mvQTL tests show the same pattern of deviation from accurate FPR control as vQTL tests (Figure S5), but to a lesser extent.

False positive rate of all test-procedure combinations in all scenarios, caterpillar plot

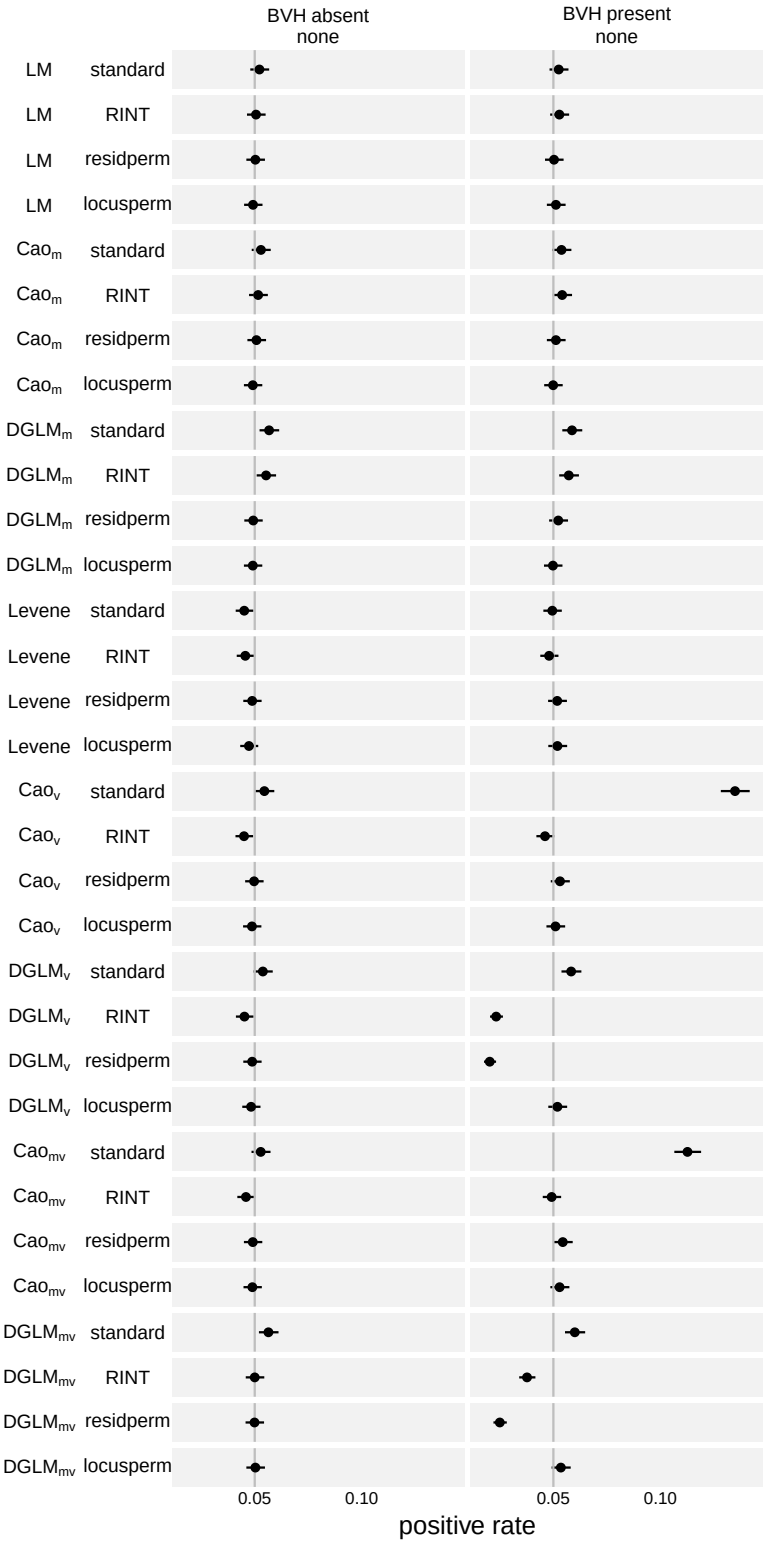
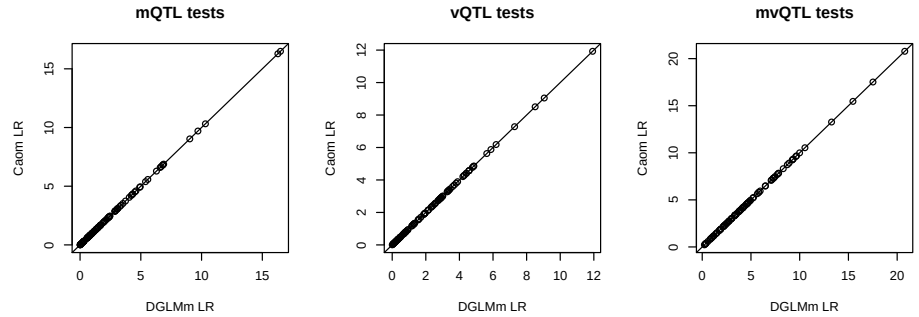


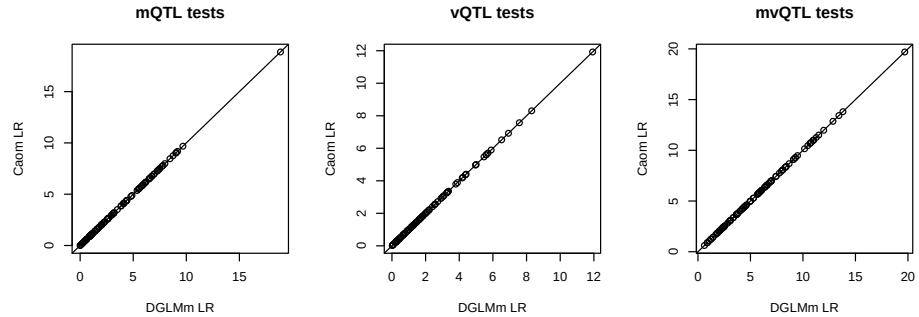
Figure S7

Cao's Profile-Likelihood Approximation is Extremely Accurate

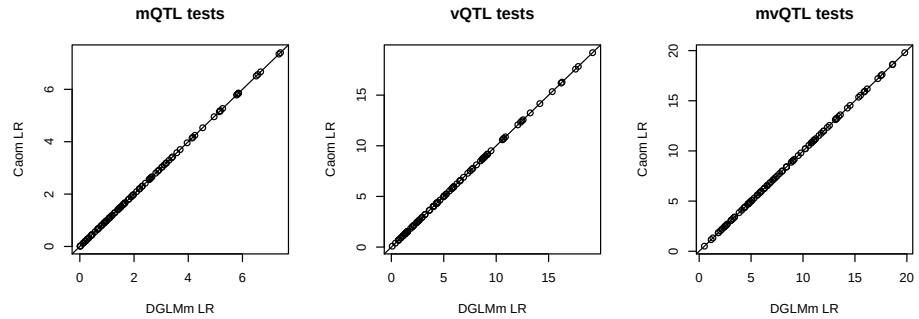
(a) null simulations



(b) mQTL simulations



(c) vQTL simulations



(d) mvQTL simulations

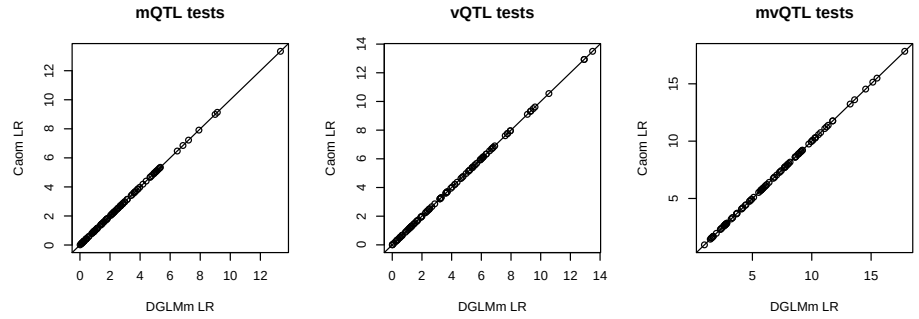


Figure S8 On simulated null loci, mQTL, vQTL, and mvQTL, Cao's method had identical likelihood ratio to the DGLM without any variance covariates. This result illustrates that although Cao's method uses a two-step profile likelihood method which is not guaranteed to attain the global maximum likelihood parameter values and the DGLM uses an iterative procedure that does guarantee this attainment, the difference between the two is negligible in this application.

Cao's Tests for All Phenotypes with BVH

These scans were conducted with the DGLM, without accounting for effects of sex and father on variance, shown by simulation to be identical to Cao's tests (Figure S8).

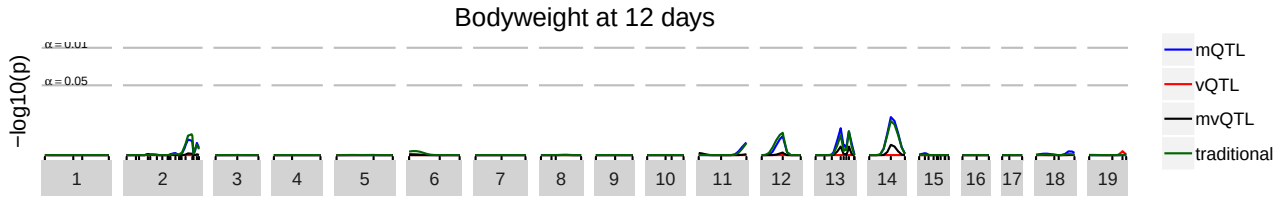


Figure S9

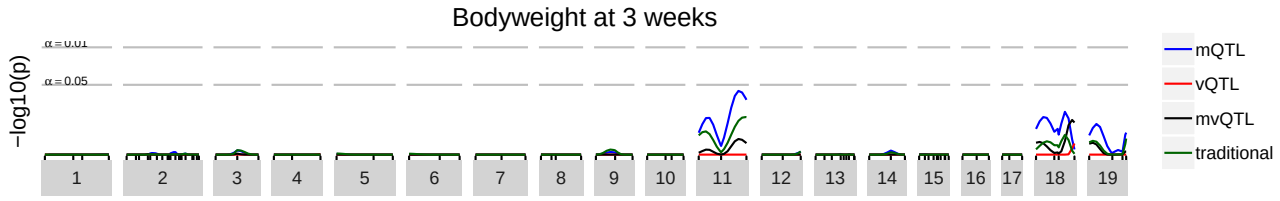


Figure S10

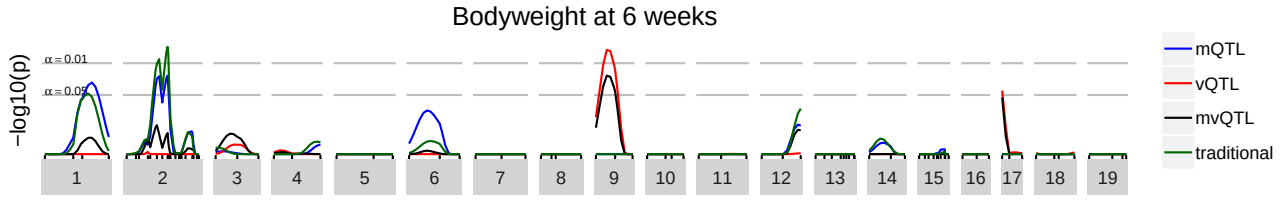


Figure S11

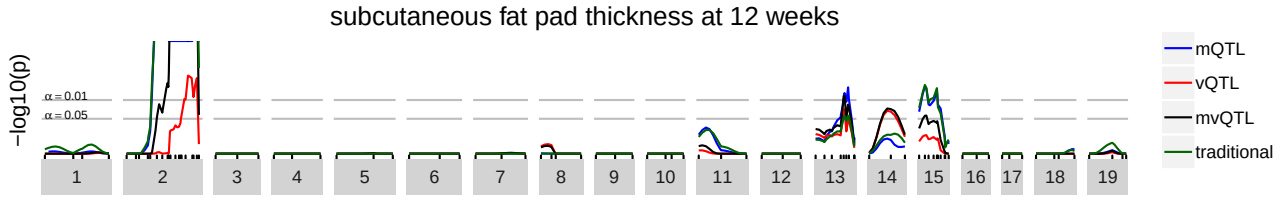


Figure S12

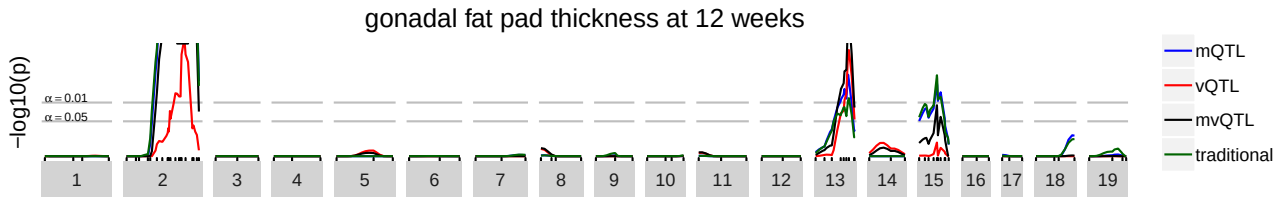


Figure S13

DGLM Tests for All Phenotypes with BVH

These scans were conducted with the DGLM, accounting for effects of sex and father on variance.

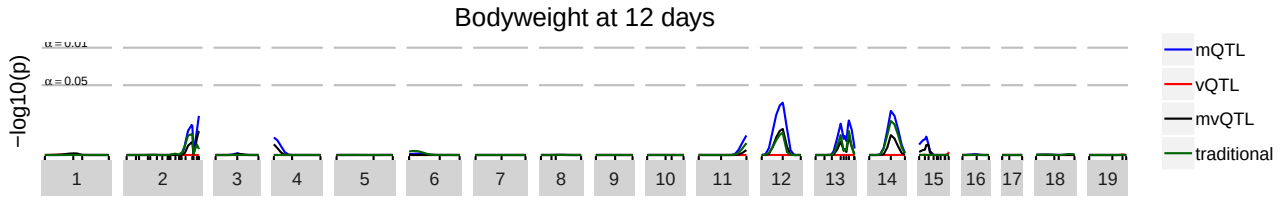


Figure S14 For bodyweight at 12 days, no QTL were identified by the SLM or any DGLM-based tests.

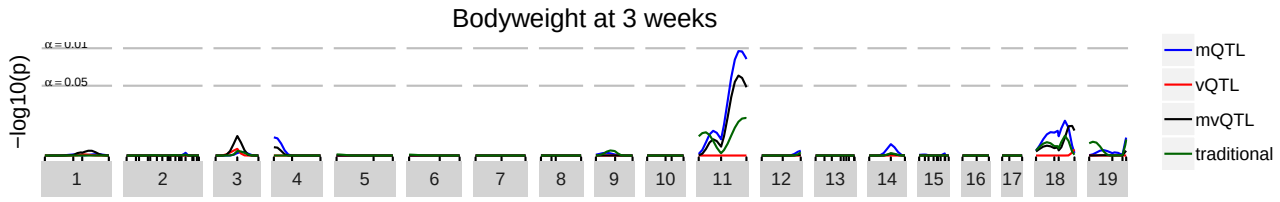


Figure S15 For bodyweight at 3 weeks, an mQTL was identified by DGLM_M, but not by SLM and no vQTL nor mvQTL were identified. Note that this scan differs slightly from that presented in the **Results** section. This scan accommodates sex as a BVH covariate, as per our "screening" procedure with all phenotypes. Given that the effect of sex on phenotype variance was minimal, we removed it from the analysis in further investigation (but kept the effects of father, which were of much greater magnitude).

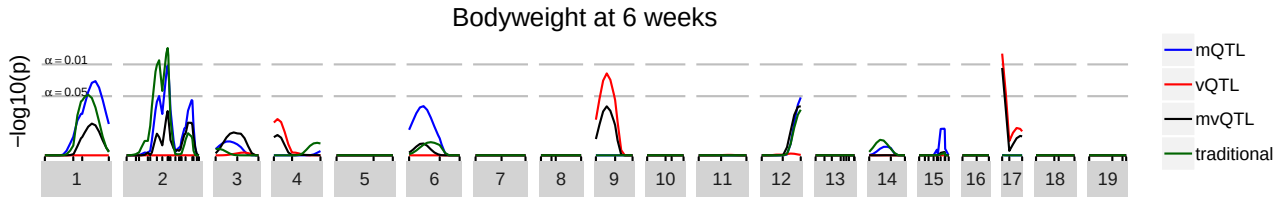


Figure S16 For bodyweight at 6 weeks, SLM identified one mQTL and DGLM_M identified the same (chromosome 2). These tests were also concordant in identifying a suggestive mQTL on chromosome 1. DGLM_V identified a suggestive vQTL on chromosome 9 and a vQTL on chromosome 17.

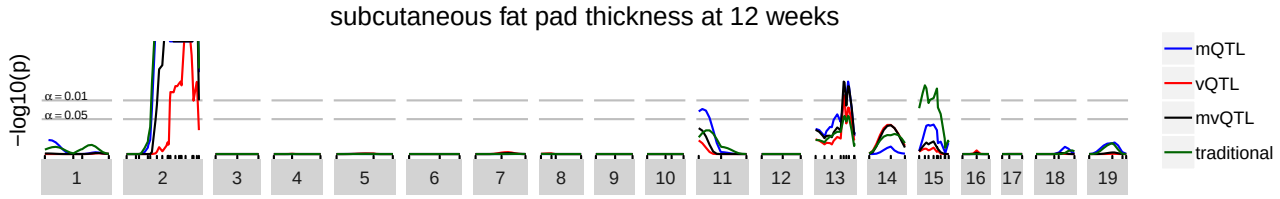


Figure S17 For subcutaneous fat pad thickness at 12 weeks, all tests were highly statistically significant on chromosome 2. On chromosome 13, all tests except SLM were statistically significant. And on chromosome 15, only SLM was statistically significant, reflecting an "un-discovery" due to accommodation of BVH, suggesting that the association identified by SLM was due to a few highly influential observations that were from high-variance sires. Note that the vertical scale is truncated at 10^{-4} for clarity.

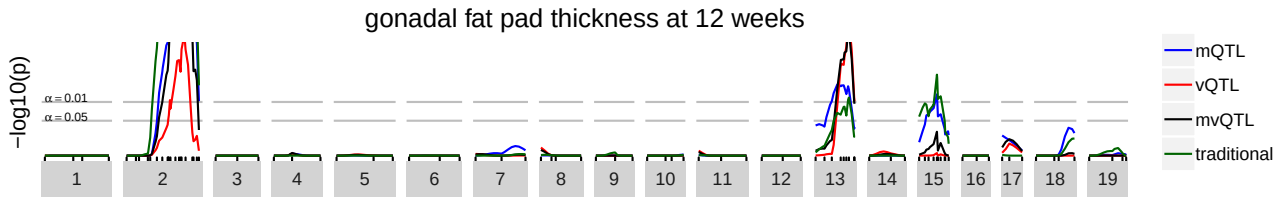


Figure S18 For gonadal fat pad thickness at 12 weeks, all tests were highly statistically significant on chromosome 2. On chromosome 13, all test were statistically significant. An on chromosome 15, both mQTL tests (SLM and DGLM_M) were statistically significant. Note that the vertical scale is truncated at 10^{-4} for clarity.