Sparse simultaneous signal detection for identifying genetically controlled disease genes

Sihai Dave Zhao*

Department of Statistics,

University of Illinois at Urbana-Champaign,

T. Tony Cai

Department of Statistics,

The Wharton School, University of Pennsylvania,

Thomas P. Cappola

Penn Cardiovascular Institute and Department of Medicine,

Perelman School of Medicine, University of Pennsylvania,

Kenneth B. Margulies

Penn Cardiovascular Institute and Department of Medicine,

Perelman School of Medicine, University of Pennsylvania,

and

Hongzhe Li

Department of Biostatistics and Epidemiology,

Perelman School of Medicine, University of Pennsylvania

Abstract

Genome-wide association studies (GWAS) and differential expression analyses have had limited success in finding genes that cause complex diseases such as heart failure (HF), a leading cause of death in the United States. This paper proposes a new statistical approach that integrates GWAS and expression quantitative trait loci (eQTL) data to identify important HF genes. For such genes, genetic variations that perturb its expression are also likely to influence disease risk. The proposed method thus tests for the presence of simultaneous signals: SNPs that are associated with the gene's expression as well as with disease. An analytic expression for the p-value

is obtained, and the method is shown to be asymptotically adaptively optimal under certain conditions. It also allows the GWAS and eQTL data to be collected from different groups of subjects, enabling investigators to integrate public resources with their own data. Simulation experiments show that it can be more powerful than standard approaches and also robust to linkage disequilibrium between variants. The method is applied to an extensive analysis of HF genomics and identifies several genes with biological evidence for being functionally relevant in the etiology of HF. It is implemented in the R package `ssa`.

Keywords: eQTL; GWAS; Higher criticism; Integrative genomics

1   Introduction

1.1   Genetically regulated disease genes

This paper proposes a new method for identifying genes under genetic control that are likely to be functionally relevant to disease processes. This is of great interest because genome-wide association study (GWAS) results have revealed that the majority of disease-associated single nucleotide polymorphisms (SNPs) lie in non-coding regions of the genome (Hindorff et al. 2009). These SNPs likely regulate the expression of a set of downstream genes (Nicolae et al. 2010), and identifying these downstream genes can lead to better understanding of disease biology as well as potential targets for drug discovery. Expression quantitative trait loci (eQTL) studies measure the association between SNPs and expression levels of both *cis*- and *trans*-genes. This paper proposes identifying genetically regulated disease genes by integrating GWAS and eQTL study results. It is assumed that the GWAS and eQTL data are collected from different sets of subjects, and that only study summary statistics are available.

This work is motivated by an ongoing study of the genomics of human heart failure conducted by Cappola, Margulies, and colleagues at the Myocardial Applied Genomics Network (MAGNet, www.med.upenn.edu/magnet); see Section 4.1 for a detailed description. Heart failure has been shown to be a heritable trait (Lee et al. 2006), but many of the causal genes that mediate the functions of disease variants remain unknown. To address this, the MAGNet consortium collected genotype and gene expression data from left ventricular free wall tissue of a separate set of 313 subjects with and without heart failure. These eQTL data can be used to characterize genetic regulation of gene expression in the human heart, specifically between causal genes and disease variants. The MAGNet eQTL data will be integrated with GWAS results from the Penn Heart Failure Study (PHFS; Ky et al. 2009, Cappola et al. 2011, Ky et al. 2011, 2012), a large multi-center prospective cohort study of heart failure; see Section 4.1 for more details.
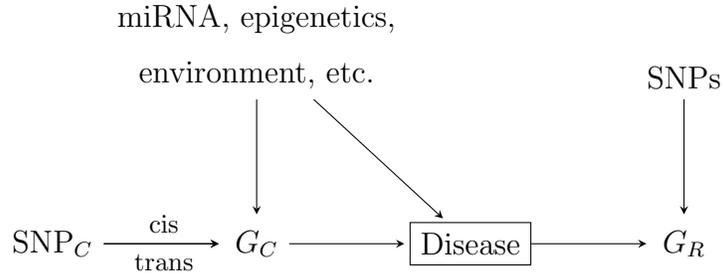
Figure 1: A simple causal model illustrating a problematic setting for differential expression analysis. SNP_C: causal SNP; G_C: causal gene; G_R: reactive gene. SNP_C can be either *cis* or *trans* to G_C. Differential expression analysis cannot distinguish between G_C, which is of interest, and G_R, which is not.

1.2   Causal vs. reactive genes

The standard strategy for identifying disease genes is a differential expression analysis. Genes with different average expression levels between cases and controls, for example, are deemed to be potentially important. However, there are two drawbacks to this approach. First, it does not utilize any SNP information, so the genes that it identifies may not be regulated by disease variants. More importantly, it cannot distinguish between causal genes, whose expression changes cause disease, and reactive genes, whose expression changes are caused by disease. This is depicted in Figure 1. Reactive genes are not of biological interest.

To illustrate this problem, a differential expression analysis was conducted using the MAGNet heart failure data. The expression data were normalized, batch-corrected, and quality-controlled, which left 13,219 transcripts; see Section 4.1 for the specifics of the pre-processing. As Figure 2 illustrates, a standard analysis using `limma` (Smyth 2005, Ritchie et al. 2015) found 8,245 differentially expressed transcripts after controlling the false discovery rate at 5%. This is more than half of all measured transcripts, making downstream biological validation implausible. Furthermore, these results likely contain numerous reactive genes. New methods are needed to further narrow the list of putative causal genes.
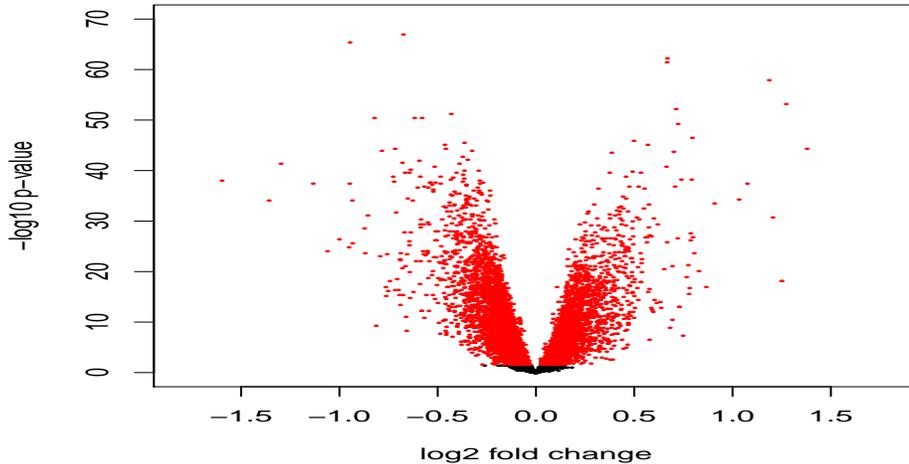
Figure 2: Differentially expressed genes in heart failure. Red points are genes that are significant after controlling the false discovery rate at 5%.

1.3 Simultaneous signal detection

This paper presents a new statistical method for identifying important disease genes by integrating eQTL and GWAS results. As in the heart failure problem described above, the focus is on settings where the GWAS and eQTL studies are conducted on separate sets of subjects and only summary statistics are available. Many existing GWAS-eQTL integration methods are not applicable under these conditions. For example, Xiong et al. (2012) combine differential expression and SNP association test statistics, but differential expression cannot be assessed because the eQTL dataset contains no outcome information. Huang (2014) and Zhao et al. (2014) combine expression levels and genotypes in a mediation analysis framework, but their regression models require that genotype, expression, and outcome measurements all be available from the same subjects. Gamazon et al. (2015) impute gene expression data for subjects in the GWAS dataset using genotype information, learning the imputation models from the eQTL data, but their approach requires access to individual-level genotype data from the GWAS study.

   Instead, motivated by Figure 1, this paper proposes testing each gene for whether there are any SNPs in the genome that are associated both with the gene's expression, using the eQTL data, and with disease, using the GWAS data. This can be done using only summary statistics, which can come from independent samples. Each significant SNP association,

whether with expression or with disease, is termed a "signal", and the method detects SNPs with simultaneous signals. The statistical problem is more formally stated in Section 2.1.

The rationale is that SNPs can be viewed as perturbations of the underlying biological systems, especially the gene regulatory networks underlying complex diseases. Therefore for a disease-causing gene, any genetic variation that perturbs its expression is also likely to influence disease risk. Furthermore, unlike differential expression, the proposed approach is able to differentiate causal genes G_R from reactive genes G_C in Figure 1, because G_R and SNP_C are independent conditional on disease so G_R should not exhibit any simultaneous signals. One caveat is that this proposed approach may fail to identify some genetically regulated disease genes, if those genes are regulated only by SNPs that have no marginal association with the outcome. On the other hand, if there is at least one regulating SNP that does have a marginal association, those genes will still be detectable by the proposed method.

This simultaneous detection approach has been previously proposed in the statistical genetics literature, where it is also known as colocalization testing, as it tests for SNP-expression and SNP-disease associations that colocalize to the same SNPs (He et al. 2013, Ware et al. 2013, Giambartolomei et al. 2014). However, very few existing methods have been studied in a rigorous statistical framework. Most are variations of a two-stage procedure (Chen et al. 2008, Emilsson et al. 2008, Nicolae et al. 2010): the identification stage uses fixed significance thresholds to define indicator variables for whether each SNP is non-null in the GWAS and eQTL studies, and the enrichment stage tests for independence between these indicators using a one-tailed hypergeometric test. Significant positive dependence indicates the presence of simultaneous signals. However, it is unclear how the significance thresholds in the identification stage should be chosen. For example, in their study of whether disease-associated SNPs are more likely to be also associated with gene expression, Nicolae et al. (2010) used three different p-value thresholds of 10^{-4}, 10^{-6}, and 10^{-8} to define expression-associated SNPs. Their qualitative conclusions were found to be consistent across the three choices, but this may not always be the case. Closely related to colocalization is the approach of Zhu et al. (2016), who propose a method for using summary statistics from separate GWAS and eQTL studies to test whether a given gene

6

mediates the effect of a causal variant on the outcome. However, for each gene they consider only the top outcome-associated *cis*-SNP as the putative causal variant. This will fail to detect genes that function by mediating the effects of *trans*-SNPs.

More recently developed genome-wide methods can avoid this problem. Bayesian procedures such as (He et al. 2013, Giambartolomei et al. 2014) typically first define a latent indicator for whether colocalized signals exist. They then model the joint distribution of the observed GWAS and eQTL summary statistics for each SNP, conditional on the latent indicator. Finally, given a prior for the latent indicator, they compute the posterior probability of colocalization. The frequentist GPA method of Chung et al. (2014) posits that the observed test statistics arise from four types of SNPs: those not associated with either the disease or gene expression, those associated with one but not the other, and those associated with both. This gives a four-group mixture model for the test statistics, with the last group corresponding to colocalized signals. After making parametric assumptions on the distribution of the test statistics for non-null SNPs, the model is fit using the EM algorithm and a generalized likelihood ratio test is used to assess whether there are more colocalized SNPs than expected by chance. These methods thus do not require arbitrary thresholds, but do make rather restrictive assumptions. Furthermore, little is known about the theoretical properties of both these and the two-stage methods.

This paper proposes a one-step approach for simultaneous signal detection that does not require any thresholds or priors. A simple closed-form approximation to its *p*-value is derived, making it exceedingly computationally efficient and especially suitable for unbiased genome-wide applications. Under certain conditions the proposed method is asymptotically adaptively optimal for detecting any possible configuration of simultaneous signals. Importantly, it can integrate GWAS and eQTL data from different sources, which allows investigators to leverage public data resources in their own studies. Finally, in addition to being used to detect single disease-associated genes under genetic control, the method is easily extended to detect gene sets that may be related to disease.

Section 2 formalizes the simultaneous detection problem, introduces the proposed method, and describes its properties. Simulation results are discussed in Section 3 and Section 4 conducts an in-depth analysis of the PHFS GWAS and MAGNet eQTL studies, including

a gene-set enrichment analysis. Additional extensions are discussed in Section 5.

2 Methods

2.1 Statistical formulation and previous work

Simultaneous signal detection is conducted one gene at a time. The observed data consist only of summary test statistics U_i, for the SNP-disease association, and V_i, for the SNP-expression association. The U_i and V_i are available from the GWAS and eQTL studies, respectively, which are assumed to have been conducted using independent samples.

For a given gene, define unobserved signal indicators $X_i, Y_i \in \{0, 1\}$ to indicate whether the ith SNP, $i = 1, \ldots, n$, is truly associated with the disease or the gene's expression, respectively, where n is the total number of typed SNPs in the genome. Significant GWAS and eQTL SNPs are usually rare, or sparse, so very few of the X_i and Y_i equal 1. The U_i and V_i are asusmed to follow

$$
\begin{aligned}
U_i \mid X_i = 0 \sim F_0^U, \quad U_i \mid X_i = 1 \sim F_i^U, \quad F_i^U \leq F_0^U \\
V_i \mid Y_i = 0 \sim F_0^V, \quad V_i \mid Y_i = 1 \sim F_i^V, \quad F_i^V \leq F_0^V, \quad U_i \perp\!\!\!\perp V_i \mid X_i, Y_i,
\end{aligned}
\tag{1}
$$

where the F_0^U and F_0^V are null distributions, which may be unknown, and the F_i^U and F_i^V are unknown alternative distributions. The test statistics are assumed to be stochastically larger under the alternatives, which is reasonable for two-sided tests. For example, one possibility is to take $U_i = |Z_i^U|$ and $V_i = |Z_i^V|$, where Z_i^U and Z_i^V are Z-scores obtained from GWAS and eQTL studies, respectively, using linear or logistic regressions. Finally, for all i the U_i and V_i are independent conditional on X_i and Y_i because the GWAS and eQTL data arise from separate subjects. The setting where the two datasets include overlapping subjects is left for future work.

Under model 1, let ϵ denote the proportion of simultaneous signals, i.e., SNPs with $X_i = Y_i = 1$. The simultaneous signal detection problem is to test

$$
H_0 : \epsilon = 0 \quad vs. \quad H_A : \epsilon > 0
\tag{2}
$$

using the observed $(U_i, V_i), i = 1, \ldots, n$. Rejecting H_0 indicates that the expression of the

gene being tested is regulated by SNPs which are also associated with disease, which by Figure 1 suggests that the gene is likely to be functionally relevant.

In the statistical literature there has been a great deal of recent work on signal detection, such as the normal mixture detection problem (Ingster 1997, 1998, Donoho & Jin 2004): given Z-scores $Z_i, i = 1, \ldots, n$, test

$$H_0 : Z_i \sim N(0,1), \quad vs. \quad H_A : Z_i \sim (1-\epsilon)N(0,1) + \epsilon N(\mu,1), \epsilon > 0. \tag{3}$$

The proposed simultaneous detection problem (2) is a generalization of (3). In most large-scale genomics studies the proportion of signal ϵ is small and the signal strength μ is not very large. This has been termed the "rare and weak" regime by Donoho & Jin (2004), and is of considerable interest. It has been shown that in this regime there exist tests for (3) that are asymptotically adaptively optimal: they do not require knowledge of the unknown parameters, yet still asymptotically perform as well as the likelihood ratio test (Ingster 2002a,b, Donoho & Jin 2004). In particular, Donoho & Jin (2004) showed that this is true of the higher criticism statistic of Tukey:

$$HC = \sup_z n^{1/2} \frac{|\hat{F}(z) - \Phi(z)|}{[\Phi(z)\{1 - \Phi(z)\}]^{1/2}}, \tag{4}$$

where $\hat{F}(z)$ is the empirical distribution function of the Z_i and $\Phi(z)$ is the null distribution of the Z_i. Jager & Wellner (2007) showed that other goodness-of-fit tests can have similar properties.

Recent research has focused on finding adaptively optimal procedures while relaxing the distributional assumptions of (3). Cai et al. (2011) considered heteroscedastic normal mixtures and Cai & Wu (2014) studied mixtures of arbitrary distributions. Hall & Jin (2008, 2010) allowed for the Z_i to have certain dependency structures. Arias-Castro et al. (2011) and Mukherjee et al. (2015) considered detecting non-zero regression parameters in linear and logistic regression, respectively. However, most work in this area has centered on detection of signal in a single sequence of test statistics, while here the goal is to detect simultaneous signals using two sequences of test statistics.

2.2 Proposed method

To test whether V_i for a given gene and U_i share any simultaneous signals, recall from model (1) that the U_i and V_i are assumed to be stochastically larger when the signal indicators X_i and Y_i equal 1, respectively. Thus if SNP i is truly simultaneously associated with both the disease and the gene's expression, then both U_i and V_i should be large, so it is reasonable to define the statistic $T_i = U_i \wedge V_i$. Intuitively, the simultaneous signal detection (2) null should be rejected if at least one SNP has an observed large value of T_i, so the proposed test statistic is

$$M = \max_{i=1,\ldots,n} T_i. \tag{5}$$

A large value of M would imply that the gene is functionally relevant for disease. One caveat is that to achieve the best power, U_i and V_i should be on roughly the same scale, meaning that the null variances of U_i and V_i should be comparable.

This formulation reduces (2) to a signal detection problem for a single sequence of test statistics T_i. While this type of problem has been thoroughly studied, as mentioned previously, existing tests such as HC (4) cannot be used because they require knowledge of the null distribution of T_i. Here T_i has a composite null: when SNP i is not a simultaneous signal it can still be non-null in either the GWAS or the eQTL study, and the null of T_i will depend on one of the unknown alternative distributions F_i^U or F_i^V. In some cases it may be possible to estimate the null of T_i, but estimation is usually difficult, complicates the procedure, and may have poor asymptotic properties.

Obtaining accurate p-values for M is difficult, once again because of the composite null of T_i. A permutation procedure is instead proposed: the SNP labels of the U_i can be randomly permuted while fixing the labels of the V_i, which removes simultaneous associations. Then for each permutation M can be recalculated and the permutation null distribution can be used to calculate a p-value. In fact, this p-value can be obtained without any actual permutation. By definition it is the proportion of permutations in which the recalculated max statistic is at least as large as the observed M. This is equal to the probability that at least one of the U_i with magnitude at least M is permuted such that it overlaps with one of the V_i with magnitude at least M. Then if there are k SNPs such that $U_i \geq M$ and

m SNPs such that $V_i \geq M$, the permutation p-value equals

$$1 - \binom{m}{0}\binom{n-m}{k}\binom{n}{k}^{-1}. \tag{6}$$

Thus the proposed procedure does not require separate identification and enrichment steps, is tuning parameter- and prior-free, and is extremely simple to compute. It is available in the R package `ssa` and is easily scalable to large GWAS studies, calculating p-values in seconds even with tens of millions of SNPs.

One caveat is that the permutation null does not exactly reproduce the true simultaneous detection null. In some of the permutations, some non-null U_i will be permuted to overlap with non-null V_i. However, the proposed M tends to be larger under permutation than under the null, which gives conservative inference. To be more precise, let g be any permutation of the indices $i = 1, \ldots, n$, where $g(i)$ is the index to which i is mapped and $g^{-1}(i)$ is the index which is mapped to i. Using this notation, the proposed permutation test calculates $T_i^g = U_{g^{-1}(i)} \wedge V_i$ and $M^g = \max_i T_i^g$ for each g. Let $S_i^U = 1 - F_i^U$ and $S_0^U = 1 - F_0^U$ and define S_i^V and S_0^V similarly.

Theorem 1 *Let M^0 be the max statistic (5) calculated under H_0 of (2), the true simultaneous detection null. Under model (1), if $U_i \perp\!\!\!\perp U_{i'}$ and $V_i \perp\!\!\!\perp V_{i'}$ for $i \neq i'$, then for any permutation g, $\mathrm{P}(M^0 \geq t) \leq \mathrm{P}(M^g \geq t)$.*

Assuming that both the U_i and V_i are independent across i is reasonable if the SNPs come from different linkage disequilibrium blocks. This can be achieved using linkage disequilibrium pruning, a common pre-processing step in statistical genetics. On the other hand, Section 2.4 argues that proposed permutation procedure is actually fairly robust to correlation arising from linkage disequilibrium, in that it still maintains type I error. This is verified in simulations in Section 3.2.

Theorem 1 indicates that the permutation p-value (6) is conservative, but in typical genomics applications it will usually not be overly conservative. Suppose a proportion π_U of the GWAS signals and π_V of the eQTL signals are non-null. Then random permutation will give a proportion $\epsilon = \pi_U \pi_V$ of simultaneous signals. But since GWAS and eQTL signals are typically rare, π_U and π_V are usually very small so ϵ is usually nearly zero, recovering the true simultaneous detection null.

11

2.3 Asymptotic justification

This section analyzes the asymptotic testing performance of M, which reveals that under certain conditions it has the same adaptive optimality properties as the higher criticism statistic (4). Another consequence of this analysis is a quantitative characterization of how many simultaneous signals must exist, and how strong they must be, before detection is possible for a given total number of SNPs; see Figure 3. This can be especially useful for study design.

These theoretical results are derived for the following special case of model (1). Let m_U and m_V be the sample sizes of the GWAS and eQTL studies, respectively, and suppose to each SNP i there correspond Z-scores $Z_i^U \sim N(m_U^{1/2}\mu_i, 1)$ for the SNP-disease association and $Z_i^V \sim N(m_V^{1/2}\nu_i, 1)$ for the SNP-expression association. Non-significant associations have μ_i and ν_i are equal to zero, and here the significant associations will be modeled as following $m_U^{1/2}\mu_i \sim (1-a)N(-\mu, \sigma_0^2) + aN(\mu, \sigma_0^2)$ and $\mu_V^{1/2}\nu_i \sim (1-b)N(-\nu, \tau_0^2) + bN(\nu, \tau_0^2)$ for some mixture proportions a and b. Letting $U_i = |Z_i^U|$ and $V_i = |Z_i^V|$, the null and alternative distributions from model (1) become

$$F_0^U, F_0^V \sim |N(0,1)|, \quad F_i^U \sim |N(\mu, \sigma^2)|, \quad F_i^V \sim |N(\nu, \tau^2)|, \tag{7}$$

where $\sigma^2 = \sigma_0^2 + 1$ and $\tau^2 = \tau_0^2 + 1$.

The asymptotics in this section apply to the total the number of SNPs n, such that n is assumed to tend toward infinity. This is meaningful because in practice n is typically very large. In this setting, if ϵ, μ, and ν were fixed with n, any reasonable test would asymptotically perfectly separate H_0 and H_A. Instead, a more meaningful comparison between tests can be obtained by allowing the parameters to vary with n such that H_A approaches H_0. Thus similar to Donoho & Jin (2004) and Cai et al. (2011), let

$$\mu = \mu_n = r_\mu(\log n)^{1/2}, \quad \nu = \nu_n = r_\nu(\log n)^{1/2}, \quad \epsilon = \epsilon_n = n^{-\beta}, \beta \in (1/2, 1], \tag{8}$$

where r_μ, and r_ν are positive constants and the subscripts n make the dependence on the total number of SNPs explicit. This calibration of ϵ_n formalizes the notion, described in Section 2.1, that simultaneous signals tend to be sparse in GWAS and eQTL studies. This parameterization relates the asymptotics in n to the usual asymptotics in sample sizes m_U

and m_V, since above it was assumed that the average of the $m_U^{1/2}\mu_i$ and the $m_V^{1/2}\nu_i$ behave like μ_n and ν_n, respectively.

To study the asymptotic properties of using the proposed statistic M (5) to detect simultaneous signals, define the following asymptotic test:

$$\phi_M(T_1,\ldots,T_n) = I[M \geq \{(1+\delta)\log n\}^{1/2}], \tag{9}$$

where $\delta > 0$. The critical function ϕ_M is a function of the observed data that gives the probability of rejecting the null. To motivate (9), define p_{1n} to be the proportion of SNPs associated with neither disease or expression ($X_i = 0$ and $Y_i = 0$), p_{2n} to be the proportion associated with disease but not expression ($X_i = 1$ and $Y_i = 0$), and p_{3n} is the proportion associated with expression but not disease ($X_i = 0$ and $Y_i = 1$). The X_i and Y_i are the signal indicators from model (1). When there are no simultaneous signals, $p_{1n} + p_{2n} + p_{3n} = 1$. Since GWAS and eQTL signals are sparse, calibrate $p_{2n}, p_{3n} \leq n^{-1/2}$. Then M would roughly behave like

$$\max[\{\log(np_{1n})\}^{1/2}, \{2\log(np_{2n})\}^{1/2}, \{2\log(np_{3n})^{1/2}\}] \leq (\log n)^{1/2},$$

since intuitively the maximum of p variables behaves like $(2\log p)^{1/2}$ when they are distributed like $|N(0,1)|$, and like $(\log n)^{1/2}$ when they are distributed like $|N(0,1)| \wedge |N(0,1)|$. Thus $(\log n)^{1/2}$ is the appropriate critical value for M.

The performance of a test with critical function ϕ can be measured using the sum of its type I and type II errors: $S_{H_0,H_A}(\phi) = E_{H_0}\phi + E_{H_A}(1 - \phi)$, which depends on the test statistic and on the true values of the parameters under H_0 and H_A of (2). The detection boundary separates the region of the parameter space where $S_{H_0,H_A}(\phi) \to 1$ for all tests ϕ, called the undetectable region, from the region where there exists a test ϕ with $S_{H_0,H_A}(\phi) \to 0$.

Under this framework, under certain conditions test (9) is asymptotically adaptively optimal among all possible tests based on T_i. In other words, it can attain zero error everywhere in the interior of the detectable region.

Theorem 2 *Assume that $p_{2n} = n^{-\alpha_2}$ and $p_{3n} = n^{-\alpha_3}$ with $\alpha_2, \alpha_3 \geq 1/2$. Under model (1), the distributional assumptions (7), and the calibrations (8), when $r_\mu \vee r_\nu \geq 1$, the simulta-*
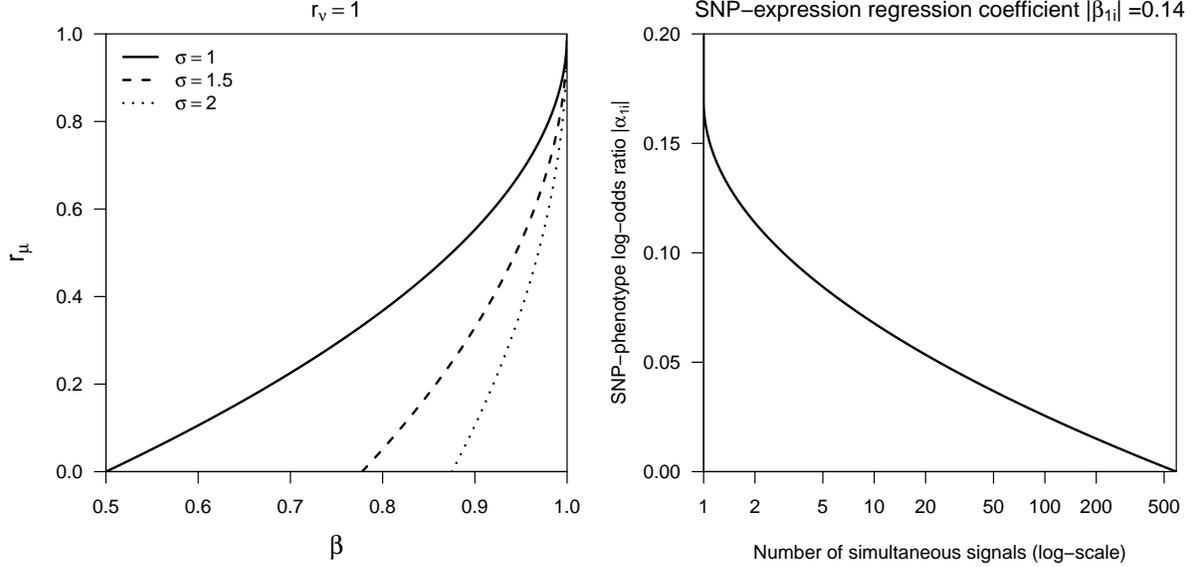
Figure 3: Simultaneous signal detection boundaries (10). The detectable regions lie above the lines and the undetectable regions lie below. The right panel plots the detection boundary in terms of a more interpretable set of parameters; see text for details.

neous detection boundary for any test based on T_i is characterized by

$$
\begin{aligned}
\rho(\beta, r_\mu, \sigma, r_\nu, \tau) &= 1 - \beta = 0, & 1 \le r_\mu \wedge r_\nu, \\
\rho(\beta, r_\mu, \sigma, r_\nu, \tau) &= 1 - \beta - \tfrac{1}{2}\left(\tfrac{1-r_\mu}{\sigma}\right)^2 = 0, & r_\mu < 1 \le r_\nu, \\
\rho(\beta, r_\mu, \sigma, r_\nu, \tau) &= 1 - \beta - \tfrac{1}{2}\left(\tfrac{1-r_\nu}{\tau}\right)^2 = 0, & r_\nu < 1 \le r_\mu.
\end{aligned}
\tag{10}
$$

When $\rho(\beta, r_\mu, \sigma, r_\nu, \tau) > 0$, $S_{H_0,H_A}(\phi_M) \to 0$ for ϕ_M (9) based on M. Otherwise when $\rho(\beta, r_\mu, \sigma, r_\nu, \tau) < 0$, $S_{H_0,H_A}(\phi) \to 1$ for any critical function ϕ.

The detection boundary is plotted in the left panel of Figure 3. The condition that $r_\mu \vee r_\nu \ge 1$ assumes either the GWAS signals or the eQTL signals, or both, are sufficiently large. This is usually satisfied because SNP-expression associations can be quite strong, especially between a gene and its *cis*-SNPs.

The motivating heart failure data, described in detail in Section 4.1, can be used to give a more interpretable illustration of the detection boundary (10). Let Y be heart failure status, X be the expression of a given gene, and S_i be the genotype of SNP i under additive coding. Xie et al. (2011) and Bentkus et al. (2007) gave formulas for calculating μ_n, σ, ν_n,

14

and τ in terms of the parameters of the models

$$\text{logit P}(Y \mid S_i) = \alpha_{0i} + \alpha_{1i}S_i, \quad X = \beta_{0i} + \beta_{1i}S_i + N(0, s^2),$$

the disease prevalence, the GWAS case-control sampling fraction, the total number of SNPs, and the study sample sizes. Inverting their formulas and using calibrations (8) leads to an expression for the detection boundary in terms of the regression model parameters.

The right panel of Figure 3 plots the boundary for parameter values estimated from the heart failure data analyzed in Section 4. For example, for a certain SNP and gene in the MAGNet eQTL data, $|\hat{\beta}_{1i}| = 0.14$ and $\hat{s} = 0.21$, and the figure shows that for this gene the proposed statistic can detect roughly 3 or more simultaneous signals if $\alpha_{1i} \geq 0.10$. In fact, $\hat{\alpha}_{1i} = 0.25$ for that SNP in the PHFS GWAS data. These values can be shown to satisfy the condition $r_\mu \vee r_\nu \geq 1$ of Theorem 2, which suggests that the proposed M may be nearly optimal for simultaneous signal detection in this dataset. Figure 3 is also useful for designing simultaneous signal detection studies.

2.4 Linkage disequilibrium

So far it has been assumed that the U_i are independent across i, as are the V_i. However, this assumption is frequently violated due to linkage disequilibrium between adjacent SNPs. On the other hand, the following arguments suggest that the proposed permutation p-value (6) is fairly robust to linkage disequilibrium.

Consider the set of SNPs with non-null U_i, V_i, and $U_{g^{-1}(i)}$ for some permutation g. To show the conservativeness of the permutation procedure in Theorem 1, the proof requires the distribution of the maximum of the T_i over these SNPs to be invariant to permutation. When the SNPs are independent this is clearly true. Under linkage disequilibrium, SNPs are only "weakly dependent", in the sense that the proportion of very highly correlated SNPs is low. For example, Dawson et al. (2002) showed that the average r^2 between SNPs separated by more than 25kb is already below 0.3. There is recent work showing that in certain cases, the maximum of a sequence of this type of weakly dependent variables has the same asymptotic distribution as if the variables were independent (Cai et al. 2013).

Another requirement of the proof of Theorem 1 is independence between the SNPs with non-null U_i or V_i when no simultaneous signals exist. This is reasonable because the

disease-associated SNPs and the expression-associated SNPs are likely not close together in the genome. The remainder of the proof should hold if the maximum of the T_i over these SNPs is independent of the maximum of the T_i over all SNPs with U_i and V_i both null. This also seems plausible because the latter is the maximum of a very large set of null SNPs, most of which will be physically distant from the non-null SNPs. Thus the permutation p-values may remain conservative under suitable conditions on the correlation structure. This is in fact demonstrated in simulations in Section 3.2.

3 Simulations

3.1 Independent test statistics

Test statistics U_i and V_i were independently generated for $n = 100,000$ SNPs. Null U_i and V_i were generated from $|N(0,1)|$, non-null $U_i \sim |N(\mu_i, 1)|$, and non-null $V_i \sim |N(\nu_i, 1)|$. Under H_A of (2) the non-null SNPs were positioned to give $n\epsilon$ simultaneous signals. The various μ_i and ν_i were generated randomly from $N(a, 1)$ and $N(b, 1)$, respectively, for different values of a and b. Different simulation settings considered different numbers of non-null signals in the U_i and V_i, different numbers of simultaneous signals, and different a and b. In each setting the positions of all non-null U_i and V_i, as well as the values of μ_i and ν_i, were generated once and then fixed across replications.

The proposed max statistic M (5) was used to test for simultaneous signals. The permutation procedure (6) was implemented, and to assess its conservativeness the true null distribution of M was also used to calculate p-values. The proposed method was compared to the usual two-stage procedure and the GPA method of Chung et al. (2014), described in Section 1.3. There is no standard for what threshold to use in the identification step of the two-stage method, so 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, and 10^{-6} were all implemented. The method of He et al. (2013) was also considered for comparison, but its p-value is calculated by fixing the eQTL profiles and randomly swapping the cases and controls in the GWAS dataset. Permuting the case-control status in the GWAS data does not reflect the true null (2) of no simultaneous signals, which can lead to inflated type I error.

Table 1 reports the average type I errors, over 1,000 simulations, of the various methods,

which were conducted at a nominal $\alpha = 0.05$. To put the values of a and b into context, recall that the μ_i and ν_i were generated from $N(a, 1)$ and $N(b, 1)$; Z-scores equal to 2.5, 3, 3.5, and 4 correspond to p-values of 1.2×10^{-2}, 3×10^{-3}, 5×10^{-5}, and 6×10^{-5}, respectively. All methods controlled the type I error at the nominal level, though GPA and the two-stage method with stringent thresholds were both very conservative. The permutation p-value (6) was indeed conservative compared to the true p-value of M, but not exceedingly so.

Table 2 reports the average powers corresponding to the type I errors from Table 1. In general, increasing the number of non-null signals in each sequence reduced the power of all methods, while increasing the number of simultaneous signals and/or the signal strengths increased power. Among the various methods, the proposed procedure had the most power, with the true p-value giving slightly more power than the permutation p-value. The performance of the two-stage approach heavily depended on the p-value threshold. Though it performed well at some thresholds, for example 10^{-3}, the optimal threshold is unknown in practice. GPA had very low power in about half of the simulations. This is because it requires estimating the parameters of a four-component mixture model, as described in Section 1.3. Settings with few non-null or simultaneous signals correspond to scenarios with few observations from one or more of the mixture components, making parameter estimation difficult.

3.2 Linkage disequilibrium

To study the effect of linkage disequilibrium on the performance of the simultaneous signal detection methods, GWAS and eQTL data were simulated using real genotype data from the MAGNet heart failure study analyzed in Section 4. These data consist of 347,019 SNPs under additive coding for 136 controls and 177 cases; see Section 4.1 for more details.

To simulate GWAS data, genotypes were generated by randomly sampling 136 subjects with replacement from the MAGNet control group. Let $\boldsymbol{S}^G$ denote the resulting $136 \times 347,019$ matrix. To simulate eQTL data, genotypes were generated by randomly sampling 177 subjects with replacement from the MAGNet cases, giving a $136 \times 347,019$ matrix $\boldsymbol{S}^E$. Outcomes were simulated according to the linear models $\boldsymbol{Y}^G = \boldsymbol{S}^G \boldsymbol{\alpha} + \boldsymbol{\epsilon}^G$ and $\boldsymbol{Y}^E = \boldsymbol{S}^E \boldsymbol{\beta} + \boldsymbol{\epsilon}^E$, where $\boldsymbol{\epsilon}^G$ and $\boldsymbol{\epsilon}^E$ were independent vectors of $N(0, 0.2^2)$ random errors.

Under H_0 the non-zero components of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ were placed such that every SNP was associated only with $\boldsymbol{Y}^G$ or only with $\boldsymbol{Y}^E$. Under H_A the non-zero components were placed to give $n\epsilon$ SNPs simultaneously associated with both GWAS and eQTL outcomes. All but 10 components of the coefficient vector $\boldsymbol{\alpha}$ were set to zero; the non-zero ones were simulated by first drawing values from $N(a, 0.1^2)$ and then randomly multiplying the value by -1 with probability 0.5. The $\boldsymbol{\beta}$ was generated similarly except that the non-zero components were drawn from $N(b, 0.1^2)$. Different simulation settings considered different values of a and b. In each setting the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ were generated once and then fixed across all replications.

It was suggested in Section 2.4 that the proposed permutation procedure should remain valid under linkage disequilibrium as long as non-null SNPs are independent. To simulate this condition, under H_0 the non-null SNPs were randomly scattered across the genome; after being placed, their positions were kept fixed in all replications. Additional simulations that study violations of this condition, as well as consider different numbers of non-zero components of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, are reported in the Supplementary Material.

Simultaneous signal detection methods were used to test H_0 against H_A. These were applied to the GWAS and eQTL marginal test statistics U_i and V_i, obtained by taking the absolute values of the Z-statistics of the marginal regressions of $\boldsymbol{Y}^G$ on the ith column of $\boldsymbol{S}^G$ and $\boldsymbol{Y}^E$ on the ith column of $\boldsymbol{S}^E$, respectively. Missing genotypes were imputed using the average minor allele dosage for the corresponding SNP and then fast marginal regressions were performed using large matrix multiplication (Sikorska et al. 2013).

Table 3 reports the average type I errors and powers over 1,000 simulations. The true p-value of the max test statistic M (5) cannot be calculated here because the true SNP correlation structure is unknown. The proposed permutation p-value (6) was indeed robust to linkage disequilibrium, giving conservative p-values under these simulations and all simulations in the Supplementary Material. This supports the arguments in Section 2.4. In contrast, no other method was able to control the type I error except the two-stage approach with restrictive p-value thresholds. The proposed procedure also had the most power among all methods.

4 Genomics of heart failure

4.1 Description of the heart failure data

Heart failure occurs when the heart is unable to pump enough blood to supply the body's demands and affects roughly 5.8 million Americans (Roger 2013). In the past two decades, modern high-throughput biology has transformed our understanding of the genetic and genomic basis of heart failure, but the translation of these findings into new treatments has not proceeded as quickly as hoped (Mudd & Kass 2008, Creemers et al. 2011). As described in Section 1.2, simple differential expression analyses sometimes identify more than half of all measured genes. These results are difficult to interpret and validate, and furthermore many of the identified genes may be reactive rather than causal. New analyses are needed to narrow the list of findings by prioritizing the ones that are more likely to be functional.

To this end, the proposed simultaneous signal detection procedure was applied to identify genes involved in the biological mechanisms of heart failure. GWAS results were obtained from the Penn Heart Failure Study (PHFS), a large prospective study of patients recruited from the University of Pennsylvania, Case Western Reserve University, and the University of Wisconsin between 2003 and 2012. Study details have been reported elsewhere (Ky et al. 2009, Cappola et al. 2011, Ky et al. 2011, 2012). Genotype data were collected from 1,586 controls and 2,027 cases using the Illumina OmniExpress Plus array.

Heart failure eQTL data were obtained from the MAGNet eQTL study. Left ventricular free-wall tissue was collected from hearts of 177 patients with advanced heart failure who were undergoing transplantation and from 136 donor hearts without heart failure. Genotype data were collected using using Affymetrix Genome-Wide SNP Array 6.0 and only markers in Hardy-Weinberg equilibrium with minor allele frequencies above 15% were considered. Gene expression data were collected using Affymetrix GeneChip ST1.1 arrays, normalized using RMA (Irizarry et al. 2003), and batch-corrected using ComBat (Johnson et al. 2007). Probesets expression levels were averaged at the transcript level and only those with high expression, specifically with RMA values at least 4.8 in all samples, were considered, leaving 13,219 transcripts.

4.2 Results

GWAS summary statistics were calculated controlling for age, gender, and the first two principal components of the genotypes. SNPs were imputed using 1000 Genomes Project data (1000 Genomes Project Consortium 2010). Summary statistics for the MAGNet eQTL data were conducted controlling for age and gender, using the fast marginal regression algorithm of Sikorska et al. (2013). Only data from normal heart tissue were used; see Section 4.3 for a detailed discussion about choosing the appropriate tissue for this analysis. No population stratification adjustment was performed, as all subjects were Caucasian. All analyses were performed with genotypes under additive coding.

Simultaneous signal detection tests were applied to each of the 13,219 transcripts to test for colocalization between the GWAS and eQTL summary statistics. Only the 347,019 SNPs genotyped or imputed in both the GWAS and the eQTL study were used. For each transcript, the GWAS and eQTL test statistics were converted to Z-scores, and the U_i and V_i were taken to be their absolute values. The proposed method, along with the two-stage approach and GPA (Chung et al. 2014), were applied to the (U_i, V_i).

Table 4 reports the results of the proposed method and contains genes with permutation p-values (6) less than 10^{-3}. Almost all of the genes were highly differentially expressed between normal and failing heart tissue, which is significant because data from failing tissue were never used in this analysis. Furthermore, the biological validity of these genes enjoys significant literature support. In general they are involved in three classes of biological processes: heart muscle strength and contraction, angiogenesis, and inflammation.

Multiple testing correction for the simultaneous signal detection p-values is difficult because of the unknown correlation structure between the genes and the fact that the PHFS GWAS dataset was used to test for each of the genes. A Bonferroni correction for 13,219 tests would thus be too conservative. Instead, the eigenvalue ratio function of Galwey (2009) was used to estimate an effective number of tests:

$$M_{eff} = \frac{(\sum_l \lambda_l^{1/2})^2}{\sum_l \lambda_l},$$

where λ_l is the lth-largest eigenvalue of the correlation matrix of the gene expression values. Bonferroni correction was then done using M_{eff} instead of the total number of tests.

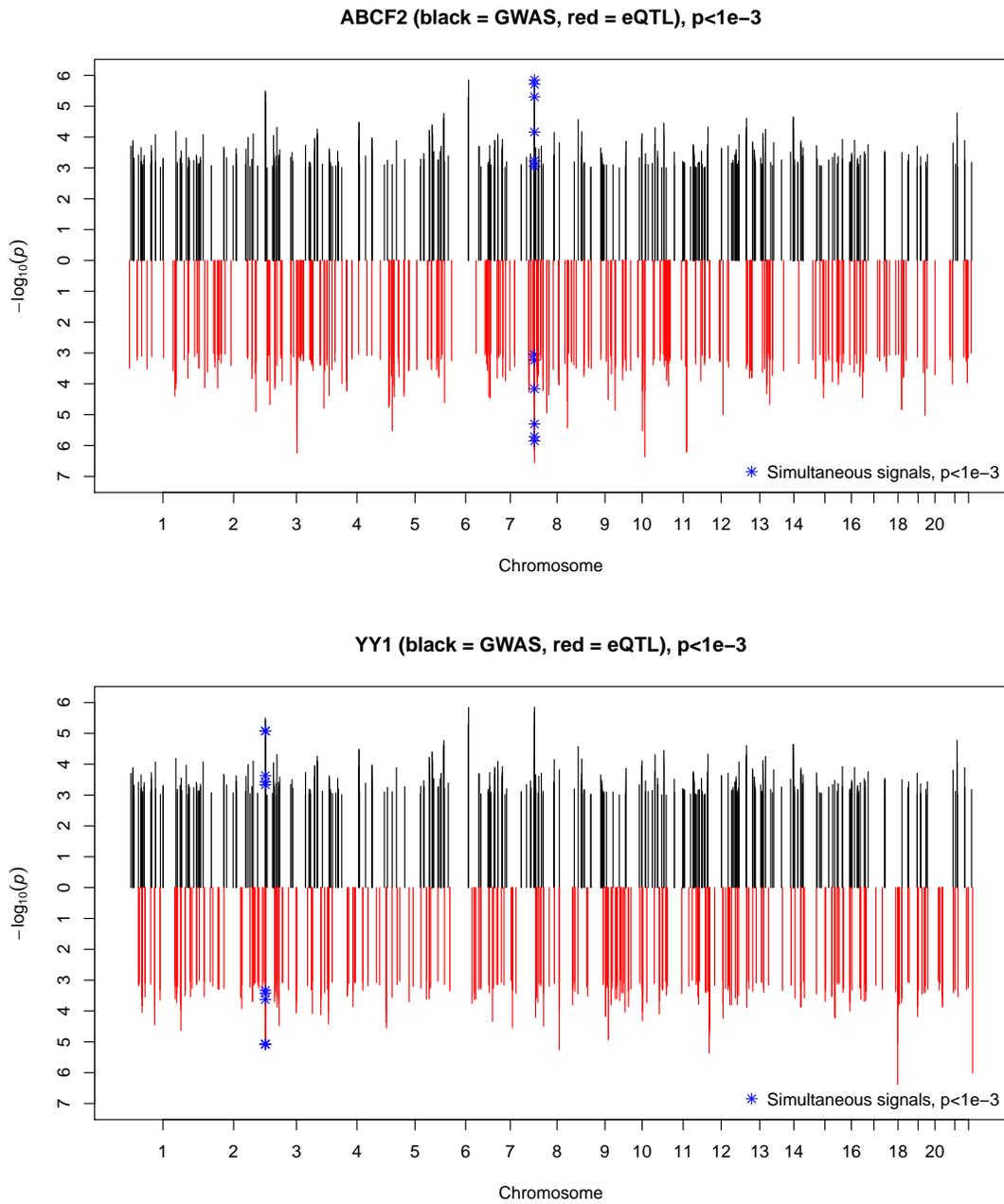Figure 4: Manhattan plots of the GWAS and eQTL p-values for the genes ABCF2 and YY1, which pass multiple testing correction for simultaneous signal detection. The upper half of each plot corresponds to GWAS results and the bottom half corresponds to eQTL results; only p-values less than 10^{-3} are plotted. Stars indicate possible positions of simultaneous signals, where both GWAS and eQTL p-values are less than 10^{-3}.

In the eQTL data $M_{eff} = 181.82$, giving a Bonferroni threshold of $0.05/182 = 2.75 \times 10^{-4}$. Two genes in Table 4 pass this threshold, ABCF2 and YY1, and Manhattan plots for the GWAS and eQTL p-values of these two genes are shown in Figure 4. ABCF2 is an ATP-binding cassette transporter, which have been found to protect against cardiac hypertrophy by promoting angiogenesis (Higashikuni et al. 2012, Maher et al. 2014). YY1 is a transcription factor which in experiments on rat cardiomyocytes was found to repress expression of α-myosin heavy chain, which is responsible for heart muscle contraction (Sucharov et al. 2003, Mariner et al. 2004). The plot for YY1 indicates the presence of SNPs on chromosome 2 that both regulate YY1 expression and are associated with heart failure, while the YY1 gene itself is located on chromosome 14. This shows that the proposed method can detect *trans*-regulatory relationships. The remaining genes in Table 4 may pass other forms of multiple testing correction, such as false discovery rate control, and more research into multiple simultaneous signal detection is necessary.

To compare with the results of the proposed method, Table 5 reports the discoveries of the other simultaneous signal detection methods that pass the 2.75×10^{-4} Bonferroni threshold. Two-stage approaches with thresholds of 10^{-2}, 10^{-3}, and 10^{-4} were not considered because they were unable to control the type I error rate under linkage disequilibrium in simulations. GPA was also unable to control the type I error: though Table 5 reports that it found ZNF266 to exhibit simultaneous signals in the heart failure data, this gene was highly non-significant when tested using the two-stage approach regardless of the p-value threshold. This suggests that ZNF266 is a false positive. The two-stage approach with a threshold of 10^{-5} found the same two genes found by the proposed procedure. This is reassuring, but the two-stage methods are still highly dependent on choosing the proper threshold, as Table 5 shows that the 10^{-6} threshold made no discoveries.

4.3   Choosing the correct eQTL tissue

While the MAGNet consortium measured eQTL data from both normal and failing heart tissue, in the above analysis the V_i were calculated using only normal hearts. Using normal, rather than failing, heart tissue is appropriate because it more accurately reflects the true regulatory relationships between SNPs and gene expression. Gene expression patterns in

the failing hearts are likely to be influenced by many other factors, such as the patients' medication histories and heart failure disease processes, so the causal model from Figure 1 likely no longer holds.

To illustrate the importance of selecting the correct tissue, the PHFS GWAS results were integrated with eQTL data from lymphoblastoid cell lines (LCLs) collected by Duan et al. (2008), instead of with the MAGNet eQTL data from normal heart tissue. This serves as a negative control experiment, as LCLs are not immediately relevant to cardiovascular disease.

Table 6 reports the results. Some of the genes detected in this negative control may in fact be important for heart failure, as long as the genetic regulation of these genes in LCLs is similar to their regulation in heart tissue. Indeed, two of the top four genes, UBE2D2 and JTB, were differentially expressed in the MAGNet eQTL data. However, without additional heart tissue-specific expression data, from this analysis alone it is impossible to tell which of the detected genes are important.

The MAGNet Consortium's study is unique because it was able to collect eQTL data from live human heart tissue. In general, however, genomics data from relevant tissue may be difficult to obtain, for example when studying diseases affecting the heart or the brain. As mentioned previously, the proposed method can integrate GWAS and eQTL datasets collected from different groups of subjects. This enables individual investigators to leverage public resources such as the Genotype-Tissue Expression project (Lonsdale et al. 2013), from which eQTL data from multiple tissue types are available, in combination with their own GWAS results.

4.4   Gene set enrichment analysis

The proposed simultaneous signal detection test has so far been applied to one gene at a time. To derive more functional insight it can be extended to gene sets. Let V_{ij} denote the test statistic for association between SNP i and the jth gene of a gene set. Then Gene Set Enrichment Analysis (GSEA; Mootha et al. 2003, Subramanian et al. 2005) can be applied to the max statistics $M_j = \max_i(U_i \wedge V_{ij})$ proposed in (5). Given a gene set $\mathcal{S}$, this can be

done with the Kolmogorov-Smirnov statistic

$$\sup_x \left| \frac{1}{s} \sum_{j \in \mathcal{S}} I(M_j \leq x) - \frac{1}{s'} \sum_{j \in \mathcal{S}^c} I(M_j \leq x) \right|, \tag{11}$$

where s and s' are the number of genes in $\mathcal{S}$ and $\mathcal{S}^c$, respectively. This amounts to testing whether the distribution of the M_j differs between genes in $\mathcal{S}$ and genes in $\mathcal{S}^c$.

This analysis was applied to gene sets from Gene Ontology (Ashburner et al. 2000) containing at least 10 genes, specifically, 5,023 Biological Process terms and 936 Molecular Function terms. Table 7 reports the most significant findings, and Figures 5 and 6 depict all Gene Ontology terms that are connected to these findings through any path. A number of the identified Biological Process gene sets relate to chromatin structure and centromere assembly. For example, the CENP-A histone H3-like centromeric protein A has been found to be critical in cardiac stem cells (McGregor et al. 2014). The top Molecular Function gene sets are involved in processes such as unfolded protein binding, protein kinase regulation and kinase regulator activity, implying that protein quality control may play a role in cardiac homeostasis (Wang & Robbins 2006).

5   Discussion

The asymptotic adaptive optimality of the proposed test was established assuming that the test statistics U_i and V_i are absolute values of normal random variables (7). Additional issues arise when this does not hold. Mukherjee et al. (2015) studied detecting non-zero regression coefficients in logistic regression and found that the detection boundary can be different from the boundary for Gaussian outcomes if too many SNPs have very low minor allele frequencies. As another example, Delaigle et al. (2011) studied signal detection for a single sequence of t-statistics and found that the detection boundary depends on the rate at which the sample size grows with the number of tests. Additional work is necessary to characterize the detection boundary for testing (2) when U_i and V_i do not follow (7).

Even when the distribution assumptions (7) hold, results from the signal detection problem for normal mean mixtures (3) under dependence (Hall & Jin 2008, 2010) suggest that M is likely no longer optimal when the test statistics are correlated across i. Additional

Figure 5: Directed acyclic graph of all Biological Process nodes connected by some path to the most significant results from the simultaneous signal GSEA analysis. Yellow: least significant; Red: most significant; Rectangles: top GSEA results.
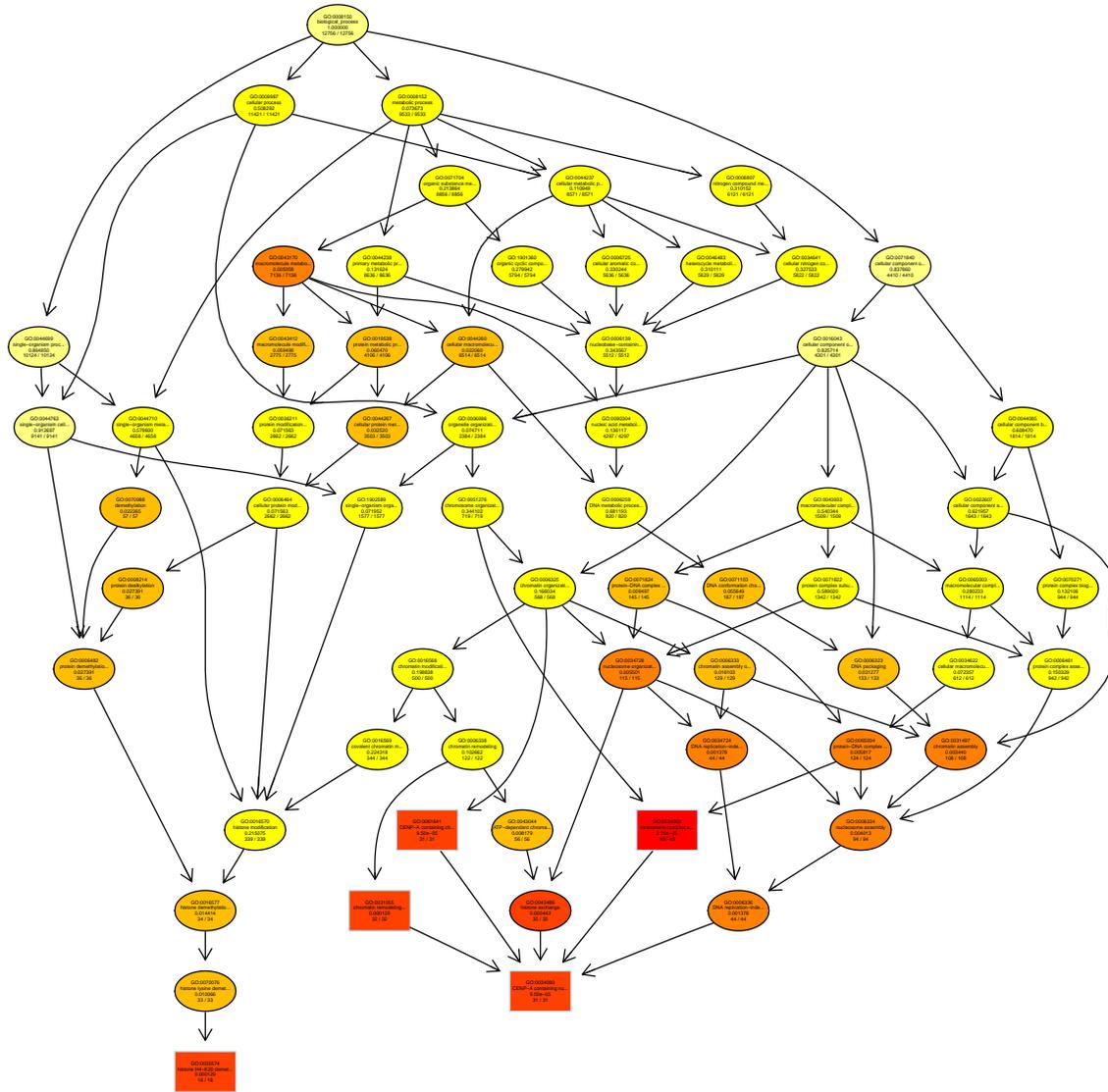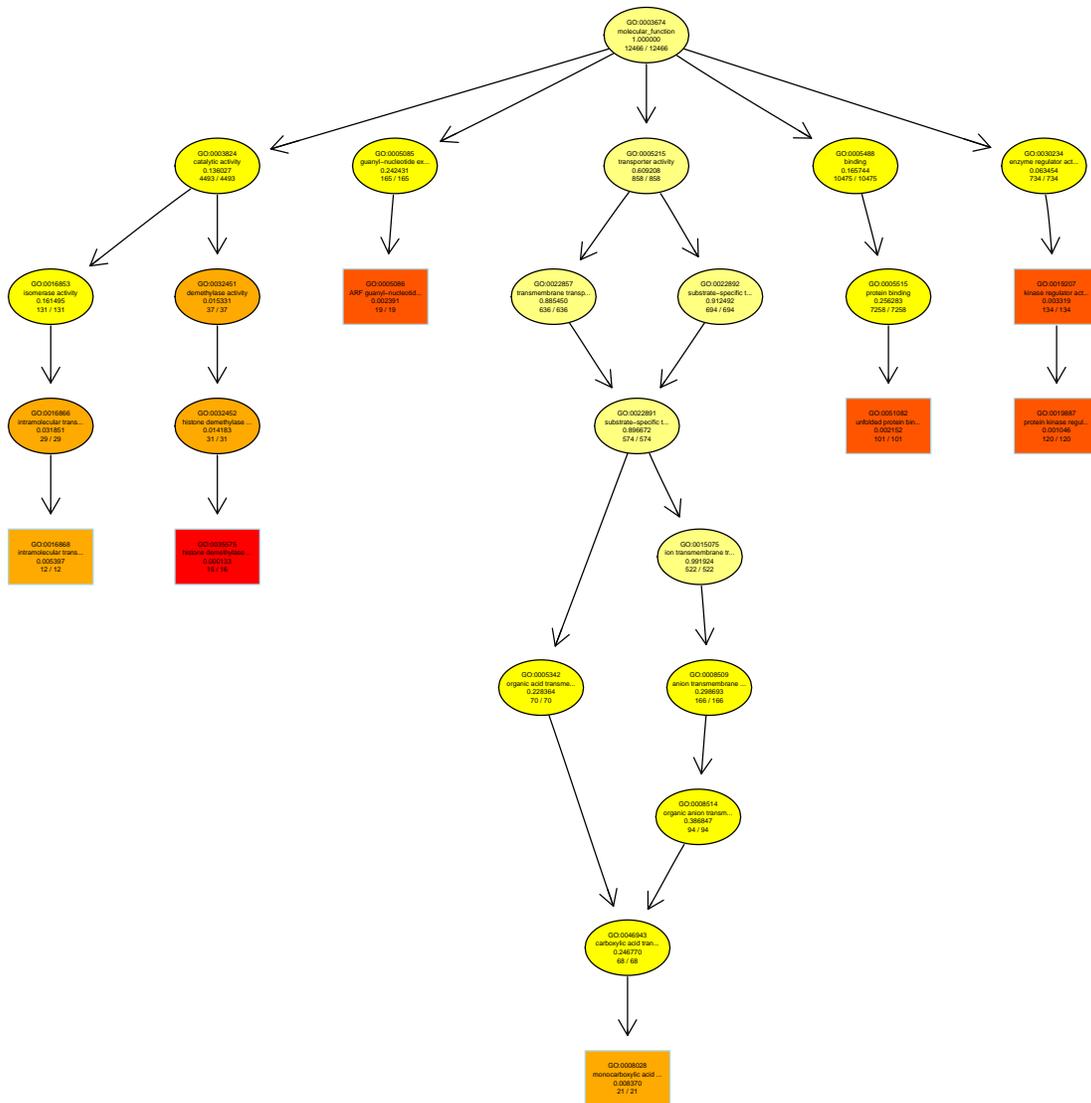
Figure 6: Directed acyclic graph of all Molecular Function nodes connected by some path to the most significant results from the simultaneous signal GSEA analysis. Yellow: least significant; Red: most significant; Rectangles: top GSEA results.

research is necessary to determine how best to incorporate information about the linkage disequilibrium structure into a simultaneous signal detection method. On the other hand, simulations in Section 3.2 indicate that M is still effective in a sizable portion of the detectable region even under dependence.

Alternative detection procedures that take advantage of additional biological information may be more powerful than the proposed approach. For example, it is known that *cis*-eQTL signals tend to be stronger than *trans*-eQTL signals. The optimal approach to incorporating this information into a simultaneous signal detection test statistic remains to be determined, and is an important direction for future research.

The proposed method can be extended to settings involving more than two sequences of test statistics. For example, to detect simultaneous signals across K sequences $\{U_i^k\}$, $k = 1, \ldots, K$, the max test statistic (5) can be extended to $M = \max_i(U_i^1 \wedge \ldots \wedge U_i^K)$. To obtain a p-value for this M the indices of each of the K sequences can be permuted independently. It may also be possible to obtain an analytic expression for the permutation p-value with multiple sequences, analogous to (6). In some cases it may instead be of interest to detect simultaneous signals between one sequence U_i and any of a set of sequences $\{V_i^k\}$, $k = 1, \ldots, K$. This can be done by defining $V_i = V_i^1 \vee \ldots \vee V_i^K$ and applying (6) to the (U_i, V_i). Since Theorem 1 requires no parametric distributional assumptions, it would still hold for these U_i and V_i, and the permutation p-value would remain conservative.

Finally, while this paper has only considered the problem of detecting simultaneous signals, it is sometimes also of interest to identify such signals. This requires the development of a multiple testing method. Consider a procedure that identifies the ith SNP as a simultaneous signal if $T_i \geq (\log n)^{1/2}$. Whenever simultaneous signals are detectable using T_i, this procedure will also achieve asymptotically perfect identification. It would be interesting to develop other identification procedures that control the false discovery rate. The fact that the null hypothesis is composite poses a major difficulty in developing such a multiple testing procedure.

SUPPLEMENTARY MATERIAL

Supplementary file: Contains additional linkage disequilibrium simulations and a detailed discussion of the results, as well as proofs of all theorems. (.pdf file)

References

1000 Genomes Project Consortium (2010), 'A map of human genome variation from population-scale sequencing', *Nature* **467**(7319), 1061–1073.

Arias-Castro, E., Candès, E. J. & Plan, Y. (2011), 'Global testing under sparse alternatives: Anova, multiple comparisons and the higher criticism', *The Annals of Statistics* **39**(5), 2533–2556.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T. et al. (2000), 'Gene ontology: tool for the unification of biology', *Nature genetics* **25**(1), 25–29.

Bentkus, V., Jing, B.-Y., Shao, Q.-M., Zhou, W. et al. (2007), 'Limiting distributions of the non-central t-statistic and their applications to the power of t-tests under non-normality', *Bernoulli* **13**(2), 346–364.

Cai, T. T., Jeng, X. J. & Jin, J. (2011), 'Optimal detection of heterogeneous and heteroscedastic mixtures', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(5), 629–662.

Cai, T. T., Liu, W. & Xia, Y. (2013), 'Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings', *Journal of the American Statistical Association* **108**(501), 265–277.

Cai, T. T. & Wu, Y. (2014), 'Optimal detection for sparse mixtures against a given null distribution', *IEEE Trans. Inf. Theory* **60**(4), 2217–2232.

Cappola, T. P., Matkovich, S. J., Wang, W., van Booven, D., Li, M., Wang, X., Qu, L., Sweitzer, N. K., Fang, J. C., Reilly, M. P. et al. (2011), 'Loss-of-function dna sequence variant in the clcnka chloride channel implicates the cardio-renal axis in interindividual heart failure risk variation', *Proceedings of the National Academy of Sciences* **108**(6), 2456–2461.

Chen, Y., Zhu, J., Lum, P., Yang, X., Pinto, S., MacNeil, D., Zhang, C., Lamb, J., Edwards, S., Sieberts, S. et al. (2008), 'Variations in DNA elucidate molecular networks that cause disease', *Nature* **452**(7186), 429–435.

Chung, D., Yang, C., Li, C., Gelernter, J. & Zhao, H. (2014), 'Gpa: A statistical approach to prioritizing gwas results by integrating pleiotropy and annotation.', *PLoS genetics* **10**(11), e1004787.

Creemers, E. E., Wilde, A. A. & Pinto, Y. M. (2011), 'Heart failure: advances through genomics', *Nature Reviews Genetics* **12**(5), 357–362.

Dawson, E., Abecasis, G. R., Bumpstead, S., Chen, Y., Hunt, S., Beare, D. M., Pabial, J., Dibling, T., Tinsley, E., Kirby, S., Carter, D., Papaspyridonos, M., Livingstone, S., Ganskell, R., Lõhmussaar, E., Zernant, J., Tõnisson, N., Remm, M., Mägi, R., Puurand, T., vilo, J., Kurg, A., Rice, K., Deloukas, P., Mott, R., Metspalu, A., Bentley, D. R., Cardon, L. R. & Dunham, I. (2002), 'A first-generation linkage disequilibrium map of human chromosome 22', *Nature* **418**(6897), 544–548.

Delaigle, A., Hall, P. & Jin, J. (2011), 'Robustness and accuracy of methods for high dimensional data analysis based on student's t-statistic', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(3), 283–301.

Donoho, D. & Jin, J. (2004), 'Higher criticism for detecting sparse heterogeneous mixtures', *The Annals of Statistics* **32**(3), 962–994.

Duan, S., Huang, R. S., Zhang, W., Bleibel, W. K., Roe, C. A., Clark, T. A., Chen, T. X., Schweitzer, A. C., Blume, J. E., Cox, N. J. et al. (2008), 'Genetic architecture of transcript-level variation in humans', *The American Journal of Human Genetics* **82**(5), 1101–1113.

Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G., Gunnarsdottir, S. et al. (2008), 'Genetics of gene expression and its effect on disease', *Nature* **452**(7186), 423–428.

Galwey, N. W. (2009), 'A new measure of the effective number of tests, a practical tool for comparing families of non-independent significance tests', *Genetic epidemiology* **33**(7), 559–568.

Gamazon, E. R., Wheeler, H. E., Shah, K. P., Mozaffari, S. V., Aquino-Michaels, K., Carroll, R. J., Eyler, A. E., Denny, J. C., Nicolae, D. L., Cox, N. J. et al. (2015), 'A gene-based association method for mapping traits using reference transcriptome data', *Nature Genetics* **47**(9), 1091–1098.

Giambartolomei, C., Vukcevic, D., Schadt, E. E., Franke, L., Hingorani, A. D., Wallace, C. & Plagnol, V. (2014), 'Bayesian test for colocalisation between pairs of genetic association studies using summary statistics', *PLoS genetics* **10**(5), e1004383.

Hall, P. & Jin, J. (2008), 'Properties of higher criticism under strong dependence', *The Annals of Statistics* pp. 381–402.

Hall, P. & Jin, J. (2010), 'Innovated higher criticism for detecting sparse signals in correlated noise', *The Annals of Statistics* **38**(3), 1686–1732.

He, X., Fuller, C. K., Song, Y., Meng, Q., Zhang, B., Yang, X. & Li, H. (2013), 'Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS', *The American Journal of Human Genetics* **92**(5), 667–680.

Higashikuni, Y., Sainz, J., Nakamura, K., Takaoka, M., Enomoto, S., Iwata, H., Tanaka, K., Sahara, M., Hirata, Y., Nagai, R. et al. (2012), 'The atp-binding cassette transporter abcg2 protects against pressure overload–induced cardiac hypertrophy and heart failure by promoting angiogenesis and antioxidant response', *Arteriosclerosis, thrombosis, and vascular biology* **32**(3), 654–661.

Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S. & Manolio, T. A. (2009), 'Potential etiologic and functional implications of genome-wide association loci for human diseases and traits', *Proceedings of the National Academy of Sciences* **106**(23), 9362–9367.

Huang, Y.-T. (2014), 'Integrative modeling of multiple genomic data from different types of genetic association studies', *Biostatistics* **15**(4), 587–602.

Ingster, Y. I. (1997), 'Some problems of hypothesis testing leading to infinitely divisible distributions', *Mathematical Methods of Statistics* **6**(1), 47–69.

Ingster, Y. I. (1998), 'Minimax detection of a signal for l_n^p-balls', *Mathematical Methods of Statistics* **7**(4), 401–428.

Ingster, Y. I. (2002*a*), 'Adaptive detection of a signal of growing dimension, i', *Mathematical Methods of Statistics* **10**, 395–421.

Ingster, Y. I. (2002*b*), 'Adaptive detection of a signal of growing dimension, ii', *Mathematical Methods of Statistics* **11**(1), 37–68.

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., Speed, T. P. et al. (2003), 'Exploration, normalization, and summaries of high density oligonucleotide array probe level data', *Biostatistics* **4**(2), 249–264.

Jager, L. & Wellner, J. A. (2007), 'Goodness-of-fit tests via phi-divergences', *The Annals of Statistics* **35**(5), 2018–2053.

Johnson, W. E., Li, C. & Rabinovic, A. (2007), 'Adjusting batch effects in microarray expression data using empirical bayes methods', *Biostatistics* **8**(1), 118–127.

Ky, B., French, B., Levy, W. C., Sweitzer, N. K., Fang, J. C., Wu, A. H., Goldberg, L. R., Jessup, M. & Cappola, T. P. (2012), 'Multiple biomarkers for risk prediction in chronic heart failure', *Circulation: Heart Failure* **5**(2), 183–190.

Ky, B., French, B., McCloskey, K., Rame, J. E., McIntosh, E., Shahi, P., Dries, D. L., Tang, W. W., Wu, A. H., Fang, J. C. et al. (2011), 'High-sensitivity st2 for prediction of adverse outcomes in chronic heart failure', *Circulation: Heart Failure* **4**(2), 180–187.

Ky, B., Kimmel, S. E., Safa, R. N., Putt, M. E., Sweitzer, N. K., Fang, J. C., Sawyer, D. B. & Cappola, T. P. (2009), 'Neuregulin-1β is associated with disease severity and adverse outcomes in chronic heart failure', *Circulation* **120**(4), 310–317.

Lee, D. S., Pencina, M. J., Benjamin, E. J., Wang, T. J., Levy, D., O'Donnell, C. J., Nam, B.-H., Larson, M. G., D'Agostino, R. B. & Vasan, R. S. (2006), 'Association of parental heart failure with risk of heart failure in offspring', *New England Journal of Medicine* **355**(2), 138–147.

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., Foster, B., Moser, M., Karasik, E., Gillard, B., Ramsey, K., Sullivan, S., Bridge, J., Magazine, H., Syron, J., Fleming, J., Siminoff, L., Traino, H., Mosavel, M., Barker, L., Jewell, S., Rohrer, D., Maxim, D., Filkins, D., Harbach, P., Cortadillo, E., Berghuis, B., Turner, L., Hudson, E., Feenstra, K., Sobin, L., Robb, J., Branton, P., Korzeniewski, G., Shive, C., Tabor, D., Qi, L., Groch, K., Nampally, S., Buia, S., Zimmerman, A., Smith, A., Burges, R., Robinson, K., Valentino, K., Bradbury, D., Cosentino, M., Diaz-Mayoral, N., Kennedy, M., Engel, T., Williams, P., Erickson, K., Ardlie, K., Winckler, W., Getz, G., DeLuca, D., MacArthur, D., Kellis, M., Thomson, A., Young, T., Gelfand, e., Donovan, M., Meng, Y., Grant, G., Mash, D., Marcus, Y., Basile, M., Liu, J., Zhu, J., Tu, Z., Cox, N., Nicolae, D., Gamazon, E., Im, H. K., Konkashbaev, A., Pritchard, J., Stevens, M., Flutre, T., Wen, X., Dermitzakis, E., Lappalainen, T., Guigo, R., Monlong, J., Sammeth, M., Koller, D., Battle, A., Mostafavi, S., McCarthy, M., Rivas, M., Maller, J., Rusyn, I., Nobel, A., Wright, F., Shabalin, A., Feolo, M., Sharopova, N., Sturcke, A., Paschal, J., Anderson, J., Wilder, E., Derr, L., Green, E., Struewing, J., Temple, G., Volpi, S., Boyer, J., Thomson, E., Guyer, M., Ng, C., Abdallah, A., Colantuoni, D., Insel, T., Koester, S., Little, A., Bender, P., Lehner, T., Yao, Y., Compton, C., Vaught, J., Sawyer, S., Lockhart, N., Demchok, J. & Moore, H. (2013), 'The genotype-tissue expression (GTEx) project', *Nature Genetics* **45**(6), 580–585.

Maher, T. J., Ren, Y., Li, Q., Braunlin, E., Garry, M. G., Sorrentino, B. P. & Martin, C. M. (2014), 'Atp-binding cassette transporter abcg2 lineage contributes to the cardiac vasculature after oxidative stress', *American Journal of Physiology-Heart and Circulatory Physiology* **306**(12), H1610–H1618.

Marat, A. L. & McPherson, P. S. (2010), 'Variants of dennd1b associated with asthma in

children', *N Engl J Med* **363**(10), 988–989.

Mariner, P. D., Luckey, S. W., Long, C. S., Sucharov, C. C. & Leinwand, L. A. (2004), 'Yin yang 1 represses α-myosin heavy chain gene expression in pathologic cardiac hypertrophy', *Biochemical and biophysical research communications* **326**(1), 79–86.

McGregor, M., Hariharan, N., Joyo, A., Margolis, R. L. & Sussman, M. (2014), 'Cenp-a is essential for cardiac progenitor cell proliferation', *Cell Cycle* **13**(5), 739–748.

Mootha, V. K., Lindgren, C. M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., N., H., Daly, M., Patterson, N., Mesirov, J., Golub, T., Tamayo, P., Spiegelman, B., Lander, E., Hirschhorn, J., Altshuler, D. & Groop, L. (2003), 'Pgc-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes', *Nature Genetics* **34**(3), 267–273.

Morey, L., Santanach, A., Blanco, E., Aloia, L., Nora, E. P., Bruneau, B. G. & Di Croce, L. (2015), 'Polycomb regulates mesoderm cell fate-specification in embryonic stem cells through activation and repression mechanisms', *Cell stem cell* **17**(3), 300–315.

Mudd, J. O. & Kass, D. A. (2008), 'Tackling heart failure in the twenty-first century', *Nature* **451**(7181), 919–928.

Mukherjee, R., Pillai, N. S., Lin, X. et al. (2015), 'Hypothesis testing for high-dimensional sparse binary regression', *The Annals of Statistics* **43**(1), 352–381.

Nicolae, D. L., Gamazon, E., Zhang, W., Duan, S., Dolan, M. E. & Cox, N. J. (2010), 'Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS', *PLoS Genetics* **6**(4), e1000888.

Oka, T., Hikoso, S., Yamaguchi, O., Taneike, M., Takeda, T., Tamai, T., Oyabu, J., Murakawa, T., Nakayama, H., Nishida, K. et al. (2012), 'Mitochondrial dna that escapes from autophagy causes inflammation and heart failure', *Nature* **485**(7397), 251–255.

Petrovic, A., Pasqualato, S., Dube, P., Krenn, V., Santaguida, S., Cittaro, D., Monzani, S., Massimiliano, L., Keller, J., Tarricone, A. et al. (2010), 'The mis12 complex is a protein

interaction hub for outer kinetochore assembly', *The Journal of cell biology* **190**(5), 835–852.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. & Smyth, G. K. (2015), 'limma powers differential expression analyses for RNA-sequencing and microarray studies', *Nucleic Acids Research* **43**(7), e47.

Roger, V. L. (2013), 'Epidemiology of heart failure', *Circulation research* **113**(6), 646–659.

Ross, R. S. & Borg, T. K. (2001), 'Integrins and the myocardium', *Circulation research* **88**(11), 1112–1119.

Sack, M. N., Smith, R. M. & Opie, L. H. (2000), 'Tumor necrosis factor in myocardial hypertrophy and ischaemiaan anti-apoptotic perspective', *Cardiovascular research* **45**(3), 688–695.

Sikorska, K., Lesaffre, E., Groenen, P. F. & Eilers, P. H. (2013), 'GWAS on your notebook: fast semi-parallel linear and logistic regression for genome-wide association studies', *BMC bioinformatics* **14**(1), 166.

Sin, N., Meng, L., Wang, M. Q., Wen, J. J., Bornmann, W. G. & Crews, C. M. (1997), 'The anti-angiogenic agent fumagillin covalently binds and inhibits the methionine aminopeptidase, metap-2', *Proceedings of the National Academy of Sciences* **94**(12), 6099–6103.

Smyth, G. K. (2005), Limma: linear models for microarray data, *in* 'Bioinformatics and computational biology solutions using R and Bioconductor', Springer, pp. 397–420.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. & Mesirov, J. (2005), 'Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles', *Proceedings of the National Academy of Sciences of the United States of America* **102**(43), 15545–15550.

Sucharov, C. C., Mariner, P., Long, C., Bristow, M. & Leinwand, L. (2003), 'Yin yang 1 is increased in human heart failure and represses the activity of the human α-myosin heavy chain promoter', *Journal of Biological Chemistry* **278**(33), 31233–31239.

Wang, X. & Robbins, J. (2006), 'Heart failure and protein quality control', *Circulation research* **99**(12), 1315–1328.

Ware, J. S., Petretto, E. & Cook, S. A. (2013), 'Integrative genomics in cardiovascular medicine', *Cardiovascular research* **97**(4), 623–630.

Wei, L., Yuan, M., Zhou, R., Bai, Q., Zhang, W., Zhang, M., Huang, Y. & Shi, L. (2015), 'Microrna-101 inhibits rat cardiac hypertrophy by targeting rab1a', *Journal of cardiovascular pharmacology* **65**(4), 357–363.

Xie, J., Cai, T. T. & Li, H. (2011), 'Sample size and power analysis for sparse signal recovery in genome-wide association studies', *Biometrika* **98**(2), 273–290.

Xiong, Q., Ancona, N., Hauser, E. R., Mukherjee, S. & Furey, T. S. (2012), 'Integrating genetic and gene expression evidence into genome-wide association analysis of gene sets', *Genome research* **22**(2), 386–397.

Zhang, W.-M., Popova, S. N., Bergman, C., Velling, T., Gullberg, M. K. & Gullberg, D. (2002), 'Analysis of the human integrin α11 gene (itga11) and its promoter', *Matrix biology* **21**(6), 513–523.

Zhao, S. D., Cai, T. T. & Li, H. (2014), 'More powerful genetic association testing via a new statistical framework for integrative genomics', *Biometrics* **70**(4), 881–890.

Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M. R., Powell, J. E., Montgomery, G. W., Goddard, M. E., Wray, N. R., Visscher, P. M. et al. (2016), 'Integration of summary data from gwas and eqtl studies predicts complex trait gene targets', *Nature genetics* .

Table 1: Average type I errors (%) at nominal $\alpha = 0.05$ over 1,000 replications for independent test statistics

Simulation setting					Methods							
#U	#V	$n\epsilon$	a	b	True	Perm	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	GPA
10	10	5	2.5	3.0	4.0	4.2	3.7	0.4	0.2	0.0	0.0	0.0
10	50	5	2.5	3.0	4.5	4.1	3.3	0.6	0.2	0.0	0.0	0.0
10	100	5	2.5	3.0	4.6	4.5	3.2	0.8	0.3	0.0	0.0	0.0
50	50	5	2.5	3.0	4.6	4.5	3.0	0.9	0.3	0.0	0.0	0.0
50	100	5	2.5	3.0	4.5	4.3	3.1	1.1	0.3	0.0	0.0	0.0
100	100	5	2.5	3.0	4.5	3.9	3.0	0.9	0.4	0.0	0.0	0.1
10	10	10	2.5	3.0	4.0	4.2	3.7	0.4	0.2	0.0	0.0	0.0
10	50	10	2.5	3.0	4.5	4.1	3.3	0.6	0.2	0.0	0.0	0.0
10	100	10	2.5	3.0	4.6	4.5	3.2	0.8	0.3	0.0	0.0	0.0
50	50	10	2.5	3.0	4.6	4.5	3.0	0.9	0.3	0.0	0.0	0.0
50	100	10	2.5	3.0	4.5	4.3	3.1	1.1	0.3	0.0	0.0	0.0
100	100	10	2.5	3.0	4.5	3.9	3.0	0.9	0.4	0.0	0.0	0.1
10	10	5	3.5	4.0	4.1	4.1	3.6	0.4	0.2	0.0	0.0	0.0
10	50	5	3.5	4.0	4.2	4.1	3.2	0.7	0.3	0.0	0.0	0.0
10	100	5	3.5	4.0	4.8	4.4	3.6	0.9	0.6	0.0	0.0	0.2
50	50	5	3.5	4.0	4.7	3.2	3.7	1.3	0.5	0.0	0.0	0.9
50	100	5	3.5	4.0	4.0	3.0	3.5	1.6	0.7	0.0	0.0	1.0
100	100	5	3.5	4.0	4.3	2.3	3.1	1.3	1.1	0.0	0.0	1.0
10	10	10	3.5	4.0	4.1	4.1	3.6	0.4	0.2	0.0	0.0	0.0
10	50	10	3.5	4.0	4.2	4.1	3.2	0.7	0.3	0.0	0.0	0.0
10	100	10	3.5	4.0	4.8	4.4	3.6	0.9	0.6	0.0	0.0	0.2
50	50	10	3.5	4.0	4.7	3.2	3.7	1.3	0.5	0.0	0.0	0.9
50	100	10	3.5	4.0	4.0	3.0	3.5	1.6	0.7	0.0	0.0	1.0
100	100	10	3.5	4.0	4.3	2.3	3.1	1.3	1.1	0.0	0.0	1.0

$\#U, \#V$ = number of non-null U_i, V_i; $n\epsilon$ = number of simultaneous signals; a, b = means used to generate μ_i, ν_i; True = true p-value of M; Perm = permutation p-value (6); 10^{-x} = two-stage approach with p-value threshold of 10^{-x}; GPA = method of Chung et al. (2014).

Table 2: Average powers (%) at nominal $\alpha = 0.05$ over 1,000 replications for independent test statistics

Simulation setting					Methods							
#U	#V	$n\epsilon$	a	b	True	Perm	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	GPA
10	10	5	2.5	3.0	58.4	58.4	11.2	23.0	25.3	6.2	0.9	0.0
10	50	5	2.5	3.0	52.6	52.2	10.8	21.6	20.4	5.0	0.5	0.0
10	100	5	2.5	3.0	16.3	16.0	6.6	4.3	3.5	0.2	0.0	0.0
50	50	5	2.5	3.0	58.1	57.8	11.9	29.4	28.6	9.5	1.7	0.1
50	100	5	2.5	3.0	23.3	22.0	8.0	7.3	6.1	0.7	0.0	0.2
100	100	5	2.5	3.0	22.4	20.5	5.9	5.0	7.9	1.2	0.2	2.6
10	10	10	2.5	3.0	88.7	88.6	25.5	62.5	58.4	25.9	8.6	0.0
10	50	10	2.5	3.0	73.7	73.6	22.8	46.4	41.1	11.9	3.8	0.0
10	100	10	2.5	3.0	69.8	70.4	19.0	42.1	37.1	14.0	4.4	0.4
50	50	10	2.5	3.0	77.1	76.9	25.8	55.0	44.1	12.8	3.0	1.0
50	100	10	2.5	3.0	59.1	57.6	16.8	37.0	29.0	7.5	2.0	4.5
100	100	10	2.5	3.0	93.3	92.4	26.3	81.3	75.6	38.5	14.7	75.4
10	10	5	3.5	4.0	97.4	97.4	22.4	87.2	85.3	55.4	25.2	0.7
10	50	5	3.5	4.0	96.6	95.7	24.1	86.9	83.4	51.6	20.3	16.7
10	100	5	3.5	4.0	69.8	68.5	16.3	49.9	42.9	13.7	3.1	5.9
50	50	5	3.5	4.0	97.3	96.7	23.5	90.0	88.3	59.2	28.3	96.3
50	100	5	3.5	4.0	78.3	74.1	16.5	59.7	54.1	22.8	5.5	70.2
100	100	5	3.5	4.0	73.2	62.0	12.7	43.3	52.8	22.5	7.1	43.4
10	10	10	3.5	4.0	100.0	100.0	62.3	99.9	98.6	86.7	58.2	3.7
10	50	10	3.5	4.0	99.9	99.8	60.2	99.2	96.2	74.5	40.8	35.1
10	100	10	3.5	4.0	99.3	99.2	52.4	98.0	94.8	72.2	36.9	38.6
50	50	10	3.5	4.0	100.0	100.0	61.5	99.7	97.8	77.6	43.9	99.6
50	100	10	3.5	4.0	98.7	98.0	52.1	98.5	92.3	62.7	28.6	99.8
100	100	10	3.5	4.0	100.0	100.0	49.7	99.9	99.6	95.1	75.5	99.9

#U, #V = number of non-null U_i, V_i; $n\epsilon$ = number of simultaneous signals; a, b = means used to generate μ_i, ν_i; True = true p-value of M; Perm = permutation p-value (6); 10^{-x} = two-stage approach with p-value threshold of 10^{-x}; GPA = method of Chung et al. (2014).

Table 3: Average type I errors and powers (%) at nominal $\alpha = 0.05$ over 1,000 replications under linkage disequilibrium

Setting			Methods						
$n\epsilon$	a	b	Perm	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	GPA
Type I error									
5	0.1	0.2	2.7	13.5	13.9	10.1	4.0	1.3	23.9
10	0.1	0.2	2.7	13.5	13.9	10.1	4.0	1.3	23.9
5	0.2	0.3	3.0	13.5	14.7	12.0	7.1	2.0	24.8
10	0.2	0.3	3.0	13.5	14.7	12.0	7.1	2.0	24.8
Power									
5	0.1	0.2	84.2	20.7	28.5	40.5	64.8	76.7	25.2
10	0.1	0.2	86.6	24.8	38.8	60.9	78.1	79.4	30.7
5	0.2	0.3	72.1	21.4	26.7	37.3	49.9	63.1	28.3
10	0.2	0.3	82.9	24.4	39.0	61.5	71.7	76.7	35.5

$n\epsilon$ = number of simultaneous signals; a, b = means used to generate $\boldsymbol{\alpha}, \boldsymbol{\beta}$; True = true p-value of M; Perm = permutation p-value (6); 10^{-x} = two-stage approach with p-value threshold of 10^{-x}; GPA = method of Chung et al. (2014).

Table 4: Genes with simultaneous signal detection p-values less than 10^{-3} using PHFS GWAS and MAGNet eQTL data, using the proposed method (6)

Gene	SS p	DE p	Annotation
ABCF2	2.6e-5	9.2e-9	ATP-binding cassette transporter; angiogenesis (Higashikuni et al. 2012)
YY1	1.8e-4	3.0e-3	Transcription factor; represses heart muscle contraction (Sucharov et al. 2003)
NSL1	2.9e-4	2.0e-10	MIS12 kinetochore complex (Petrovic et al. 2010)
ITGA11	6.0-4	0.1	Integrin; myocardial extracellular matrix; myocardial strength and plasticity (Ross & Borg 2001, Zhang et al. 2002)
DENND1B	6.1e-4	6.2e-15	Release of cytokines; myocardial contractile performance (Sack et al. 2000, Marat & McPherson 2010)
METAP1	6.6e-4	1.4e-13	Molecular target of angiogenesis (Sin et al. 1997)
PCGF5	7.5e-4	6.6e-6	Polycomb group; regulate heart development (Morey et al. 2015)
RUSC2	8.3e-4	1.1e-6	Interacts with RAB1A, RAB1B; causes cardiac hypertrophy (Wei et al. 2015)
MAP1LC3B2	8.6e-4	9.6e-15	Microtubule-associated protein; inflammation (Oka et al. 2012)

SS p: simultaneous signal detection p-value; DE p: differential expression in MAGNet eQTL data.

Table 5: Genes with simultaneous signal detection p-values less than 2.75×10^{-4}, using existing methods

	10^{-5}		10^{-6}		GPA	
Gene	p	Gene	p	Gene	p
YY1	6.8e-13	—	—	ZNF266	6.4e-7
ABCF2	6.8e-12	—	—	—	—

10^{-x} = two-stage approach with p-value threshold of 10^{-x}; GPA = method of Chung et al. (2014); p = simultaneous signal detection p value; — = no finding.

Table 6: Genes with simultaneous signal detection p-values less than 10^{-3} using PHFS GWAS and LCL eQTL data

Gene	SS p	DE p
UBE2D2	2.88e-6	1.35e-7
TOMM7	1.44e-5	0.70
JTB	9.51e-5	1.19e-35
RAB13	9.51e-5	0.38

SS p: simultaneous signal detection p-value; DE p: differential expression p-value in MAG-Net eQTL data.

Table 7: Top five GSEA results using (11), with gene sets were derived from Gene Ontology (GO)

GO term	p-value
Biological process	
Centromere complex assembly	2.3e-5
CENP-A containing nucleosome assembly	5.0e-5
CENP-A containing chromatin organization	9.5e-5
Histone H4-K20 demethylation	1.3e-4
Chromatin remodeling at centromere	1.4e-4
Molecular function	
Histone demethylase activity	1.3e-4
Protein kinase regulator activity	1.1e-3
Unfolded protein binding	2.2e-3
ARF guanyl-nucleotide exchange factor activity	2.4e-4
Kinase regulator activity	3.3e-3