DEPARTMENT OF STATISTICS
University of Wisconsin
1300 University Avenue
Madison, WI 53706

TECHNICAL REPORT NO. 1157
April 22, 2010

Causal Model Selection Tests in Systems Genetics [1]

Elias Chaibub Neto[2]

Mark P Keller[3]

Aimee T Broman[3]

Alan D Attie[3]

Brian S Yandell[2,4]

[1]This work was supported by CNPq Brazil (ECN); NIDDK grants DK66369, DK58037 and DK06639 (ADA, MPK, BSY, ECN); and by NIGMS grants PA02110 and GM069430-01A2 (BSY).
[2]Department of Statistics, University of Wisconsin - Madison.
[3]Department of Biochemistry, University of Wisconsin - Madison.
[4]Department of Horticulture, University of Wisconsin - Madison.

Causal model selection tests in systems genetics

Elias Chaibub Neto, Mark P Keller, Aimee T Broman,
Alan D Attie, Brian S Yandell

Abstract

We develop a novel hypothesis testing framework for causal inference among pairs of phenotypes in the context of segregating populations. Our model selection test extends Vuong's test to the case of three misspecified models, to handle the full range of possible causal relationships among a pair of traits, namely, causal, reactive or independent models. The ability to properly address misspecified models for systems genetics is key since in general any two phenotypes may be part of a complex network that is grossly oversimplified by the pairwise models. We evaluate and compare our test against the BIC model selection criterium and to another causality inference test in simulation studies using data generated from simple and complex networks and from models affected by measurement error. While our causal model selection test is less powered than alternative approaches when one of the competing models is correctly specified, it makes mistakes at much lower rates than the other approaches under model misspecification and in the presence of measurement error. Finally, we apply and compare these three approaches using a mice intercross data set.

Introduction

A key objective of biomedical research is to unravel the biochemical mechanisms underlying complex disease traits such as type 2 diabetes. Integration of genetic information with genomic, proteomic and metabolomic data has been used to infer causal relationships among phenotypes (Schadt et al. 2005; Li et al. 2006; Kulp and Jagalur 2006; Chen et al. 2006; Zhu et al. 2007; Aten et al. 2008; Liu et al. 2008; Chaibub Neto et al. 2008, 2009; Winrow et al. 2009; Millstein et al. 2009). Current approaches for causal inference in systems genetics can be classified into whole network scoring methods (Li et al. 2006, Zhu et al. 2007, Liu et al. 2008 and Chaibub Neto et al. 2008, 2009, Winrow et al. 2009) or pairwise methods, which focus on the inference of causal relationships among pairs of phenotypes (Schadt et al. 2005, Li et al. 2006, Kulp and Jagalur 2006, Chen et al. 2006 and Aten et al. 2008, Millstein et al. 2009). In this paper we address the latter pairwise problem.

We propose a formal hypothesis testing framework for the inference of causal relationships between pairs of phenotypes. Our approach accommodates multiple quantitative trait loci (QTL), additive and interactive covariates. Given a pair of phenotypes, (Y_1, Y_2), we consider three models: (M_1) Y_1 drives Y_2; (M_2) Y_1 is driven by Y_2; and (M_3) there is no causal relationship between Y_1 and Y_2 and their correlation is a consequence of latent causal phenotypes, common causal QTLs or of common environmental effects (see Figure 1).

Our causal model selection test (CMST) extends Vuong's (1989) asymptotic test, based on Kullback-Leibler information criterium, to the comparison of three models. It uses penalized log-likelihood ratios scaled by their standard errors, as test statistics. The main novelty of the CMST in systems genetics is that it is designed to perform model selection among misspecified models. That is, the true data generating process need not be one of the models under consideration. Accounting for the misspecification of the models is key. In general, any two phenotypes of interest are embedded in a complex network and are affected by many other phenotypes not considered in the grossly simplified (and thus misspecified) pairwise models.

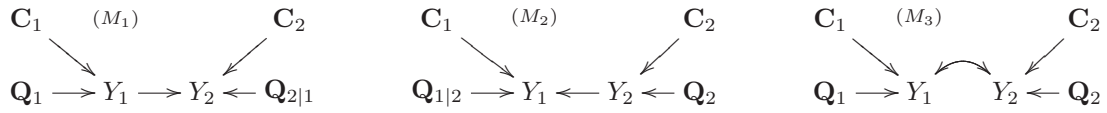

Figure 1. Pairwise causal models. M_1, M_2 and M_3 represent, respectively, the causal, reactive and independent models. Following Li et al. (2006) and Chaibub Neto et al. (2009), QTL detection for the putative reactive trait is done using the putative causal trait as a covariate, whereas standard unconditional mapping analysis is performed to detect QTLs for the putative causal trait. Here $\mathbf{C}$ is a notation for an arbitrary set of covariates and $\mathbf{Q}_{1|2}$ represents the set of QTLs identified in the mapping analysis of Y_1 conditional in Y_2. Note further, that the $\mathbf{Q}$ and $\mathbf{C}$ sets may have QTLs and covariates in common, even though it is not represented in the figure.

The CMST tests the null hypothesis that the selected model is no closer to the true model than the other models. Generally, when the models are misspecified, the standard error of a penalized likelihood ratio is high, shrinking the test statistic value towards zero, avoiding the rejection of the null hypothesis. We evaluate the statistical properties of the CMST in simulation studies where: (1) the true model corresponds to one of our fitted models, i.e., one of our pairwise models is correctly specified; and (2) the true model does not match any of our fitted models, so that M_1, M_2 and M_3 are misspecified. Simulations show that the CMST is conservative test and has error rates that are well controlled under both model misspecification and correct specification.

Previous work on the pairwise problem provides in silico evidence of good power to recover the correct causal relationship when data is simulated from small networks (Schadt et al. 2005, Chen et al. 2007, Aten et al. 2008, Millsteim et al. 2009). In this paper we evaluate and compare the performance of CMST with the BIC model selection criterium and the CIT approach (Millstein et al. 2009) using data simulated from toy and more realistic networks, as well as from small networks subjected to measurement error. Although the BIC and CIT approaches are better powered than the CMST when one of the fitted models is correctly specified, our simulations show that these two approaches can show high error rates under model misspecification. The CMST approach, on the other hand, is designed to forfeit making calls in these adverse situations, and out-performed the BIC and CIT methods.

Large measurement error often leads to conditional independence relationships that are inconsistent with the true causal model. In other words, the faithfulness condition (Spirtes et al. 2000) is violated. This problem is more serious for methods that use a single QTL, common to both phenotypes, as an anchor for causal inference such as the CIT (Millstein et al. 2009) and the approach proposed by Schadt et al. (2005). Theory suggests, and our simulations confirm, that the measurement error problem is reduced with models allowing for different sets of QTLs for each phenotype.

As with most methods for causal inference in the context of eukaryotic genetics, our approach relies on the fact that genetic variation precedes phenotypic variation, and on the fact that Mendelian randomization of alleles provides a mechanism to eliminate the effects of confounding. Both conditions need to be met in order to justify causal claims between QTLs and phenotypes. Causal inference among phenotypes, on the other hand, is justified by conditional independence relations under Markov properties (Li et al. 2006, Chaibub Neto et al. 2009). In the remainder of this paper we will adopt the standard conditional independence notation, $A \perp\!\!\!\perp B \mid C$, as a shorthand for A is independent of B conditional on C (the symbol $\not\perp\!\!\!\perp$ stands for not independent).

Results

Simulation studies

We assessed and compared the empirical error rates of the CMST, BIC and CIT approaches in three distinct simulation studies: 1) with data generated from toy networks composed of two or three phenotype and QTL nodes; 2) with data generated from a more realistic network composed of 30 nodes whose mapping patterns give rise to a QTL hotspot; and 3) with data generated from toy networks affected by measurement error.

In the first simulation study we considered 7 distinct toy networks shown in the Supplement Figure 8. Two out of the 7 networks correspond to the fitted models on M_1 and M_3 Figure 1, and allows us to investigate the case where one of our fitted models is correctly specified. Four of the networks do not correspond to any of our fitted models and allows us to investigate the effects of model misspecification. We also performed simulations under the null model where the phenotypes are uncorrelated. For each one of the 7 toy models we investigated 8 different simulation settings amounting to a total of 56 distinct simulation experiments. A detailed description of the results and the choice of parameters for the simulation experiments is given in the Supplement. Overall, all approaches perform well when the true model corresponds or is close to one of our fitted models, but the BIC and CIT approaches outperform the CMST, since they are better powered than the CMST, and although they show higher error rates than the CMST, the error rates were still adequately low. However, when the true model is more complex and our fitted models are misspecified, the CMST still manages to keep error rates below the adopted significance levels, while the BIC and CIT can show considerably higher error rates.

In the second simulation study we generated data from the more realistic network depicted in Figure 9 in the Supplement. In accordance with the results from first simulation study, CMST made calls and mistakes at lower rates then CIT and BIC. However, this more complex model represents a much more challenging simulation setting in part because the majority of the pairs of phenotypes do not correspond to our simplified fitted models. The top node of this network is, by design, causal to all other nodes; all methods inferred this at high rates. However, inference for pairs of phenotypes located within the network are more difficult because noise propagates from top to bottom of a network and there may be several latent variables (other nodes) that are not specified in the fitted models (Supplement Figure XX). For all methods, the proportion of false discoveries were high when calls were made in the middle of the network, especially for CIT and BIC. CMST made few calls, which resulted in an error rate below the nominal significance level and acceptable FDR overall (see Supplement Table 33).

In the third simulation study we investigate the effects of measurement error using the networks in Supplement Figure 11. Measurement error has been pointed out as an important drawback for causal inference in systems genetics (Schadt et al. 2005, Rockman 2008, Li et al. 2010). If the causal trait is poorly measured in comparison with the reactive trait, model selection approaches may incorrectly support the reversed causal direction. The intuition is that the measurement of the reactive trait may be a better measurement of the causal trait than its own measurement in this situation (Rockman 2008). In the Methods section we provide a formal description of the effects of measurement error in terms of violations of the faithfulness assumption (Spirtes et al. 2000). Using the toy network in the Supplement Figure 11a, we show how the conditional independence relations induced by the measurement error are unfaithful to the causal graph generating the data, but are consistent with the graph with reversed causal relation between phenotypes. However, in situations such as Figure 11b, where the two phenotypes are directly affected by distinct sets of QTLs, the conditional independence relations induced by the measurement error are not consistent with either the causal, reactive or independence models. Therefore, the impact of measurement error is reduced. A detailed description of the results and the choice of parameters for the simulation experiments is given in the Supplement. With no measurement error, all approaches perform well for data generated under models (a) and (b) in Supplement Figure 11, specially at higher R^2 and sample size. Overall, the results are better for the model (b). In the presence of measurement error, all

approaches perform better under data simulated from model (b) than from model (a). Nonetheless, the CIT and specially the BIC tend to make mistakes at higher levels than CMST. Under data generated from model (a) the CMST approach was still affected by measurement error, although to a lesser degree than the BIC and CIT. However, under the more favorable conditions of model (b) the CMST approach made very few calls, and avoided M_2 calls to a great extent.

An application to liver expression data of a mice intercross

We illustrate the application of the CMST, BIC and CIT approaches to a liver gene expression data set derived from a F2 intercross between inbred lines $C3H/HeJ$ and $C57BL/6J$ (Ghazalpour et al. 2006, Wang et al. 2006). The data set is composed of genotype data on 1,065 markers and liver expression data on the 3,421 public available transcripts from 135 female mice.

QTL mapping analysis detected an eQTL hotspot located at 29.06cM (see supplementary Figure 12) on chromosome 2. We investigated the causality architecture of this chromosome, that is, we tested each one of the genes physically located on chromosome 2 against the 78 expression traits composing the hotspot. We selected chromosome 2 for this study because it shows the strongest genetic linkages in this data set.

Figure 2a presents the results for the BIC approach. The red bars represent the number of traits causally affected by the local trait, i.e., the number of M_1 calls. The blue bars show the number of M_2 calls, that is, the number of traits affecting the local trait. The black bars show the number of M_3 calls, i.e., the number of traits associated with the local trait due to common genetic factors, latent variables and environmental factors. For the BIC approach we forfeit making a call when the difference between the BIC scores of the best and second best models corresponds to an approximate Bayes factor smaller than 10. The bars are stacked over each other in the sequence red, blue and black. The plot shows many M_1 and M_2 relationships widespread over the chromosome. Figure 15 in the Supplement shows the results for much higher Bayes factors thresholds. Although, the number of calls decrease as we increase the threshold, we still see a considerable amount of M_1 and M_2 calls over the entire chromosome.

The CMST approach, on the other hand, shows a completely different picture (Figure 2b). The vast majority of tests were not significant at a 0.05 significance level. Only *Pscdbp*, a transcription factor physically located at 32.4cM (58.4Mb) and mapping to a QTL located at 28.07cM, stands out showing 12 out of 77 significant M_1 calls. *Pscdbp* encodes a protein modulating the activation of alternate open reading frame genes.

Figure 2c presents the results for the CIT approach. Because the CIT can only be applied to traits that map to a common locus, many fewer tests were actually performed in this case. Nonetheless, we still observe a fair amount of significant M_1, M_2 and M_3 calls across the entire chromosome.

It is interesting to note that the locations pinpointed by the CMST approach were also detected by the CIT approach and, of course, by the BIC approach (recall that the CMST attaches a p-value to the model selected by the BIC criterion). Furthermore, the BIC and CIT approaches detected a greater number of M_1 calls on those regions.

We also tested each one of the 78 traits composing the hotspot against each other. The results depicted in Figure 3 show that the CMST detects causal relationships at much lower rates than the BIC and CIT in this situation too. Because all 78 traits co-map to the same genomic region, this time we could apply the CIT to most pairs of traits, and it actually detected causal relationships at slightly higher rates than the BIC.

Once a set of causal relations have been inferred, it is useful to examine the empirical mapping patterns (see Figure 14 in the Supplement) to verify that they agree with what is expected with the causal call. Sometimes, we find evidence for a causal call, but the mapping pattern includes contradictions. We are investigating formal methods to check for consistent mapping patterns. Manual inspection of the 12 traits reactive to *Pscdbp* found no problems.

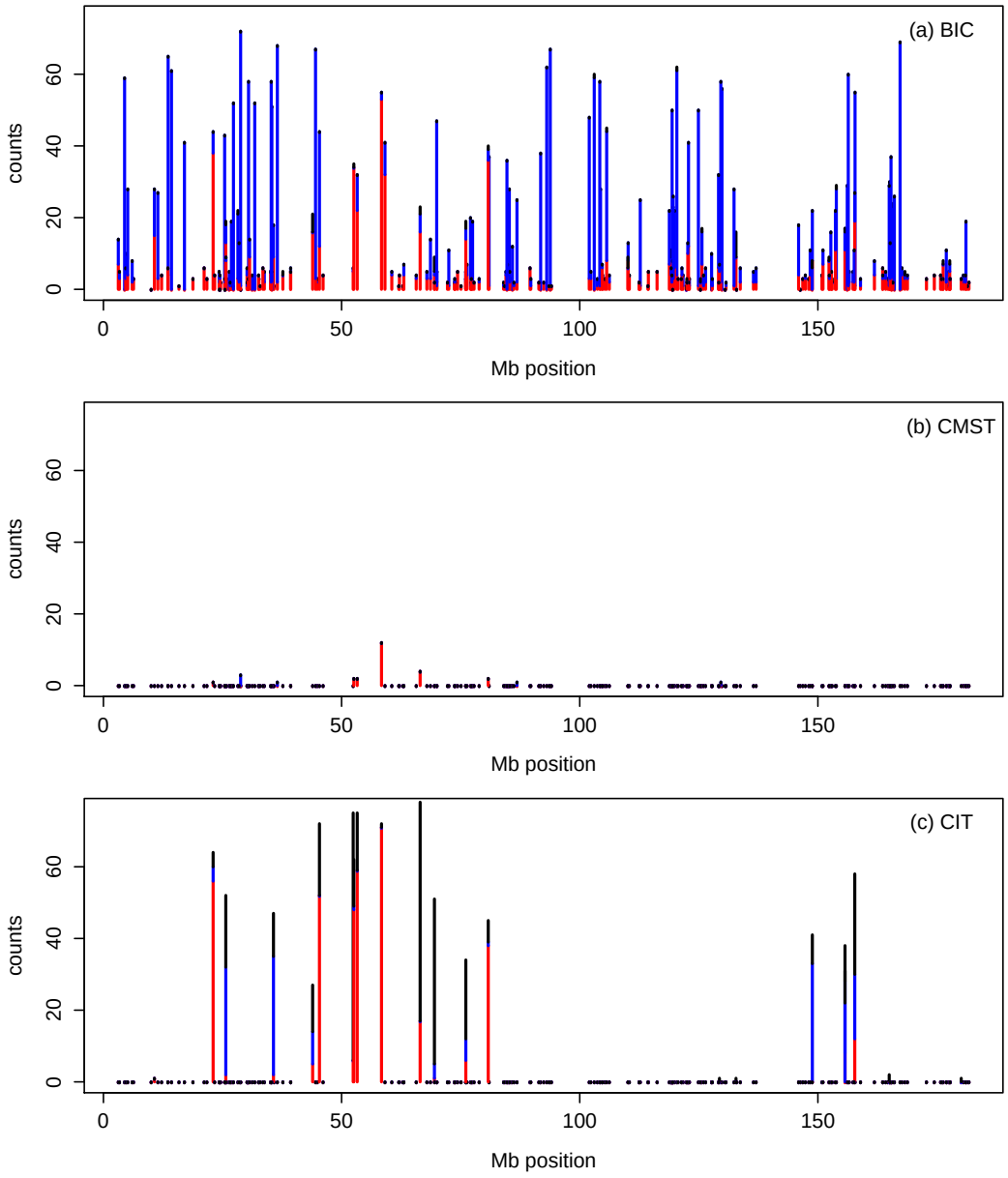

Figure 2. Causal architectures restricted to the transcripts composing the hotspot on chromosome 2. Panels (a), (b) and (c) present the causal architectures uncovered by the BIC, CMST and CIT approaches, respectively. The red, blue and black bars represent, respectively, the numbers of M_1, M_2, M_3 calls, according to approximate Bayes factors for the BIC approach, and significance p-values for the CMST and CIT. We adopted a Bayes factor threshold of 10 and significance level 0.05. The bars are stacked over each other in the sequence red, blue and black.

We constructed a causal network using the QTLnet algorithm (Chaibub Neto et al. 2009) with *Pscdbp* and the 12 transcripts detected by the CMST approach (Figure 4). *Pscdbp* is at the top of the network driving the expression of the remainder transcripts.

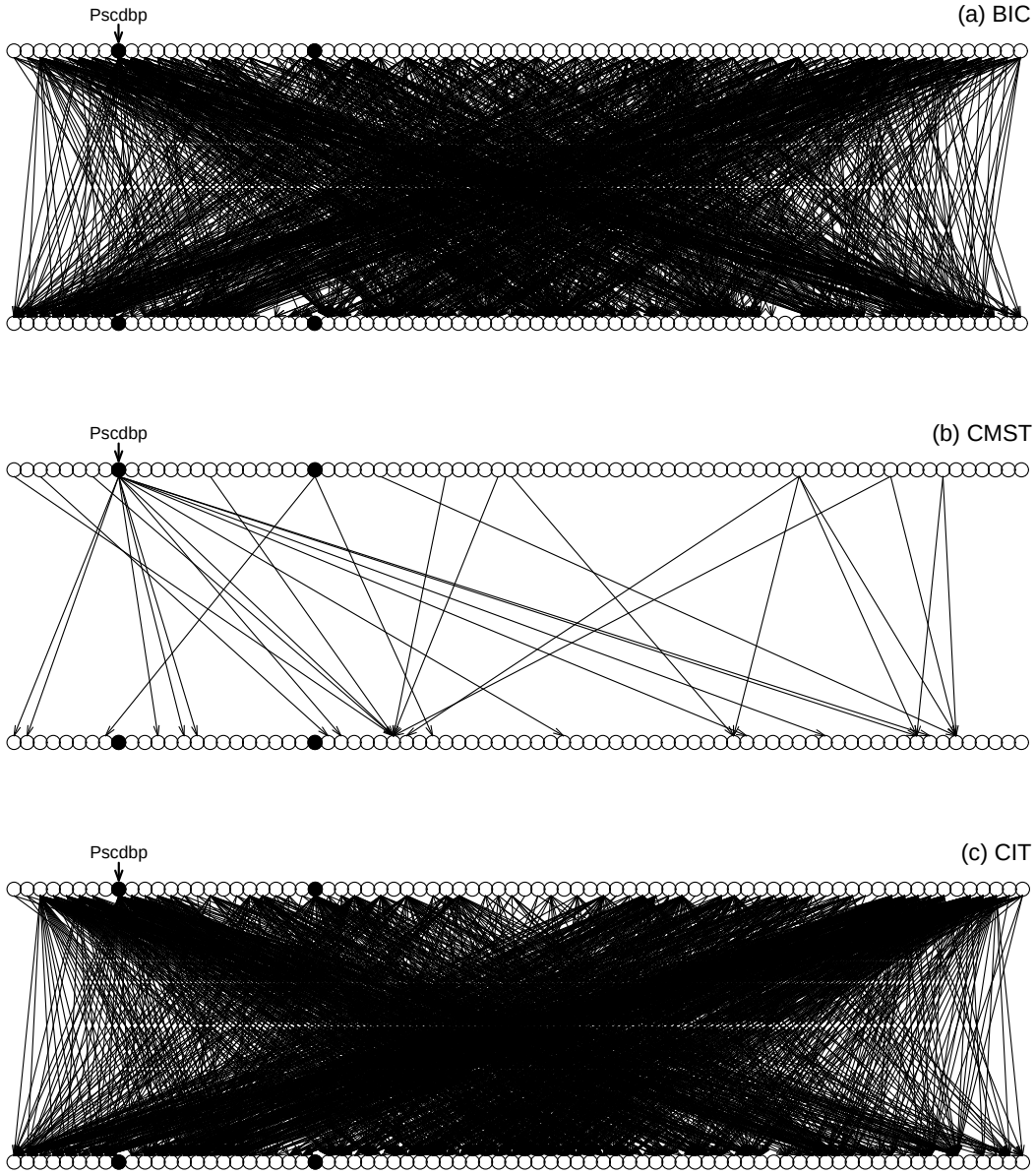

Figure 3. Inferred causal relationships among all pairs of traits composing the hotspot on chromosome 2. Panel (b) presents the results for the CMST. The dots represent the traits and the arrows show the inferred causal relationships among pairs of traits. The BIC and CIT approaches infer causal relationships at much higher rates than the CMST. Out of the 3,003 possible pairs of traits derived from the 78 traits composing the hotspot, the CMST detected only 30 causal relationships, whereas the BIC and CIT detected 1,248 and 1,413 relationships, respectively. The black dots represent local traits physically located close to the hotspot peak that show significant linkages.

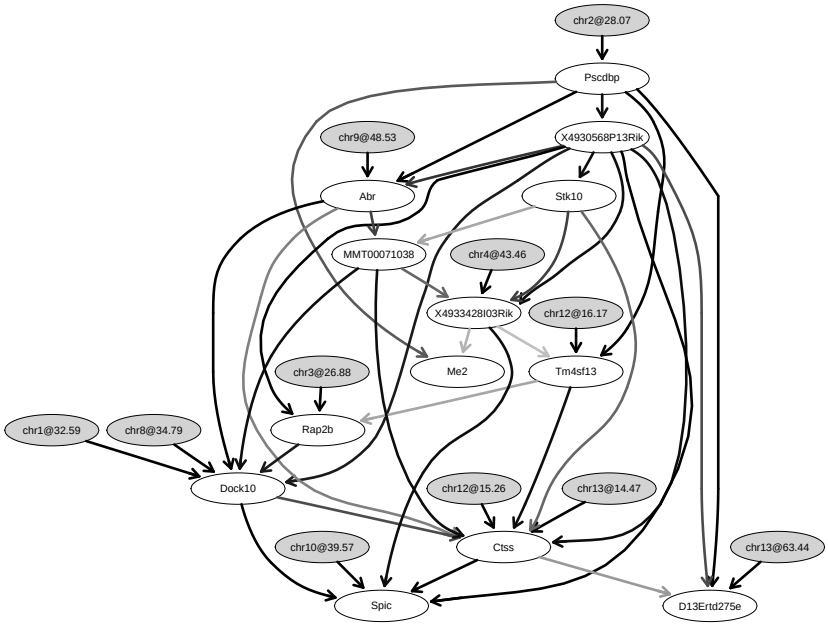

Figure 4. Causal network constructed with the QTLnet algorithm. This algorithm learns the structure of a mixed Bayesian network composed of phenotypes and QTL nodes. It jointly infers the causal phenotype network and the associated genetic architecture, and is able to sort out the direct and indirect effects of QTLs and phenotypes. This network approach supports the CMST results that *Pscdbp* causally affects the 12 remainder transcripts. QTLnet is implemented with a Metropolis-Hastings algorithm. Diagnostic plots and measures (see Supplement) support the convergence of the Markov chain.

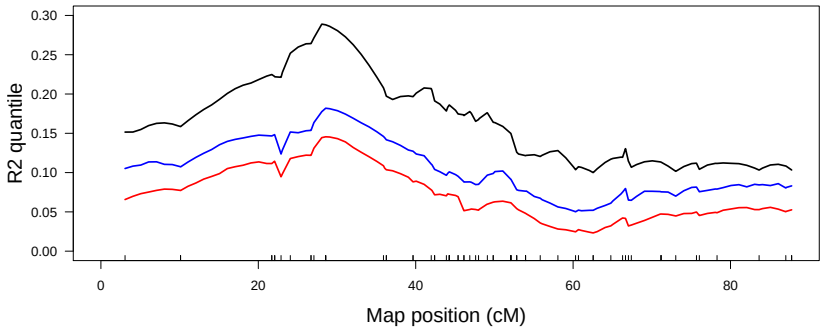

Figure 5. R^2 quantiles based on 78 traits composing the eQTL hotspot on chromosome 2. The black, blue and red curves represent, respectively, the 1th, 10th and 39th highest R^2 values, among the 78 transcripts. The R^2 values were computed based on simple regression models, where we regress the phenotype on the QTL, and measure the proportion of the phenotype variance explained by the QTL. Overall, the QTLs explain a fairly modest amount of the total variability of the phenotypes.

It is important to notice that the sample size in this study is only 135, and that the genetic signal is, in general, modest as seen from Figure 5. Nonetheless, it is reassuring to see that the region with highest signal coincides with the region where the CMST made more significant calls. Note, however, that this is not the case for the BIC and CIT approaches.

The BIC approach made more calls than the other approaches and showed many M_1 and M_2 calls widespread over the chromosome (Figure 2a). We still see many calls over the entire chromosome even when we adopt Bayes factor thresholds as high as 10,000 (Figure 15 in the Supplement). Because the Bayes factor is a relative measure of how well two different models fit the data, it is possible to get very high Bayes factors when both models have bad fits but one of them is much worse than the other. All BIC plots show many calls in regions where no causal signal was found with the CMST approach.

These analysis results are consistent with the expected behavior of the CMST approach and with our simulation results. For example, the simulation results for model (d) with $n = 135$ and $R^2 = 0.25$ show power and conditional power around 0.22 at a 0.05 significance level (see Table 15 in the Supplement). For $R^2 = 0.125$ the power and conditional power quickly decrease to 0.04 and 0.05, respectively. The R^2 associated with *Pscdbp* and the QTL at 28.07cM is 0.24, and we got 12 significant M_1 calls out of 77 tests giving an approximate power of 0.16. Also, 70 out of 77 traits showed an M_1 relationship with *Pscdbp* (with BF=1), giving an approximate conditional power of 0.17. The CMST approach returned "no calls" most of the time and the few calls made by the CMST were located in the genomic regions showing the highest signal levels. The BIC and CIT approaches detected calls at higher rates then the CMST in these regions, in accordance with our simulation results that these approaches are better powered than the CMST. However the BIC and CIT approaches also detected many calls across the whole chromosome and in the situation depicted in Figure 3, that are probably false positives as suggested by our simulations.

Discussion

In his pioneering work, Schadt et al. (2005) proposed an AIC model selection approach to infer the causal direction between pairs of phenotypes. The authors perform genome scans for the phenotypes and whenever two correlated phenotypes co-map to a same genomic region, this common QTL is used to infer the causal direction. The underlying assumption is that either: (a) the reactive phenotype Y_2 maps to Q indirectly via the causal phenotype Y_1; (b) the other way around; (c) or both phenotypes map to Q and their correlation is explained by Q or by the effects of other loci and/or environmental effects not included in the model. Model (c) is denoted the independence model. Mathematically, model (c) represents three distribution/likelihood equivalent models: (c_1) $Q \rightarrow Y_1 \rightarrow Y_2$; (c_2) $Q \rightarrow Y_1 \leftarrow Y_2$; and (c_3) $Q \rightarrow Y_1 \leftrightarrow Y_2$, where the double arrow represents statistical association without a causal direction. Note that only (c_3) represents a true independence causal relation. However, since the causal models c_1 and c_2 cannot be distinguish from c_3 using the data, they are all lumped together under the name independence model. The authors evaluated the performance of their approach in a simulation study with data generated from a toy network composed by one QTL and two phenotype nodes.

Chen et al. (2007) proposed a non-parametric empirical Bayes approach to estimate the posterior probability that a transcript Y_1 has a causal regulatory effect on Y_2. The authors consider triplets (Q_1, Y_1, Y_2) where Q_1 is a local-QTL for Y_1 and a distal-QTL to Y_2, and estimate the posterior probability that $Q_1 \rightarrow Y_1 \rightarrow Y_2$, based on three testable conditional independence relations that hold if and only if Y_1 causally affects Y_2, and there is no confounding due to latent variables affecting both phenotypes. This single QTL approach provides a direct quantification of the uncertainty associated with the causal relationship and allows for the estimation of false discovery rate. The performance of the approach was assessed with data simulated from a mixture of six different toy networks composed by a single QTL and two or three phenotype nodes.

Aten et al. (2008) considered a model selection approach based on local structural equation models (SEM) of pairs of phenotypes and multiple markers. SEM focus in the modeling of covariance relationships

and represent a widely used tool for the evaluation of the fit of causal models. The SEM framework allows for the explicit modeling of a latent variable whenever the phenotypes are affected by non-overlapping sets of markers. The method's performance was evaluated in two simulation studies with data simulated from a small network containing six phenotype nodes and a latent variable, in the first, and from a two-phenotype toy network, in the second.

Millstein et al. (2009) proposed an intersection-union causality test, for single QTL models, in which a number of equivalence and conditional F tests are conservatively combined in a single test. In a nutshell, the authors consider a series of regression models and use equivalence tests to test whether a regression coefficients are different from zero, and conditional F-tests to test if some of the regression coefficients are equal to zero. The null hypothesis for the causal and reactive models are then constructed as unions of the simpler equivalence and conditional F-test null hypothesis. No formal test is conducted for the independence model and the decision rule adopted by the authors is: (1) to make a causal call if the causal p-value is less than a significance threshold α and the reactive p-value is greater than α; (2) to make a reactive call if it is the other way around; (3) to make an independence call if both p-values are greater than α; and (4) to make a "no call" if both p-values are less than α. Note that they make a "no call" when both causal and reactive models are well supported by their tests. The authors evaluated the performance of their method in a simulation study using six different types of toy networks involving one QTL and two or three phenotype nodes.

Sun et al. (2007) were the first to apply Vuong's model selection test in the context of causal inference in systems genetics. The authors employ a variant of Schadt's test to determine the causal ordering between transcription factors and distal (trans) genes using a local (cis) expression trait as a causal anchor. Explicitly, they compare models (a) $GC \rightarrow TA \rightarrow GT$; (b) $GC \rightarrow GT \rightarrow TA$; and (c) $GT \leftarrow GC \rightarrow TA$, where GC, TA and GT represent, respectively, the expression level of the cis-trait, the transcription factor activity and the expression level of the trans-trait. Nonetheless the authors perform three separate Vuong's test to compare their models and do not account for the correlation structure among the three test statistics. Interesting the authors do not make any mention about the usefulness of Vuong's tests to deal with model misspecification and justify their adoption of this test as a way to perform a likelihood ratio test comparing regression models with different sets of covariates.

In this paper we propose a novel hypothesis test for causal inference in systems genetics. We assess the significance of a selected model by testing the null hypothesis that the selected model is not significantly closer to the true model than the other models, according to the Kullback-Leibler distance. Our test is designed to be conservative and to forfeit making calls when our fitted models are misspecified. We evaluated the performance of pairwise approaches in a more realistic settings where the true model does not correspond to any of our fitted models, and as expected, our simulation studies showed that the more realistic settings represents a much more challenging inference problem. The BIC and CIT approaches showed high error rates in these more challenging settings, whereas our method produced mistakes at considerably lower rates. As expected, all approaches perform well when one of the fitted models is correctly specified, although the CMST approach shows lower power. We point out, however, that situations where our fitted models are misspecified are likely the rule and not the exception, so that the eventual advantage of the BIC and CIT approaches under correct specification, is largely contra-balanced by their higher error rates under model misspecification. Furthermore, we can always use the CMST approach to detect the most reliable regions, followed by investigation of the candidates detected by the BIC approach on those regions.

Contrary to previous approaches that mostly focus on single QTL models with additive genetic effects, our multiple QTL approach also handles dominance genetic effects, additive covariates such as sex and batch-effects and genotype by additive covariate interactions (we restricted our simulations to additive effect for the sake of comparison with the CIT approach, that is only implemented for additive genetic effects). Previous studies provide in-silico evidence that multiple QTL models improve the power to detect causal relationships (Aten et al. 2008). An extension to incorporate epistatic effects detected

under a model selection framework (Manichaikul et al. 2009) is possible, but computationally expensive and yet to be done.

In a similar spirit to Li et al. (2006) and following Chaibub Neto et al. (2009) the detection of QTLs for the putative reactive phenotype is done conditional on the putative causal phenotype. By allowing the phenotypes to have distinct sets of QTLs, instead of relying in a single common QTL as a causal anchor, our approach is often able to distinguish among the distribution/likelihood equivalent models c_1, c_2 and c_3 of Schadt et al. (2005). When both phenotypes are affected by exactly the same set of QTLs, our approach is unable to distinguish among then. Our independence model (M_3) corresponds to a seemingly unrelated regression (SUR) model (Zellner 1962). In a nutshell, the SUR model is an extension of the multivariate regression model that, in addition to imposing a correlation structure to the error terms, also allows the independent variables to be affected by different sets of dependent variables. This attractive feature of SUR models has been explored in the context of multiple trait QTL mapping (Verzilli et al 2005, Banerjee et al 2008). Models M_1 and M_2 are homogeneous conditional Gaussian regression (HCGR) models (Chaibub Neto et al. 2009). The main difference between a SUR model and the HCGR model is that in SUR modeling we assume that the correlation structure between the phenotypes is a consequence of a correlation between the error terms, induced by latent variables and uncontrolled environmental factors, whereas in the HCGR model the correlation between the phenotypes is defined by the causal relationships between the phenotypes, as determined by the directed graph structure.

Measurement error has been pointed out as an important drawback for causal inference in systems genetics. Rockman (2008) argues that there is good biological reason to be concerned about this problem since regulatory molecules are often present in the cell at lower concentrations than the structural protein transcripts they target. The problem is that the latter may be easier to measure with good precision, than the low abundance regulatory transcripts that might be present at the threshold of detection. Rockman suggests going back to the early microarray experimental designs with technical replicates as a way to estimate and incorporate measurement error in the statistical modeling. We re-cast the measurement error problem as a violation of the faithfulness assumption (Spirtes et al. 2000) of causal inference theory. We argue that although the measurement error problem may affect the CMST, it is likely a more important problem to approaches that use a single QTL common to both phenotypes, as an anchor for causal inference, such as the CIT and the approach proposed by Schadt et al. (2005). Our simulations provide in silico evidence that the CMST provide some protection against measurement error, specially when the phenotypes are directly affected by distinct sets of QTLs.

Genome-wide application of our causal model selection tests involves a large number of hypothesis tests. Unfortunately, false discovery rate (FDR) control mechanisms will likely fail to work with misspecified models. To see why, consider Vuong's model selection test (see Supplement) comparing only models M_1 and M_2 (for the sake of simplicity) applied to data generated from model (f), in Figure 8, that shows a causal relationship from Y_1 to Y_2, but is also affected by the confounder variable Y_5 and by Q_5. In our simulations using $R^2 = 0.25$ and $n = 135$, the empirical distribution of the test statistic Z_{12} is symmetric around zero (Figure 13 in the Supplement), so the number of false and true discoveries corresponding, respectively, to the left and right tails (M_2 and M_1 calls, respectively) of this distribution will tend to be equally likely, no matter what significance threshold we adopt. Consequently, the FDR will tend to be around 0.5 for any significance threshold we choose.

We point out, however, that although the FDR of the CMST was sometimes high for models (c), (e) and (f), the actual number of significant calls tended to be rather small so that the error rates were kept low. Therefore, the high FDRs are not a big concern in these cases. On the other hand, the FDR was usually low for models (b), (d) and (g). Finally, empirical estimates of the FDR associated with a network, i.e., the proportion of edges in the network that are mistakes, will depend on the complexity of the network. Our simulations from the network on Figure 9 show considerably higher FDR than the simulations from a mixture of the 7 different models in Figure 8.

Our simulations show that the CMST has adequate power and low error and false discovery rates

for pairs of nodes showing a type (d) model, where the correlation between the phenotypes is due to the causal effect of the driver phenotype into the target phenotype, and there is no confounding due to common ancestors, i.e., latent nodes affecting both phenotypes. For example, our simulations from the network on Figure 9 show that CMST mostly recover the causal relationships between Y_1 and the remaining 29 other phenotypes. These results suggest that in practice putative phenotype drivers showing almost exclusively M_1 relationships with the other phenotypes represent our best candidates. On the other hand, we should be suspicious about traits showing just a few significant calls for both M_1 and M_2 relationships.

Although our model selection test can handle model misspecification in very broad terms, i.e., the true and fitted models may even belong to different parametric families, in our applications model misspecification simply means that the fitted models fail to incorporate important causal covariates such as other parent phenotype nodes in the true phenotype network. Whole network methods such as Zhu et al. (2007, 2008), Liu et al. (2008), Winrow et al. (2009) and Chaibub Neto et al. (2009) can potentially incorporate these covariates and, in principle, are immune to this type of problem. Furthermore, contrary to the pairwise approaches that can only determine whether a phenotype is upstream or downstream another one, the whole network approaches can tell whether the causal relations between two phenotypes are direct or indirect. The drawback of network approaches is that they tend to be computationally expensive and sensitive to latent variables.

The CMST approach is currently implemented for inbred line crosses. Extension to outbred populations involving mixed effects models is yet to be done. Finally, although in this paper we focused in a systems genetics application, our extension of Vuong's model selection tests to the comparison of three models is actually valid to the comparison of any three non-nested, overlapping and not observationally identical models, under the fairly general conditions given by Vuong (1989).

Methods

Causal model selection tests

In this section we present our causal model selection test (CMST). Its derivation is technical and is left to the Supplement. In a nutshell, the MST is a extension of Vuong's model selection test (Vuong 1989) to the comparison of three possibly misspecified models, M_1, M_2 and M_3. The test is based on the Kullback-Leibler information criterium (Kullback 1959) and is derived from the null hypothesis that none of the three models is closer to the true model, i.e.,

$$H_0: \quad E^0\left[\log\frac{f_1}{f_2}\right] = 0 \text{ and } E^0\left[\log\frac{f_1}{f_3}\right] = 0 \text{ and } E^0\left[\log\frac{f_2}{f_3}\right] = 0,$$

where $f_u = f_u(\mathbf{y} \mid \mathbf{x}_u\,;\,\boldsymbol{\theta}_{u*})$ represents the probability density function of model M_u for $u = 1, 2, 3$; $\boldsymbol{\theta}_{u*}$ represents the parameter value that minimizes the Kullback-Leibler distance from f_u to the true model; and E^0 represents the expectation with respect to the true joint distribution. Under H_0,

$$\mathbf{Z} = \text{diag}(\hat{\boldsymbol{\Sigma}})^{-\frac{1}{2}}\,\mathbf{L}\tilde{\mathbf{R}}/\sqrt{n} \longrightarrow^d N_3(\mathbf{0}\,,\,\boldsymbol{\rho})\,, \tag{1}$$

where $\boldsymbol{\rho} = \text{diag}(\boldsymbol{\Sigma})^{-\frac{1}{2}}\,\boldsymbol{\Sigma}\,\text{diag}(\boldsymbol{\Sigma})^{-\frac{1}{2}}$ is the correlation matrix. $\mathbf{L}\tilde{\mathbf{R}} = (L\tilde{R}_{12}\,,\,L\tilde{R}_{13}\,,\,L\tilde{R}_{23})^T$ is a vector of penalized log-likelihood statistics, $L\tilde{R}_{uv} = LR_{uv} - D_{uv}$, where

$$LR_{uv} = \sum_{i=1}^{n} \log\frac{\hat{f}_{u,i}}{\hat{f}_{v,i}}, \tag{2}$$

and $\hat{f}_{u,i} = f_u(\mathbf{y}_i \mid \mathbf{x}_{ui}\,;\,\hat{\boldsymbol{\theta}}_u)$; $\hat{\boldsymbol{\theta}}_u$ is the maximum likelihood estimates of $\boldsymbol{\theta}_u$; $D_{uv} = k_{f_u} - k_{f_v}$ or $D_{uv} = (k_{f_u} - k_{f_v})(\log n)/2$ for the AIC and BIC penalties, respectively, where k_{f_u} and k_{f_v} represent the number

of parameters of f_u and f_v, respectively. And

$$\hat{\mathbf{\Sigma}} = \begin{pmatrix} \hat{\sigma}_{12.12} & \hat{\sigma}_{12.13} & \hat{\sigma}_{12.23} \\ \hat{\sigma}_{12.13} & \hat{\sigma}_{13.13} & \hat{\sigma}_{13.23} \\ \hat{\sigma}_{12.23} & \hat{\sigma}_{13.23} & \hat{\sigma}_{23.23} \end{pmatrix}, \tag{3}$$

where

$$\hat{\sigma}_{uv.st} = \frac{1}{n}\sum_{i=1}^{n}\left[\log\frac{f_{u,i}}{f_{v,i}}\log\frac{f_{s,i}}{f_{t,i}}\right] - \left[\frac{1}{n}\sum_{i=1}^{n}\log\frac{f_{u,i}}{f_{v,i}}\right]\left[\frac{1}{n}\sum_{i=1}^{n}\log\frac{f_{s,i}}{f_{t,i}}\right]. \tag{4}$$

Note that the elements of the test statistic vector $\mathbf{Z}$

$$Z_{uv} = \frac{L\tilde{R}_{uv}}{\sqrt{n\,\hat{\sigma}_{uv.uv}}} = \frac{-(BIC_u - BIC_v)/2}{\sqrt{n\,\hat{\sigma}_{uv.uv}}}, \tag{5}$$

correspond to penalized log-likelihood ratios scaled by their standard errors, or equivalently, to a scaled contrast of BIC scores, if we adopt the BIC penalty. (Similarly, if we adopt the AIC penalty, Z_{uv} corresponds to a scaled contrast of AIC scores.) Note that because the scaling factors in our test statistics are positive, performing model selection using the BIC criterium is equivalent to performing model selection using the CMST test statistics. Explicitly, the data supports model: M_1 if $BIC_1 < BIC_2$ and $BIC_1 < BIC_3$, or equivalently, if $z_{12} > 0$ and $z_{13} > 0$; M_2 if $BIC_1 > BIC_2$ and $BIC_2 < BIC_3$, or equivalently, if $z_{12} < 0$ and $z_{23} > 0$; M_3 if $BIC_1 > BIC_3$ and $BIC_2 > BIC_3$, or equivalently, if $z_{13} < 0$ and $z_{23} < 0$. The advantage of the CMST is that it goes one step further and provides a formal measure of significance for the penalized log-likelihood scores that accounts for possible mis-specifications of the models.

Given that the data supports model M_u better, the CMST allows us to test the null hypothesis, $H_0^{M_u}$, that model M_u is not significantly closer to the true model, according to the Kullback-Liebler distance, than the other models. Note that given the best model, the relative closeness of the other models to the true model has no further impact on the model selection call. Therefore, we derive the rejection region for the null hypothesis from the appropriate asymptotic bivariate marginal null distribution. For example, we make a call in favor of M_1 if z_{12} and z_{13} are both positive and we test if the call is statistically significant using the asymptotic marginal null distribution $(Z_{12}, Z_{13})^T \sim N_2(\mathbf{0}\,,\,\boldsymbol{\rho}_{12.13})$. Table 1 presents our CMST tests. The details of the derivation of these tests are given in the Supplement.

H_0	Test statistic distr. under H_0	Rejection region	P-value
$H_0^{M_1}$	$(Z_{12}, Z_{13})^T \sim N_2(\mathbf{0}, \boldsymbol{\rho}_{12.13})$	$z_{12} > c^{\alpha}_{12.13}\,,\; z_{13} > c^{\alpha}_{12.13}$	$p^{MS}_{M_1} = p_{12.13}$
$H_0^{M_2}$	$(Z_{12}, Z_{23})^T \sim N_2(\mathbf{0}, \boldsymbol{\rho}_{12.23})$	$z_{12} < -c^{\alpha}_{12.23}\,,\; z_{23} > c^{\alpha}_{12.23}$	$p^{MS}_{M_2} = p_{12.23}$
$H_0^{M_3}$	$(Z_{13}, Z_{23})^T \sim N_2(\mathbf{0}, \boldsymbol{\rho}_{13.23})$	$z_{13} < -c^{\alpha}_{13.23}\,,\; z_{23} < -c^{\alpha}_{13.23}$	$p^{MS}_{M_3} = p_{13.23}$

Table 1. Model selection tests. Here $c^{\alpha}_{uv.st} = (\chi^2_{2,\alpha}\,\lambda_{uv.st}/2)^{1/2}$ corresponds to the critical value of the test of level α based on the joint distributions of (Z_{uv}, Z_{st}); $\lambda_{uv.st}$ corresponds to the first eigenvalue of the bivariate marginal correlation matrix $\boldsymbol{\rho}_{uv.st}$; $\chi^2_{2,\alpha}$ corresponds to the αth upper quantile of a chi-squared distribution with two degrees of freedom; and
$p_{uv.st} = \max\left\{P(\chi^2_2 \geq 2z^2_{uv}/\lambda_{uv.st})\,,\, P(\chi^2_2 \geq 2z^2_{st}/\lambda_{uv.st})\right\}$.

Measurement error and unfaithfulness

A common assumption in the causal graphical models literature is that of the faithfulness of a probability distribution to a graph (Spirtes et al. 2000). This assumption implies that given a graph and a probability distribution associated with it, all the conditional independence relations spanned by a probability distribution must match the d-separation (Pearl 1988, 2000) relations predicted from the graph structure.

Unfaithfulness of a probability distribution is often (but not always) a consequence of a very particular parametrization of the model generating the data. Suppose, for example, the data is generated from the following model

$$Q \rightarrow Y_1 \rightarrow Y_2 \ , \tag{6}$$

where Q is a discrete random variable with variance σ_q^2, $(Y_1 \mid q) \sim N(\beta_{1q}\,, \sigma_1^2)$ and $(Y_2 \mid y_1, q) \sim N(\beta_{2q} + \beta_{21}\, y_1\,, \sigma_2^2)$. In this case $Cov(Y_2, Q) = (\beta_{2q} + \beta_{21}\beta_{1q})\sigma_q^2$ and we see that if $\beta_{2q} = -\beta_{21}\beta_{1q}$, then $Cov(Y_2, Q) = 0$ and the data shows the probabilistic independence $Y_2 \perp\!\!\!\perp Q$, whereas d-separation clearly predicts that $Y_2 \not\perp\!\!\!\perp Q$.

Unfaithfulness can affect model selection when the probability distribution of the data is unfaithful to the structure of the data generating process but is faithful to the structure of a different causal model. For instance, application of d-separation to graph 6 yields the following conditional (in)dependence relations: $Y_1 \not\perp\!\!\!\perp Q$, $Y_1 \not\perp\!\!\!\perp Y_2$, $Y_2 \not\perp\!\!\!\perp Q$, $Y_1 \not\perp\!\!\!\perp Q \mid Y_2$, $Y_2 \not\perp\!\!\!\perp Q \mid Y_1$ and $Y_1 \not\perp\!\!\!\perp Y_2 \mid Q$. However, when $\beta_{2q} = -\beta_{21}\beta_{1q}$ we have that $Y_2 \perp\!\!\!\perp Q$ (although the other five (in)dependence relations are still preserved). The problem is that the (in)dependence relations $Y_1 \not\perp\!\!\!\perp Q$, $Y_1 \not\perp\!\!\!\perp Y_2$, $Y_2 \perp\!\!\!\perp Q$, $Y_1 \not\perp\!\!\!\perp Q \mid Y_2$, $Y_2 \not\perp\!\!\!\perp Q \mid Y_1$ and $Y_1 \not\perp\!\!\!\perp Y_2 \mid Q$, are consistent with graph $Q \rightarrow Y_1 \leftarrow Y_2$, so that model selection criteria will tend to favor this model with reversed causal relationship between Y_1 and Y_2.

It has been argued in the literature (Meek 1995, Spirtes et al 2000) that this type of faithfulness violation is probably not important in practice since it requires very specific parameterizations and the set of such finely-tuned parameterizations has measure zero. Others have argued, however, that there are situations in which such finely-tuned parameterizations are feasible (for example, Shipley (2000) points out homeostasis mechanisms balancing counter-acting causal effects as likely examples).

A second, and probably far more widespread form of unfaithfulness is caused by measurement error. It has long been recognized that measurement error can potentially lead model selection approaches to the inference of inverted causal direction (Schadt et al. 2005, Rockman 2008, Li et al. 2010). We argue here that the measurement error problem can be framed in the context of a faithfulness violation problem, where measurement error provides a mechanism to the generation of conditional independence relations that do not match the d-separation relations implied by the graphical model generating the data.

To illustrate the problem consider the toy network on Figure 11a. The random variables Y^* and Y represent, respectively, the true and measured values of the traits. Further, suppose that the reactive trait Y_2^* is well measured so that Y_2 is a good approximation to Y_2^*, whereas Y_1 is a poor measurement of Y_1^*. Because Y_1 is poorly measured in comparison with Y_2, it follows that Y_2 is a better measurement of Y_1^* than Y_1 itself (Rockman 2008). The consequence is that Y_2 has more information about Q_1 than Y_1, and given the value of Y_2, Y_1 does not gives us further information about Q_1. Formally, this implies that $Y_1 \perp\!\!\!\perp Q_1 \mid Y_2$, what disagrees with one of the relations predicted by the application of d-separation to the graph in Figure 11a: i.e., $Y_1 \not\perp\!\!\!\perp Q_1$, $Y_2 \not\perp\!\!\!\perp Q_1$, $Y_1 \not\perp\!\!\!\perp Y_2$, $Y_1 \not\perp\!\!\!\perp Q_1 \mid Y_2$, $Y_2 \not\perp\!\!\!\perp Q_1 \mid Y_1$ and $Y_1 \not\perp\!\!\!\perp Y_2 \mid Q_1$. Moreover, in the presence of measurement error, the conditional independence relations generated by the data (i.e., $Y_1 \not\perp\!\!\!\perp Q_1$, $Y_2 \not\perp\!\!\!\perp Q_1$, $Y_1 \not\perp\!\!\!\perp Y_2$, $Y_1 \perp\!\!\!\perp Q_1 \mid Y_2$, $Y_2 \not\perp\!\!\!\perp Q_1 \mid Y_1$ and $Y_1 \not\perp\!\!\!\perp Y_2 \mid Q_1$) exactly match the conditional independence relations predicted by application of d-separation to the causal graph $Q_1 \rightarrow Y_2 \rightarrow Y_1$, and model M_2 provides a better fit to the data than M_1.

However, the measurement error problem gets alleviated in situations where the phenotypes are directly affected by different sets of QTLs, such as in Figure 11b. The set of conditional independence relations in Figure 11b does not match the set of conditional independence relations associated with any of the causal graphs: (i) $Q_1 \rightarrow Y_1 \rightarrow Y_2 \leftarrow Q_2$; (ii) $Q_1 \rightarrow Y_1 \leftarrow Y_2 \leftarrow Q_2$; and (iii) $Q_1 \rightarrow Y_1 \leftrightarrow Y_2 \leftarrow Q_2$.

(For example, in Figure 11b we have that $Y_1 \not\perp\!\!\!\perp Q_2 \mid Y_2$, $Y_2 \not\perp\!\!\!\perp Q_1$ and $Y_2 \not\perp\!\!\!\perp Q_1 \mid Y_1$. In the causal graph (i) we have $Y_1 \not\perp\!\!\!\perp Q_2 \mid Y_2$, $Y_2 \not\perp\!\!\!\perp Q_1$ and $Y_2 \perp\!\!\!\perp Q_1 \mid Y_1$. In causal graph (ii) $Y_1 \perp\!\!\!\perp Q_2 \mid Y_2$, $Y_2 \perp\!\!\!\perp Q_1$ and $Y_2 \not\perp\!\!\!\perp Q_1 \mid Y_1$. And in causal graph (iii) we have $Y_1 \perp\!\!\!\perp Q_2 \mid Y_2$, $Y_2 \perp\!\!\!\perp Q_1$ and $Y_2 \perp\!\!\!\perp Q_1 \mid Y_1$.)

References

1. Aten JE, Fuller TF, Lusis AJ, Horvath S (2008) Using genetic markers to orient the edges in quantitative trait networks: the NEO software. *BMC Sys Biol* **2**: 34.

2. Banerjee S, Yandell BS, Yi N (2008) Bayesian QTL mapping for multiple traits. *Genetics* **179**: 2275-2289.

3. Broman K, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**: 889-890.

4. Chaibub Neto E, Ferrara C, Attie AD, Yandell BS (2008) Inferring causal phenotype networks from segregating populations. *Genetics* **179**: 1089-1100.

5. Chaibub Neto E, Keller MP, Attie AD, Yandell BS (2009) Causal graphical models in system genetics: a unified framework for joint inference of causal network and genetic architecture for correlated phenotypes. *Annals of Applied Statistics* (in press).

6. Chen LS, Emmert-Streib F, Storey JD (2007) Harnessing naturally randomized transcription to infer regulatory relationships among genes. *Genome Biology* **8**: R219.

7. Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. *Genetics* **138**: 963-971.

8. Geweke J (1992) Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4* (ed JM Bernado, JO Berger, AP Dawid and AFM Smith). Clarendon Press, Oxford, UK.

9. Ghazalpour A, Doss S, Zhang B, Wang S, Plaisier C, et al. (2006) Integrating Genetic and Network Analysis to Characterize Genes Related to Mouse Weight. *PLoS Genetics* **2**(8): e130.

10. Haley C, Knott S (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**: 315-324.

11. Kass RE and Raftery A (1995) Bayes factors. *Journal of the American Statistical Association* **90**: 773-795.

12. Kulp DC, Jagalur M (2006) Causal inference of regulator-target pairs by gene mapping of expression phenotypes. *BMC Genomics* **7**: 125.

13. Kullback S (1959) Information theory and statistics. John Wiley and Sons. New York.

14. Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185-199.

15. Li R, Tsaih SW, Shockley K, Stylianou IM, Wergedal J, Paigen B, Churchill GA (2006) Structural model analysis of multiple quantitative traits. *PLoS Genetics* **2**: e114.

16. Li Y, Tesson BM, Churchill GA, Jansen RC (2010) Critical preconditions for causal inference in genome-wide association studies (manuscript).

17. Liu B, de la Fuente A, Hoeschele I (2008) Gene network inference via structural equation modeling in genetical genomics experiments. *Genetics* **178**: 1763-1776.

18. Manichaikul A, Moon JY, Sen S, Yandell BS, Broman KW (2009) A model selection approach for the identification of quantitative trait loci in experimental crosses. *Genetics* **181**: 1077-1086.

19. Meek C (1995) Strong completeness and faithfulness in Bayesian networks. *Proceedings of the Eleventh Conference on Uncertainty in AI*. Norman Kauffman. San Francisco. 411-418.

20. Millstein J, Zhang B, Zhu J, Schadt EE (2009) Disentangling molecular relationships with a causal inference test. *BMC Genetics* **10**: 23 doi:10.1186/1471-2156-10-23.

21. Pearl J (1988) Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann.

22. Pearl J (2000) Causality: models, reasoning and inference. Cambridge University Press.

23. Raftery AE, Lewis SM (1992). One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Statistical Science* **7**, 493-497.

24. Raftery AE, Lewis SM (1995). The number of iterations, convergence diagnostics and generic Metropolis algorithms. In Practical Markov Chain Monte Carlo (W.R. Gilks, D.J. Spiegelhalter and S. Richardson, eds.). London, U.K.: Chapman and Hall.

25. Rockman MV (2008) Reverse engineering the genotype-phenotype map with natural genetic variation. *Nature* **456**: 738-744.

26. Sawa T (1978) Information criteria for discriminating among alternative regression models. *Econometrica* **46**: 1273-1291.

27. Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D, Sieberts SK, Monks S, Reitman M, Zhang C, Lum PY, Leonardson A, Thieringer R, Metzger JM, Yang L, Castle J, Zhu H, Kash SF, Drake TA, Sachs A, Lusis AJ (2005) An integrative genomics approach to infer causal associations between gene expression and disease. *Nature Genetics* **37**: 710-717.

28. Shipley, B (2002) Cause and Correlation in Biology. Cambridge University Press.

29. Spirtes P, Glymour C, Scheines R (2000) Causation, prediction and search. MIT Press.

30. Srivastrava VK, Giles DEA (1987) Seemingly unrelated regression equation models. Marcel Dekker. New York.

31. Sun W, Yu T, Li KC (2007) Detection of eQTL modules mediated by activity levels of transcription factors. *Bioinformatics* **23**: 2290-2297.

32. Verzilli CJ, Stallard N, Whittaker JC (2005) Bayesian modeling of multivariate quantitative traits using seemingly unrelated regression. *Genetic Epidemiology* **28**: 313325.

33. Vuong QH (1989) Likelihood ratio tests for model selection and non-nested hypothesis. *Econometrica* **57**: 307-333.

34. Wang S, Yehya N, Schadt EE, Wang H, Drake TA, Lusis AJ (2006) Genetic and genomic analysis of a fat mass trait with complex inheritance reveals marked sex specificity. *PLoS Genetics* **2**: e15.

35. Winrow CJ, Williams DL, Kasarskis A, Millstein J, Laposky AD, Yang HS, Mrazek K, Zhou L, Owens JR, Radzicki D, Preuss F, Schadt EE, Shimomura K, Vitaterna MH, Zhang C, Koblan KS, Renger JJ, Turek FW (2009) Uncovering the genetic landscape for multiple sleep-wake traits. *PLoS ONE* **4**: e5161.

36. Zellner A (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J. Am. Stat. Assoc.* **57**: 348368.

37. Zhu J, Wiener MC, Zhang C, Fridman A, Minch E, et al. (2007) Increasing the Power to Detect Causal Associations by Combining Genotypic and Expression Data in Segregating Populations. *PLoS Computational Biology* **3**(4): e69. doi:10.1371/journal.pcbi.0030069

38. Zhu J, Zhang B, Smith EN, Drees B, Brem RB, Kruglyak L, Bumgarner RE, Schadt EE (2008) Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nature Genetics* **40**: 854-861.

Supplement for:

Causal model selection tests in systems genetics

Elias Chaibub Neto, Mark P Keller, Aimee T Broman,
Alan D Attie, Brian S Yandell

Vuong's model selection tests

Vuong (1989) proposed asymptotic likelihood ratio tests for model selection of two nested, non-nested and non-nested overlapping models, applicable in situations where one or more competing models may be misspecified. He showed that under fairly general conditions the asymptotic distribution of the likelihood ratio statistic converges to a weighted sum of chi-square distributions or to a normal distribution depending on whether the models are observationally identical or not. Two models are observationally identical if their probability densities are the same, when evaluated at the respective pseudo-true parameter values, i.e., $f_1(\mathbf{y}\,;\,\boldsymbol{\theta}_{1*}) = f_2(\mathbf{y}\,;\,\boldsymbol{\theta}_{2*})$ for almost all $\mathbf{y}$, where the pseudo-true parameter values, $\boldsymbol{\theta}_{k*}$, corresponds to the parameter value that minimizes the Kullback-Leibler distance from the true model (Sawa 1978).

Our focus is model selection for non-nested Gaussian regression models with potentially overlapping but distinct sets of covariates. Furthermore, we mostly consider models that are not observationally identical. (Vuong points out that in the normal linear regression case, two overlapping models are observationally identical if and only if the pseudo-true parameters associated with the variables specific to each regression are simultaneously null. In other words, regression models are observationally identical if and only if they have the same set of covariates. In our applications, when the phenotypes map to different sets of QTLs the models are not observationally identical.) Hence we will only present here, Voung's model selection test for non-nested, not observationally identical, and overlapping models.

Vuong's test derives from the Kullback-Leibler (1959) Information Criterion (KLIC) that measures the closeness of a probability model to the true distribution generating the data. Formally, let $\{f(\mathbf{y} \mid \mathbf{x}\,;\,\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ represent a parametric family of conditional models. Then

$$\begin{aligned} KLIC(h^0; f) &= E^0\left[\log h^0(\mathbf{y} \mid \mathbf{x})\right] - E^0\left[\log f(\mathbf{y} \mid \mathbf{x}\,;\,\boldsymbol{\theta}_*)\right] \\ &= \int_{\mathbf{x}} \left[\int_{\mathbf{y}} h^0(\mathbf{y} \mid \mathbf{x}) \, \log \frac{h^0(\mathbf{y} \mid \mathbf{x})}{f(\mathbf{y} \mid \mathbf{x}\,;\,\boldsymbol{\theta}_*)} \, d\mathbf{y}\right] h^0(\mathbf{x})\, d\mathbf{x}, \end{aligned} \tag{7}$$

where E^0 represents the expectation with respect to the true joint distribution $h^0(\mathbf{y}, \mathbf{x}) = h^0(\mathbf{y} \mid \mathbf{x})h^0(\mathbf{x})$, and $\boldsymbol{\theta}_*$ is the parameter value that minimizes the KLIC distance from f to the true model (Sawa 1978). Note that f need not belong to the same parametric family as h^0.

A model $f_1(\mathbf{y} \mid \mathbf{x}_1\,;\,\boldsymbol{\theta}_{1*})$, denoted f_1 for short, is regarded as a better approximation to the true model $h^0(\mathbf{y} \mid \mathbf{x})$, than the alternative model $f_2(\mathbf{y} \mid \mathbf{x}_2\,;\,\boldsymbol{\theta}_{2*})$ if and only if $KLIC(h^0; f_1) < KLIC(h^0; f_2)$, or alternatively, $E^0[\log f_1] > E^0[\log f_2]$ (Sawa 1978). Vuong's model selection test is based on later criterion and the test hypotheses are defined as

$$H_0: \quad E^0\left[\log \frac{f_1}{f_2}\right] = 0, \qquad H_1: \quad E^0\left[\log \frac{f_1}{f_2}\right] > 0, \qquad H_2: \quad E^0\left[\log \frac{f_1}{f_2}\right] < 0. \tag{8}$$

The null hypothesis is that models f_1 and f_2 are equally close to the true distribution. H_1 means that f_1 is better than f_2 and conversely for H_2. The notation $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}$ makes explicit that the models f_1 and f_2 can have different sets of covariates from the true set of covariates $\mathbf{x}$.

The quantity $E^0[\log f_1] - E^0[\log f_2]$ is unknown, but Vuong (1989) showed that under fairly general conditions

$$\frac{1}{n} LR_{12}(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) \longrightarrow^{a.s.} E^0\left[\log \frac{f_1}{f_2}\right], \quad \hat{\sigma}_{12.12} \longrightarrow^{a.s.} Var^0\left[\log \frac{f_1}{f_2}\right] = \sigma_{12.12}, \tag{9}$$

where $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ are the maximum likelihood estimates and

$$LR_{12}(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) = \sum_{i=1}^{n} \log \frac{\hat{f}_{1,i}}{\hat{f}_{2,i}} \tag{10}$$

is the log likelihood ratio statistic, where $\hat{f}_{1,i} = f_1(\mathbf{y}_i \mid \mathbf{x}_{1i}\,; \hat{\boldsymbol{\theta}}_1)$, and the variance is

$$\hat{\sigma}_{12.12} = \frac{1}{n} \sum_{i=1}^{n} \left[\log \frac{\hat{f}_{1,i}}{\hat{f}_{2,i}}\right]^2 - \left[\frac{1}{n} \sum_{i=1}^{n} \log \frac{\hat{f}_{1,i}}{\hat{f}_{2,i}}\right]^2 . \tag{11}$$

Vuong's model selection statistic for overlapping, non-nested and not observationally identical models (see Vuong's condition (6.4) on page 320 and Theorem 5.1 on page 318) converges to

$$\text{under } H_0\text{:} \quad LR_{12}(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)/\sqrt{n\hat{\sigma}_{12.12}} \longrightarrow^{d} N(0,1), \tag{12}$$

$$\text{under } H_1\text{:} \quad LR_{12}(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)/\sqrt{n\hat{\sigma}_{12.12}} \longrightarrow^{a.s.} +\infty, \tag{13}$$

$$\text{under } H_2\text{:} \quad LR_{12}(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)/\sqrt{n\hat{\sigma}_{12.12}} \longrightarrow^{a.s.} -\infty. \tag{14}$$

This test is based on the unadjusted log likelihood ratio statistic. However, competing models may have different dimensions, requiring a complexity penalty. The penalized log-likelihood ratio is then

$$\tilde{LR}_{12}(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) = LR_{12}(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) - D_{12}, \tag{15}$$

where $D_{12} = k_{f_1} - k_{f_2}$ or $D_{12} = (k_{f_1} - k_{f_2})(\log n)/2$ for the AIC and BIC penalties, respectively, and k_{f_1} and k_{f_2} represent the number of parameters of f_1 and f_2, respectively. Because the penalty term divided by $n^{1/2}$ converges to zero, $n^{-1/2}\tilde{LR}_{12}(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)/\sqrt{\hat{\sigma}_{12.12}}$ has the same asymptotic properties as $n^{-1/2}LR_{12}(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)/\sqrt{\hat{\sigma}_{12.12}}$ and we can use the adjusted log likelihood ratio for the model selection test (Vuong 1989).

Model selection tests for the comparison of three models

We now present our causal model selection test (CMST). In our applications we consider three models M_1, M_2 and M_3. Simple application of three separate model selection tests, namely, $f_1 \times f_2$, $f_1 \times f_3$ and $f_2 \times f_3$ can be misleading, since the three test statistics are dependent (for example, LR_{12} share its numerator with LR_{13} and its denominator is the numerator of LR_{23}). In order to adjust for the dependency among the test statistics, we consider a multivariate extension of Vuong's model selection tests, as follows.

Under the same general regularity conditions of Vuong (1989), the empirical covariance

$$\hat{\sigma}_{12.13} = \frac{1}{n} \sum_{i=1}^{n} \left[\log \frac{\hat{f}_{1,i}}{\hat{f}_{2,i}} \log \frac{\hat{f}_{1,i}}{\hat{f}_{3,i}}\right] - \left[\frac{1}{n} \sum_{i=1}^{n} \log \frac{\hat{f}_{1,i}}{\hat{f}_{2,i}}\right] \left[\frac{1}{n} \sum_{i=1}^{n} \log \frac{\hat{f}_{1,i}}{\hat{f}_{3,i}}\right] \tag{16}$$

converges to

$$\hat{\sigma}_{12.13} \longrightarrow^{a.s.} Cov^0\left[\log \frac{f_1}{f_2}, \log \frac{f_1}{f_3}\right] = \sigma_{12.13} . \tag{17}$$

Thus the empirical covariance matrix converges, $\hat{\mathbf{\Sigma}} \longrightarrow^{a.s.} \mathbf{\Sigma}$, where

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{12.12} & \sigma_{12.13} & \sigma_{12.23} \\ \sigma_{12.13} & \sigma_{13.13} & \sigma_{13.23} \\ \sigma_{12.23} & \sigma_{13.23} & \sigma_{23.23} \end{pmatrix} . \tag{18}$$

It follows from the multivariate central limit and Slutsky's theorems that under the null hypothesis

$$H_0 : \quad E^0 \left[\log \frac{f_1}{f_2} \right] = 0 \text{ and } E^0 \left[\log \frac{f_1}{f_3} \right] = 0 \text{ and } E^0 \left[\log \frac{f_2}{f_3} \right] = 0$$

we have that

$$\mathbf{Z} = \text{diag}(\hat{\mathbf{\Sigma}})^{-\frac{1}{2}} \mathbf{L\tilde{R}}/\sqrt{n} \longrightarrow^{d} N_3(\mathbf{0}, \boldsymbol{\rho}) , \tag{19}$$

where $\mathbf{L\tilde{R}} = (L\tilde{R}_{12}, L\tilde{R}_{13}, L\tilde{R}_{23})^T$ and $\boldsymbol{\rho} = \text{diag}(\mathbf{\Sigma})^{-\frac{1}{2}} \mathbf{\Sigma}\, \text{diag}(\mathbf{\Sigma})^{-\frac{1}{2}}$ is the correlation matrix

$$\boldsymbol{\rho} = \begin{pmatrix} 1 & \rho_{12.13} & \rho_{12.23} \\ \rho_{12.13} & 1 & \rho_{13.23} \\ \rho_{12.23} & \rho_{13.23} & 1 \end{pmatrix} . \tag{20}$$

Note that the elements of the test statistic vector $\mathbf{Z}$, e.g., Z_{12}

$$Z_{12} = \frac{L\tilde{R}_{12}}{\sqrt{n}\,\hat{\sigma}_{12.12}} = \frac{-(BIC_1 - BIC_2)/2}{\sqrt{n}\,\hat{\sigma}_{12.12}}, \tag{21}$$

correspond to penalized log-likelihood ratios scaled by their standard errors, or equivalently, to a scaled contrast of BIC scores, if we adopt the BIC penalty. (Similarly, if we adopt the AIC penalty, Z_{12} corresponds to a scaled contrast of AIC scores.) Note that because the scaling factors in our test statistics are positive, performing model selection using the BIC criterium is equivalent to performing model selection using the CMST test statistics. Explicitly, the data supports model: M_1 if $BIC_1 < BIC_2$ and $BIC_1 < BIC_3$, or equivalently, if $z_{12} > 0$ and $z_{13} > 0$; M_2 if $BIC_1 > BIC_2$ and $BIC_2 < BIC_3$, or equivalently, if $z_{12} < 0$ and $z_{23} > 0$; M_3 if $BIC_1 > BIC_3$ and $BIC_2 > BIC_3$, or equivalently, if $z_{13} < 0$ and $z_{23} < 0$. The advantage of the CMST is that it goes one step further and provides a formal measure of significance for the penalized likelihood scores that accounts for possible mis-specifications of the models.

Suppose, for example, that the data supports model M_1 better, according to the BIC model selection criterium. The CMST allows us to test the null hypothesis, $H_0^{M_1}$, that model M_1 is not significantly closer to the true model, according to the Kullback-Liebler distance, than models M_2 and M_3. Note that given the best model, the relative closeness of the other models to the true model has no further impact on the model selection call, that is, given that the data supports model M_1 better, we are indifferent about the relative closeness of models M_2 and M_3 to the to the true model. Therefore, we derive the rejection region for the null hypothesis from the appropriate asymptotic bivariate marginal null distribution. For example, we select M_1 if z_{12} and z_{13} are both positive and we test if the call is statistically significant using the asymptotic marginal null distribution $(Z_{12}, Z_{13})^T \sim N_2(\mathbf{0}, \boldsymbol{\rho}_{12.13})$, where

$$\boldsymbol{\rho}_{12.13} = \begin{pmatrix} 1 & \rho_{12.13} \\ \rho_{12.13} & 1 \end{pmatrix} . \tag{22}$$

The marginal bivariate distributions of our test statistics have nice geometric properties. For instance, because the statistics have the same variance (equal to 1), the major axis of the confidence ellipse will be along the 45^o line through the mean for positively correlated random variables, and along the line at right angles to the 45^o line through the mean for negatively correlated random variables (see Figures 6a

and 6b). Furthermore, because the half-length of the major axis of a confidence ellipse with probability $1-\alpha$ is given by $(\chi^2_{2,\alpha}\,\lambda)^{1/2}$, where λ is the first eigenvalue of the correlation matrix and $\chi^2_{2,\alpha}$ is the upper αth quantile of a chi-square distribution with two degrees of freedom, we have that the critical value (the point where the first principal component axis first touches the confidence ellipse) is $(\chi^2_{2,\alpha}\,\lambda/2)^{1/2}$ (see Figure 6a).

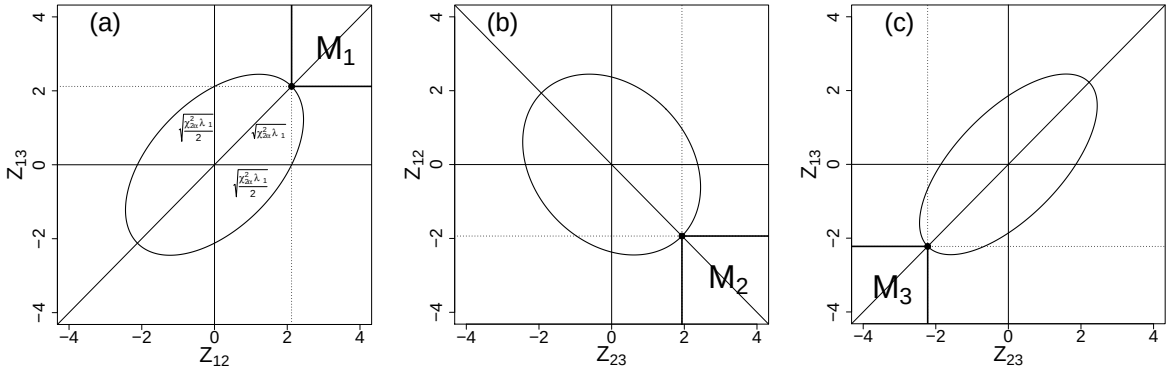

Figure 6. Rejection regions for $H_0^{\mathcal{M}_1}$, $H_0^{\mathcal{M}_2}$ and $H_0^{\mathcal{M}_3}$. Panel (a) shows the rejection region for $H_0^{\mathcal{M}_1}$ (the M_1 region). Note that because f_1 is located in the numerator of both Z_{12} and Z_{13} we have that $\rho_{12.13} > 0$, and the major axis of the confidence ellipse is along the 45^o line through $\mathbf{0}$. Panel (b) shows the rejection region for $H_0^{\mathcal{M}_2}$. Here $\rho_{12.23} < 0$ since f_2 is located in the denominator of Z_{12} and at the numerator of Z_{23}. The major axis then is then along the line at right angle with the 45^o line through the mean. Panel (c) shows the rejection region for $H_0^{\mathcal{M}_3}$. Because f_3 is located in the denominator of both Z_{13} and Z_{23} we have that $\rho_{13.23} > 0$. The figure depicts the 95% confidence regions associated with the bivariate marginal distributions of a tri-variate normal with mean $\mathbf{0}$ and correlation matrix with off-diagonal entries $\rho_{12.13} = 0.5$, $\rho_{12.23} = -0.25$ and $\rho_{13.23} = 0.65$.

From the relationship between the quadratic form of a p-dimensional multivariate normal, $\mathbf{X} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$, and the upper αth percentile of a chi-square distribution with p degrees of freedom, χ^2_p, we have that the equation $\mathbf{x}^t\mathbf{\Sigma}^{-1}\mathbf{x} = \chi^2_{2,\alpha}$ corresponds to an ellipsoide countour of a bivariate normal centered at zero with probability $1-\alpha$. Since $P(\chi^2_p \geq \chi^2_{p,\alpha}) = \alpha$ we have that $P(\chi^2_p \geq \mathbf{x}^t\mathbf{\Sigma}^{-1}\mathbf{x}) = \alpha$. In Figure 7a we illustrate how the p-value associated with the asymptotic marginal null distribution of (Z_{12}, Z_{13}) is computed. Setting $(Z_{12}, Z_{13}) = (z_{12}, z_{12})$ and to $(Z_{12}, Z_{13}) = (z_{13}, z_{13})$ the observed quadratic forms reduce, respectively, to $2z^2_{12}/\lambda_{12.13}$ and $2z^2_{13}/\lambda_{12.13}$ and the p-value testing $H_0^{\mathcal{M}_1}$ is given by the maximum of $P(\chi^2_2 \geq 2z^2_{12}/\lambda_{12.13})$ and $P(\chi^2_2 \geq 2z^2_{13}/\lambda_{12.13})$. Table 2 presents the null hypothesis, null distributions, rejection regions and p-values for the CMST tests associated with the causal, reactive and independence models.

Operationally, our procedure is composed of two steps: (1) check the sign of the test statistics to select one of the models (or equivalently, use the penalized likelihood criterion to select among the models); and (2) test the null hypothesis that the selected model is not closer to the true model than the other models.

Note that we should not pay attention to the CMST p-values of unselected models. Because our model selection tests are directional and our p-values correspond to the probability mass outside a confidence ellipse it is possible that the CMST p-value of a unselected model to be smaller than the CMST p-value of the selected model. For example, in Figure 7b we have that the p-value for $H_0^{\mathcal{M}_3}$ is smaller than the p-value for $H_0^{\mathcal{M}_1}$ (the CMST test associated with the actual selected model, M_1). However, the observed test statistics, $z_{13} = 2.5$ and $z_{23} = 3$, are positive and are in the wrong direction of the rejection region

H_0	Test statistic distr. under H_0	Rejection region	P-value
$H_0^{M_1}$	$(Z_{12}, Z_{13})^T \sim N_2(\mathbf{0}, \boldsymbol{\rho}_{12.13})$	$z_{12} > c^{\alpha}_{12.13}\,,\; z_{13} > c^{\alpha}_{12.13}$	$p^{MS}_{M_1} = p_{12.13}$
$H_0^{M_2}$	$(Z_{12}, Z_{23})^T \sim N_2(\mathbf{0}, \boldsymbol{\rho}_{12.23})$	$z_{12} < -c^{\alpha}_{12.23}\,,\; z_{23} > c^{\alpha}_{12.23}$	$p^{MS}_{M_2} = p_{12.23}$
$H_0^{M_3}$	$(Z_{13}, Z_{23})^T \sim N_2(\mathbf{0}, \boldsymbol{\rho}_{13.23})$	$z_{13} < -c^{\alpha}_{13.23}\,,\; z_{23} < -c^{\alpha}_{13.23}$	$p^{MS}_{M_3} = p_{13.23}$

Table 2. Model selection tests. Here $c^{\alpha}_{uv.st} = (\chi^2_{2,\alpha}\,\lambda_{uv.st}/2)^{1/2}$ corresponds to the critical value of the test of level α based on the joint distributions of (Z_{uv}, Z_{st}); $\lambda_{uv.st}$ corresponds to the first eigenvalues of the bivariate marginal correlation matrix $\boldsymbol{\rho}_{uv.st}$; $\chi^2_{2,\alpha}$ corresponds to the αth upper quantile of a chi-squared distribution with two degrees of freedom; and $p_{uv.st} = \max\left\{P(\chi^2_2 \geq 2z^2_{uv}/\lambda_{uv.st})\,,\, P(\chi^2_2 \geq 2z^2_{st}/\lambda_{uv.st})\right\}$.

of $H_0^{M_3}$, where both test statistics need to be negative.

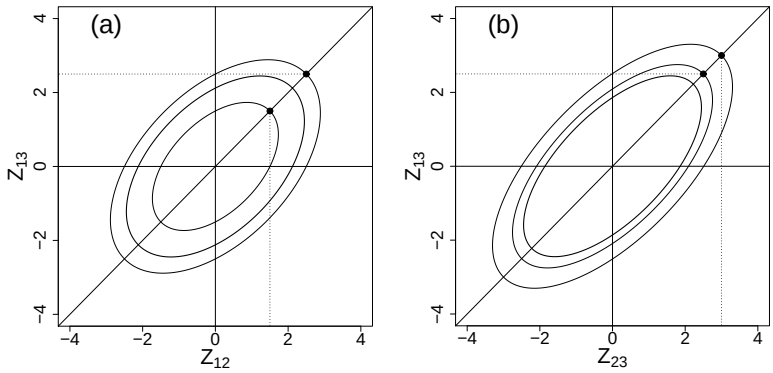

Figure 7. Panel (a) illustrates the computation of the p-value for $H_0^{M_1}$. The observed values of the test statistics, $z_{12} = 1.5$ and $z_{13} = 2.5$, are indicated by the dashed lines. Setting $(Z_{12}, Z_{13}) = (z_{12}, z_{12})$ the observed quadratic form reduces to $2z^2_{12}/\lambda_{12.13}$ and the p-value $P(\chi^2_2 \geq 2z^2_{12}/\lambda_{12.13}) = 0.223$ corresponds to the probability mass outside the inner ellipsoid contour. Setting $(Z_{12}, Z_{13}) = (z_{13}, z_{13})$ we obtain the p-value $P(\chi^2_2 \geq 2z^2_{13}/\lambda_{12.13}) = 0.016$ corresponding to the probability outside the outer ellipsoid. The p-value for $H_0^{M_1}$ is given by $\max\{\, P(\chi^2_2 \geq 2z^2_{12}/\lambda_{12.13})\,,\, P(\chi^2_2 \geq 2z^2_{13}/\lambda_{12.13})\,\}$ corresponding to the probability mass outside the inner ellipsoid contour. The probability mass outside the middle (and thicker) ellipsoid is 0.05. Panel (b) illustrates the computation of the p-value for $H_0^{M_3}$ in the invalid situation that M_3 is not the selected model. In this case, the p-value for the $H_0^{M_3}$ (probability mass outside the middle ellipsoid contour) is 0.023 (and is smaller than the p-value for $H_0^{M_1}$, 0.223). Note, however, the observed test statistics, $z_{13} = 2.5$ and $z_{23} = 3$, are positive and in the wrong direction of the rejection region of $H_0^{M_3}$, where both test statistics need to be negative.

Finally, in our applications it will sometimes be the case that two (or all three) models are likelihood equivalent (for example, when both phenotypes are affected by exactly the same set of QTLs). In these situations, the results presented in this section are no longer valid since likelihood equivalent models are a simple re-parametrization of each other, i.e., $f_1(.\,;\boldsymbol{\theta}_1) = f_2(.\,;\boldsymbol{\theta}_2)$ for all $\boldsymbol{\theta}_1 = h(\boldsymbol{\theta}_2)$, and thus are observationally identical (recall that observationally identical models satisfy $f_1(.\,;\boldsymbol{\theta}_{1*}) = f_2(.\,;\boldsymbol{\theta}_{2*})$ where $\boldsymbol{\theta}_{1*}$ is the parameter values that minimize the Kullback-Liebler distance between f_1 and the density

of the true model). Nonetheless, because the likelihood ratio statistic of likelihood equivalent models is always 1, and likelihood equivalent models always have the same dimension, we have that our test statistic have a degenerated point mass distribution concentrated at 0, and the p-value is trivially 1. Therefore, whenever two (or all three) models are tied in terms of a penalized likelihood model selection criterium, our approach cannot make a significant call in favor of one of the models (in accordance with our intuition that if two models are the same, then none of then can be closer to the truth).

Simulation studies

Simulation study 1: assessing empirical error rates using toy networks

In this section we evaluate and compare the performance of the CMST to the BIC and CIT approaches in terms of empirical error rates. We considered the 7 models listed on Figure 8 in this simulation study.

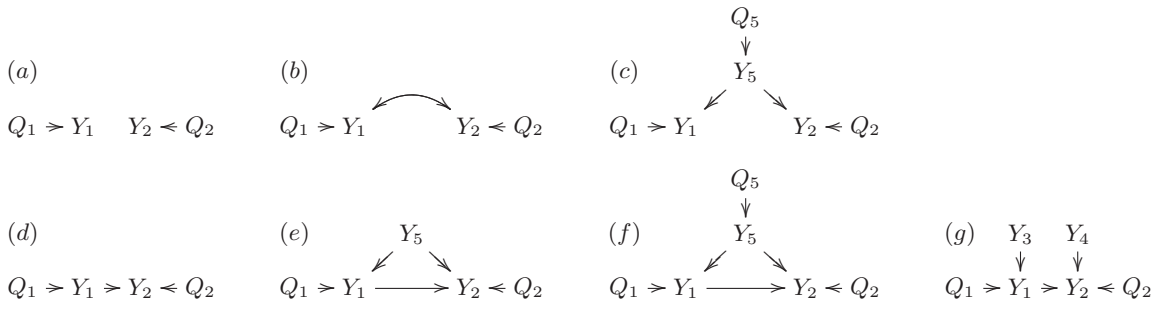

Figure 8. Toy network models used in the simulation study. We generate data from models (a)-(g) but fit models M_1, M_2 and M_3 from Figure 1. Model (a) represents the null model with uncorrelated phenotypes. Models (b) and (d) corresponds to the fitted models M_3 and M_1, respectively. These two models allows us to investigate the case where one of the fitted models is correctly specified. Models (c), (e), (f) and (g) do not correspond to any of our fitted models and allows us to investigate the effects of model misspecification.

For each model we investigated 8 different simulation settings comprising all combinations between $R^2 = \{0.125, 0.25, 0.5, \text{range}\}$ and $n = \{135, 500\}$, where R^2 represents the amount of the variability of Y_1 explained by the QTLs. For a fixed R^2 value, and residual variances and phenotype effects set to 1, we computed the corresponding additive genetic effect, and then used these parameter values to generate the simulated data sets. For the "R^2 = range" case, each one of the data sets were generated using a different combination of parameter values, with residual variances and additive genetic effects ranging from 0.2 to 2, and phenotype effects ranging from 0.5 to 1.5. For each one of the 56 different simulations we generated 10,000 F_2 crosses with 505 markers unequally spaced across 5 chromosomes of length 100cM. We adopted a QTL mapping threshold of 4.

We investigated the performance of the CMST and CIT for 10 different significance levels varying from 0.1 to 0.01, and the BIC performance using Bayes factors thresholds of 1, 3 and 10. In order to directly compare our results to the CIT and BIC approaches we kept track of empirical power and error rates. When the true model shows a M_k causal relationship, the power is given by the number of significant M_k calls, m_k^*, divided by the total number of simulations, t. The error rates are given by m_j^*/t, for $j \neq k$.

The conditional nature of our tests implies that we select a model and then test whether it is closer to the true model than other models. Therefore, it seems more appropriate to assess the CMST performance

conditional on the selected model. In addition to the empirical power and error rates described above, we also kept track of conditional power and conditional error rates for the CMST. When the true model has a M_k causal relationship, the conditional power is given by m_k^*/m_k, where m_k corresponds to number of times that model M_k was selected. The conditional error rates are given by m_j^*/m_j, for $j \neq k$.

For the BIC approach, m_k^* represents the number of interesting calls. A call is considered interesting when the difference between the BIC scores of the best and second best models corresponds to an approximate Bayes factor (Kass and Raftery 1995) greater than a fixed threshold (chosen as 1, 3 or 10 in this study).

We also kept track of the false discovery rate (FDR) for all three approaches. The FDR is given by the number of mistakes divided by the number of significant (interesting) calls. The number of mistakes corresponds to the number of times we incorrectly detect models M_2 and M_3 when the true model is M_1, and models M_1 and M_2 when the true relation is M_3.

Tables 3-30 present the results for all 56 simulation studies. Overall, the simulation results show that although the CMST is a low powered test, the conditional error rates (and consequently the unconditional ones) were almost always well controlled, i.e., were below the nominal significance levels. Except for the simulation with $R^2 = 0.5$ for model (e), where the conditional error rates for model M_3 were slightly above the nominal levels (see Tables 19 and 21), the conditional error rates (the cER columns) were always below the nominal significance levels (the level columns in tables 3-30). The unconditional error rates were below the nominal significance levels in all 56 simulation settings.

The CIT approach on the other hand, showed higher power in almost all simulation settings, but also higher error rates. The error rates were specially high for models (e) and (f) (see Tables 19, 21, 23 and 25 in the Supplement). Furthermore, the CIT approach tends to favor the independence model as we adopt stricter significance levels. This phenomenon is a consequence of the decision rule behind the CIT approach, where we: (1) make a call in favor of model M_1 if the p-value derived from a intersection-union test for a causal relationship (M_1) is lower than the nominal significance level α, and the p-value for the reactive relationship (M_2) is higher than α; (2) make a call in favor of model M_2 if it is the other way around; (3) make a call in favor of model M_3 if the causal and reactive p-values are greater than α (hence as α decreases it gets more likely to make a call in favor of model M_3); and (4) forfeit to make a call if the causal and reactive p-values are lower than α. Inspection of the fifth columns in Tables **??** and **??** illustrates this phenomenon. Nonetheless, when the true model corresponds to one of our fitted models, such as model (d), which corresponds to our model M_1, the CIT approach has much higher power then the CMST and error rates that, although higher than the CMST, are still relatively low (see Tables 15 and 17 in the Supplement). On a technical note, the CIT requires that both phenotypes map to the same QTL, but often times the phenotypes map to nearby but not precisely the same QTL. Instead of having to decide which QTL to use as a causal anchor, we use the joint LOD profile to determine the QTL to be used as a causal anchor, as described in the Supplement.

The BIC approach showed higher power and higher error rates than the CMST as well (see the even numbered tables among Tables 4-30). The error rates were specially high for models (f) (Tables 24 and 26) and (c) to a lesser degree (Table 12).

		CMST						
		ER	ER	ER		cER	cER	cER
R^2	level	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/m_1	m_2^*/m_2	m_3^*/m_3
	0.10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.125	0.09	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.08	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.06	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.05	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.04	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.03	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.02	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.01	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.10	0.0000	0.0001	0.0001	1.0000	0.0000	0.0003	0.0004
0.25	0.09	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.08	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.06	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.05	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.04	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.03	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.02	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.01	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.10	0.0002	0.0001	0.0000	1.0000	0.0006	0.0003	0.0000
0.5	0.09	0.0000	0.0001	0.0000	1.0000	0.0000	0.0003	0.0000
	0.08	0.0000	0.0001	0.0000	1.0000	0.0000	0.0003	0.0000
	0.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.06	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.05	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.04	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.03	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.02	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.01	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
range	0.09	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.08	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.06	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.05	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.04	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.03	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.02	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.01	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 3. Simulation results for the CMST with data generated from model (a) with n=135. Model (a) corresponds to the null model, where the phenotypes are uncorrelated. As expected, the CMST made very few significant calls and has very low error rates (ER) and conditional error rates (cER). Note, however, it does not correspond to any of the type of relations that we seek to infer with our fitted models, not even to M_3, since M_3 assumes that the phenotypes are correlated. Hence whenever the false discovery proportion (FDR) is either 0 or 1.

	BF $R^2 = 0.125$			BF $R^2 = 0.25$			BF $R^2 = 0.5$			BF $R^2 = \text{range}$		
	1	3	10	1	3	10	1	3	10	1	3	10
m_1^*/t	0.4197	0.0324	0.0196	0.3586	0.0251	0.0044	0.3457	0.0459	0.0068	0.3679	0.0425	0.0092
m_2^*/t	0.4130	0.0327	0.0196	0.3617	0.0239	0.0027	0.3475	0.0429	0.0062	0.3548	0.0386	0.0083
m_3^*/t	0.1673	0.0030	0.0001	0.2796	0.0100	0.0009	0.3069	0.0290	0.0037	0.2773	0.0205	0.0028
FDR	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 4. Simulation results for the BIC approach with data generated from model (a) with n=135. The BIC makes many interesting calls using a Bayes factor (BF) of 10.

		CMST						
		ER	ER	ER		cER	cER	cER
R^2	level	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/m_1	m_2^*/m_2	m_3^*/m_3
	0.10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.125	0.09	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.08	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.06	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.05	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.04	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.03	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.02	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.01	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.25	0.09	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.08	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.06	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.05	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.04	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.03	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.02	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.01	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.10	0.0001	0.0001	0.0000	1.0000	0.0003	0.0003	0.0000
0.5	0.09	0.0001	0.0001	0.0000	1.0000	0.0003	0.0003	0.0000
	0.08	0.0001	0.0001	0.0000	1.0000	0.0003	0.0003	0.0000
	0.07	0.0001	0.0001	0.0000	1.0000	0.0003	0.0003	0.0000
	0.06	0.0001	0.0001	0.0000	1.0000	0.0003	0.0003	0.0000
	0.05	0.0001	0.0000	0.0000	1.0000	0.0003	0.0000	0.0000
	0.04	0.0001	0.0000	0.0000	1.0000	0.0003	0.0000	0.0000
	0.03	0.0001	0.0000	0.0000	1.0000	0.0003	0.0000	0.0000
	0.02	0.0001	0.0000	0.0000	1.0000	0.0003	0.0000	0.0000
	0.01	0.0001	0.0000	0.0000	1.0000	0.0003	0.0000	0.0000
	0.10	0.0002	0.0001	0.0001	1.0000	0.0006	0.0003	0.0004
range	0.09	0.0002	0.0001	0.0001	1.0000	0.0006	0.0003	0.0004
	0.08	0.0001	0.0001	0.0001	1.0000	0.0003	0.0003	0.0004
	0.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.06	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.05	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.04	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.03	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.02	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.01	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 5. Model (a), n=500

	BF $R^2 = 0.125$			BF $R^2 = 0.25$			BF $R^2 = 0.5$			BF $R^2 =$ range		
	1	3	10	1	3	10	1	3	10	1	3	10
m_1^*/t	0.3684	0.0084	0.0006	0.3619	0.0197	0.0015	0.3500	0.0447	0.0060	0.3545	0.0371	0.0073
m_2^*/t	0.3580	0.0082	0.0009	0.3492	0.0195	0.0017	0.3446	0.0406	0.0060	0.3676	0.0351	0.0054
m_3^*/t	0.2736	0.0012	0.0000	0.2889	0.0072	0.0001	0.3054	0.0257	0.0023	0.2779	0.0169	0.0018
FDR	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 6. Model (a), n=500

		CMST						
		ER	ER	PW		cER	cER	cPW
R^2	level	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/m_1	m_2^*/m_2	m_3^*/m_3
	0.10	0.0021	0.0015	0.0068	0.3462	0.0064	0.0043	0.0310
0.125	0.09	0.0015	0.0012	0.0056	0.3253	0.0046	0.0035	0.0255
	0.08	0.0012	0.0010	0.0044	0.3333	0.0037	0.0029	0.0200
	0.07	0.0010	0.0009	0.0036	0.3455	0.0030	0.0026	0.0164
	0.06	0.0009	0.0006	0.0029	0.3409	0.0027	0.0017	0.0132
	0.05	0.0006	0.0004	0.0026	0.2778	0.0018	0.0012	0.0118
	0.04	0.0004	0.0004	0.0021	0.2759	0.0012	0.0012	0.0096
	0.03	0.0002	0.0003	0.0015	0.2500	0.0006	0.0009	0.0068
	0.02	0.0001	0.0001	0.0007	0.2222	0.0003	0.0003	0.0032
	0.01	0.0000	0.0001	0.0003	0.2500	0.0000	0.0003	0.0014
	0.10	0.0001	0.0000	0.1001	0.0010	0.0009	0.0000	0.1276
0.25	0.09	0.0001	0.0000	0.0927	0.0011	0.0009	0.0000	0.1182
	0.08	0.0001	0.0000	0.0849	0.0012	0.0009	0.0000	0.1082
	0.07	0.0001	0.0000	0.0746	0.0013	0.0009	0.0000	0.0951
	0.06	0.0001	0.0000	0.0650	0.0015	0.0009	0.0000	0.0829
	0.05	0.0000	0.0000	0.0554	0.0000	0.0000	0.0000	0.0706
	0.04	0.0000	0.0000	0.0449	0.0000	0.0000	0.0000	0.0573
	0.03	0.0000	0.0000	0.0347	0.0000	0.0000	0.0000	0.0442
	0.02	0.0000	0.0000	0.0228	0.0000	0.0000	0.0000	0.0290
	0.01	0.0000	0.0000	0.0102	0.0000	0.0000	0.0000	0.0130
	0.10	0.0000	0.0001	0.4008	0.0003	0.0000	0.0039	0.4235
0.5	0.09	0.0000	0.0001	0.3826	0.0003	0.0000	0.0039	0.4043
	0.08	0.0000	0.0001	0.3632	0.0003	0.0000	0.0039	0.3838
	0.07	0.0000	0.0001	0.3430	0.0003	0.0000	0.0039	0.3625
	0.06	0.0000	0.0000	0.3156	0.0000	0.0000	0.0000	0.3334
	0.05	0.0000	0.0000	0.2847	0.0000	0.0000	0.0000	0.3008
	0.04	0.0000	0.0000	0.2522	0.0000	0.0000	0.0000	0.2665
	0.03	0.0000	0.0000	0.2141	0.0000	0.0000	0.0000	0.2262
	0.02	0.0000	0.0000	0.1713	0.0000	0.0000	0.0000	0.1810
	0.01	0.0000	0.0000	0.1056	0.0000	0.0000	0.0000	0.1116
	0.10	0.0009	0.0003	0.0923	0.0129	0.0057	0.0019	0.1763
range	0.09	0.0008	0.0002	0.0875	0.0113	0.0050	0.0013	0.1671
	0.08	0.0006	0.0001	0.0826	0.0084	0.0038	0.0006	0.1579
	0.07	0.0003	0.0000	0.0774	0.0039	0.0019	0.0000	0.1480
	0.06	0.0002	0.0000	0.0713	0.0028	0.0013	0.0000	0.1363
	0.05	0.0002	0.0000	0.0648	0.0031	0.0013	0.0000	0.1238
	0.04	0.0001	0.0000	0.0584	0.0017	0.0006	0.0000	0.1116
	0.03	0.0001	0.0000	0.0487	0.0021	0.0006	0.0000	0.0932
	0.02	0.0000	0.0000	0.0399	0.0000	0.0000	0.0000	0.0763
	0.01	0.0000	0.0000	0.0266	0.0000	0.0000	0.0000	0.0508

Table 7. Model (b), n=135

	BF $R^2 = 0.125$			BF $R^2 = 0.25$			BF $R^2 = 0.5$			BF $R^2 =$ range		
	1	3	10	1	3	10	1	3	10	1	3	10
m_1^*/t	0.3674	0.2389	0.1688	0.1083	0.0618	0.0306	0.0279	0.0140	0.0061	0.1891	0.0898	0.0518
m_2^*/t	0.3869	0.2521	0.1778	0.1064	0.0575	0.0300	0.0258	0.0130	0.0051	0.1897	0.0879	0.0498
m_3^*/t	0.2457	0.1589	0.0784	0.7852	0.6470	0.4741	0.9464	0.9000	0.8250	0.6212	0.4586	0.3275
FDR	0.7543	0.7555	0.8155	0.2148	0.1557	0.1134	0.0536	0.0292	0.0134	0.3788	0.2792	0.2367

Table 8. Model (b), n=135

		CMST						
		ER	ER	PW		cER	cER	cPW
R^2	level	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/m_1	m_2^*/m_2	m_3^*/m_3
	0.10	0.0001	0.0000	0.2293	0.0004	0.0019	0.0000	0.2555
0.125	0.09	0.0001	0.0000	0.2165	0.0005	0.0019	0.0000	0.2413
	0.08	0.0001	0.0000	0.2004	0.0005	0.0019	0.0000	0.2233
	0.07	0.0001	0.0000	0.1865	0.0005	0.0019	0.0000	0.2078
	0.06	0.0001	0.0000	0.1696	0.0006	0.0019	0.0000	0.1890
	0.05	0.0001	0.0000	0.1517	0.0007	0.0019	0.0000	0.1691
	0.04	0.0000	0.0000	0.1363	0.0000	0.0000	0.0000	0.1519
	0.03	0.0000	0.0000	0.1177	0.0000	0.0000	0.0000	0.1312
	0.02	0.0000	0.0000	0.0961	0.0000	0.0000	0.0000	0.1071
	0.01	0.0000	0.0000	0.0677	0.0000	0.0000	0.0000	0.0754
	0.10	0.0000	0.0000	0.4442	0.0000	0.0000	0.0000	0.4656
0.25	0.09	0.0000	0.0000	0.4280	0.0000	0.0000	0.0000	0.4486
	0.08	0.0000	0.0000	0.4121	0.0000	0.0000	0.0000	0.4320
	0.07	0.0000	0.0000	0.3931	0.0000	0.0000	0.0000	0.4120
	0.06	0.0000	0.0000	0.3734	0.0000	0.0000	0.0000	0.3914
	0.05	0.0000	0.0000	0.3529	0.0000	0.0000	0.0000	0.3699
	0.04	0.0000	0.0000	0.3316	0.0000	0.0000	0.0000	0.3476
	0.03	0.0000	0.0000	0.3038	0.0000	0.0000	0.0000	0.3184
	0.02	0.0000	0.0000	0.2710	0.0000	0.0000	0.0000	0.2840
	0.01	0.0000	0.0000	0.2239	0.0000	0.0000	0.0000	0.2347
	0.10	0.0000	0.0000	0.6310	0.0000	0.0000	0.0000	0.6454
0.5	0.09	0.0000	0.0000	0.6186	0.0000	0.0000	0.0000	0.6327
	0.08	0.0000	0.0000	0.6035	0.0000	0.0000	0.0000	0.6172
	0.07	0.0000	0.0000	0.5862	0.0000	0.0000	0.0000	0.5995
	0.06	0.0000	0.0000	0.5708	0.0000	0.0000	0.0000	0.5838
	0.05	0.0000	0.0000	0.5495	0.0000	0.0000	0.0000	0.5620
	0.04	0.0000	0.0000	0.5264	0.0000	0.0000	0.0000	0.5384
	0.03	0.0000	0.0000	0.4940	0.0000	0.0000	0.0000	0.5053
	0.02	0.0000	0.0000	0.4536	0.0000	0.0000	0.0000	0.4640
	0.01	0.0000	0.0000	0.3925	0.0000	0.0000	0.0000	0.4015
	0.10	0.0000	0.0001	0.2808	0.0004	0.0000	0.0011	0.3589
range	0.09	0.0000	0.0001	0.2697	0.0004	0.0000	0.0011	0.3447
	0.08	0.0000	0.0000	0.2581	0.0000	0.0000	0.0000	0.3299
	0.07	0.0000	0.0000	0.2453	0.0000	0.0000	0.0000	0.3135
	0.06	0.0000	0.0000	0.2355	0.0000	0.0000	0.0000	0.3010
	0.05	0.0000	0.0000	0.2197	0.0000	0.0000	0.0000	0.2808
	0.04	0.0000	0.0000	0.2038	0.0000	0.0000	0.0000	0.2605
	0.03	0.0000	0.0000	0.1848	0.0000	0.0000	0.0000	0.2362
	0.02	0.0000	0.0000	0.1617	0.0000	0.0000	0.0000	0.2067
	0.01	0.0000	0.0000	0.1279	0.0000	0.0000	0.0000	0.1635

Table 9. Model (b), n=500

	BF $R^2 = 0.125$			BF $R^2 = 0.25$			BF $R^2 = 0.5$			BF $R^2 =$ range		
	1	3	10	1	3	10	1	3	10	1	3	10
m_1^*/t	0.0515	0.0255	0.0107	0.0222	0.0103	0.0039	0.0111	0.0051	0.0022	0.0946	0.0418	0.0177
m_2^*/t	0.0512	0.0273	0.0127	0.0238	0.0110	0.0042	0.0112	0.0047	0.0023	0.0928	0.0403	0.0197
m_3^*/t	0.8973	0.8189	0.6795	0.9540	0.9044	0.8001	0.9777	0.9499	0.8905	0.8126	0.7090	0.5903
FDR	0.1027	0.0606	0.0333	0.0460	0.0230	0.0100	0.0223	0.0102	0.0050	0.1874	0.1037	0.0596

Table 10. Model (b), n=500

		CIT				CMST						
		ER	ER	PW		ER	ER	PW		cER	cER	cPW
R^2	level	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/m_1	m_2^*/m_2	m_3^*/m_3
	0.10	0.0820	0.0890	0.0094	0.9481	0.0085	0.0090	0.0000	1.0000	0.0247	0.0257	0.0000
0.125	0.09	0.1054	0.0937	0.0094	0.9551	0.0069	0.0078	0.0000	1.0000	0.0200	0.0223	0.0000
	0.08	0.1265	0.1030	0.0094	0.9608	0.0054	0.0070	0.0000	1.0000	0.0157	0.0200	0.0000
	0.07	0.1358	0.1124	0.0141	0.9464	0.0039	0.0046	0.0000	1.0000	0.0113	0.0132	0.0000
	0.06	0.1710	0.1522	0.0141	0.9583	0.0024	0.0033	0.0000	1.0000	0.0070	0.0094	0.0000
	0.05	0.1944	0.1710	0.0258	0.9341	0.0016	0.0027	0.0000	1.0000	0.0046	0.0077	0.0000
	0.04	0.2225	0.2131	0.0328	0.9300	0.0011	0.0014	0.0000	1.0000	0.0032	0.0040	0.0000
	0.03	0.2576	0.2553	0.0703	0.8795	0.0002	0.0007	0.0000	1.0000	0.0006	0.0020	0.0000
	0.02	0.3162	0.2881	0.1218	0.8323	0.0001	0.0003	0.0000	1.0000	0.0003	0.0009	0.0000
	0.01	0.3185	0.2857	0.3091	0.6615	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.10	0.0670	0.0725	0.0022	0.9842	0.0161	0.0165	0.0031	0.9127	0.0404	0.0408	0.0166
0.25	0.09	0.0827	0.0818	0.0026	0.9846	0.0134	0.0154	0.0027	0.9137	0.0336	0.0381	0.0145
	0.08	0.0940	0.0914	0.0026	0.9863	0.0110	0.0136	0.0025	0.9074	0.0278	0.0336	0.0134
	0.07	0.1081	0.1097	0.0032	0.9855	0.0085	0.0112	0.0020	0.9078	0.0215	0.0279	0.0107
	0.06	0.1276	0.1292	0.0038	0.9852	0.0070	0.0094	0.0017	0.9061	0.0177	0.0234	0.0091
	0.05	0.1597	0.1543	0.0055	0.9829	0.0055	0.0078	0.0012	0.9172	0.0139	0.0194	0.0064
	0.04	0.1934	0.1879	0.0115	0.9706	0.0036	0.0056	0.0011	0.8932	0.0091	0.0139	0.0059
	0.03	0.2332	0.2328	0.0231	0.9528	0.0021	0.0037	0.0008	0.8788	0.0053	0.0092	0.0043
	0.02	0.2861	0.2838	0.0523	0.9160	0.0013	0.0018	0.0003	0.9118	0.0033	0.0045	0.0016
	0.01	0.3307	0.3239	0.1511	0.8125	0.0008	0.0007	0.0001	0.9375	0.0020	0.0017	0.0005
	0.10	0.0210	0.0224	0.0131	0.7677	0.0034	0.0035	0.0871	0.0739	0.0195	0.0196	0.1351
0.5	0.09	0.0254	0.0264	0.0159	0.7648	0.0033	0.0033	0.0814	0.0755	0.0189	0.0185	0.1262
	0.08	0.0309	0.0303	0.0192	0.7610	0.0032	0.0030	0.0735	0.0783	0.0183	0.0168	0.1140
	0.07	0.0358	0.0337	0.0234	0.7484	0.0026	0.0029	0.0670	0.0764	0.0149	0.0162	0.1039
	0.06	0.0432	0.0447	0.0277	0.7600	0.0020	0.0022	0.0594	0.0665	0.0115	0.0123	0.0922
	0.05	0.0505	0.0541	0.0356	0.7459	0.0017	0.0015	0.0529	0.0575	0.0097	0.0084	0.0820
	0.04	0.0664	0.0679	0.0453	0.7480	0.0010	0.0009	0.0439	0.0418	0.0057	0.0050	0.0681
	0.03	0.0837	0.0874	0.0635	0.7293	0.0008	0.0003	0.0370	0.0291	0.0046	0.0017	0.0573
	0.02	0.1189	0.1201	0.0925	0.7210	0.0008	0.0003	0.0269	0.0396	0.0046	0.0017	0.0417
	0.01	0.1804	0.1758	0.1688	0.6786	0.0003	0.0001	0.0146	0.0268	0.0017	0.0006	0.0227
	0.10	0.0858	0.0749	0.0929	0.6337	0.0178	0.0177	0.0281	0.5577	0.0618	0.0593	0.1015
range	0.09	0.0910	0.0839	0.1007	0.6346	0.0155	0.0159	0.0261	0.5455	0.0538	0.0533	0.0943
	0.08	0.1007	0.0891	0.1075	0.6383	0.0138	0.0141	0.0235	0.5430	0.0482	0.0472	0.0848
	0.07	0.1094	0.0948	0.1174	0.6348	0.0123	0.0124	0.0216	0.5346	0.0429	0.0418	0.0780
	0.06	0.1205	0.1087	0.1283	0.6411	0.0109	0.0107	0.0196	0.5255	0.0381	0.0361	0.0707
	0.05	0.1368	0.1198	0.1411	0.6453	0.0096	0.0090	0.0178	0.5124	0.0335	0.0303	0.0642
	0.04	0.1519	0.1430	0.1550	0.6555	0.0079	0.0076	0.0153	0.5049	0.0276	0.0256	0.0551
	0.03	0.1701	0.1656	0.1812	0.6495	0.0060	0.0057	0.0135	0.4643	0.0209	0.0192	0.0489
	0.02	0.1959	0.1905	0.2186	0.6387	0.0047	0.0042	0.0107	0.4541	0.0164	0.0142	0.0388
	0.01	0.2164	0.2190	0.2970	0.5945	0.0027	0.0024	0.0074	0.4080	0.0094	0.0081	0.0268

Table 11. Model (c), n=135

	BF $R^2 = 0.125$			BF $R^2 = 0.25$			BF $R^2 = 0.5$			BF R^2 = range		
	1	3	10	1	3	10	1	3	10	1	3	10
m_1^*/t	0.4814	0.3695	0.3027	0.4025	0.3253	0.2548	0.1756	0.1213	0.0811	0.3334	0.2578	0.1989
m_2^*/t	0.4884	0.3752	0.3040	0.4079	0.3284	0.2589	0.1799	0.1249	0.0799	0.3454	0.2601	0.1983
m_3^*/t	0.0302	0.0115	0.0036	0.1897	0.1211	0.0671	0.6445	0.5320	0.3973	0.3211	0.2341	0.1511
FDR	0.9698	0.9849	0.9940	0.8103	0.8437	0.8844	0.3555	0.3164	0.2884	0.6789	0.6887	0.7244

Table 12. Model (c), n=135

		CIT				CMST						
		ER	ER	PW		ER	ER	PW		cER	cER	cPW
R^2	level	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/m_1	m_2^*/m_2	m_3^*/m_3
	0.10	0.0460	0.0430	0.0007	0.9925	0.0006	0.0007	0.0064	0.1688	0.0025	0.0029	0.0123
0.125	0.09	0.0533	0.0494	0.0007	0.9935	0.0004	0.0007	0.0056	0.1642	0.0017	0.0029	0.0108
	0.08	0.0593	0.0560	0.0007	0.9942	0.0004	0.0006	0.0043	0.1887	0.0017	0.0025	0.0083
	0.07	0.0664	0.0643	0.0007	0.9949	0.0004	0.0006	0.0032	0.2381	0.0017	0.0025	0.0062
	0.06	0.0805	0.0756	0.0007	0.9957	0.0002	0.0005	0.0026	0.2121	0.0008	0.0020	0.0050
	0.05	0.0966	0.0877	0.0007	0.9963	0.0001	0.0004	0.0018	0.2174	0.0004	0.0016	0.0035
	0.04	0.1145	0.1072	0.0009	0.9957	0.0001	0.0002	0.0013	0.1875	0.0004	0.0008	0.0025
	0.03	0.1419	0.1334	0.0016	0.9941	0.0000	0.0001	0.0009	0.1000	0.0000	0.0004	0.0017
	0.02	0.1817	0.1714	0.0041	0.9886	0.0000	0.0001	0.0005	0.1667	0.0000	0.0004	0.0010
	0.01	0.2483	0.2357	0.0189	0.9623	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000	0.0004
	0.10	0.0013	0.0015	0.0005	0.8387	0.0000	0.0000	0.1325	0.0000	0.0000	0.0000	0.1628
0.25	0.09	0.0017	0.0017	0.0005	0.8649	0.0000	0.0000	0.1232	0.0000	0.0000	0.0000	0.1514
	0.08	0.0020	0.0022	0.0005	0.8889	0.0000	0.0000	0.1155	0.0000	0.0000	0.0000	0.1419
	0.07	0.0030	0.0029	0.0005	0.9167	0.0000	0.0000	0.1062	0.0000	0.0000	0.0000	0.1305
	0.06	0.0037	0.0034	0.0005	0.9306	0.0000	0.0000	0.0971	0.0000	0.0000	0.0000	0.1193
	0.05	0.0044	0.0046	0.0005	0.9438	0.0000	0.0000	0.0875	0.0000	0.0000	0.0000	0.1075
	0.04	0.0058	0.0053	0.0005	0.9541	0.0000	0.0000	0.0766	0.0000	0.0000	0.0000	0.0941
	0.03	0.0079	0.0065	0.0005	0.9643	0.0000	0.0000	0.0653	0.0000	0.0000	0.0000	0.0802
	0.02	0.0108	0.0093	0.0006	0.9691	0.0000	0.0000	0.0526	0.0000	0.0000	0.0000	0.0646
	0.01	0.0181	0.0178	0.0006	0.9825	0.0000	0.0000	0.0374	0.0000	0.0000	0.0000	0.0460
	0.10	0.0000	0.0000	0.0005	0.0000	0.0004	0.0010	0.3073	0.0045	0.0046	0.0115	0.3723
0.5	0.09	0.0000	0.0000	0.0005	0.0000	0.0004	0.0009	0.2967	0.0044	0.0046	0.0103	0.3595
	0.08	0.0000	0.0000	0.0005	0.0000	0.0003	0.0009	0.2852	0.0042	0.0034	0.0103	0.3455
	0.07	0.0000	0.0000	0.0005	0.0000	0.0003	0.0009	0.2738	0.0044	0.0034	0.0103	0.3317
	0.06	0.0000	0.0000	0.0005	0.0000	0.0003	0.0005	0.2622	0.0030	0.0034	0.0057	0.3177
	0.05	0.0000	0.0000	0.0005	0.0000	0.0002	0.0005	0.2451	0.0028	0.0023	0.0057	0.2969
	0.04	0.0000	0.0000	0.0005	0.0000	0.0002	0.0004	0.2327	0.0026	0.0023	0.0046	0.2819
	0.03	0.0000	0.0000	0.0005	0.0000	0.0002	0.0003	0.2146	0.0023	0.0023	0.0034	0.2600
	0.02	0.0000	0.0000	0.0005	0.0000	0.0002	0.0002	0.1913	0.0021	0.0023	0.0023	0.2318
	0.01	0.0000	0.0000	0.0005	0.0000	0.0001	0.0000	0.1602	0.0006	0.0011	0.0000	0.1941
	0.10	0.0311	0.0296	0.0065	0.9030	0.0029	0.0043	0.1597	0.0431	0.0155	0.0231	0.2655
range	0.09	0.0335	0.0309	0.0074	0.8969	0.0025	0.0035	0.1537	0.0376	0.0134	0.0188	0.2556
	0.08	0.0365	0.0331	0.0080	0.8966	0.0021	0.0027	0.1454	0.0320	0.0112	0.0145	0.2418
	0.07	0.0399	0.0359	0.0092	0.8922	0.0019	0.0026	0.1395	0.0312	0.0102	0.0139	0.2320
	0.06	0.0442	0.0380	0.0107	0.8851	0.0017	0.0019	0.1326	0.0264	0.0091	0.0102	0.2205
	0.05	0.0487	0.0423	0.0123	0.8809	0.0013	0.0016	0.1241	0.0228	0.0070	0.0086	0.2064
	0.04	0.0542	0.0461	0.0146	0.8732	0.0009	0.0015	0.1150	0.0204	0.0048	0.0080	0.1912
	0.03	0.0614	0.0523	0.0176	0.8662	0.0004	0.0013	0.1037	0.0161	0.0021	0.0070	0.1724
	0.02	0.0703	0.0614	0.0241	0.8453	0.0003	0.0009	0.0913	0.0130	0.0016	0.0048	0.1518
	0.01	0.0875	0.0801	0.0345	0.8292	0.0002	0.0003	0.0717	0.0069	0.0011	0.0016	0.1192

Table 13. Model (c), n=500

	BF $R^2 = 0.125$			BF $R^2 = 0.25$			BF $R^2 = 0.5$			BF $R^2 =$ range		
	1	3	10	1	3	10	1	3	10	1	3	10
m_1^*/t	0.2365	0.1590	0.1006	0.0943	0.0558	0.0305	0.0874	0.0527	0.0293	0.1917	0.1360	0.0917
m_2^*/t	0.2441	0.1629	0.1028	0.0918	0.0491	0.0248	0.0871	0.0532	0.0312	0.1913	0.1337	0.0899
m_3^*/t	0.5193	0.3876	0.2450	0.8139	0.7062	0.5478	0.8255	0.7498	0.6116	0.6171	0.5189	0.3970
FDR	0.4807	0.4538	0.4536	0.1861	0.1293	0.0917	0.1745	0.1238	0.0900	0.3829	0.3420	0.3139

Table 14. Model (c), n=500

		CIT				CMST						
		PW	ER	ER		PW	ER	ER		cPW	cER	cER
R^2	level	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/m_1	m_2^*/m_2	m_3^*/m_3
	0.10	0.6982	0.0761	0.0380	0.1405	0.1007	0.0001	0.0000	0.0010	0.1269	0.0012	0.0000
0.125	0.09	0.7041	0.0752	0.0473	0.1483	0.0892	0.0001	0.0000	0.0011	0.1124	0.0012	0.0000
	0.08	0.7101	0.0744	0.0583	0.1575	0.0764	0.0001	0.0000	0.0013	0.0964	0.0012	0.0000
	0.07	0.7244	0.0727	0.0676	0.1623	0.0634	0.0000	0.0000	0.0000	0.0799	0.0000	0.0000
	0.06	0.7270	0.0651	0.1006	0.1856	0.0528	0.0000	0.0000	0.0000	0.0665	0.0000	0.0000
	0.05	0.7270	0.0659	0.1209	0.2044	0.0398	0.0000	0.0000	0.0000	0.0501	0.0000	0.0000
	0.04	0.7134	0.0558	0.1623	0.2341	0.0289	0.0000	0.0000	0.0000	0.0364	0.0000	0.0000
	0.03	0.6881	0.0533	0.2113	0.2777	0.0207	0.0000	0.0000	0.0000	0.0261	0.0000	0.0000
	0.02	0.6492	0.0431	0.2823	0.3339	0.0124	0.0000	0.0000	0.0000	0.0157	0.0000	0.0000
	0.01	0.5714	0.0304	0.3888	0.4232	0.0046	0.0000	0.0000	0.0000	0.0058	0.0000	0.0000
	0.10	0.8626	0.0039	0.0037	0.0088	0.3358	0.0000	0.0000	0.0000	0.3429	0.0000	0.0000
0.25	0.09	0.8748	0.0041	0.0044	0.0096	0.3148	0.0000	0.0000	0.0000	0.3214	0.0000	0.0000
	0.08	0.8873	0.0050	0.0050	0.0111	0.2925	0.0000	0.0000	0.0000	0.2987	0.0000	0.0000
	0.07	0.8991	0.0044	0.0066	0.0121	0.2672	0.0000	0.0000	0.0000	0.2728	0.0000	0.0000
	0.06	0.9101	0.0054	0.0081	0.0146	0.2437	0.0000	0.0000	0.0000	0.2488	0.0000	0.0000
	0.05	0.9202	0.0054	0.0110	0.0175	0.2170	0.0000	0.0000	0.0000	0.2216	0.0000	0.0000
	0.04	0.9306	0.0060	0.0135	0.0205	0.1847	0.0000	0.0000	0.0000	0.1886	0.0000	0.0000
	0.03	0.9368	0.0054	0.0201	0.0265	0.1524	0.0000	0.0000	0.0000	0.1556	0.0000	0.0000
	0.02	0.9360	0.0060	0.0334	0.0404	0.1152	0.0000	0.0000	0.0000	0.1176	0.0000	0.0000
	0.01	0.9262	0.0044	0.0591	0.0641	0.0757	0.0000	0.0000	0.0000	0.0773	0.0000	0.0000
	0.10	0.9078	0.0000	0.0082	0.0089	0.5214	0.0000	0.0000	0.0000	0.5284	0.0000	0.0000
0.5	0.09	0.9161	0.0000	0.0088	0.0096	0.5041	0.0000	0.0000	0.0000	0.5108	0.0000	0.0000
	0.08	0.9245	0.0000	0.0098	0.0105	0.4883	0.0000	0.0000	0.0000	0.4948	0.0000	0.0000
	0.07	0.9328	0.0000	0.0111	0.0117	0.4669	0.0000	0.0000	0.0000	0.4731	0.0000	0.0000
	0.06	0.9384	0.0000	0.0136	0.0143	0.4454	0.0000	0.0000	0.0000	0.4514	0.0000	0.0000
	0.05	0.9439	0.0000	0.0152	0.0159	0.4192	0.0000	0.0000	0.0000	0.4248	0.0000	0.0000
	0.04	0.9493	0.0000	0.0178	0.0184	0.3898	0.0000	0.0000	0.0000	0.3950	0.0000	0.0000
	0.03	0.9564	0.0001	0.0207	0.0213	0.3589	0.0000	0.0000	0.0000	0.3637	0.0000	0.0000
	0.02	0.9586	0.0001	0.0258	0.0264	0.3167	0.0000	0.0000	0.0000	0.3209	0.0000	0.0000
	0.01	0.9583	0.0000	0.0352	0.0355	0.2572	0.0000	0.0000	0.0000	0.2606	0.0000	0.0000
	0.10	0.8678	0.0065	0.0340	0.0447	0.3731	0.0004	0.0000	0.0011	0.4349	0.0111	0.0000
range	0.09	0.8758	0.0064	0.0375	0.0477	0.3561	0.0004	0.0000	0.0011	0.4150	0.0111	0.0000
	0.08	0.8854	0.0056	0.0402	0.0492	0.3385	0.0001	0.0000	0.0003	0.3946	0.0028	0.0000
	0.07	0.8932	0.0056	0.0449	0.0536	0.3217	0.0000	0.0000	0.0000	0.3749	0.0000	0.0000
	0.06	0.8981	0.0051	0.0511	0.0589	0.3027	0.0000	0.0000	0.0000	0.3529	0.0000	0.0000
	0.05	0.8998	0.0042	0.0589	0.0655	0.2804	0.0000	0.0000	0.0000	0.3268	0.0000	0.0000
	0.04	0.8992	0.0033	0.0688	0.0742	0.2592	0.0000	0.0000	0.0000	0.3021	0.0000	0.0000
	0.03	0.8961	0.0031	0.0813	0.0861	0.2325	0.0000	0.0000	0.0000	0.2709	0.0000	0.0000
	0.02	0.8863	0.0027	0.0991	0.1031	0.2022	0.0000	0.0000	0.0000	0.2356	0.0000	0.0000
	0.01	0.8596	0.0015	0.1340	0.1362	0.1571	0.0000	0.0000	0.0000	0.1831	0.0000	0.0000

Table 15. Model (d), n=135

	BF $R^2 = 0.125$			BF $R^2 = 0.25$			BF $R^2 = 0.5$			BF $R^2 =$ range		
	1	3	10	1	3	10	1	3	10	1	3	10
m_1^*/t	0.8953	0.8192	0.7456	0.9805	0.9570	0.9114	0.9868	0.9683	0.9201	0.9502	0.9122	0.8532
m_2^*/t	0.0970	0.0605	0.0399	0.0105	0.0066	0.0029	0.0045	0.0025	0.0006	0.0399	0.0228	0.0141
m_3^*/t	0.0077	0.0031	0.0011	0.0089	0.0037	0.0011	0.0086	0.0031	0.0011	0.0099	0.0043	0.0013
FDR	0.1047	0.0720	0.0522	0.0195	0.0107	0.0044	0.0132	0.0058	0.0019	0.0498	0.0289	0.0178

Table 16. Model (d), n=135

		CIT				CMST						
		PW	ER	ER		PW	ER	ER		cPW	cER	cER
R^2	level	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/m_1	m_2^*/m_2	m_3^*/m_3
	0.10	0.8930	0.0001	0.0010	0.0012	0.5021	0.0000	0.0000	0.0000	0.5055	0.0000	0.0000
0.125	0.09	0.9043	0.0001	0.0010	0.0012	0.4831	0.0000	0.0000	0.0000	0.4864	0.0000	0.0000
	0.08	0.9164	0.0001	0.0010	0.0012	0.4653	0.0000	0.0000	0.0000	0.4684	0.0000	0.0000
	0.07	0.9267	0.0001	0.0010	0.0012	0.4455	0.0000	0.0000	0.0000	0.4485	0.0000	0.0000
	0.06	0.9380	0.0001	0.0010	0.0012	0.4238	0.0000	0.0000	0.0000	0.4267	0.0000	0.0000
	0.05	0.9496	0.0002	0.0010	0.0013	0.3992	0.0000	0.0000	0.0000	0.4019	0.0000	0.0000
	0.04	0.9609	0.0001	0.0011	0.0013	0.3700	0.0000	0.0000	0.0000	0.3725	0.0000	0.0000
	0.03	0.9708	0.0001	0.0011	0.0012	0.3395	0.0000	0.0000	0.0000	0.3418	0.0000	0.0000
	0.02	0.9783	0.0001	0.0016	0.0017	0.3016	0.0000	0.0000	0.0000	0.3036	0.0000	0.0000
	0.01	0.9885	0.0000	0.0024	0.0024	0.2507	0.0000	0.0000	0.0000	0.2524	0.0000	0.0000
	0.10	0.8915	0.0000	0.0003	0.0003	0.6361	0.0000	0.0000	0.0000	0.6396	0.0000	0.0000
0.25	0.09	0.9022	0.0000	0.0003	0.0003	0.6173	0.0000	0.0000	0.0000	0.6207	0.0000	0.0000
	0.08	0.9117	0.0000	0.0003	0.0003	0.5978	0.0000	0.0000	0.0000	0.6010	0.0000	0.0000
	0.07	0.9231	0.0000	0.0003	0.0003	0.5790	0.0000	0.0000	0.0000	0.5821	0.0000	0.0000
	0.06	0.9336	0.0000	0.0003	0.0003	0.5581	0.0000	0.0000	0.0000	0.5611	0.0000	0.0000
	0.05	0.9439	0.0000	0.0003	0.0003	0.5335	0.0000	0.0000	0.0000	0.5364	0.0000	0.0000
	0.04	0.9554	0.0000	0.0003	0.0003	0.5019	0.0000	0.0000	0.0000	0.5046	0.0000	0.0000
	0.03	0.9675	0.0000	0.0003	0.0003	0.4676	0.0000	0.0000	0.0000	0.4701	0.0000	0.0000
	0.02	0.9773	0.0001	0.0003	0.0004	0.4234	0.0000	0.0000	0.0000	0.4257	0.0000	0.0000
	0.01	0.9874	0.0000	0.0004	0.0004	0.3641	0.0000	0.0000	0.0000	0.3661	0.0000	0.0000
	0.10	0.9038	0.0000	0.0001	0.0001	0.7433	0.0000	0.0000	0.0000	0.7456	0.0000	0.0000
0.5	0.09	0.9110	0.0000	0.0001	0.0001	0.7293	0.0000	0.0000	0.0000	0.7316	0.0000	0.0000
	0.08	0.9214	0.0000	0.0001	0.0001	0.7134	0.0000	0.0000	0.0000	0.7156	0.0000	0.0000
	0.07	0.9316	0.0000	0.0001	0.0001	0.6953	0.0000	0.0000	0.0000	0.6975	0.0000	0.0000
	0.06	0.9408	0.0000	0.0001	0.0001	0.6771	0.0000	0.0000	0.0000	0.6792	0.0000	0.0000
	0.05	0.9504	0.0000	0.0001	0.0001	0.6551	0.0000	0.0000	0.0000	0.6571	0.0000	0.0000
	0.04	0.9596	0.0000	0.0001	0.0001	0.6274	0.0000	0.0000	0.0000	0.6294	0.0000	0.0000
	0.03	0.9693	0.0000	0.0001	0.0001	0.5941	0.0000	0.0000	0.0000	0.5959	0.0000	0.0000
	0.02	0.9782	0.0000	0.0001	0.0001	0.5466	0.0000	0.0000	0.0000	0.5483	0.0000	0.0000
	0.01	0.9896	0.0000	0.0001	0.0001	0.4751	0.0000	0.0000	0.0000	0.4766	0.0000	0.0000
	0.10	0.8926	0.0022	0.0073	0.0105	0.6069	0.0001	0.0000	0.0002	0.6239	0.0119	0.0000
range	0.09	0.9020	0.0021	0.0082	0.0113	0.5931	0.0000	0.0000	0.0000	0.6097	0.0000	0.0000
	0.08	0.9095	0.0021	0.0089	0.0120	0.5774	0.0000	0.0000	0.0000	0.5936	0.0000	0.0000
	0.07	0.9158	0.0018	0.0108	0.0135	0.5605	0.0000	0.0000	0.0000	0.5762	0.0000	0.0000
	0.06	0.9241	0.0014	0.0123	0.0146	0.5417	0.0000	0.0000	0.0000	0.5569	0.0000	0.0000
	0.05	0.9351	0.0013	0.0137	0.0158	0.5225	0.0000	0.0000	0.0000	0.5372	0.0000	0.0000
	0.04	0.9456	0.0011	0.0153	0.0170	0.4992	0.0000	0.0000	0.0000	0.5132	0.0000	0.0000
	0.03	0.9560	0.0009	0.0173	0.0188	0.4702	0.0000	0.0000	0.0000	0.4834	0.0000	0.0000
	0.02	0.9622	0.0004	0.0210	0.0217	0.4317	0.0000	0.0000	0.0000	0.4438	0.0000	0.0000
	0.01	0.9646	0.0002	0.0266	0.0271	0.3742	0.0000	0.0000	0.0000	0.3847	0.0000	0.0000

Table 17. Model (d), n=500

	BF $R^2 = 0.125$			BF $R^2 = 0.25$			BF $R^2 = 0.5$			BF $R^2 =$ range		
	1	3	10	1	3	10	1	3	10	1	3	10
m_1^*/t	0.9933	0.9823	0.9571	0.9946	0.9862	0.9662	0.9969	0.9918	0.9747	0.9878	0.9722	0.9456
m_2^*/t	0.0030	0.0017	0.0004	0.0005	0.0002	0.0001	0.0002	0.0001	0.0000	0.0085	0.0057	0.0030
m_3^*/t	0.0037	0.0011	0.0005	0.0049	0.0017	0.0004	0.0029	0.0005	0.0001	0.0037	0.0011	0.0005
FDR	0.0067	0.0028	0.0009	0.0054	0.0019	0.0005	0.0031	0.0006	0.0001	0.0122	0.0070	0.0037

Table 18. Model (d), n=500

		CIT				CMST						
		PW	ER	ER		PW	ER	ER		cPW	cER	cER
R^2	level	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/m_1	m_2^*/m_2	m_3^*/m_3
	0.10	0.6626	0.0149	0.1287	0.1782	0.0735	0.0031	0.0000	0.0405	0.1008	0.0138	0.0000
0.125	0.09	0.6843	0.0122	0.1355	0.1775	0.0654	0.0028	0.0000	0.0411	0.0897	0.0124	0.0000
	0.08	0.6965	0.0136	0.1463	0.1867	0.0575	0.0025	0.0000	0.0417	0.0788	0.0111	0.0000
	0.07	0.7046	0.0122	0.1585	0.1950	0.0495	0.0017	0.0000	0.0333	0.0679	0.0075	0.0000
	0.06	0.7222	0.0108	0.1667	0.1973	0.0420	0.0015	0.0000	0.0346	0.0576	0.0067	0.0000
	0.05	0.7182	0.0095	0.1870	0.2148	0.0341	0.0010	0.0000	0.0286	0.0467	0.0044	0.0000
	0.04	0.7263	0.0068	0.2060	0.2266	0.0267	0.0006	0.0000	0.0221	0.0365	0.0027	0.0000
	0.03	0.7276	0.0068	0.2263	0.2426	0.0188	0.0003	0.0000	0.0157	0.0258	0.0013	0.0000
	0.02	0.7222	0.0041	0.2615	0.2689	0.0122	0.0000	0.0000	0.0000	0.0168	0.0000	0.0000
	0.01	0.6545	0.0027	0.3401	0.3438	0.0057	0.0000	0.0000	0.0000	0.0078	0.0000	0.0000
	0.10	0.4765	0.0017	0.2361	0.3329	0.0728	0.0004	0.0009	0.0176	0.0921	0.0026	0.0172
0.25	0.09	0.5017	0.0017	0.2441	0.3288	0.0672	0.0004	0.0008	0.0176	0.0850	0.0026	0.0153
	0.08	0.5279	0.0013	0.2555	0.3273	0.0611	0.0004	0.0007	0.0177	0.0774	0.0026	0.0134
	0.07	0.5537	0.0017	0.2683	0.3278	0.0551	0.0003	0.0006	0.0161	0.0698	0.0019	0.0115
	0.06	0.5873	0.0024	0.2777	0.3229	0.0483	0.0003	0.0005	0.0164	0.0611	0.0019	0.0096
	0.05	0.6162	0.0020	0.2928	0.3236	0.0421	0.0002	0.0003	0.0118	0.0532	0.0013	0.0057
	0.04	0.6390	0.0013	0.3086	0.3266	0.0347	0.0002	0.0002	0.0114	0.0440	0.0013	0.0038
	0.03	0.6461	0.0013	0.3324	0.3406	0.0293	0.0002	0.0001	0.0102	0.0371	0.0013	0.0019
	0.02	0.6259	0.0010	0.3640	0.3683	0.0205	0.0001	0.0001	0.0097	0.0259	0.0006	0.0019
	0.01	0.5863	0.0010	0.4117	0.4131	0.0116	0.0001	0.0001	0.0169	0.0147	0.0006	0.0019
	0.10	0.2942	0.0001	0.4924	0.6261	0.0285	0.0067	0.0293	0.5583	0.0701	0.0225	0.0999
0.5	0.09	0.3098	0.0001	0.5028	0.6188	0.0255	0.0059	0.0269	0.5628	0.0627	0.0198	0.0917
	0.08	0.3257	0.0000	0.5152	0.6127	0.0231	0.0053	0.0237	0.5568	0.0568	0.0178	0.0807
	0.07	0.3441	0.0000	0.5306	0.6066	0.0202	0.0048	0.0214	0.5649	0.0496	0.0161	0.0728
	0.06	0.3591	0.0000	0.5479	0.6041	0.0181	0.0039	0.0188	0.5567	0.0444	0.0131	0.0640
	0.05	0.3708	0.0001	0.5660	0.6042	0.0155	0.0031	0.0163	0.5562	0.0380	0.0104	0.0554
	0.04	0.3809	0.0002	0.5855	0.6059	0.0126	0.0024	0.0129	0.5487	0.0309	0.0081	0.0438
	0.03	0.3758	0.0000	0.6106	0.6190	0.0095	0.0020	0.0095	0.5476	0.0235	0.0067	0.0325
	0.02	0.3525	0.0000	0.6442	0.6463	0.0065	0.0013	0.0066	0.5486	0.0160	0.0044	0.0226
	0.01	0.3049	0.0000	0.6948	0.6950	0.0044	0.0009	0.0036	0.5056	0.0109	0.0030	0.0123
	0.10	0.4319	0.0020	0.3380	0.4405	0.0753	0.0045	0.0043	0.1049	0.1280	0.0260	0.0578
range	0.09	0.4458	0.0017	0.3468	0.4388	0.0700	0.0040	0.0039	0.1017	0.1190	0.0231	0.0524
	0.08	0.4661	0.0014	0.3549	0.4332	0.0644	0.0036	0.0033	0.0969	0.1096	0.0208	0.0444
	0.07	0.4858	0.0014	0.3641	0.4293	0.0583	0.0029	0.0028	0.0892	0.0992	0.0167	0.0376
	0.06	0.5041	0.0014	0.3753	0.4276	0.0524	0.0025	0.0026	0.0889	0.0891	0.0144	0.0349
	0.05	0.5173	0.0010	0.3898	0.4304	0.0459	0.0023	0.0024	0.0931	0.0781	0.0133	0.0323
	0.04	0.5339	0.0014	0.4017	0.4302	0.0398	0.0018	0.0022	0.0915	0.0677	0.0104	0.0296
	0.03	0.5424	0.0014	0.4193	0.4368	0.0318	0.0014	0.0020	0.0969	0.0540	0.0081	0.0269
	0.02	0.5319	0.0014	0.4488	0.4584	0.0247	0.0009	0.0014	0.0855	0.0419	0.0052	0.0188
	0.01	0.4966	0.0010	0.4986	0.5015	0.0158	0.0005	0.0011	0.0920	0.0269	0.0029	0.0148

Table 19. Model (e), n=135

	BF $R^2 = 0.125$			BF $R^2 = 0.25$			BF $R^2 = 0.5$			BF $R^2 =$ range		
	1	3	10	1	3	10	1	3	10	1	3	10
m_1^*/t	0.7586	0.6749	0.5808	0.7902	0.7093	0.6059	0.4070	0.3247	0.2446	0.7031	0.6223	0.5279
m_2^*/t	0.2349	0.1709	0.1227	0.1573	0.1062	0.0695	0.2992	0.2260	0.1607	0.2078	0.1531	0.1133
m_3^*/t	0.0066	0.0030	0.0016	0.0524	0.0341	0.0186	0.2938	0.2253	0.1526	0.0891	0.0609	0.0363
FDR	0.2414	0.2049	0.1763	0.2098	0.1652	0.1269	0.5930	0.5816	0.5616	0.2969	0.2559	0.2209

Table 20. Model (e), n=135

		CIT				CMST						
		PW	ER	ER		PW	ER	ER		cPW	cER	cER
R^2	level	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/m_1	m_2^*/m_2	m_3^*/m_3
	0.10	0.0420	0.0019	0.4394	0.9131	0.0346	0.0000	0.0004	0.0114	0.0432	0.0000	0.0035
0.125	0.09	0.0474	0.0016	0.4456	0.9041	0.0306	0.0000	0.0003	0.0097	0.0382	0.0000	0.0026
	0.08	0.0499	0.0009	0.4540	0.9011	0.0262	0.0000	0.0003	0.0113	0.0327	0.0000	0.0026
	0.07	0.0538	0.0012	0.4618	0.8959	0.0213	0.0000	0.0003	0.0139	0.0266	0.0000	0.0026
	0.06	0.0620	0.0017	0.4688	0.8835	0.0172	0.0000	0.0003	0.0171	0.0215	0.0000	0.0026
	0.05	0.0689	0.0012	0.4787	0.8745	0.0141	0.0000	0.0002	0.0140	0.0176	0.0000	0.0017
	0.04	0.0801	0.0011	0.4908	0.8600	0.0109	0.0000	0.0000	0.0000	0.0136	0.0000	0.0000
	0.03	0.0965	0.0006	0.5056	0.8398	0.0078	0.0000	0.0000	0.0000	0.0097	0.0000	0.0000
	0.02	0.1242	0.0012	0.5238	0.8087	0.0048	0.0000	0.0000	0.0000	0.0060	0.0000	0.0000
	0.01	0.1737	0.0019	0.5524	0.7614	0.0026	0.0000	0.0000	0.0000	0.0032	0.0000	0.0000
	0.10	0.0358	0.0015	0.4877	0.9318	0.0646	0.0004	0.0061	0.0914	0.0845	0.0073	0.0356
0.25	0.09	0.0371	0.0016	0.4951	0.9305	0.0584	0.0004	0.0054	0.0903	0.0764	0.0073	0.0315
	0.08	0.0377	0.0015	0.5044	0.9306	0.0512	0.0003	0.0048	0.0906	0.0670	0.0055	0.0280
	0.07	0.0394	0.0013	0.5141	0.9290	0.0443	0.0002	0.0038	0.0828	0.0580	0.0036	0.0222
	0.06	0.0418	0.0014	0.5231	0.9261	0.0363	0.0001	0.0033	0.0856	0.0475	0.0018	0.0193
	0.05	0.0435	0.0011	0.5345	0.9249	0.0289	0.0000	0.0022	0.0707	0.0378	0.0000	0.0129
	0.04	0.0493	0.0009	0.5450	0.9172	0.0232	0.0000	0.0017	0.0683	0.0304	0.0000	0.0099
	0.03	0.0539	0.0015	0.5608	0.9125	0.0167	0.0000	0.0009	0.0511	0.0218	0.0000	0.0053
	0.02	0.0636	0.0012	0.5810	0.9015	0.0100	0.0000	0.0005	0.0476	0.0131	0.0000	0.0029
	0.01	0.0757	0.0009	0.6146	0.8905	0.0043	0.0000	0.0002	0.0444	0.0056	0.0000	0.0012
	0.10	0.0859	0.0003	0.2825	0.7669	0.0944	0.0036	0.0284	0.2532	0.1397	0.0727	0.1164
0.5	0.09	0.0885	0.0004	0.2935	0.7685	0.0826	0.0034	0.0271	0.2697	0.1222	0.0687	0.1111
	0.08	0.0959	0.0002	0.3048	0.7608	0.0745	0.0034	0.0256	0.2802	0.1102	0.0687	0.1049
	0.07	0.1035	0.0002	0.3152	0.7528	0.0667	0.0031	0.0231	0.2820	0.0987	0.0626	0.0947
	0.06	0.1119	0.0003	0.3288	0.7463	0.0575	0.0028	0.0212	0.2945	0.0851	0.0566	0.0869
	0.05	0.1246	0.0002	0.3430	0.7337	0.0496	0.0026	0.0190	0.3034	0.0734	0.0525	0.0779
	0.04	0.1367	0.0004	0.3618	0.7260	0.0403	0.0023	0.0180	0.3350	0.0596	0.0465	0.0738
	0.03	0.1476	0.0004	0.3844	0.7227	0.0321	0.0017	0.0160	0.3554	0.0475	0.0343	0.0656
	0.02	0.1632	0.0001	0.4187	0.7196	0.0227	0.0014	0.0133	0.3930	0.0336	0.0283	0.0545
	0.01	0.1939	0.0000	0.4669	0.7066	0.0107	0.0008	0.0100	0.5023	0.0158	0.0162	0.0410
	0.10	0.1019	0.0017	0.3955	0.7959	0.0701	0.0018	0.0107	0.1513	0.1018	0.0184	0.0632
range	0.09	0.1058	0.0015	0.4000	0.7914	0.0643	0.0018	0.0096	0.1506	0.0934	0.0184	0.0567
	0.08	0.1135	0.0010	0.4048	0.7814	0.0569	0.0015	0.0087	0.1520	0.0826	0.0153	0.0514
	0.07	0.1208	0.0009	0.4113	0.7734	0.0522	0.0013	0.0079	0.1498	0.0758	0.0133	0.0466
	0.06	0.1279	0.0015	0.4175	0.7661	0.0467	0.0011	0.0073	0.1525	0.0678	0.0112	0.0431
	0.05	0.1375	0.0012	0.4230	0.7551	0.0398	0.0010	0.0063	0.1550	0.0578	0.0102	0.0372
	0.04	0.1502	0.0010	0.4307	0.7419	0.0335	0.0007	0.0059	0.1646	0.0486	0.0071	0.0348
	0.03	0.1681	0.0010	0.4400	0.7240	0.0273	0.0006	0.0049	0.1677	0.0396	0.0061	0.0289
	0.02	0.1904	0.0010	0.4523	0.7042	0.0205	0.0006	0.0038	0.1767	0.0298	0.0061	0.0224
	0.01	0.2241	0.0012	0.4756	0.6802	0.0118	0.0006	0.0022	0.1918	0.0171	0.0061	0.0130

Table 21. Model (e), n=500

	BF $R^2 = 0.125$			BF $R^2 = 0.25$			BF $R^2 = 0.5$			BF $R^2 =$ range		
	1	3	10	1	3	10	1	3	10	1	3	10
m_1^*/t	0.8013	0.6906	0.5488	0.7718	0.6241	0.4661	0.6972	0.5514	0.4338	0.7203	0.6047	0.4833
m_2^*/t	0.0837	0.0492	0.0252	0.0553	0.0329	0.0202	0.0511	0.0346	0.0240	0.1025	0.0694	0.0471
m_3^*/t	0.1151	0.0706	0.0366	0.1729	0.0914	0.0546	0.2517	0.1534	0.0969	0.1772	0.1154	0.0722
FDR	0.1987	0.1479	0.1013	0.2282	0.1661	0.1383	0.3028	0.2542	0.2180	0.2797	0.2341	0.1979

Table 22. Model (e), n=500

		CIT				CMST						
		PW	ER	ER		PW	ER	ER		cPW	cER	cER
R^2	level	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/m_1	m_2^*/m_2	m_3^*/m_3
	0.10	0.1347	0.3568	0.4128	0.8511	0.0118	0.0107	0.0000	0.4756	0.0313	0.0289	0.0000
0.125	0.09	0.1296	0.3436	0.4405	0.8581	0.0098	0.0089	0.0000	0.4759	0.0260	0.0241	0.0000
	0.08	0.1233	0.3329	0.4632	0.8658	0.0076	0.0076	0.0000	0.5000	0.0202	0.0205	0.0000
	0.07	0.1233	0.3103	0.4965	0.8674	0.0058	0.0065	0.0000	0.5285	0.0154	0.0176	0.0000
	0.06	0.1145	0.2920	0.5337	0.8782	0.0048	0.0049	0.0000	0.5052	0.0127	0.0132	0.0000
	0.05	0.1076	0.2775	0.5683	0.8871	0.0035	0.0032	0.0000	0.4776	0.0093	0.0086	0.0000
	0.04	0.1026	0.2568	0.6048	0.8936	0.0023	0.0022	0.0000	0.4889	0.0061	0.0059	0.0000
	0.03	0.0925	0.2335	0.6507	0.9053	0.0017	0.0015	0.0000	0.4688	0.0045	0.0041	0.0000
	0.02	0.0755	0.1926	0.7174	0.9234	0.0010	0.0008	0.0000	0.4444	0.0027	0.0022	0.0000
	0.01	0.0585	0.1334	0.8024	0.9411	0.0004	0.0001	0.0000	0.2000	0.0011	0.0003	0.0000
	0.10	0.1061	0.5879	0.1975	0.8810	0.0284	0.0323	0.0000	0.5322	0.0582	0.0640	0.0000
0.25	0.09	0.1049	0.5800	0.2191	0.8840	0.0252	0.0279	0.0000	0.5255	0.0516	0.0552	0.0000
	0.08	0.1016	0.5705	0.2414	0.8888	0.0213	0.0250	0.0000	0.5401	0.0436	0.0495	0.0000
	0.07	0.1018	0.5610	0.2638	0.8901	0.0187	0.0212	0.0000	0.5315	0.0382	0.0419	0.0000
	0.06	0.1020	0.5466	0.2888	0.8912	0.0164	0.0178	0.0000	0.5206	0.0335	0.0352	0.0000
	0.05	0.1023	0.5290	0.3178	0.8922	0.0136	0.0149	0.0000	0.5211	0.0280	0.0294	0.0000
	0.04	0.1000	0.5059	0.3565	0.8961	0.0110	0.0111	0.0000	0.5023	0.0226	0.0220	0.0000
	0.03	0.0968	0.4748	0.4031	0.9007	0.0083	0.0081	0.0000	0.4939	0.0171	0.0161	0.0000
	0.02	0.0879	0.4234	0.4745	0.9108	0.0051	0.0056	0.0000	0.5234	0.0105	0.0111	0.0000
	0.01	0.0768	0.3428	0.5761	0.9229	0.0027	0.0031	0.0000	0.5345	0.0056	0.0062	0.0000
	0.10	0.0241	0.7215	0.1377	0.9727	0.0949	0.0143	0.0000	0.1308	0.1355	0.0491	0.0000
0.5	0.09	0.0263	0.7290	0.1399	0.9706	0.0881	0.0136	0.0000	0.1334	0.1258	0.0467	0.0000
	0.08	0.0257	0.7347	0.1440	0.9716	0.0786	0.0127	0.0000	0.1388	0.1122	0.0436	0.0000
	0.07	0.0280	0.7423	0.1485	0.9696	0.0707	0.0117	0.0000	0.1416	0.1009	0.0401	0.0000
	0.06	0.0298	0.7472	0.1556	0.9681	0.0606	0.0103	0.0000	0.1447	0.0865	0.0353	0.0000
	0.05	0.0297	0.7519	0.1631	0.9686	0.0535	0.0085	0.0000	0.1378	0.0763	0.0294	0.0000
	0.04	0.0319	0.7543	0.1717	0.9667	0.0453	0.0069	0.0000	0.1327	0.0647	0.0239	0.0000
	0.03	0.0344	0.7529	0.1845	0.9646	0.0368	0.0057	0.0000	0.1348	0.0525	0.0197	0.0000
	0.02	0.0359	0.7465	0.2033	0.9636	0.0272	0.0040	0.0000	0.1286	0.0389	0.0138	0.0000
	0.01	0.0309	0.7169	0.2479	0.9690	0.0153	0.0023	0.0000	0.1314	0.0218	0.0080	0.0000
	0.10	0.1191	0.4857	0.2012	0.8523	0.1075	0.0193	0.0001	0.1527	0.1889	0.0658	0.0084
range	0.09	0.1194	0.4880	0.2125	0.8544	0.1012	0.0176	0.0001	0.1485	0.1780	0.0600	0.0084
	0.08	0.1205	0.4921	0.2231	0.8558	0.0958	0.0150	0.0001	0.1365	0.1684	0.0514	0.0084
	0.07	0.1204	0.4956	0.2373	0.8589	0.0894	0.0132	0.0001	0.1299	0.1571	0.0452	0.0084
	0.06	0.1239	0.4994	0.2512	0.8584	0.0840	0.0113	0.0001	0.1199	0.1476	0.0387	0.0084
	0.05	0.1268	0.4988	0.2699	0.8584	0.0754	0.0098	0.0001	0.1163	0.1326	0.0336	0.0084
	0.04	0.1255	0.4996	0.2923	0.8632	0.0673	0.0084	0.0001	0.1124	0.1183	0.0288	0.0084
	0.03	0.1263	0.4973	0.3159	0.8655	0.0588	0.0066	0.0001	0.1026	0.1034	0.0226	0.0084
	0.02	0.1201	0.4860	0.3539	0.8749	0.0483	0.0047	0.0001	0.0907	0.0848	0.0161	0.0084
	0.01	0.1078	0.4580	0.4173	0.8903	0.0353	0.0036	0.0001	0.0951	0.0621	0.0123	0.0084

Table 23. Model (f), n=135

	BF			BF			BF			BF		
	$R^2 = 0.125$			$R^2 = 0.25$			$R^2 = 0.5$			$R^2 =$ range		
	1	3	10	1	3	10	1	3	10	1	3	10
m_1^*/t	0.5026	0.4143	0.3398	0.4896	0.4176	0.3469	0.7004	0.6444	0.5769	0.6512	0.5891	0.5220
m_2^*/t	0.4934	0.4113	0.3386	0.5069	0.4329	0.3547	0.2908	0.2373	0.1869	0.3351	0.2759	0.2245
m_3^*/t	0.0040	0.0013	0.0004	0.0034	0.0020	0.0008	0.0088	0.0036	0.0015	0.0137	0.0087	0.0049
FDR	0.4974	0.4990	0.4994	0.5104	0.5102	0.5062	0.2996	0.2721	0.2462	0.3488	0.3258	0.3054

Table 24. Model (f), n=135

		CIT				CMST						
		PW	ER	ER		PW	ER	ER		cPW	cER	cER
R^2	level	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/m_1	m_2^*/m_2	m_3^*/m_3
	0.10	0.0443	0.6930	0.0794	0.9458	0.0048	0.0054	0.0000	0.5294	0.0092	0.0113	0.0000
0.125	0.09	0.0461	0.6944	0.0875	0.9444	0.0043	0.0048	0.0000	0.5275	0.0083	0.0101	0.0000
	0.08	0.0465	0.6958	0.0971	0.9446	0.0036	0.0045	0.0000	0.5556	0.0069	0.0095	0.0000
	0.07	0.0475	0.6943	0.1071	0.9440	0.0033	0.0036	0.0000	0.5217	0.0064	0.0076	0.0000
	0.06	0.0493	0.6925	0.1189	0.9427	0.0028	0.0033	0.0000	0.5410	0.0054	0.0069	0.0000
	0.05	0.0518	0.6869	0.1346	0.9407	0.0024	0.0025	0.0000	0.5102	0.0046	0.0053	0.0000
	0.04	0.0563	0.6780	0.1543	0.9366	0.0023	0.0018	0.0000	0.4390	0.0044	0.0038	0.0000
	0.03	0.0628	0.6639	0.1822	0.9309	0.0016	0.0014	0.0000	0.4667	0.0031	0.0029	0.0000
	0.02	0.0694	0.6371	0.2248	0.9254	0.0011	0.0010	0.0000	0.4762	0.0021	0.0021	0.0000
	0.01	0.0843	0.5795	0.2962	0.9122	0.0003	0.0004	0.0000	0.5714	0.0006	0.0008	0.0000
	0.10	0.0062	0.5693	0.2465	0.9925	0.0064	0.0025	0.0000	0.2809	0.0108	0.0065	0.0000
0.25	0.09	0.0071	0.5767	0.2481	0.9915	0.0053	0.0020	0.0000	0.2740	0.0090	0.0052	0.0000
	0.08	0.0069	0.5843	0.2506	0.9919	0.0041	0.0017	0.0000	0.2931	0.0069	0.0044	0.0000
	0.07	0.0083	0.5905	0.2520	0.9903	0.0027	0.0012	0.0000	0.3077	0.0046	0.0031	0.0000
	0.06	0.0097	0.5974	0.2540	0.9888	0.0021	0.0008	0.0000	0.2759	0.0035	0.0021	0.0000
	0.05	0.0095	0.6039	0.2572	0.9891	0.0015	0.0007	0.0000	0.3182	0.0025	0.0018	0.0000
	0.04	0.0107	0.6116	0.2598	0.9879	0.0012	0.0005	0.0000	0.2941	0.0020	0.0013	0.0000
	0.03	0.0114	0.6191	0.2638	0.9873	0.0008	0.0003	0.0000	0.2727	0.0014	0.0008	0.0000
	0.02	0.0150	0.6269	0.2685	0.9835	0.0005	0.0001	0.0000	0.1667	0.0008	0.0003	0.0000
	0.01	0.0186	0.6312	0.2809	0.9801	0.0003	0.0001	0.0000	0.2500	0.0005	0.0003	0.0000
	0.10	0.0028	0.4816	0.4422	0.9970	0.0207	0.0048	0.0000	0.1882	0.0374	0.0120	0.0000
0.5	0.09	0.0031	0.4868	0.4427	0.9967	0.0190	0.0043	0.0000	0.1845	0.0343	0.0108	0.0000
	0.08	0.0032	0.4933	0.4432	0.9966	0.0168	0.0034	0.0000	0.1683	0.0303	0.0085	0.0000
	0.07	0.0027	0.4981	0.4445	0.9971	0.0156	0.0024	0.0000	0.1333	0.0282	0.0060	0.0000
	0.06	0.0034	0.5039	0.4454	0.9964	0.0120	0.0021	0.0000	0.1489	0.0217	0.0053	0.0000
	0.05	0.0029	0.5096	0.4465	0.9970	0.0100	0.0015	0.0000	0.1304	0.0181	0.0038	0.0000
	0.04	0.0028	0.5155	0.4475	0.9971	0.0079	0.0008	0.0000	0.0920	0.0143	0.0020	0.0000
	0.03	0.0025	0.5205	0.4493	0.9974	0.0059	0.0006	0.0000	0.0923	0.0107	0.0015	0.0000
	0.02	0.0026	0.5276	0.4499	0.9973	0.0046	0.0005	0.0000	0.0980	0.0083	0.0013	0.0000
	0.01	0.0030	0.5333	0.4517	0.9970	0.0017	0.0000	0.0000	0.0000	0.0031	0.0000	0.0000
	0.10	0.0551	0.3108	0.2594	0.9119	0.1066	0.0044	0.0007	0.0457	0.1542	0.0185	0.0148
range	0.09	0.0572	0.3166	0.2621	0.9101	0.1016	0.0038	0.0007	0.0424	0.1470	0.0160	0.0148
	0.08	0.0579	0.3221	0.2663	0.9103	0.0943	0.0035	0.0005	0.0407	0.1364	0.0147	0.0105
	0.07	0.0598	0.3288	0.2709	0.9094	0.0871	0.0031	0.0005	0.0397	0.1260	0.0130	0.0105
	0.06	0.0616	0.3340	0.2759	0.9083	0.0801	0.0027	0.0004	0.0373	0.1159	0.0114	0.0084
	0.05	0.0620	0.3421	0.2823	0.9096	0.0710	0.0018	0.0003	0.0287	0.1027	0.0076	0.0063
	0.04	0.0655	0.3529	0.2894	0.9074	0.0616	0.0013	0.0002	0.0238	0.0891	0.0055	0.0042
	0.03	0.0695	0.3628	0.2987	0.9049	0.0537	0.0010	0.0001	0.0201	0.0777	0.0042	0.0021
	0.02	0.0725	0.3746	0.3137	0.9047	0.0436	0.0006	0.0000	0.0136	0.0631	0.0025	0.0000
	0.01	0.0810	0.3898	0.3340	0.8994	0.0306	0.0002	0.0000	0.0065	0.0443	0.0008	0.0000

Table 25. Model (f), n=500

	BF $R^2 = 0.125$			BF $R^2 = 0.25$			BF $R^2 = 0.5$			BF $R^2 =$ range		
	1	3	10	1	3	10	1	3	10	1	3	10
m_1^*/t	0.5195	0.4249	0.3353	0.5918	0.4751	0.3610	0.5548	0.4051	0.2740	0.7080	0.6162	0.5156
m_2^*/t	0.4760	0.3778	0.2837	0.3872	0.2733	0.1726	0.4002	0.2584	0.1481	0.2434	0.1763	0.1182
m_3^*/t	0.0045	0.0019	0.0006	0.0210	0.0082	0.0026	0.0450	0.0131	0.0041	0.0485	0.0221	0.0107
FDR	0.4805	0.4719	0.4588	0.4082	0.3721	0.3267	0.4452	0.4012	0.3571	0.2920	0.2435	0.1999

Table 26. Model (f), n=500

		CIT				CMST						
		PW	ER	ER		PW	ER	ER		cPW	cER	cER
R^2	level	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/m_1	m_2^*/m_2	m_3^*/m_3
	0.10	0.6732	0.0786	0.0560	0.1667	0.0963	0.0001	0.0000	0.0010	0.1217	0.0012	0.0000
0.125	0.09	0.6794	0.0786	0.0708	0.1803	0.0837	0.0000	0.0000	0.0000	0.1057	0.0000	0.0000
	0.08	0.6864	0.0794	0.0872	0.1953	0.0728	0.0000	0.0000	0.0000	0.0920	0.0000	0.0000
	0.07	0.6872	0.0833	0.1051	0.2151	0.0607	0.0000	0.0000	0.0000	0.0767	0.0000	0.0000
	0.06	0.6848	0.0809	0.1276	0.2334	0.0481	0.0000	0.0000	0.0000	0.0608	0.0000	0.0000
	0.05	0.6926	0.0755	0.1525	0.2477	0.0369	0.0000	0.0000	0.0000	0.0466	0.0000	0.0000
	0.04	0.6872	0.0669	0.1852	0.2684	0.0265	0.0000	0.0000	0.0000	0.0334	0.0000	0.0000
	0.03	0.6630	0.0576	0.2397	0.3096	0.0174	0.0000	0.0000	0.0000	0.0220	0.0000	0.0000
	0.02	0.6218	0.0451	0.3097	0.3633	0.0099	0.0000	0.0000	0.0000	0.0125	0.0000	0.0000
	0.01	0.5276	0.0288	0.4374	0.4691	0.0038	0.0000	0.0000	0.0000	0.0048	0.0000	0.0000
	0.10	0.8647	0.0057	0.0048	0.0120	0.3199	0.0000	0.0000	0.0000	0.3279	0.0000	0.0000
0.25	0.09	0.8758	0.0059	0.0053	0.0126	0.2998	0.0000	0.0000	0.0000	0.3073	0.0000	0.0000
	0.08	0.8907	0.0069	0.0055	0.0137	0.2772	0.0000	0.0000	0.0000	0.2841	0.0000	0.0000
	0.07	0.9010	0.0076	0.0071	0.0161	0.2563	0.0000	0.0000	0.0000	0.2627	0.0000	0.0000
	0.06	0.9126	0.0088	0.0086	0.0188	0.2303	0.0000	0.0000	0.0000	0.2360	0.0000	0.0000
	0.05	0.9208	0.0086	0.0122	0.0221	0.2051	0.0000	0.0000	0.0000	0.2102	0.0000	0.0000
	0.04	0.9296	0.0090	0.0164	0.0266	0.1803	0.0000	0.0000	0.0000	0.1849	0.0000	0.0000
	0.03	0.9372	0.0082	0.0242	0.0334	0.1476	0.0000	0.0000	0.0000	0.1513	0.0000	0.0000
	0.02	0.9365	0.0078	0.0357	0.0444	0.1147	0.0000	0.0000	0.0000	0.1176	0.0000	0.0000
	0.01	0.9262	0.0059	0.0607	0.0671	0.0730	0.0000	0.0000	0.0000	0.0748	0.0000	0.0000
	0.10	0.9065	0.0000	0.0077	0.0084	0.5199	0.0000	0.0000	0.0000	0.5276	0.0000	0.0000
0.5	0.09	0.9150	0.0000	0.0085	0.0092	0.5024	0.0000	0.0000	0.0000	0.5099	0.0000	0.0000
	0.08	0.9233	0.0000	0.0089	0.0096	0.4846	0.0000	0.0000	0.0000	0.4919	0.0000	0.0000
	0.07	0.9304	0.0000	0.0104	0.0110	0.4655	0.0000	0.0000	0.0000	0.4724	0.0000	0.0000
	0.06	0.9395	0.0000	0.0110	0.0116	0.4441	0.0000	0.0000	0.0000	0.4507	0.0000	0.0000
	0.05	0.9474	0.0000	0.0126	0.0131	0.4189	0.0000	0.0000	0.0000	0.4251	0.0000	0.0000
	0.04	0.9559	0.0000	0.0145	0.0149	0.3920	0.0000	0.0000	0.0000	0.3978	0.0000	0.0000
	0.03	0.9610	0.0000	0.0174	0.0178	0.3597	0.0000	0.0000	0.0000	0.3650	0.0000	0.0000
	0.02	0.9643	0.0000	0.0216	0.0219	0.3193	0.0000	0.0000	0.0000	0.3241	0.0000	0.0000
	0.01	0.9612	0.0000	0.0320	0.0322	0.2558	0.0000	0.0000	0.0000	0.2596	0.0000	0.0000
	0.10	0.8530	0.0093	0.0283	0.0421	0.2484	0.0001	0.0000	0.0004	0.3273	0.0023	0.0000
range	0.09	0.8636	0.0090	0.0312	0.0445	0.2343	0.0001	0.0000	0.0004	0.3087	0.0023	0.0000
	0.08	0.8736	0.0090	0.0349	0.0478	0.2201	0.0001	0.0000	0.0005	0.2900	0.0023	0.0000
	0.07	0.8803	0.0090	0.0415	0.0542	0.2040	0.0000	0.0000	0.0000	0.2688	0.0000	0.0000
	0.06	0.8898	0.0085	0.0452	0.0569	0.1858	0.0000	0.0000	0.0000	0.2448	0.0000	0.0000
	0.05	0.8964	0.0069	0.0526	0.0622	0.1699	0.0000	0.0000	0.0000	0.2238	0.0000	0.0000
	0.04	0.8974	0.0063	0.0595	0.0683	0.1538	0.0000	0.0000	0.0000	0.2027	0.0000	0.0000
	0.03	0.8948	0.0061	0.0719	0.0802	0.1347	0.0000	0.0000	0.0000	0.1775	0.0000	0.0000
	0.02	0.8906	0.0040	0.0896	0.0951	0.1134	0.0000	0.0000	0.0000	0.1495	0.0000	0.0000
	0.01	0.8625	0.0037	0.1279	0.1324	0.0861	0.0000	0.0000	0.0000	0.1134	0.0000	0.0000

Table 27. Model (g), n=135

	BF $R^2 = 0.125$			BF $R^2 = 0.25$			BF $R^2 = 0.5$			BF $R^2 =$ range		
	1	3	10	1	3	10	1	3	10	1	3	10
m_1^*/t	0.8961	0.8214	0.7509	0.9764	0.9534	0.9095	0.9853	0.9666	0.9179	0.9336	0.8841	0.8163
m_2^*/t	0.0980	0.0624	0.0406	0.0153	0.0070	0.0029	0.0049	0.0023	0.0009	0.0547	0.0341	0.0240
m_3^*/t	0.0059	0.0027	0.0005	0.0083	0.0033	0.0016	0.0097	0.0036	0.0010	0.0117	0.0052	0.0018
FDR	0.1039	0.0735	0.0519	0.0236	0.0107	0.0049	0.0147	0.0061	0.0021	0.0664	0.0426	0.0307

Table 28. Model (g), n=135

		CIT				CMST						
		PW	ER	ER		PW	ER	ER		cPW	cER	cER
R^2	level	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/t	m_2^*/t	m_3^*/t	FDR	m_1^*/m_1	m_2^*/m_2	m_3^*/m_3
	0.10	0.8834	0.0000	0.0006	0.0007	0.5060	0.0000	0.0000	0.0000	0.5099	0.0000	0.0000
0.125	0.09	0.8953	0.0000	0.0006	0.0007	0.4873	0.0000	0.0000	0.0000	0.4910	0.0000	0.0000
	0.08	0.9059	0.0000	0.0006	0.0007	0.4685	0.0000	0.0000	0.0000	0.4721	0.0000	0.0000
	0.07	0.9176	0.0002	0.0007	0.0011	0.4446	0.0000	0.0000	0.0000	0.4480	0.0000	0.0000
	0.06	0.9286	0.0002	0.0007	0.0010	0.4208	0.0000	0.0000	0.0000	0.4240	0.0000	0.0000
	0.05	0.9399	0.0000	0.0012	0.0013	0.3967	0.0000	0.0000	0.0000	0.3997	0.0000	0.0000
	0.04	0.9515	0.0000	0.0012	0.0013	0.3682	0.0000	0.0000	0.0000	0.3710	0.0000	0.0000
	0.03	0.9635	0.0002	0.0012	0.0015	0.3358	0.0000	0.0000	0.0000	0.3384	0.0000	0.0000
	0.02	0.9756	0.0004	0.0017	0.0021	0.2937	0.0000	0.0000	0.0000	0.2959	0.0000	0.0000
	0.01	0.9872	0.0001	0.0034	0.0035	0.2409	0.0000	0.0000	0.0000	0.2427	0.0000	0.0000
	0.10	0.8916	0.0000	0.0002	0.0002	0.6370	0.0000	0.0000	0.0000	0.6407	0.0000	0.0000
0.25	0.09	0.9024	0.0000	0.0002	0.0002	0.6205	0.0000	0.0000	0.0000	0.6241	0.0000	0.0000
	0.08	0.9126	0.0000	0.0002	0.0002	0.6029	0.0000	0.0000	0.0000	0.6064	0.0000	0.0000
	0.07	0.9254	0.0000	0.0002	0.0002	0.5821	0.0000	0.0000	0.0000	0.5854	0.0000	0.0000
	0.06	0.9354	0.0000	0.0002	0.0002	0.5638	0.0000	0.0000	0.0000	0.5670	0.0000	0.0000
	0.05	0.9442	0.0000	0.0002	0.0002	0.5371	0.0000	0.0000	0.0000	0.5402	0.0000	0.0000
	0.04	0.9535	0.0000	0.0002	0.0002	0.5102	0.0000	0.0000	0.0000	0.5131	0.0000	0.0000
	0.03	0.9631	0.0000	0.0002	0.0002	0.4779	0.0000	0.0000	0.0000	0.4806	0.0000	0.0000
	0.02	0.9746	0.0000	0.0002	0.0002	0.4359	0.0000	0.0000	0.0000	0.4384	0.0000	0.0000
	0.01	0.9873	0.0000	0.0003	0.0003	0.3775	0.0000	0.0000	0.0000	0.3797	0.0000	0.0000
	0.10	0.8979	0.0000	0.0003	0.0003	0.7452	0.0000	0.0000	0.0000	0.7474	0.0000	0.0000
0.5	0.09	0.9083	0.0000	0.0003	0.0003	0.7327	0.0000	0.0000	0.0000	0.7349	0.0000	0.0000
	0.08	0.9177	0.0000	0.0003	0.0003	0.7173	0.0000	0.0000	0.0000	0.7195	0.0000	0.0000
	0.07	0.9282	0.0000	0.0003	0.0003	0.7006	0.0000	0.0000	0.0000	0.7027	0.0000	0.0000
	0.06	0.9381	0.0000	0.0003	0.0003	0.6807	0.0000	0.0000	0.0000	0.6827	0.0000	0.0000
	0.05	0.9471	0.0000	0.0003	0.0003	0.6581	0.0000	0.0000	0.0000	0.6601	0.0000	0.0000
	0.04	0.9561	0.0000	0.0003	0.0003	0.6322	0.0000	0.0000	0.0000	0.6341	0.0000	0.0000
	0.03	0.9676	0.0000	0.0003	0.0003	0.5975	0.0000	0.0000	0.0000	0.5993	0.0000	0.0000
	0.02	0.9787	0.0000	0.0003	0.0003	0.5528	0.0000	0.0000	0.0000	0.5545	0.0000	0.0000
	0.01	0.9897	0.0000	0.0003	0.0003	0.4858	0.0000	0.0000	0.0000	0.4873	0.0000	0.0000
	0.10	0.8801	0.0022	0.0070	0.0103	0.4930	0.0000	0.0000	0.0000	0.5327	0.0000	0.0000
range	0.09	0.8907	0.0023	0.0075	0.0109	0.4776	0.0000	0.0000	0.0000	0.5160	0.0000	0.0000
	0.08	0.9008	0.0021	0.0082	0.0113	0.4637	0.0000	0.0000	0.0000	0.5010	0.0000	0.0000
	0.07	0.9132	0.0018	0.0100	0.0127	0.4481	0.0000	0.0000	0.0000	0.4842	0.0000	0.0000
	0.06	0.9221	0.0016	0.0108	0.0133	0.4297	0.0000	0.0000	0.0000	0.4643	0.0000	0.0000
	0.05	0.9303	0.0016	0.0133	0.0158	0.4101	0.0000	0.0000	0.0000	0.4431	0.0000	0.0000
	0.04	0.9368	0.0018	0.0164	0.0191	0.3886	0.0000	0.0000	0.0000	0.4199	0.0000	0.0000
	0.03	0.9442	0.0018	0.0189	0.0214	0.3612	0.0000	0.0000	0.0000	0.3903	0.0000	0.0000
	0.02	0.9511	0.0016	0.0227	0.0250	0.3295	0.0000	0.0000	0.0000	0.3560	0.0000	0.0000
	0.01	0.9568	0.0016	0.0309	0.0329	0.2786	0.0000	0.0000	0.0000	0.3010	0.0000	0.0000

Table 29. Model (g), n=500

	BF $R^2 = 0.125$			BF $R^2 = 0.25$			BF $R^2 = 0.5$			BF $R^2 =$ range		
	1	3	10	1	3	10	1	3	10	1	3	10
m_1^*/t	0.9924	0.9828	0.9574	0.9943	0.9865	0.9667	0.9970	0.9923	0.9773	0.9765	0.9525	0.9156
m_2^*/t	0.0031	0.0014	0.0003	0.0012	0.0003	0.0001	0.0003	0.0001	0.0000	0.0176	0.0090	0.0065
m_3^*/t	0.0045	0.0016	0.0007	0.0045	0.0020	0.0006	0.0027	0.0011	0.0004	0.0059	0.0015	0.0003
FDR	0.0076	0.0030	0.0010	0.0057	0.0023	0.0007	0.0030	0.0012	0.0004	0.0235	0.0108	0.0074

Table 30. Model (g), n=500

As expected, when data is generated from the null model (a), that do not correspond to any of the fitted models, the CMST made very few significant calls (see Tables 3 and 5). Note that in this case any call correspond to a mistake and the false discovery proportion is 1, even though the error rates were very low.

Models (b) and specially (d) and (g) were the less challenging for all methods, showing high power, low error rates and low false discovery proportions (see Tables 7-10, 15-18 and 27-30). Note that models (b) and (d) correspond to simulation settings under correctly specified models. It is interesting to notice that the CMST is less powered to detect a relationship of type M_3 than M_1. For example, consider the results for models (b) and (d), that correspond to our M_3 and M_1 fitted models (see Tables 7, 9, 15 and 17). However, since in practice the real interest is to discover causal relationships, failing to detect uninteresting M_3 relationships is not a concern.

Models (c) and specially (f) were the most challenging for all methods showing the lowest power and high error rates. Nonetheless, the CMST still managed to keep the error rates controlled at the nominal levels, but showed very low power in these situations. For example, in Table 11 we see that the power to detect a M_3 relation when $R^2 = 0.125$ was 0, whereas the error rates, although very low, were higher than 0. So in this case the false discovery proportion was 1. In other words, the very few significant calls were M_1 or M_2, instead of M_3. The CMST also showed high false discovery proportions for model (f), but in this case too, the error rates where low.

Model (g) showed performance comparable to model (d) (see Tables 15, 17, 27 and 29), suggesting that missing to include a common latent variable (models c, e and f) has a stronger effect than missing covariates that are specific to each phenotype.

Overall, these results show than when the true model corresponds (or is close) to one of our fitted models, such as in model (d), the BIC and CIT approaches outperform the CMST, since they are better powered and show low error rates. However, when the true model is more complex and our fitted models are misspecified, then CMST detects false positives at much lower rates than the BIC and CIT.

So far we have considered the performances of each simulated model, separately. Following Chen et al. (2007) we now investigate the performances of the BIC, CIT and CMST approaches using the results from all 7 models together. In other words, we now view this simulation study as a genetics of gene expression experiment with pairwise causal relationships simulated from a mixture of the 7 different models on Figure 8 with a uniform prior probability on the models. Note that, now, the FDR corresponds to the number of incorrect calls (across all 7 models) divided by the number of calls (across all 7 models). Tables 31 and 32 present the results for the CIT, CMST and BIC approaches, respectively.

	n=135						n=500					
	CIT			CMST			CIT			CMST		
level	PW	ER	FDR	PW	ER	FDR	PW	ER	FDR	PW	ER	FDR
0.1	0.4500	0.2595	0.3821	0.1553	0.0080	0.0568	0.3758	0.2433	0.3927	0.3110	0.0029	0.0106
0.09	0.4577	0.2690	0.3846	0.1465	0.0071	0.0535	0.3809	0.2474	0.3934	0.3008	0.0026	0.0100
0.08	0.4654	0.2793	0.3872	0.1373	0.0062	0.0503	0.3859	0.2519	0.3946	0.2899	0.0024	0.0094
0.07	0.4734	0.2904	0.3908	0.1276	0.0053	0.0461	0.3914	0.2566	0.3954	0.2784	0.0021	0.0085
0.06	0.4809	0.3067	0.3959	0.1173	0.0044	0.0423	0.3970	0.2620	0.3968	0.2664	0.0018	0.0079
0.05	0.4872	0.3234	0.4022	0.1062	0.0037	0.0388	0.4028	0.2682	0.3988	0.2526	0.0015	0.0069
0.04	0.4924	0.3442	0.4106	0.0947	0.0028	0.0337	0.4094	0.2756	0.4012	0.2376	0.0013	0.0064
0.03	0.4957	0.3708	0.4219	0.0821	0.0021	0.0285	0.4167	0.2850	0.4048	0.2198	0.0011	0.0056
0.02	0.4940	0.4064	0.4397	0.0676	0.0014	0.0237	0.4253	0.2979	0.4101	0.1975	0.0008	0.0050
0.01	0.4918	0.4523	0.4642	0.0487	0.0008	0.0190	0.4380	0.3182	0.4184	0.1663	0.0005	0.0038

Table 31. Simulation results for the CIT and CMST approaches using the results from all 7 models together. PW represents the overall power, given the number of correct calls divided by the total number of tests. ER represents the overall error rates given by the number of mistakes divided by total number of tests.

When considering all seven models together we see that the CMST has much lower FDRs than CIT and BIC. Note that this result does not contradict the high FDRs observed for the CMST under models

	n=135			n=500		
BF	PW	ER	FDR	PW	ER	FDR
1	0.6830	0.4146	0.4026	0.8215	0.2959	0.2958
3	0.6143	0.2109	0.2762	0.7493	0.1075	0.1431
10	0.5372	0.1502	0.2358	0.6600	0.0621	0.0987

Table 32. Simulation results for the BIC approach using the results from all 7 models together. PW represents the overall power, given the number of correct calls divided by the total number of tests. ER represents the overall error rates given by the number of mistakes divided by total number of tests.

(c), (e) and (f), since the number of calls was rather low for these models (recall that these models showed low power and low error rates). Models (d) and (g), on the other hand, showed very low FDRs and reasonable power, so that under these models most of the calls were correct.

Simulation study 2: a realistic network

In this section we compare the performance of the BIC, CIT and CMST approaches with data simulated from a network composed of 30 phenotype nodes, that generates a QTL hotspot pattern. We assessed and compared the performances of the CMST, BIC and CIT approaches under two different simulation settings: 1) strong phenotype and genetic effects; and 2) strong phenotype effects and weak genetic effects. In each setting, we generated 100 data sets according to the network on Figure 9. Each data set corresponds to a F_2 population of 500 individuals. We generated measurements on 30 phenotypes and 1919 markers unequally distributed across 19 autosomes of length 100cM (101 markers per chromosome). We adopted a QTL mapping threshold of 4 determined by a permutation test (Churchill and Doerge 1994). Each phenotype is directly affected by a single QTL. The phenotype-to-phenotype effects varied from 1 to 1.2. Additive genetic effects varied from 1 to 1.2 in simulation setting 1, and from 0.4 to 0.6 in simulation setting 2. The residual variance was set to 1.

We adopted strong phenotype-to-phenotype effects in order to make most phenotypes map to the QTL directly affecting Y_1, generating a QTL hotspot pattern. In practice, we are often interested in determining the causal drivers generating a QTL hotspot. Furthermore, in order to be able to compare our approach to the CIT we need to enforce that most phenotypes map to common QTLs, since the CIT approach requires that both phenotypes map to the same QTL, and cannot be applied otherwise.

Figure 10 presents the simulation results. For each of the three approaches under study, we kept track of: (i) the proportion of calls (for the BIC approach we make a call when the difference between the BIC scores of the best and second best models corresponds to an approximate Bayes factor (Kass and Raftery 1995) greater than 10. For the CIT and CMST approaches we make a call when the tests are significant at an 0.05 level); (ii) the error rate, defined as the number of mistakes divided by the number of tests; and (iii) the false discovery proportion, defined as the number of mistakes divided by the number of calls.

Figure 10a shows that the BIC and CIT (to a lesser degree) make calls at much higher rates than the CMST. Unlike the BIC and CIT, the CMST concentrate its calls in situations where one of our fitted models is correctly specified (or close to correctly specified), such as the tests involving Y_1 against all other phenotypes (far left regions of Figure 10a). Figure 10b shows that the CMST have much smaller error rates than the BIC and CIT. These results indicate that the CMST achieve low error rates by forfeiting to make calls when our fitted models are misspecified like at the bottom of the network. Under model misspecification, the BIC and CIT approaches tended to make errors at much higher rates than the CMST.

Inspection of Figure 10c shows high false discovery proportions for all three approaches, specially for the CIT. We point out, however, that the CMST error rate (Figure 10b and e) was always low, so that the high false discovery proportion simply means that many of the few calls made by the CMST are mistakes. The BIC and CIT approaches, on the other hand, showed high error rates and high false discovery proportions. Inspection of Figures 10d-f reveals similar patterns to the Figures 10a-c.

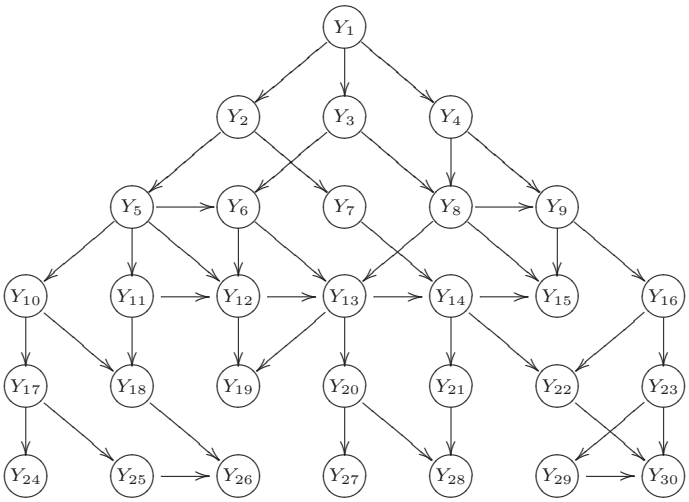

Figure 9. Network composed of 30 phenotype and QTL nodes (only phenotype nodes shown). Of the 435 pairs, 196 are of type M_1 and 239 are of type M_3. The M_1 type represents direct relationships such as $Y_1 \to Y_2$, as well as, indirect relationships such as $Y_1 \to Y_4 \to Y_9 \to Y_{16}$, where there is an indirect path connecting Y_1 to Y_{16}. Examples of M_3 relationships are the pairs (Y_2, Y_3) whose correlation arises due to the common cause Y_1 ($Y_2 \leftarrow Y_1 \to Y_3$), and the pair (Y_{24}, Y_{30}) that have the common ancestors Y_1, Y_2 and Y_5 adding to their correlation. Because the causal flow in the network always go from the nodes with smaller numbers to the ones with big numbers there are no M_2 relations.

Observe that all approaches perform well when one of the fitted models is correctly specified. The BIC and CIT out-perform the CMST and PCT in these situations since they made calls at higher rates and showed equally low error rates. However, under the model misspecification situations, the BIC and CIT showed high error rates, whereas the CMST did not make calls most of the time in these less favorable situations, and showed low error rates.

So far we have considered the results for each pair of nodes separately. Table 33 presents results based on the whole network.

	BIC			CIT			CMST		
Simulation	Calls	ER	FDR	Calls	ER	FDR	Calls	ER	FDR
(1)	0.817	0.392	0.480	0.488	0.346	0.702	0.200	0.039	0.189
(2)	0.710	0.494	0.694	0.669	0.507	0.754	0.057	0.023	0.352

Table 33. Whole network results. The proportion of calls (Calls) is given by the number of significant (interesting) calls (edges in the network) divided by 435. The error rate (ER) represents the number of incorrectly detected edges divided by 435. The FDR is given by the number of incorrect edges in the network divided by the number of detected edges. Results were computed using $\alpha = 0.05$ and BF=10, and represent averages across 100 simulations.

In accordance with the previous results, CMST made calls and mistakes at lower rates then CIT and BIC. Observe, however, that error and false discovery rates were considerably higher in this simulation than in the one presented in the previous section (Tables 31, 32 and 33). The present network represents a much more challenging simulation setting in part because the majority of the pairs of phenotypes represent models of type (f) and (c), with fewer pairs representing a type (d) model (e.g., the pair

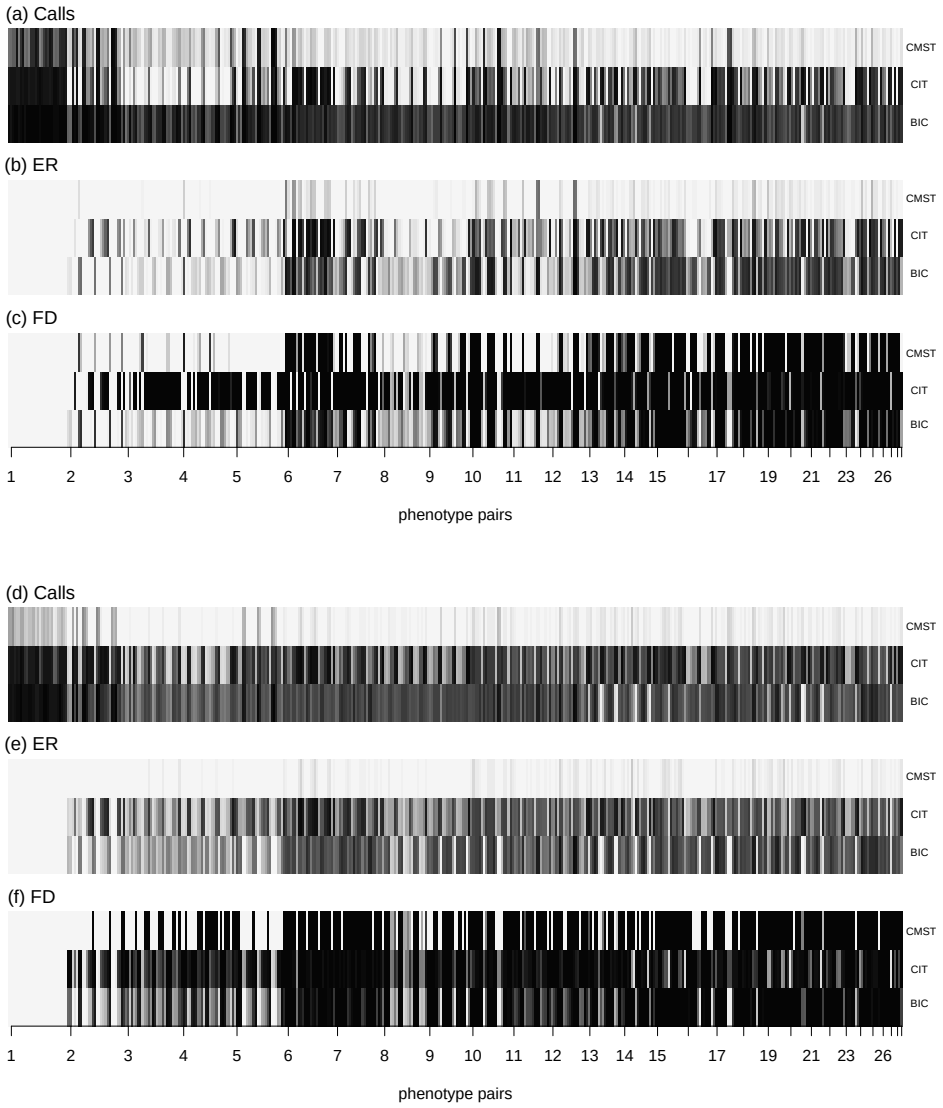

Figure 10. Simulation results. Top and bottom panels summarize the results from simulation settings 1 and 2, respectively. The x-axis represents all 435 phenotype pairs ordered from left to right in the following way: $(1,2)$, $(1,3)$, ..., $(1,30)$, $(2,3)$, $(2,4)$, ..., $(2,30)$, ..., $(29,30)$. Panels (a) and (d) present the average proportion of calls; panels (b) and (e) present the average error rates; and panels (c) and (f) present the average false discovery proportions. The results were based on 100 simulations. The grey scale goes from 0 to 1 corresponding, respectively, to light grey and black.

(Y_1, Y_{30}) is a type (d) model, whereas phenotype pairs (Y_{12}, Y_{15}) and (Y_{16}, Y_{21}), represent type (f) and (c) models, repectively). Furthermore, because noise propagates from top to the bottom of a network, inference for pairs of phenotypes located at the bottom of the network will likely be harder. Finally, QTL mapping patterns will likely be more complex in complex networks.

Simulation study 3: toy networks subjected to measurement error

Measurement error has been pointed out as an important drawback for causal inference in systems genetics (Schadt et al. 2005, Rockman 2008, Li et al. 2010). If the causal trait is poorly measured in comparison with the reactive trait, model selection approaches may incorrectly support the reversed causal direction. The intuition is that the measurement of the reactive trait may be a better measurement of the causal trait than its own measurement in this situation (Rockman 2008). In the Methods section we provide a formal description of the effects of measurement error in terms of violations of the faithfulness assumption (Spirtes et al. 2000). Using the toy network in Figure 11a, we show how the conditional independence relations induced by the measurement error are unfaithful to the causal graph generating the data, but are consistent with the graph with reversed causal relation between phenotypes. However, in situations such as Figure 11b, where the two phenotypes are directly affected by distinct sets of QTLs, the conditional independence relations induced by the measurement error are not consistent with either the causal, reactive or independence models (see Methods for details). Therefore, the measurement error problem should be reduced in these situations.

Here we illustrate the above points with a simulation study assessing and comparing the performance of the CMST methods against the BIC and CIT approaches using data generated from the toy networks of Figure 11. We set the measurement error of Y_2 to zero, i.e, $\delta_2^2 = 0$. The measurement error levels for Y_1 were set to $\delta_1^2 = \{0, 1, 3\}$ and $\sigma_1^2 = \sigma_2^2 = 1$. We adopted two distinct phenotype effects levels (1 and 2), and for each level, we considered, as before, the 8 different simulation settings comprising all combinations between $R^2 = \{0.125, 0.25, 0.5, \text{range}\}$ and $n = \{135, 500\}$. Here R^2 represents the amount of the variability of Y_1 explained by the QTLs, without measurement error. For each one of the 16 simulation settings, we generated 1,000 F_2 crosses with 505 markers unequally spaced across 5 chromosomes of length 100cM, and adopted a QTL mapping threshold of 4.

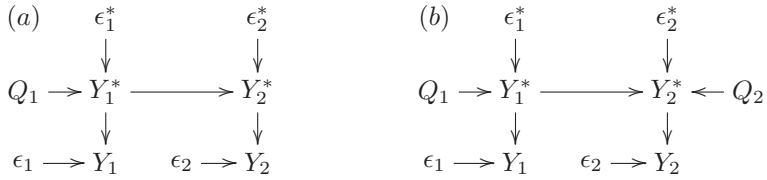

Figure 11. Measurement error model. The random variables Y^* and Y represent, respectively, the true value of a trait and it's measured value (subjected to error). The error terms $\epsilon_k^* \sim N(0, \sigma_k^2)$ and $\epsilon_k \sim N(0, \delta_k^2)$, $k = 1, 2$, represent the residual and measurement errors, respectively. Y_1 and Y_2 represent the causal and reactive traits, respectively. The true causal relationship on this example is M_1.

Tables 34 and 35 present the frequency that the correct (M_1) and reversed (M_2) causal directions were inferred out of 1,000 simulations. With no measurement error, all approaches perform well for data generated under models (a) and (b) in Figure 11, specially at higher R^2 and sample size. Overall, the results are better for the model (b). In the presence of measurement error, all approaches perform better under data simulated from model (b) than from model (a). Nonetheless, the CIT and specially the BIC tend to make mistakes at higher levels than CMST. Under data generated from model (a) the CMST approach was still affected by measurement error, although to a lesser degree than the BIC and CIT. However, under the more favorable conditions of model (b) the CMST approach made very few calls, and avoided M_2 calls to a great extent.

			n=135						n=500					
				(a)			(b)			(a)			(b)	
R^2	ME	Call	BIC	CIT	CMST	BIC	CIT	CMST	BIC	CIT	CMST	BIC	CIT	CMST
0.125	0	M_1	583	104	10	756	76	41	969	787	326	991	759	379
		M_2	97	12	0	89	7	0	25	0	0	3	0	0
	1	M_1	202	12	1	370	9	3	442	86	5	652	108	13
		M_2	200	12	1	197	4	1	414	83	7	121	55	0
	3	M_1	75	2	0	178	1	0	154	1	0	298	2	1
		M_2	211	2	0	247	0	0	731	86	31	442	61	1
0.25	0	M_1	916	517	132	965	446	199	974	918	467	997	912	538
		M_2	66	9	0	17	7	0	24	0	0	1	0	0
	1	M_1	406	81	1	667	76	33	419	5	1	523	17	4
		M_2	382	105	2	160	56	0	392	5	0	43	1	1
	3	M_1	106	7	0	311	8	5	258	0	0	497	0	0
		M_2	613	39	7	426	21	5	580	137	11	136	59	0
0.5	0	M_1	934	906	325	981	853	403	988	943	577	995	944	679
		M_2	50	0	0	4	0	0	7	0	0	1	0	0
	1	M_1	350	55	3	414	184	19	396	0	0	541	0	15
		M_2	364	67	0	127	7	0	402	0	0	19	0	0
	3	M_1	215	3	0	401	25	6	263	0	0	499	0	17
		M_2	549	294	24	212	98	0	482	5	0	19	0	0
range	0	M_1	873	601	188	936	532	260	986	913	478	995	872	535
		M_2	66	5	0	29	0	0	13	0	0	0	0	0
	1	M_1	357	84	7	566	96	46	437	20	3	538	24	13
		M_2	363	116	5	154	46	1	384	37	0	51	18	0
	3	M_1	173	8	0	345	11	11	240	0	1	478	0	10
		M_2	501	124	20	323	62	6	566	102	21	127	59	0

Table 34. Frequency of correct (M_1) and reversed (M_2) causal directions out of 1,000 simulations when the phenotype effect, β_{21}, was set to 1. The CIT and CMST results were based on $\alpha = 0.05$. The BIC results were based on BF=1.

			n=135						n=500					
				(a)			(b)			(a)			(b)	
R^2	ME	Call	BIC	CIT	CMST	BIC	CIT	CMST	BIC	CIT	CMST	BIC	CIT	CMST
0.125	0	M_1	520	124	6	755	122	23	939	813	155	992	834	393
		M_2	156	19	0	147	19	1	59	14	0	7	8	0
	1	M_1	106	28	0	156	12	1	92	6	0	481	7	1
		M_2	453	69	4	419	58	5	884	574	154	474	542	25
	3	M_1	64	1	0	102	1	0	30	0	0	214	0	0
		M_2	435	5	2	410	7	1	945	358	121	753	321	41
0.25	0	M_1	868	565	74	950	609	258	971	928	382	1000	947	620
		M_2	119	21	2	48	16	0	26	6	0	0	1	0
	1	M_1	128	28	0	338	49	1	120	0	0	641	0	0
		M_2	798	431	60	602	349	15	834	519	135	221	345	0
	3	M_1	62	1	0	183	1	0	39	0	0	405	0	0
		M_2	856	119	19	748	118	15	934	702	280	469	590	11
0.5	0	M_1	934	872	220	986	929	602	986	943	566	998	944	773
		M_2	62	8	1	9	3	0	13	0	0	1	0	0
	1	M_1	135	3	0	679	15	22	197	0	0	670	0	22
		M_2	794	674	122	215	360	3	680	219	48	70	5	0
	3	M_1	78	0	0	494	0	15	49	0	0	471	0	3
		M_2	885	661	179	412	525	24	895	637	238	271	197	0
range	0	M_1	861	673	159	941	653	309	976	931	493	997	922	620
		M_2	79	7	0	34	9	0	23	1	0	2	0	0
	1	M_1	227	50	0	526	75	10	297	3	0	608	6	6
		M_2	600	330	36	255	165	3	560	150	31	97	76	1
	3	M_1	97	1	0	300	4	7	118	0	0	450	0	2
		M_2	700	232	61	471	166	8	787	351	88	297	199	5

Table 35. Frequency of correct (M_1) and reversed (M_2) causal directions out of 1,000 simulations when the phenotype effect, β_{21}, was set to 2. The CIT and CMST results were based on $\alpha = 0.05$. The BIC results were based on BF=1.

Linkage analysis

We performed QTL analysis using Haley-Knott regression (Haley and Knott, 1992) with the R/qtl software (Broman, et al. 2003). For the simulation studies we adopted Haldane's mapping function, genotyping error rate of 0.0001, and set the maximum distance between positions at which the genotype probabilities were calculated to 2cM. For the real data examples we adopted Kosambi's mapping function, genotyping error rate of 0.002, and set the maximum distance between positions to 1cM.

Model fitting

The test statistics of the CMST correspond to penalized log-likelihood ratios scaled by their standard errors. In this section we describe the maximum likelihood model fitting approach we used to derive the maximized log-likelihood scores. Given the sets of QTLs identified in the unconditional and conditional analysis, let $\mathbf{X}_1$ and $\mathbf{X}_{2|1}$ represent the respective genetic effects and covariates predictor matrices, and $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_{2|1}$ the respective coefficient vectors determining the additive, dominance and epistatic effects of QTLs as well as the coefficients associated with covariates. Maximum likelihood (or ordinary least squares) estimation gives $\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T\mathbf{y}_1$, $\hat{\sigma}_1^2 = RSS_1/n$, $\hat{\boldsymbol{\beta}}_{2|1} = (\mathbf{X}_{2|1}^T\mathbf{X}_{2|1})^{-1}\mathbf{X}_{2|1}^T\mathbf{y}_2$, $\hat{\sigma}_{2|1}^2 = RSS_{2|1}/n$ and the maximized likelihood of M_1 becomes $(2\pi e)^{-n}\hat{\sigma}_1^{-1}\hat{\sigma}_{2|1}^{-1}$. Similarly the maximized likelihood for M_2 is given by $(2\pi e)^{-n}\hat{\sigma}_2^{-1}\hat{\sigma}_{1|2}^{-1}$.

The SUR model M_3 does not have a closed form maximum likelihood estimator and the ML estimates need to be computed numerically by iteration of the likelihood equations (Srivastrava and Giles 1987)

$$\hat{\beta} = (\mathbf{X}^{-1}(\hat{\boldsymbol{\Sigma}}_\epsilon^{-1} \otimes \mathbf{I}_n)\mathbf{X})^{-1}\mathbf{X}^{-1}(\hat{\boldsymbol{\Sigma}}_\epsilon^{-1} \otimes \mathbf{I}_n)\mathbf{y} \ , \tag{23}$$

$$\hat{\boldsymbol{\Sigma}}_\epsilon = \frac{1}{n}\begin{pmatrix} (\mathbf{y}_1 - X_1\boldsymbol{\beta}_1)'(\mathbf{y}_1 - X_1\boldsymbol{\beta}_1) & (\mathbf{y}_1 - X_1\boldsymbol{\beta}_1)'(\mathbf{y}_2 - X_2\boldsymbol{\beta}_2) \\ (\mathbf{y}_1 - X_1\boldsymbol{\beta}_1)'(\mathbf{y}_2 - X_2\boldsymbol{\beta}_2) & (\mathbf{y}_2 - X_2\boldsymbol{\beta}_2)'(\mathbf{y}_2 - X_2\boldsymbol{\beta}_2) \end{pmatrix} \ , \tag{24}$$

where $\otimes$ represents the standard Kronecker product of two matrices. We point out that we do not use the SUR model to perform QTL mapping and simply use the QTLs detected in the unconditional analysis in the fit of the SUR model. Statistical inference is then based on a scaled version of the penalized log-likelihood ratio statistic described before.

The genetic effects predictor variables in the design matrices (e.g., $\mathbf{X}_1$ and $\mathbf{X}_{2|1}$) were constructed using the dummy variable coding of Haley-Knott regression. For a F_2 cross, the additive and dominance predictor variables associated with a QTL are given by the difference of the homozygote genotype probabilities, and by the heterozygote genotype probabilities, respectively. For a backcross, the additive effect predictor variable is computed as the difference between the homozygote and heterozygote genotype probabilities.

Supplementary figures

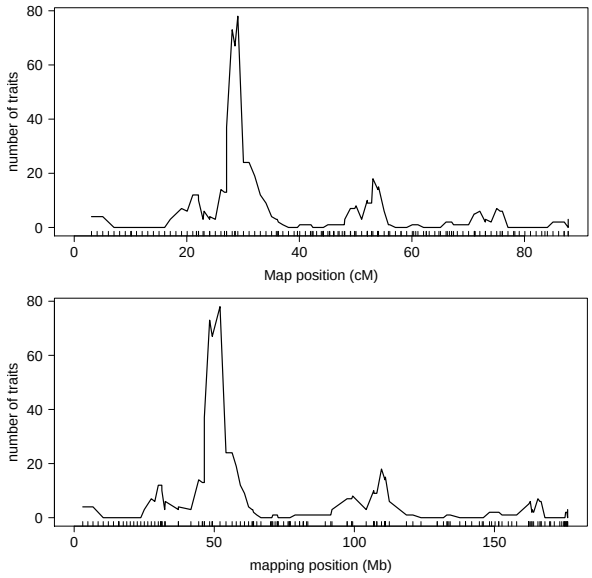

Figure 12. eQTL hotspot located at 29.06cM on chromosome 2. The y-axis represents the number of traits whose peak LOD is located within a 2cM wide window centered at the genomic location. For example, 78 transcripts have their peak within 28.06 and 30.06cM. Top and bottom panels show the results in cM and Mb, respectively.

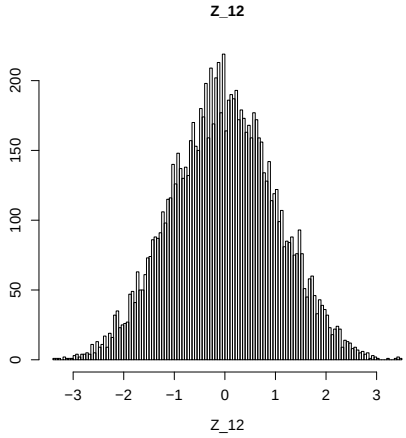

Figure 13. Observed distribution of the test statistic Z_{12} from model (f), in Figure 8, for $R^2 = 0.125$ and n=135. This distribution is fairly symmetric around 0 (maybe a little skewed to the left), so that the right and left tails have about the same probability, and for any significance level we choose, about half of our discoveries will be correct (M_1 calls in the right tail) whereas half will be mistakes (M_2 calls in the left tail). Hence the proportion of false discoveries will be about 0.5 (or a little above it), for any significance level, in this example.

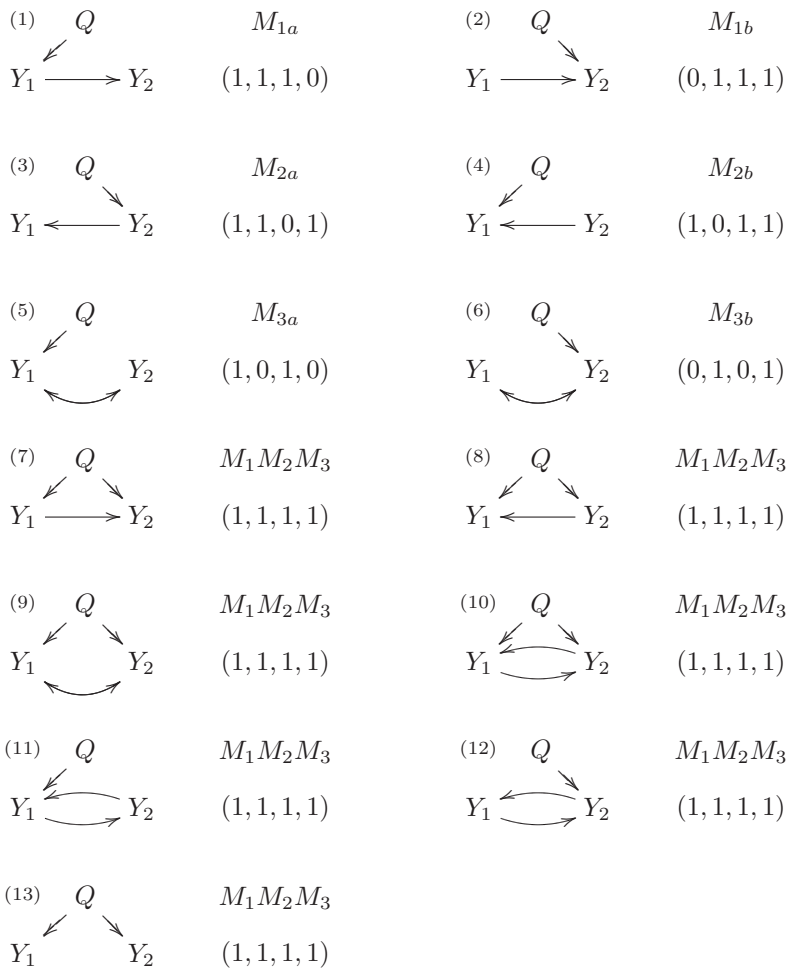

Figure 14. Mapping patterns associated with the graphical models composed of a pair of correlated phenotypes and a single QTL, to which at least one of the phenotypes maps. The mapping patterns are represented by indicator functions, $(\mathbb{1}_{Y_1 \not\perp\!\!\!\perp Q}, \mathbb{1}_{Y_2 \not\perp\!\!\!\perp Q}, \mathbb{1}_{Y_1 \not\perp\!\!\!\perp Q|Y_2}, \mathbb{1}_{Y_2 \not\perp\!\!\!\perp Q|Y_1})$, that describe the conditional independence relations between the phenotypes and QTL. For example, in model (1) we have that Y_1 and Y_2 unconditionally map to Q, hence $\mathbb{1}_{Y_1 \not\perp\!\!\!\perp Q} = 1$ and $\mathbb{1}_{Y_1 \not\perp\!\!\!\perp Q} = 1$; Y_1 still maps to Q even when we condition the mapping analysis on Y_2, and thus, $\mathbb{1}_{Y_1 \not\perp\!\!\!\perp Q|Y_2} = 1$; and Y_2 does not map to Q conditional on Y_1, i.e. $\mathbb{1}_{Y_2 \not\perp\!\!\!\perp Q|Y_1} = 0$. Models (1) and (2) agree with a M_1 causal relationship.

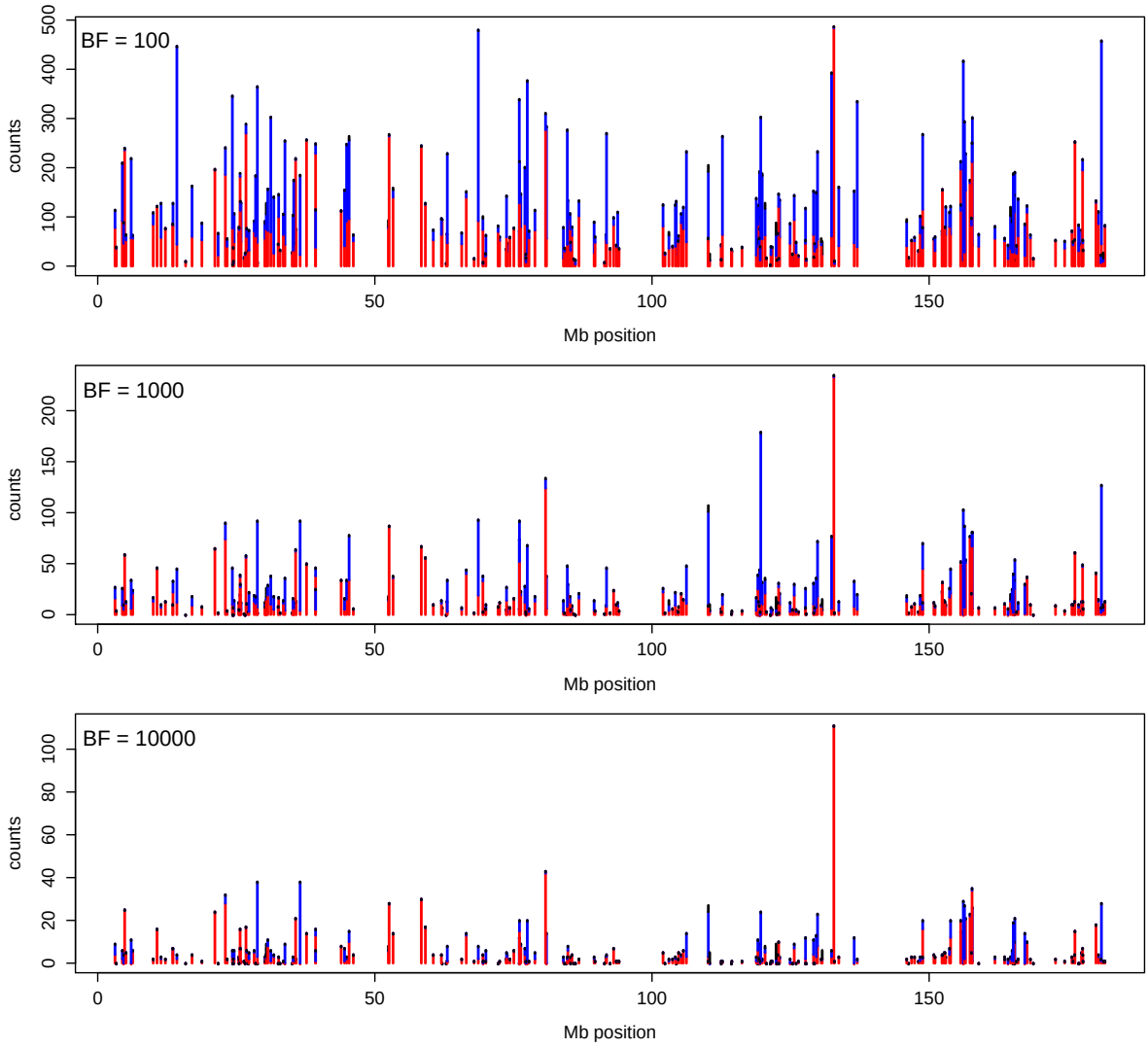

Figure 15. Causal architectures for the BIC approach based on increasing Bayes factors (BF) thresholds. Panels (a), (b) and (c) present the results using approximate BF thresholds of 100, 1000 and 10000. The red, blue and black bars represent, respectively, the numbers of M_1, M_2, M_3 calls, where a call is made when the BF comparing the best and second best models is high than a determined threshold. The no calls are not shown. Although, the number of calls decrease as we increase the BF threshold (note the different y-axis scales), we still see many M_1 and M_2 calls along the entire chromosome. Because the BF is a relative measure of how well two different models fit the data, we can still obtain very high BF when both models have bad fits but one of them is much worse than the other.

Convergence diagnostics for the QTLnet

We assess the convergence of the Markov chain with Geweke (1992) and Raftery and Lewis (1992, 1995) diagnostic measures, and by trace and autocorrelation plots. Inspection of the trace plots suggest good mixing of the Markov chain and the autocorrelation plot corroborates this finding (Figure 16).

Geweke's convergence diagnostic is based on a test for equality of the means of the first and last part of a Markov chain (we used the first 10% and the last 50% parts of the Markov chain after dropping the burnin). When the MCMC samples are drawn from the stationary distribution of the chain, the two means are equal, and Geweke's statistic is asymptotically $N(0, 1)$. Our test statistic (a standard Z-score) was 0.4845, suggesting that convergence was achieved.

The Raftery and Lewis diagnostic is a run length control diagnostic based on the accuracy of estimation of the quantile q. It estimates: (1) the size of burnin (6 for our chain); (2) the required MCMC sample size (the estimate was 6,349 and our chain had size 9,000); and (3) the dependency factor that measures the extent to which autocorrelation inflates the required sample size (the estimated dependency factor was 1.69. Values above 5 suggest problems due to high autocorrelation). These results again suggest that convergence was achieved.

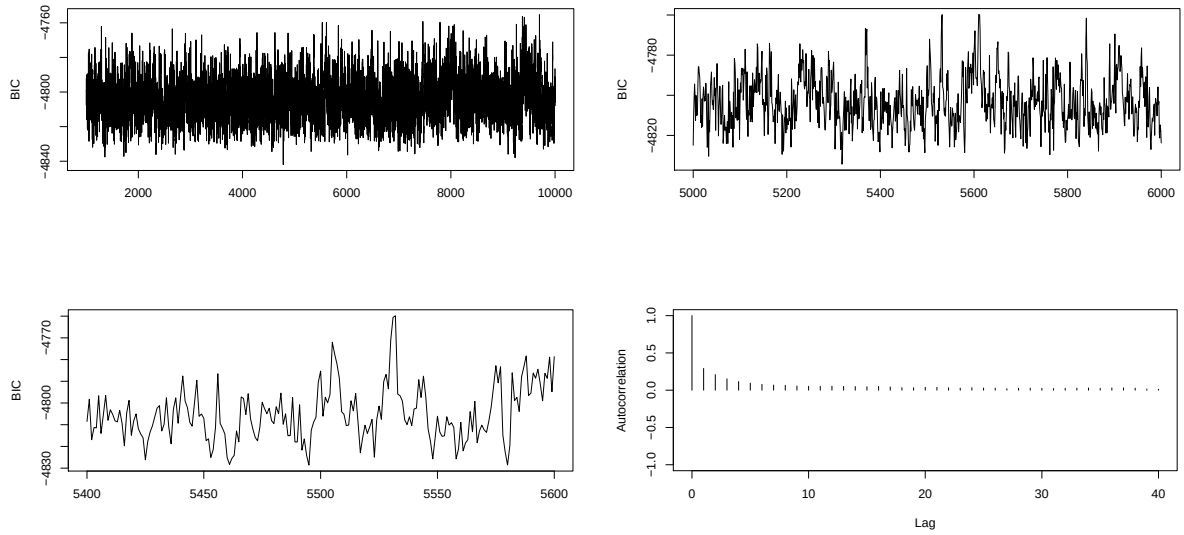

Figure 16. Trace and autocorrelation plots for the Markov chain. The y-axis on the top and left bottom panels shows the BIC for each sampled network. The top left plot displays the Markov chain for the 9,000 sampled structures. The top right and bottom left plots show windows of size 1,000 and 200, respectively. The bottom right plot shows the autocorrelation plot.

QTL selection for the CIT approach

Single QTL approaches (e.g., Millstein et al. 2009, Schadt et al. 2005), require that both phenotypes map to the same QTL. Often times the phenotypes map to nearby but not precisely the same QTL, and the investigator needs to decide which QTL to use as a causal anchor for the causality test. When testing expression traits against clinical traits, Millstein et al. and Schadt et al. suggest using the clinical trait QTL as the anchor.

We adopt a different approach. When the phenotypes map to distinct regions that are less than 5cM apart we determine the QTL position using both phenotypes, jointly, as follows. For each pair of phenotypes (Y_1,Y_2) we perform unconditional mapping analysis for Y_1 and Y_2 and conditional mapping analysis for Y_2 given Y_1. Let LOD_1 represent a LOD score for the mapping analysis of Y_1, and $LOD_{2|1}$ for the mapping analysis of Y_2 given Y_1. Since

$$\log_{10}\left\{\frac{f(y_1, y_2 \mid q)}{f(y_1, y_2)}\right\} = \log_{10}\left\{\frac{f(y_1 \mid q)}{f(y_1)}\right\} + \log_{10}\left\{\frac{f(y_2 \mid y_1, q)}{f(y_2 \mid y_1)}\right\}, \tag{25}$$

we compute the joint LOD score of (Y_1,Y_2) as $LOD_{1,2} = LOD_1 + LOD_{2|1}$ (or equivalently as $LOD_{1,2} = LOD_2 + LOD_{1|2}$). We determine the peak QTL position, λ, using the $LOD_{1,2}$ scores profile and assign the QTL to Y_1 and Y_2 if LOD_1 and LOD_2 are greater than the mapping threshold at the λ position.

Figure 17 illustrates our approach. We simulated data from a model $Q \rightarrow Y_1 \rightarrow Y_2$, with a QTL, Q, at 50cM. The blue and red curves show the (unconditional) LOD profiles of phenotypes Y_1 and Y_2, respectively. The black curve depicts the joint LOD curve, and the peak QTL position λ is given by the black vertical line. Instead of having to perform an arbitrary choice between the QTLs given by the red and blue vertical lines we use the QTL given by the black line.

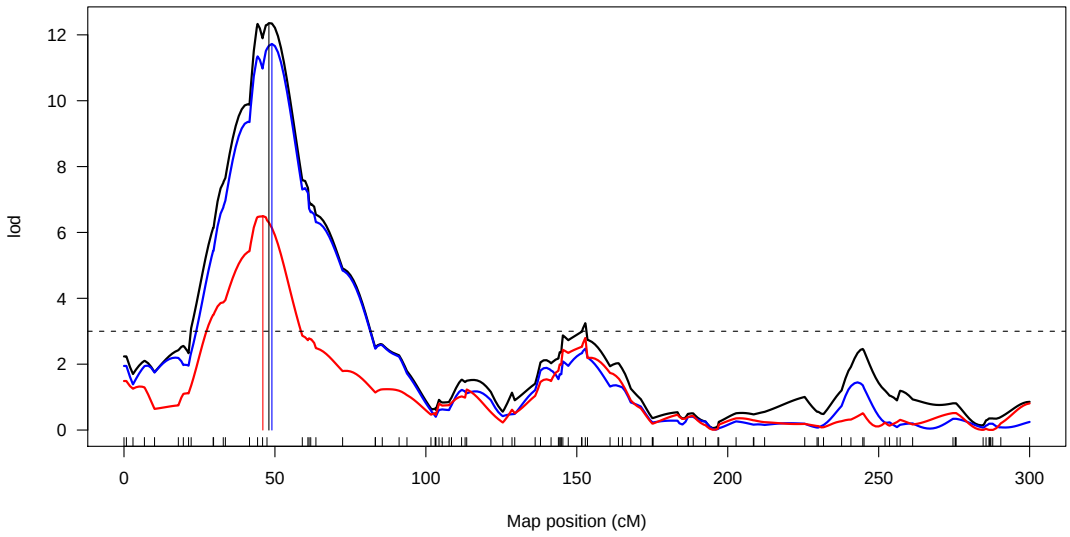

Figure 17. The black, blue and red curves represent the joint and unconditional LOD curve for phenotypes Y_1 and Y_2. The vertical lines represent the respective peak locations. We adopt the QTL represented by the black vertical line. The dashed line shows the QTL mapping threshold.

When both phenotypes map to more than one QTL, we choose the strongest QTL as the causal anchor for the CIT.