A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies

Oliver Stegle†
University of Cambridge, UK

Leopold Parts† and Richard Durbin
Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK

John Winn∗
Microsoft Research, Cambridge, UK

(Dated: May 12, 2009)

Gene expression measurements are influenced by a wide range of factors, such as the state of the cell, experimental conditions and variants in the sequence of regulatory regions. To understand the effect of a variable of interest, such as the genotype of a locus, it is important to account for variation that is due to confounding causes. Here, we present PEER (Probabilistic Estimation of Expression Residuals), a probabilistic framework for joint modeling of diverse sources of phenotypic variability. Efficient Bayesian inference makes PEER fast and tractable on large scale problems. We harness the flexibility of PEER and compare performance of latent variable methods on expression quantitative trait locus (eQTL) mapping datasets from simulations, yeast, mouse, and human. Employing Bayesian complexity control and joint modeling gives more precise estimates of the contribution of different factors and yields additional associations to measured transcript levels compared to alternative approaches. We present a threefold larger collection of cis eQTLs than previously found, as well as additional trans associations in a whole-genome eQTL scan of an outbred human population. Altogether, 30% of the tested probes show a significant genetic association in cis, and we validate that the additional eQTLs are likely to be real by replicating them in different sets of individuals. Our framework is the next step in analysis of high-dimensional phenotype data, and its application has revealed insights into genetic regulation of gene expression by demonstrating more abundant cis-acting eQTLs in human than previously shown.

I. AUTHOR SUMMARY

Gene expression is a complex phenotype. The measured expression level in an experiment can be affected by a wide range of factors - state of the cell, experimental conditions, variants in the sequence of regulatory regions, etc. To understand genotype to phenotype relationships, we need to be able to distinguish the variation that is due to the genetic state from all the confounding causes. Here, we present PEER, a probabilistic framework for understanding sources of variation in high-dimensional phenotype data. PEER allows models for different causes of variability in the observed phenotype to be quickly and flexibly combined, resulting in more accurate estimates of the effects of a variable of interest. The flexibility is demonstrated by comparison of a range of alternative models that account for confounding signal in the data. Applying PEER, we find that common genetic variation controlling gene expression levels in human is more abundant than previously thought, which has implications for a wide range of studies relating genotype to phenotype.

II. INTRODUCTION

DNA microarray technologies allow the measurement of expression levels of thousands of loci in the genome. These measurements enable exploring how a variable, such as clinical phenotype, tissue type, or genetic background, affects the transcriptional state of the sample. Recently, gene expression levels have been studied as quantitative genetic traits, investigating the effect of genotype as the primary variable. Studies have found and characterized large

† These authors contributed equally to this work.
∗Electronic address: jwinn@microsoft.com
numbers of expression quantitative trait loci (eQTLs) [1–3], exploring their complexity [2], population genetics [4, 5] and associations with disease [6, 7].

An important issue in such studies is additional variation in expression data that is not due to the genetic state, as illustrated in Figure 1. Intracellular fluctuations, environmental conditions, and experimental procedures are factors that can have a strong effect on the measured transcript levels [2, 8–10] and thereby obscure the association signal. When measured, correct estimation of the additional variation due to these known factors allows for more sensitive analysis of the genetic effect. For example, in [7], the authors reported finding additional human eQTLs when including the known factors of age, gender, and blood cell counts in the model. It is also standard practice to correct for batch effects, such as image artifacts or sample preparation differences [11].

![FIG. 1: General additive model for sources of gene expression variability.](image)

**FIG. 1:** General additive model for sources of gene expression variability. The $G \times J$ matrix $Y$ of measured gene expression levels of $G$ genes from $J$ individuals is modeled by additive contributions $\{Y^{(m)}\}_{m=1}^{M}$ from $M$ modules and observation noise $\Psi$. Here, the modules capture the signal due to primary effect of the genetic state $S$, known factors $F$ and hidden factors $X$. Other sources of variation can be present. Some examples of possible underlying sources of variation are given above the module boxes. The groupings represent some standard genetic association models commonly used.

In practice it is not possible to measure or even be aware of all potential sources of variation, but nevertheless it is important to account for them. Unobserved, hidden factors, such as cell culture conditions [12] often have an influence on large numbers of genes. We and others have proposed methods to detect and correct for such effects [9, 13, 14]. These studies demonstrated the importance of accounting for hidden factors, yielding a stronger statistical discrimination signal.

The challenge in modeling several confounding sources of variation (Figure 1) is to correctly estimate the contribution that is due to each one of them. There are open questions how to ensure that only spurious signal is eliminated by methods that account for hidden factors (see for instance discussion in [14]), and how to deal with situations when both known and hidden factors are present. In addition, there can be other sources of variability, making the problem of identifying the correct source of the signal even more severe. For example, when searching for epistatic or genotype-environment interactions, the primary effects of other known factors and hidden factors also need to be accounted for.

The key for correctly attributing expression variability is controlling the complexity of the statistical models for each source of variation. For example, the number of genotypes considered in an association scan can be enormous, and not all of them affect the expression level of every probe. Threshold values, obtained from likelihood ratio statistics or empirical p-value distributions, can be used to detect significant associations, thereby avoiding overfitting by controlling the model complexity [4, 15]. Similar measures are necessary for models of other sources of variability such as hidden factors.

In this work we present PEER (Probabilistic Estimation of Expression Residuals), a probabilistic framework to model phenotypic variability that is due to multiple sources. Our framework includes a set of commonly used models for sources of phenotypic variation (“modules”) that can be combined as needed. While previous attempts have been specific to a narrow set of underlying sources, this approach is flexible and can be adapted to a particular study design. The framework features the advantages of a probabilistic treatment, such as propagation of uncertainty between modules, and posterior distributions over model parameters. Complexity control is tackled at the level of individual modules, where parameters are regularized in a principled manner.

We explore the PEER framework in the context of detecting expression QTLs. The flexibility of combining modules makes it easy to compare the accuracy of alternative models for distinguishing genotype and hidden factor effects. We study the Variational Bayesian QTL mapper (VBQTL), a specific application of the PEER framework that explains gene expression variability by probabilistic modules for genetic effects as well as known and hidden factors. A simulation experiment contrasts VBQTL with commonly used configurations of the framework that use non-Bayesian approaches for the same task. This study highlights differences in the complexity control method with implications to
eQTL detection power, and demonstrates the necessity and difficulties of accounting for variability that confounds the genetic signal. Results on datasets from a human outbred population and crosses of inbred yeast and mouse strains show that VBQTL identifies more significant associations than alternative methods. We use VBQTL to perform a whole-genome eQTL scan on the HapMap phase 2 expression and genotype data, demonstrating the scalability of our framework to large numbers of samples and probes. We find three times more cis eQTLs than a standard association mapping method, suggesting more extensive genetic control of gene expression by common variants than previously shown.

III. METHODS

Here, we present the PEER framework for modelling diverse sources of phenotypic variability. In this exposition we focus on the gene expression levels as phenotype throughout, although PEER can be used to analyze any high-dimensional phenotype. We assume that gene expression levels are influenced by multiple additive effects (Figures 1, 2a). Models for individual sources of variation (modules) capture contributions from the signal of interest and confounding causes. We cast the full model combining all individual modules in a probabilistic setting, treating their parameters as random variables.

We perform Bayesian inference in this joint model, which is appealing for several reasons. Firstly, it allows to address possible dependencies between modules for the different sources of variation. Each module (such as the genetic association module) is learned taking distributions of parameters of all other modules (e.g. estimated effect of known covariates) into account. Propagation of parameter uncertainty between modules leads to more accurate parameter estimates [16], and avoids possible pathologies, for instance of maximum likelihood inference [17]. Secondly, Bayesian inference allows different modules to be flexibly included and combined according to the needs of the particular study. Many existing approaches can be cast as special cases of this framework, with some examples given in Figure 1. Finally, Bayesian inference allows different approximate inference schemes such as variational methods to be used [18], which makes the framework applicable for large scale high-dimensional datasets. Also, the variational approach allows the inference schedule to be specified by the user, leading to algorithms with different computational complexity and properties (see Section III C).

In the following, we present the mathematical model of our framework, and an outline of the inference procedure. We then provide three widely applicable probabilistic modules for phenotypic variation, as well as non-Bayesian alternatives. Finally, we combine the modules in different configurations to apply the framework for expression QTL mapping.

A. PEER - a Bayesian framework to account for multiple sources of variance in gene expression levels.

The observed expression matrix \( Y = \{y_{g,j}\} \) for genes \( g \in \{1, \ldots, G\} \) and individuals \( j \in \{1, \ldots, N\} \) is modeled by the sum of contributions \( Y^{(1)}, \ldots, Y^{(M)} \) from \( M \) modules and Gaussian noise with precisions \( \tau_g \) for each gene \( g \)

\[
P(y_{g,j} | y_{g,j}^{(1)}, \ldots, y_{g,j}^{(M)}, \tau_g) = \mathcal{N}(y_{g,j} | y_{g,j}^{(1)} + \cdots + y_{g,j}^{(M)}, \frac{1}{\tau_g}),
\]

where gamma priors are put on the noise precisions \( P(\tau_g) = \Gamma(\tau_g | a, b) \). The distribution of the contribution of the \( m \)th module, \( P(Y^{(m)} | \theta^{(m)}, D^{(m)}) \) depends on module-specific parameters \( \theta^{(m)} \) and potentially additional observed data \( D^{(m)} \). For example, in a standard genetic association module (Section III B), the observed data \( D^{(1)} \) are the individual SNP genotypes, and parameters \( \theta^{(1)} \) correspond to weights determining the influence of the different alleles on gene expression. The module contribution \( Y^{(1)} \) is the linear combination of the observed genotypes.

Parameter inference in the PEER framework is implemented using variational Bayesian learning [18], a generalization of the expectation maximization algorithm. An approximate distribution over model parameters is iteratively refined until convergence. In each iteration, approximate distributions of individual parameters are updated according to a specified schedule, taking the current state of all other modules and parameter distributions into account (Figure 2b-c). Choosing an approximation that factors over individual modules, the variational update equations have an intuitive interpretation:

1. The current belief of the residual dataset \( Z^{(m)} \) for module \( m \) is calculated taking the prediction form all other modules and the noise precision into account: \( z_{g,j}^{(m)} \sim \mathcal{N}(y_{g,j} - \sum_{l \neq m} (\langle y_{g,j}^{(l)} \rangle, \frac{1}{\tau_g})) \) (Figure 2b), where \( \langle \rangle \) indicate expected values under the current parameter posteriors.
FIG. 2: Bayesian network and outline of inference schedule for VBQTL. (a) The Bayesian network for the model of gene expression variation used in VBQTL (Section III C). It combines genetic (green), known factor (blue) and hidden factor (red) modules to explain the observed gene expression levels $Y$. The solid rectangles indicate that contained variables are duplicated for each gene probe ($g$), SNP ($n$) or factor ($c,k$) respectively. A similar rectangle for individuals ($j$) is omitted in this representation. The dashed rectangle indicates that the variable $b_{n,g}$ switches the contained part of the graph on or off representing the existence or lack of an association. Nodes with thick outlines ($s_{n,j}$, $f_{c,j}$ and $y_{g,j}$) are observed. (b)-(e) Update cycle of the known factors module introduced in Section III B. The red outline highlights the parts of the model that change in a step, and the thick blue arrows illustrate the flow of information. Details of these updates are presented in the text.

2. The parameters $\theta^{(m)}$ of the module are updated based on their previous states and the new $Z^{(m)}$ (Figure 2c).

3. The distribution of the module contribution $Y^{(m)}$ is then recalculated using the updated parameter values. The noise precisions $\{\tau_g\}$ are updated (Figure 2d) based on the first and second moments of all the $\{Y^{(l)}\}_{l=1}^{M}$.

4. The same procedure is in turn applied to the remaining modules in the schedule (Figure 2e) until convergence.

This iterative procedure, performed by updating local parameter distributions can be interpreted as a message passing algorithm, where sufficient statistics of parameter and data distributions are propagated across the graphical model [19]. Detailed exposition of the module update equations is given in Supplementary Methods.

B. Probabilistic models for common sources of gene expression variation

We provide implementations for a series of modules that are commonly used in the context of gene expression studies. Here, we outline their underlying probabilistic models, shown also as submodels in the Bayesian network of Figure 2a. For details on inference in these models, see Supplementary Methods.

1) Genotype effect module represents the probabilistic variant of the standard genetic association model, where some of the SNP genotypes have a linear effect on gene expression levels. The expression level $y_{g,j}^{(1)}$ of the $g$th gene probe in the $j$th individual is explained by linear effects of genotypes of $N$ SNPs $s_j = \{s_{1,j}, \ldots, s_{N,j}\}$ (Figure 2a, green plate):

$$P(y_{g,j}^{(1)} | s_j, b_g, u_g, \tau_g) = \mathcal{N}(y_{g,j}^{(1)} | \sum_{n=1}^{N} b_{n,g} \cdot (u_{n,g} s_{n,j}), \frac{1}{\tau_g})$$

$$P(b_{n,g}) = \text{Bernoulli}(b_{n,g} | \text{passe})$$

$$P(u_{n,g}) = \mathcal{N}(u_{n,g} | 0, 1).$$
The weights $u_g = \{u_{1,g}, \ldots, u_{N,g}\}$ indicate the magnitude of the effect of the SNP on the expression levels of genes $g$. The binary variables $b_g = \{b_{1,g}, \ldots, b_{N,g}\}$ determine whether the SNP effect is significant ($b_{n,g} = \text{true}$) or not ($b_{n,g} = \text{false}$). The prior probability $p_{\text{ass}}$ of an individual association controls the complexity of the model by influencing the number of significant associations, and can be interpreted as a significance threshold (Supplementary Methods). Since we treat SNPs as independent, only the strongest association for any single gene is considered, with the other $b_{n,g}$ forced to 0.

2) Known factor module accounts for the effect of known covariates $F$ of individual samples, such as environmental conditions, gender, or a population indicator. We use Bayesian regression to model the linear effect of the $C$ measured covariates in the $j$th individual, $f_j = \{f_{1,j}, \ldots, f_{C,j}\}$, where the corresponding weight vector for gene $g$ is $v_g = \{v_{g,1}, \ldots, v_{g,C}\}$ (Figure 2a, blue plate):

$$P(y_{g,j}^{(2)} | f_j, v_g, \tau_g) = \mathcal{N}(y_{g,j}^{(2)} | \sum_{c=1}^C v_{g,c} f_{c,j}, \frac{1}{\tau_g})$$

$$P(v_{g,c} | \alpha_c) = \mathcal{N}(v_{g,c} | 0, \frac{1}{\alpha_c})$$

$$P(\alpha_c) = \Gamma(\alpha_c | a_\alpha, b_\alpha).$$

The gamma prior on the inverse covariances $\alpha_c$ for weights of each factor introduces automatic relevance detection (ARD) [20, 21], driving the weights of unused factors to 0 and thereby switching them off. This provides complexity control of the model by regularizing the effective number of covariates.

3) Hidden factor module jointly infers the activation of hidden factors (such as unmeasured covariates and global effects on expression levels) and their gene-specific weights. Here, we use a probabilistic variant of the classical factor analysis model for this task. We have previously shown that this model captures hidden factors better than alternative linear models, such as probabilistic principal component analysis or independent component analysis [13]. Similarly to known factors, expression level of gene $g$ in individual $j$ is modeled by linear effects from a chosen number of $K$ hidden factors $x_j = \{x_{1,j}, \ldots, x_{K,j}\}$ (Figure 2a, red plate):

$$P(y_{g,j}^{(3)} | x_j, w_g, \tau_g) = \mathcal{N}(y_{g,j}^{(3)} | \sum_{k=1}^K w_{g,k} x_{k,j}, \frac{1}{\tau_g})$$

$$P(w_{g,k} | \beta_k) = \mathcal{N}(w_{g,k} | 0, \frac{1}{\beta_k})$$

$$P(x_{k,j}) = \mathcal{N}(x_{k,j} | 0, 1)$$

$$P(\beta_k) = \Gamma(\beta_k | a_\beta, b_\beta).$$

In contrast to the known factor module, the factor activations $X$ are unobserved random variables inferred from the expression levels. Again, the ARD prior switches unused factors off, thereby providing probabilistic complexity control [13, Results].

Additional probabilistic modules are straightforward to incorporate in the framework. PEER also allows non-Bayesian components to be included. We provide implementations of standard methods such as linear regression, principal component analysis (PCA), PCA with significance testing of the eigenvalues, and genetic association mapping with permutation testing. We also include a port to the R implementation of the surrogate variable analysis (SVA) method [9]. However, because uncertainty is not taken into account when performing inference in these modules, only point estimates are propagated to the rest of the model. These non-Bayesian modules are used mainly for comparison reasons in eQTL studies as described below. Their implementations are standard, and described in more detail in Supplementary Methods.

C. Framework configurations for expression QTL studies

The modules implemented in PEER can be combined according to the needs of the study. In the remainder of this paper, we investigate alternative configurations of PEER, combining genotype, known factor, and hidden factor modules to detect expression QTLs. To ensure comparability to published results and between alternative configurations, we use the standard genetic association model with per-gene permutation testing [4, Supplementary Methods] to find eQTLs. We train the framework, and then test for associations on a reduced dataset, where the estimated effects of known and hidden factors are subtracted out. This assures that eQTL finding performance is not confounded with the difference of the Bayesian and non-Bayesian association models.
1. **VBQTL - a probabilistic expression QTL model**

Variational Bayesian QTL mapper (VBQTL) is a configuration of the PEER framework with three modules. Probabilistic models for effects of genotype, known factors, and hidden factors outlined in previous section are used to explain gene expression variability (Figure 2a).

In experiments, we compare two alternative inference schedules of VBQTL. In iterative VBQTL (iVBQTL), the module parameters are learned using several iterations through all modules, first updating the genetic module, then known and hidden factors (Supplementary Methods). An important property of iVBQTL is that hidden factors are estimated jointly with the genetic state and known factors. This choice of schedule and the iterative learning help to ensure that variability that is due to genetic associations is not explained away by other modules (Results).

In cases where neither known nor hidden factors are correlated with the genetic state, their effect can be learned independently without running the risk of explaining away meaningful association signal. This motivates fast VBQTL (fVBQTL) that performs a single update iteration of the model, first inferring the contribution from the known and hidden factors, and then the association module. This simpler schedule can save significant computation time, since the state of the factor modules can be precalculated, and only a single iteration of the computationally more expensive association module is needed. In cases where the genetic state is approximately orthogonal to the known and hidden factors, this cheaper approximation performs equally with iVBQTL for finding genetic associations (Results).

2. **Non-Bayesian models for eQTL mapping**

We compare the VBQTL model with other configurations of the framework suited for association mapping. Similarly to VBQTL, they use three modules, estimating the effect of genotype, known factors, and hidden factors. We also consider a standard configuration that only accounts for known factor and genotype effects in the expression variability.

The inference schedule is the same for all considered methods. First, the known factor effects are estimated using the standard maximum likelihood regression module. Second, the hidden factor effects are learned. Finally, the standard genetic association module with permutation testing is run on the residuals of the estimated effects of known and hidden factors.

We use a single iteration of the framework to mimic a typical use of known and hidden factor modules as a preprocessing step on the expression dataset to account for sources of global non-genetic variability. Alternatively, the estimated hidden factors can be included in the association model, but the difference in results is negligible (data not shown). It is important to note, that since a single iteration is performed, the effect of the genotype is not taken into account when learning the hidden factors.

The differences between the alternative configurations are in the hidden factor module used, which in turn vary in the complexity control approach employed as highlighted below. Thus these alternative configurations are named after the hidden factor estimation method.

- **Standard model** explains the expression variability solely by the effects of known factors and SNP genotypes, without accounting for the hidden factors.

- **PCA** uses principal component analysis to detect hidden factors. In principle, PCA can explain all the variability in the data. Complexity is controlled by specifying the number of components to use as a parameter.

- **PCAsig** is an extension of PCA to account for hidden factors. In this model, complexity control is achieved via significance testing of eigenvalues, retaining only components that explain more variance than expected by chance at a specified significance cutoff (Supplementary Methods).

- **SVA** model extends PCAsig by accommodating a per-gene noise model, and explicitly allowing for sparse non-orthogonal components [9].

While VBQTL shares basic assumptions with the alternative configurations, there are a number of key differences. Firstly, it is a probabilistic model that operates with uncertainties in the parameter estimates as explained above. Secondly, the hidden factor module allows for non-orthogonal components, and provides probabilistic complexity control based on ARD. Thirdly, the iVBQTL schedule takes the genetic signal into account when estimating the hidden factor effect. Finally, the VBQTL model estimates a global gene-specific noise level shared across all modules, while the non-Bayesian modules either estimate noise levels separately (SVA) or assume noise-free observations (PCA, PCAsig).
IV. RESULTS

A. Simulation study highlights performance differences due to complexity control approaches.

We employed a simulated dataset in order to highlight the similarities and differences between the alternative framework configurations to find eQTLs. Our synthetic expression data combines linear effects from genetic associations (eQTLs), known factors, hidden factors, and gene-specific noise (Supplementary Methods). We use ten global factors whose influence varies significantly to simulate effects with a range of magnitudes, and treat three of them as known while keeping the remaining ones hidden. The factors represent sources of confounding variation that are encountered in the study of the real datasets. In addition to immediate cis associations between individual SNPs and genes, we chose three genes with a simulated eQTL, and simulated correlated expression levels for 15% of the genes for each, mimicking the effect of a transcriptional regulator (e.g. transcription factor or pathway component) with downstream effects. This introduces trans eQTLs, where the SNP effect is mediated by an intermediate expression level, and mimic the pivotal loci observed in several eQTL mapping studies [1, 3]. While this specific scenario may be biased in the comparative performance of different methods, it gives intuition for the results on real datasets discussed later.

1. Complexity control determines the accuracy of the hidden factor model.

We assessed the ability of the considered methods to recover the simulated confounding variability in the expression levels. All considered methods (except for the standard method) account for hidden factors. The complexity control parameters of the methods were varied to investigate their influence on the models’ performance. For methods that take the number of components in the hidden factor module as a parameter (PCA, VBQTL), performance for one to 50 hidden factors was compared. For significance testing based methods, we considered significance cutoffs $\alpha$ in the range [0.01, 0.5].

Figures 3a and 3b summarize the mean-squared error of the estimated hidden factor effects and the combined global effect of hidden factors and transcription factors. There was no difference in this scenario for the estimated contribution of known factors (data not shown). PCA was most accurate when it was allowed to learn 10 hidden factors, since in total, 7 hidden factors and 3 transcription factors were simulated. For larger number of components it overfitted, and started explaining away genetic signal, resulting in the increase in error. For a small number of components, transcription factor effects were explained away first, which increased the error in estimating the hidden factors alone. However, the estimates of the total global effects improved. PCAsig and SVA found 6 and 7 significant hidden factors for the wide range of significance cutoffs $\alpha \in [0.01, 0.5]$ tested, respectively. They failed to detect some of the weaker hidden effects that continued to mask the genetic signal, and underfit the data, performing similarly to PCA with the same number of components. While the significance testing based complexity control prevents these approaches from overfitting, only a single outcome is observed for a wide range of parameter settings, with the models settling to a rigid suboptimal solution.

fVBQTL and iVBQTL achieved the most accurate estimation of global variation. fVBQTL accurately estimated all the global effects, while iVBQTL was the sole method able to accurately account for only non-genetic effects, retaining the simulated transcriptional regulator signal. This is due to the difference in the inference schedule, where the contribution of the genetic effect is taken into account while learning hidden factors. Notably, unlike PCA, their performance did not degrade for large numbers of hidden factors in the model, exhibiting good complexity control in this scenario.

2. Hidden factor effect estimation accuracy is mirrored in eQTL finding sensitivity.

We determined the sensitivity and specificity of the considered methods for detecting the immediate and downstream confounding genetic associations. The significance of an eQTL was tested using 1000 permutations with a 1% per-gene false positive rate in the genetic association model (Supplementary Methods), in line with the approach employed in [4]. As a benchmark, the comparison includes eQTL finding using the standard method on both raw expression data (Standard), and an ideal case, where all the simulated confounding causes of variation are removed, but downstream associations maintained (Ideal).

The accuracy of the hidden factor effect estimation was indicative of the immediate eQTL finding sensitivity (Figure 3c). Specificity was consistent with the chosen false positive rate for all methods (data not shown). For cis associations, fVBQTL and iVBQTL recovered more true eQTLs compared to other methods, approaching the performance of the ideal case, and mirroring the accuracy of estimating hidden factor effects. PCA overfit when the number of components used was greater than the true number of simulated components, explaining away genetic
FIG. 3: Sensitivity of recovering simulated hidden factor effects and eQTLs for Bayesian and non-Bayesian methods. (a) Mean-squared error in estimating the contribution from the hidden factors. (b) Mean-squared error in estimating the contribution from the hidden factors and transcription factor effects. (c) Sensitivity of recovering immediate SNP associations. (d) Sensitivity of recovering downstream associations.

Seven hidden factors and three transcription factor effects were simulated. For eQTL sensitivity, standard eQTL finding on simulated data (Standard) and same data without the hidden effects (Ideal) are included as comparisons. PCAsig and SVA identify a constant number of hidden components (marked with a diamond shape), thus only one measurement (dashed line) is given.

All hidden factor methods except iVBQTL explained away simulated \textit{trans} eQTL hotspots with multiple genes participating in single association signals. (Figure 3d). This is explained by the global factor effect estimation accuracy, where iVBQTL alone refrained from explaining the hotspots away as a global factor. The standard method found nearly all the original \textit{trans} associations, actually outperforming methods that explain away confounding variability. Thus, in cases where there is true genetic signal with widespread downstream effects, its contribution needs to be taken into account to retain its relation to genotype, and avoid attributing it to a confounding global cause. This is straightforward in our framework, and is demonstrated by the good performance of iVBQTL in this scenario. iVBQTL retained the original associations, while explaining away non-genetic causes of expression variability, thus adding power to detect the weaker, masked eQTLs. The results are similar to the ones observed in the study of crosses of inbred strains below.

These results suggest that it is important to account for the confounding sources of variation in expression levels, while keeping the signal of the genetic state. Correct complexity control is required to avoid over- and underfitting in order to achieve optimal sensitivity for detecting genetic associations.
B. VBQTL finds additional expression QTLs in real datasets.

We investigated the application of our framework in the context of finding expression QTLs on real data. The configurations considered in the simulation study were applied to yeast [2], mouse [3] and human [4] eQTL mapping datasets. These represent common study designs of an outbred population (human), and a population of crosses between inbred strains (yeast, mouse). We considered 5, 15, 30, and 60 hidden factors for PCA and VBQTL, and 0.01, 0.1 and 0.3 significance cutoffs for SVA and PCAsig. Expression QTLs were detected analogously to the simulation scenario.

1. Accounting for hidden factors adds power to detect cis eQTLs in an outbred population

We applied the methods on the genotype and expression data from 90 individuals of the CEU (CEPH from Utah) HapMap phase 2 samples [4, 22]. The data consisted of genotypes of 55,000 SNPs and expression levels of 618 probes from chromosome 19 (results for three more chromosomes are similar, data not shown). The expression levels were measured in EBV-transformed lymphoblastoid cell lines of healthy individuals. The gender covariate was included as a known factor for all methods. We did not consider probes with overlapping SNPs. Following [4], an association was called to be in cis when the SNP was within 1Mb from the probe midpoint and in trans otherwise.

The standard method found the least gene probes with a cis association (27), suggesting that strong confounding sources of variation are present in this dataset. The number of identified probes with a trans association was not significantly higher than expected by chance at the chosen FPR, which is in line with previous results [4], and suggests little intrachromosomal trans regulation.

PCA, the simplest method for accounting for hidden factors, found additional associations when up to 30 principal components were used, but less for 60 components. This is expected, since there are no more than 90 degrees of freedom in this dataset, and 60 principal components accounted for over 94% of the variance (Supplementary Table XI), and hence explained away the genetic association signal.

The significance testing based methods, SVA and PCAsig both found additional associations compared to the standard method. It is remarkable that both find a constant number of significant hidden factors for the wide range $\alpha \in \{0.01, 0.1, 0.3\}$ of significance cutoffs considered, again exhibiting rigid complexity control. The performance of SVA with the 13 hidden factors found is close to performance of PCA with 15 components (50 vs. 53 probes with an association). Similarly, PCAsig with the 6 significant components performs comparably to PCA with 5 components (48 vs. 46 probes with an association). This shows the intrinsic similarity of these methods to PCA, as was also observed in the simulation scenario.

iVBQTL and fVBQTL found more probes with an association (68 and 59) than all other methods, representing an almost three fold increase in the number of genes with a cis eQTL. Complexity control assured that the performance saturated for large enough number of factors and did not degrade as for PCA.

It is important to note that the model performance depends on two aspects. First, the model complexity control, regulating the amount of variance explained, is important to ensure that meaningful signal is not attributed to a confounding source. Overfitting in case of PCA for a large number of components is an example of such an effect. Second, while alternative hidden factor modules explained similar amounts of variance, their performance differed due to the underlying model. For example, SVA and fVBQTL both explained about 60% of variance (Supplementary Table XI), yet iVBQTL identified additional associations. These findings are consistent with the simulation study results, and suggest that the additional associations found with Bayesian models are due to a more sensitive underlying model and better complexity control.

2. Accounting for hidden factors adds power to detect cis associations in crosses between inbred mouse and yeast strains.

Next, we applied the methods to two datasets of inbred strain crosses. The yeast expression dataset [2] (GEO[23] accession GSE1990 with genotypes provided by authors) contained 7084 expression measurements and 2925 genotyped loci in 112 crosses of segregating yeast strains. The mouse expression data consisted of 23,698 expression measurements for 111 F2 mouse lines, and genotypes at 137 genetic markers. An association was called to be in cis if the probe and genotyped locus were from the same chromosome, and in trans otherwise.

The relative performance of different methods was similar to their ability to detect cis eQTLs in the outbred population dataset. The absolute performance gain was significantly lower for all methods, however, suggesting that the genetic signal is stronger compared to confounding sources of variation. All factor methods identified additional associations compared to the standard method. PCA overfit for larger numbers of principal components used, explaining away genetic association signal. SVA and PCAsig found the same number of significant hidden
FIG. 4: Number of probes with an eQTL found as a function of maximum number of hidden factors for three published datasets. Significance testing based methods (PCAsig, SVA) identify the same number of factors for a wide range of cutoff values ($\alpha \in [0.01, 0.3]$), thus only one count is given (dashed lines), together with the number of factors found (diamond shape). Other methods were applied with a maximum number of 5, 15, 30 and 60 hidden factors.

3. Iterative learning scheme overcomes difficulties in detecting trans associations for crosses of inbred strains.

All methods found additional trans associations in mouse, but fewer than the standard method in yeast (Figure 4d, 4e). In yeast, the more variance was explained by the hidden factors, the less trans eQTLs were found, suggesting that the global determinants of gene expression variation were correlated with the genetic state. Indeed, the inferred hidden factor levels were correlated with genotypes of “pivotal loci” that are associated with expression levels of hundreds of genes.

The iVBQTL schedule is designed to retain the genetic signal, and found substantially more trans associations than other methods. However, it was surprising that it found less associations than the standard model. This difference was due to the association mapping method. iVBQTL uses a Bayesian association model in learning, and the permutation testing based module in the end to call associations. When we used the the computationally more intensive maximum likelihood permutation testing module in the iVBQTL model, more trans associations (1732 and 1698 for 30 and 60 factors, respectively) were found compared to the standard model, while the number of significant cis associations remained similar (1059 and 1016).

The effect of pivotal loci has been observed before, and interpreted in different ways [9, 14]. It could be that the additional variation is artefactual, and correlated to the genetic state by chance. In this case, all the original trans associations are spurious. The alternative explanation is that the genotype of these loci have real downstream effects on the expression profiles of the genes. Then the variance is not confounding the genetic signal, but in fact is a part of it, and should not be explained away.

Previous methods do not provide consistent ways of dealing with this issue. The SVA approach suggests removing
FIG. 5: Fraction of tested genes with a \textit{cis} association in individual chromosomes and effective overall false discovery rate for the HapMap CEU population (FPR=0.1%).

the effect of the primary variable first, however, they do not consider accounting for the genetic effect in their application to the same yeast dataset [9]. In [14], the application of a correction procedure also explains away \textit{trans} associations. We provide a principled approach for dealing with this situation. The iVBQTL scheduling takes the genetic state into account when learning the hidden factors, and is thus more sensitive to genetic associations. This was reflected in its ability to retain the original eQTLs and find additional ones when the same genetic association model was used in iterations and the final association calling.

C. Application of PEER recovers three times more probes with a \textit{cis} eQTL in a whole-genome scan of HapMap phase II data.

Motivated by the results of the initial study of a single human chromosome, we applied the fVBQTL configuration of PEER, learning 30 hidden factors, to the 10,000 most variable expression probes of the HapMap 2 dataset. We searched for \textit{cis} eQTLs in the original expression data (standard eQTLs) as well as the residuals of fVBQTL (VBeQTLs), using 10,000 permutations and 0.1% per-gene FPR analogously to [4].

On the CEU population, we found 1156 genes with a VBeQTL at false discovery rate (FDR) of 0.9%, and 394 genes with a standard eQTL at FDR of 2.5% (Figure 5). This result represents nearly a threefold increase in the number of genes with an association, and is consistent across chromosomes. A similar increase in associations was found for other populations (Supplementary Tables II,III).

We repeated this genome-wide experiment on pooled populations, where we included the population membership as a known factor to avoid confounding effects of population structure. Due to the increased sample size, it was possible to detect additional associations. We found 3003 genes with a VBeQTL compared to 1203 genes with a standard eQTL at the 0.1% FPR (Figure 6a and Supplementary Table IV). The VBeQTLs in the pooled sample cover 30% of all the considered probes, suggesting that the number of human genes whose expression levels are affected by common \textit{cis}-acting genetic variation may be significantly higher than previously shown [24, 25]. This additional abundance of associations suggests that detection of \textit{cis} eQTLs has not been saturated and larger sample sizes may lead to evidence of even more extensive \textit{cis} regulation by common polymorphisms.

D. Additional associations are due to increased sensitivity.

It is important to demonstrate that the additional associations found after removing the learned non-genetic factors are biologically meaningful. We provide evidence that the additional associations found in HapMap phase 2 data are real in three ways. Firstly, we investigated how many of the genes with a VBeQTL in each of the three populations individually were replicated using raw data from a pooled data set containing all populations. Note that this will only validate weak associations that occur in multiple populations – we would not expect weak population-specific associations to be replicated in the pooled data set. However, we expect many of the associations to be replicated in multiple populations [24].

The Venn diagram in Figure 6b summarizes the overlap of the set of identified genes with an eQTL. A total of 47% of the additional associations found in the CEU population were recovered using the standard method in the pooled
population. The remaining additional associations may be explained by even weaker signals that were recovered by applying fVBQTL, or as population-specific effects that do not stand out in the pooled sample.

Secondly, we evaluated to what extent the additional genes with a VBeQTL in a single population were replicated in other populations. For instance, 43% of genes with a CEU VBeQTL were replicated on the YRI population (Figure 6d), and 63% on the CHB+JPT population (Figure 6e). These overlaps are consistent with overlaps of standard eQTLs, and are similar for other populations (Supplementary Table VI), and alternative methods accounting for hidden factors.

Finally, we validated that the novel associations share similar properties with the original ones. We analyzed the distribution of the position of additional cis associations around the gene start along with the association LOD scores. The additional VBeQTLs have very similar characteristics to the standard eQTLs, being concentrated around the gene start (Figure 6c, 6f), in line with results from [24].

E. Trans associations are also found using VBQTL, while not explaining away regulatory hotspots

We tested SNPs in each of 34,143 linkage blocks as defined by recombination hotspots [26, 27] for association with the 10,000 most variable gene probes for the three HapMap populations. Since full permutation testing on the complete data is computationally infeasible, we restricted the tests to SNPs without missing values and used a conservative correction for the low number of permutations as described in Supplementary Methods.

For the CEU population and a corrected significance level which corresponds to an approximate FPR of $10^{-8}$, we found 294 standard eQTLs (FDR=1.2%) and 774 VBeQTLs (FDR=0.4%) with the same 30 factors learned as the cis study. This corresponds to 205 and 537 unique probes respectively.

To differentiate between cis and trans associations, we called an association trans if it is at least 5Mb away from the probe midpoint. Using this criterion, 84% of the standard eQTLs and 87% of the VBeQTLs were in cis. Counting only trans associations, we found 162 VBeQTLs and 126 standard eQTLs in at least one population, and 48 VBeQTLs and 32 standard eQTLs in at least two populations.

The same analysis was carried out on the remaining two HapMap populations identifying unique associations in all populations and their overlap. Similarly to cis eQTLs, the additional trans associations replicated in different populations with rates between 30% (YRI in CEU) and 57% (CEU in CHB+JPT) (Supplementary Table VII). The distribution of genomic locations of regulatory SNPs and target probes is depicted in Figure 7.
FIG. 7: Distribution of significant associations (FPR=10^{-8}) for (a) Standard eQTLs and (b) VBBeQTLs pooled over the three HapMap populations. In grey: cis associations (SNP distance to probe midpoint < 5Mb), blue: trans associations. The size of the dot is indicative of the overlap between populations (in one, two or all three populations). 13 standard and 25 VBBeQTLs are replicated in all populations; in plots, some of them overlap due to proximity of the linkage regions. The association counts below the diagrams sum up trans associations shared in at least two populations in a particular bin.

F. Interpretation of learned hidden factors.

The hidden factor modules hypothesize a set of unobserved non-genetic factors that influence the measured gene expression levels. To gain insights into their interpretation we considered correlations to known effects such as gender or population.

We applied fVBQTL to expression data from individuals of all three HapMap populations, and correlated the inferred hidden factors to the population and gender indicator variables. The correlation coefficients (Supplementary Table VIII) indicate that many of the learned latent causes are correlated with population and that one is strongly correlated with gender. This implies that the hidden factor module can recapture variance in the gene expression levels due to true underlying properties of individuals.

A recent study in yeast looked for changes in eQTLs when strains were grown in different media [28]. We applied fVBQTL to the expression data of this study (GEO accession GSE9376), without including any information about the growth condition. The first hidden factor learned was highly correlated with the indicator variable for the growth condition ($r^2 = 0.96$), demonstrating that the VBQTL model can successfully recover an environmental effect if it is present.

The global factors identified can be further analyzed for biological signals, looking for GO term over-representation in the genes that they affect. We used the ordered GO profiling method [29] to find significantly enriched GO categories for 30 genes most affected by each factor. Recent results [30] show that related linear Gaussian models find biologically relevant factors in the yeast expression dataset. We replicated these findings with our model, yielding factors enriched in biological functions, including sugar, alcohol and amino acid metabolic processes (Supplementary Information). Similar analysis in human and mouse did not show significant over-representation of GO categories, providing no evidence that the main axes of variation in the expression levels for these experiments are due to common biological function. This could be due to poor annotation of the genes, gene features not related to biological function, or more technical sources of global variation, such as cell culture conditions [12].
V. DISCUSSION

We have presented PEER, a probabilistic framework for dissecting phenotypic variation. Our framework features the flexibility to combine modules for different sources of variability, and we have provided implementations for a set of commonly used ones. Bayesian inference enables propagating uncertainty between parts of the model, and inferring posterior distributions over parameter estimates for more sensitive analysis. VBQTL, an application of the framework tailored for eQTL studies, outperformed alternative framework configurations for association finding on real datasets. In the most striking example, VBQTL found three times more eQTLs than a standard method, and 30% more compared to the best alternative in the HapMap 2 expression dataset.

Our framework advances the methodology for understanding phenotypic variation. It allows to straightforwardly combine models for explaining the observed variability. Adapting the framework and choosing which modules to include is a matter of updating the configuration, and no additional development work is needed. It is also easy to implement new modules. The only requirements for any module are the ability to update its inferred parameters based on the residual dataset from the network, and to produce a prediction of its contribution based on the parameter values. Notably, non-Bayesian modules can also be included, as we demonstrated with PCA, SVA, and regression modules.

The PEER framework employs module-level complexity control that performed well in experiments. The results of the simulation study exemplify how correct complexity control is required to avoid over- and underfitting for more sensitive analysis. Explaining away too much variability removes some signal of interest from the data, and failing to account for all sources of confounding variation decreases power to detect the real signal. When the variable of interest is correlated with many gene expression levels, its effect can be explained by the hidden factor module. The complexity control and inference schedule determine the behavior of the model in these cases. There can be no silver bullet solution that provides perfect results in any scenario with no supervision. Instead, modeling assumptions must be made explicit, and incorporated in the analysis, as is elegantly done in the Bayesian setting.

VBQTL and other configurations of the PEER framework accounting for hidden factors found additional expression QTLs in the datasets studied compared to the standard method. It is remarkable that, with only 270 samples, and looking in one tissue type, we can find significant genetic associations to 30% of the expressed genes. While similar results have been reported before, we have shown a threefold increase in the number of associations for the HapMap dataset, and analyzed their repeatability and location distribution. The replication of the additional associations in different populations suggests that they are genuine. The increase in power is due to the hidden factor module explaining away non-genetic variability, and allowing the genetic effects to stand out to a greater extent. High number of additional associations suggests that association finding studies in human have not saturated, and we expect the fraction of genes with an eQTL will increase as the number of samples increases. It may be that the expression of majority of human genes varies as a result of segregating genetic variation.

In conclusion, we believe that our framework provides a principled and accurate way to study gene expression and other high-dimensional data. Increasingly complex models combining genetic and other effects can explain significantly more of the variance in observed phenotypes, as suggested by this study and others. Our framework provides the flexibility to facilitate these richer models, for example, we have already started exploring interaction effects as an additional module of the framework. It will be interesting to see how these approaches can contribute to our understanding of human disease genetics, potentially involving intermediate phenotypes such as gene expression and other factors.

The software used in this study is freely available online at http://www.sanger.ac.uk/software/analysis/peer

VI. ACKNOWLEDGMENTS

The authors would like to thank Manolis Dermitzakis for helpful discussions and feedback, Rachel Brem, Eric Schadt and Barbara Stranger for access to their expression and genotype data, Claude Beasley for assistance, Richard Bourgon for advice, and Alexandra Nica and members of the Durbin group for comments on the manuscript.

Author contributions. The statistical models were developed by OS, LP, JW, and RD. LP and OS designed and performed the experiments. All authors contributed to writing the paper, the research was conceived by RD and JW.

Funding. This work was supported by the Wellcome Trust [grant number WT077192/Z/05/Z] and the Technical Computing Initiative (Microsoft Research). OS received funding from EPSRC and the Cambridge Gates Trust. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the
Competing interests. The authors have declared that no competing interests exist.

APPENDIX A: SUPPLEMENTARY METHODS

1. Implementation of non-Bayesian modules

a. Standard expression QTL model

To ensure a common ground when comparing different methods, we used a well established linear regression approach [15] to detect associations. For each tested SNP \( n \) with genotype \( s_{n,j} \) and gene \( g \) with expression level \( y_{g,j} \), we evaluated the log-odds (LOD) score

\[
L_{n,g} = \log_{10} \left( \prod_j \frac{P(y_{g,j} | s_{n,j}, u_{n,g})}{P(y_{g,j} | \theta_{bck})} \right),
\]

which assess how well a particular gene expression level is modeled when the genetic state is taken into account, compared to how well it is model-led by a background model ignoring the genetic effect. The terms \( u_{n,g}, \theta_{bck} \) are parameters for probe \( g \) and SNP \( n \) of the genetic and background models respectively. The probe expression levels \( y_{g,j} \) can either be the raw measurements, or residuals after subtracting the estimated effect of hidden and known factors.

Significance of an association was evaluated via permutation testing [31]. The analysis in Equation (A1) was repeated with permuted expression levels with respect to the genetic state, calculating the distribution of null log-odds scores. An eQTL was called significant if \( L_{n,g} \) was greater than \( \hat{L}_{n,g} \), the \( \delta \) tail of the null distribution for a given false positive rate (FPR) \( \delta \). The same set of permutations was used for all methods.

To account for multiple testing, we estimated a single significance threshold \( \hat{L}_g \) per gene for all tested SNPs. This was done by taking the maximum LOD score over SNPs for a given permutation and using this score distribution when estimating the \( \delta \) tail [4]. The posterior of the switch variable for the probabilistic genetic module is not used for the final tests to put all methods on equal footing.

b. PCA

PCA can be interpreted as decomposition of the gene expression matrix \( Y = (y_1...y_G) \) into a product \( UD^T \), where \( U \) is the matrix of left singular vectors, \( D \) is a diagonal matrix of singular values \( \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N \geq 0 \), and \( V \) is the matrix of right singular vectors. To apply PCA, we used \( U \) as the weight matrix \( W \), and \( DV^T \) as the latent factors \( X \). For the benchmark figures, illustrating the effect for different numbers of factors, we limited the number of learned factors to a given number \( K \) by setting \( D_{i,i} = 0 \) for \( i > K \).

c. PCAsig

PCAsig module is an extension of PCA, where complexity is controlled by retaining only components that explain more variance than expected by chance. Significance testing of PCA components in the PCAsig module was performed analogously to SVA [9], but without enforcing uniformity of the p-values. We found the variance explained by each component \( i \) by calculating the statistic \( d_i = \frac{\lambda_i^2}{\sum_{j=1}^{N} \lambda_j} \). We then permuted the columns of \( Y \) \( L \) times, calculating null statistics \( d_{1,1}, d_{1,2}, ..., d_{1,L} \) analogously. Given a cutoff value \( \alpha \), component \( i \) was deemed to be significant if the fraction of null statistics greater than \( d_i \) was less than \( \alpha \).

d. SVA

SVA package was downloaded from http://www.genomine.org/sva, and applied to datasets with default parameters, using 100 permutations and varying only the significance cutoff. The module implementation uses a Python to R bridge provided by RPy (http://rpy.sourceforge.net).
2. PEER framework

Here, we give a full self-contained treatment of the PEER framework model and inference in it.

a. Likelihood models

PEER framework. The likelihood model of PEER for observed expression levels $Y$ is

$$P(Y | Y^{(1)}, \ldots, Y^{(M)}, \tau) = \mathcal{N}(Y | Y^{(1)} + \ldots + Y^{(M)}, \Sigma),$$

where $\Sigma = \text{diag}\left\{ \frac{1}{\tau_g} \right\}$ is the diagonal matrix constructed from noise precisions $\tau_g$. The noise model is per gene, similar to a factor analysis model, where Gamma priors are put on the noise precisions $P(\tau_g) = \Gamma(\tau_g | a, b)$. In experiments we use vague Gamma prior parameters, $a = 1, b = 100$. Each of the $M$ modules itself depends on parameters $\theta^{(m)}$ and possibly other data $D^{(m)}$:

$$P(Y^{(m)} | \theta^{(m)}, D^{(m)}).$$

Genotype effect module. The expression level $y_{g,j}^{(1)}$ of the $g$th gene probe in the $j$th individual is explained by a linear effects of genotypes of $N$ SNPs $s_j = (s_1, \ldots, s_N)$:

$$P(y_{g,j}^{(1)} | s_j, b_g, u_g, \tau_g) = \mathcal{N}(y_{g,j}^{(1)} | \sum_{n=1}^{N} b_{n,g} \cdot (u_{n,g} s_{n,j}) \cdot \frac{1}{\tau_g})$$

$$P(b_{n,g}) = \text{Bernoulli}(b_{n,g} | p_{\text{pass}})$$

$$P(u_{n,g}) = \mathcal{N}(u_{n,g} | 0, 1).$$

The weight $u_g = (u_{1,g}, \ldots, u_{N,g})$ indicates the magnitude of the effect, and the binary variable $b_g = (b_{1,g}, \ldots, b_{N,g})$ determines whether it is significant (true) or not (false) based on the Bernoulli prior on the switch variable $P(b_{n,g}) = \text{Bernoulli}(b_{n,g} | p_{\text{pass}})$. When the switch variable is on, the expression level is linearly influenced by the SNP, and unaffected otherwise. The LOD score of the association model in Section A1a is closely related to the switch variable $b_{n,g}$. For a particular parameter setting, the posterior probability over the switch state $b_{n,g}$ is a monotonically increasing function of the LOD score. The exact relation is $P(b_{n,g} | y_{g,j}, s_{j,n}) = \sigma(\text{LOD score})$ where $\sigma()$ is the base 10 sigmoid function $\sigma(x) = 1/(1 + 10^{-x})$.

2) Known factor module. The effect of the measured $C$ covariates in the $j$th individual, $f_j = (f_{1,j}, \ldots, f_{C,j})$, where the weights of their effect on a gene $g$ is $v_g = (v_{g,1}, \ldots, v_{g,C})$ is modeled as:

$$P(y_{g,j}^{(2)} | f_j, v_g, \tau_g) = \mathcal{N}(y_{g,j}^{(2)} | \sum_{c=1}^{C} v_{g,c} f_{c,j} \cdot \frac{1}{\tau_g})$$

$$P(v_{g,c} | \alpha_c) = \mathcal{N}(v_{g,c} | 0, \frac{1}{\alpha_c})$$

$$P(\alpha_c) = \Gamma(\alpha_c | a_c, b_c).$$

The gamma prior on the inverse covariances for each factor introduces automatic relevance detection (ARD) [20, 21], driving the weights of unused factors to 0 and thereby switching them off. This is explained in more detail below.

3) Hidden factor module. Analogously to known factors, expression variability is modeled by a linear effects from $K$ hidden factors $X = \{x_1, \ldots, x_K\}$:
\begin{align}
P(y_{g,j}^{(3)} \mid x_j, w_g, \tau_g) &= \mathcal{N}(y_{g,j}^{(3)} \mid \sum_{k=1}^{K} w_{g,k} x_{k,j}, \frac{1}{\tau_g}) \tag{A10} \\
P(w_k, \beta_k) &= G \prod_{g=1}^{G} \mathcal{N}(w_{g,k} \mid 0, \frac{1}{\beta_k}) \tag{A11} \\
P(x_{k,j}) &= \mathcal{N}(x_{k,j} \mid 0, 1) \tag{A12} \\
P(\beta_k) &= \Gamma(\beta_k \mid a_{\beta}, b_{\beta}) \tag{A13}
\end{align}

The factor activations \( X \) are random variables that are not observed, but instead inferred from the expression levels. Again, the ARD prior allows unused factors to be switched off. This forces the model to learn factors which have a broad effect on many expression levels. In experiments we used values \( a_{\alpha} = 10^{-6}G \) and \( b_{\alpha} = G \), where \( G \) is the total number of gene probes. Similar prior settings are used for the weights of the known factors \( v_c \). We use a standard normal prior on the hidden factors \( x_{k,j} \sim \mathcal{N}(x_{k,j} \mid 0, 1) \).

b. Variational inference

As outlined in Section III A, we use variational Bayesian learning in the framework. The basic principle of variational methods \[18\] is to approximate the exact joint posterior distribution over all parameters by a factorized \( Q \)-distribution. Individual factors of the \( Q \)-distribution are refined by minimization of the KL-divergence between the exact and the approximate distributions with respect to the parameters of a single factor. This leads to an iterative algorithm, updating individual factors of the approximate distribution given the state of all others. Here, we give the factorizations and update rules for the general framework and the individual modules.

**PEER framework.** We approximate the exact joint posterior distribution over all parameters

\[
P(\{Y^{(m)}\}_{m=1}^{M}, \{\theta^{(m)}\}_{m=1}^{M}, \mathcal{D})
\]

by a factorized approximation over parameters for individual modules

\[
Q(\Theta) = \prod_{m=1}^{M} Q(\theta^{(m)})Q(Y^{(m)}).
\]

(A15)

Here we defined the abbreviation \( \mathcal{D} = \{Y, \{D^{(m)}\}_{m=1}^{M}\} \), summarizing all observed data; expression levels \( Y \) as well as module specific data \( \{D^{(m)}\}_{m=1}^{M} \). Note that as the expression contributions \( Y^{(m)} \) are not observed they are also parameters that need to be inferred. Strictly speaking they are not random variables of the model, but Gaussian messages that comprise the first and second moments of the module contribution. The distributions of parameters \( \theta^{(m)} \) for individual modules are in turn factorized. The set \( \Theta = \{\theta^{(1)}, \ldots, \theta^{(M)}\} \) denotes the set of all parameters from all modules.

The approximate \( Q \)-distributions are updated iteratively, taking the current state of all others into account. Update equations for a particular \( Q_i \) can be derived by functional minimization of the KL-divergence between \( P \) and \( Q \) with respect to \( Q_i \) which leads to

\[
\tilde{Q}(\Theta_i) \propto \exp \{\langle \log P(\mathcal{D}, \Theta) \rangle_{Q(\Theta), i \neq j} \}.
\]

(A16)

The term in the exponent is the expectation of the model log-likelihood under all other \( Q \)-distributions. Together with the expression data likelihood

\[
P(Y \mid \Theta) = \mathcal{N}(Y \mid Y^{(1)} + \cdots + Y^{(M)}, \Sigma) \prod_{m=1}^{M} P(Y^{(m)} \mid \theta^{(m)}, D^{(m)})
\]

(A17)

this allows generic update rules for all model parameters to be derived. Substituting in Equation (A15) for each \( Q(\cdot) \) we obtain the following approximate distributions...
(Approximate distributions)

\[
Q(\tau) = \prod_{g=1}^{G} \Gamma(\tau_g | \tilde{\alpha}_{\tau_g}, \tilde{\beta}_{\tau_g})
\]

(A18)

\[
Q(\mathbf{Y}(m)) = \prod_{g=1}^{G} \prod_{j=1}^{J} \mathcal{N}(y_{g,j}^{(m)} | \tilde{m}_{\mathbf{Y}(m), g}, \frac{1}{\tilde{\tau}_{\mathbf{Y}(m), g}}).
\]

(A19)

and similar factorizations for each of the modules (given below). The parameter update equations for the framework follow as

(Update rules)

\[
\tilde{\alpha}_{\tau_g} = a_{\tau} + \frac{1}{2} \sum_{j=1}^{J} \left( y_{g,j} - \frac{1}{M} \sum_{m=1}^{M} y_{g,j}^{(m)} \right)^2
\]

(A20)

\[
\tilde{\beta}_{\tau_g} = b_{\tau} + \frac{J}{2}
\]

(A21)

**Genotype effect module** The update equations for the modules introduced in Section III B follow similarly. For the modules, we give the approximate factorizations employed, and the resulting update equations that are derived in identical manner to the treatment above.

(Approximate distributions)

\[
Q(\mathbf{B}) = \prod_{n=1}^{N} \prod_{g=1}^{G} \text{Bernoulli}(b_{n,g} | \tilde{p}_{n,g})
\]

(A22)

\[
Q(\mathbf{U}) = \prod_{n=1}^{N} \prod_{g=1}^{G} \mathcal{N}(u_{n,g} | \tilde{m}_{n,g}, \tilde{\Sigma}_{n,g})
\]

(A23)

(Update rules)

\[
\tilde{\Sigma}_{n,g} = I + \langle \tau_g \rangle \langle \tau^2_n \rangle \sum_{j=1}^{J} s_{n,j} s_{n,j}
\]

(A24)

\[
\tilde{\Sigma}_{n,g}^{-1} = \tilde{\Sigma}_{n,g}^{-1} \left( \langle \tau_g \rangle \langle \tau_n \rangle \sum_{j=1}^{J} s_{n,j} \langle z_{n,j}^{(1)} \rangle n \right)
\]

(A25)

\[
\tilde{m}_{n,g} = \tilde{\Sigma}_{n,g}^{-1} \left( \langle \tau_g \rangle \langle b_{n,g} \rangle \sum_{j=1}^{J} s_{n,j} \langle z_{n,j}^{(1)} \rangle n \right)
\]

(A26)

\[
\tilde{m}_{n,g} = \frac{1}{N} \sum_{n=1}^{N} \langle b_{n,g} \rangle \langle u_{n,g} s_{n,j} \rangle
\]

(A27)

\[
\tilde{\tau}_{n,g}^{(1)} = \frac{1}{N} \sum_{n=1}^{N} \langle b_{n,g} \rangle \langle u_{n,g} s_{n,j} \rangle
\]

where we define

\[
\langle z_{n,j}^{(1)} \rangle = z_{n,j}^{(1)} - \sum_{m \neq n} \langle b_{m,g} \rangle \langle u_{m,g} \rangle s_{m,j}
\]

(A28)
and the residual expression dataset for the $m$th module

$$z_{g,j}^{(m)} = y_{g,j} - \sum_{l \neq m}^{M} y_{g,j}^{(l)}.$$  \hfill (A29)

The approximate posterior over the indicator variables can be obtained from

$$\tilde{p}_{b_{n,g}} \propto p_{b} \cdot \exp \left\{ -\frac{1}{2} \sum_{j=1}^{J} \left( (z_{g,j}^{(1)})_{n} - b_{n,g} \theta_{n,g} \right)^{2} \right\}$$

$$(1 - \tilde{p}_{b_{n,g}}) \propto (1 - p_{b}) \cdot \exp \left\{ -\frac{1}{2} \sum_{j=1}^{J} \left( (z_{g,j}^{(1)})_{n} \right)^{2} \right\},$$ \hfill (A31)

which after normalization gives rise to $\tilde{p}_{b_{n,g}}$.

The genetic module performs independent linkage tests between all SNPs and genes and hence yields one $Q$ distribution for all gene-SNP pairs. Hence when performing updates of the factor models we employ a bottleneck approximation, only considering the most explanatory SNP per gene probe $g$ to estimate the expectation value in the exponential.

**Known factor module** is identical in treatment to the **hidden factor module**, without the need for updates of the factor activations. Thus, we only present the hidden factor module here.

(A32)

\begin{align*}
Q(X) &= \prod_{j=1}^{J} \mathcal{N}(x_{j} | \tilde{m}_{x_{j}}, \tilde{\Sigma}_{x_{j}}) \hfill (A34) \\
Q(W) &= \prod_{g=1}^{G} \mathcal{N}(w_{g} | \tilde{m}_{w_{g}}, \tilde{\Sigma}_{w_{g}}) \hfill (A35) \\
Q(\beta) &= \prod_{k=1}^{K} \Gamma(\beta_{k} | \tilde{a}_{\beta_{k}}, \tilde{b}_{\beta_{k}}) \hfill (A36)
\end{align*}

\begin{align*}
\tilde{\Sigma}_{x_{j}} &= \Sigma_{x_{j}} + \langle W^{T} \text{diag} (\tau) W \rangle \hfill (A37) \\
\tilde{m}_{x_{j}} &= \tilde{\Sigma}_{x_{j}}^{-1} \langle W^{T} \rangle \text{diag} (\tau) \left( \langle z_{j}^{(3)} \rangle \right) \hfill (A38)
\end{align*}
\[ \tilde{\Sigma}_{w_g} = \text{diag} \langle \beta \rangle + \langle \tau_g \rangle \sum_{j=1}^{J} \langle x_j x_j^T \rangle \]  
(A39)

\[ \tilde{\mathbf{m}}_{w_g} = \tilde{\Sigma}_{w_g}^{-1} \langle \tau_g \rangle \sum_{j=1}^{J} \langle x_j \rangle \left( \langle z_j^{(3)} \rangle \right) \]  
(A40)

\[ \tilde{\mathbf{m}}_{y_g^{(3)}} = \sum_{k=1}^{K} \langle w_{g,k} \rangle \langle x_{j,k} \rangle \]  
(A41)

\[ \tilde{\tau}_{y_g^{(3)}} = \left[ \sum_{n=1}^{N} \langle b_{n,g}^2 \rangle \langle u_{g,n}^2 \rangle s_{n,j}^2 \right] \]  
(A42)

\[ (A43) \]

**Initialization.** The initial states of hidden factor module weights \( Q(w_g) \) and levels \( Q(x_j) \) are determined from a PCA solution, and the weights for known factors \( Q(v_g) \) are initialized to the maximum likelihood estimate. The remaining \( Q \) distributions for all modules are initialized randomly, sampling from the prior. A random initialization is possible for the factor model as well; we have not explored its consequences in this work.

3. VBQTL

Both the iterative (iVBQTL) and the fast variant (fVBQTL) of the studied algorithms use these update equations presented above. iVBQTL uses the full variational approximation with a specific update order of the \( Q(\theta_i) \) distributions (Algorithm 1b). In experiments, we used 3 iterations of the full model. Within each full iteration, the genetic module was iterated 3, known factor module 30 and hidden factor module 30 times.

To compare the eQTL detection performance of VBQTL with standard methods and previous studies, we do not directly evaluate the linkage probabilities \( P(b_{n,g}) \) which are obtained during learning. Instead, we apply the standard association model (Section A.1a) on the residuals of the known and unknown factor models after convergence similarly to the traditional methods.

fVBQTL is a faster approximate variant of iVBQTL (Algorithm 1a). Rather than performing full inference in the model, the genetic part of the model is ignored when inferring the parameters for the factor models, which can be cast as a specific update schedule.

<table>
<thead>
<tr>
<th>Algorithm 1a iVBQTL</th>
<th>Algorithm 1b iVBQTL</th>
</tr>
</thead>
<tbody>
<tr>
<td>initialize ( Q ) distributions</td>
<td>initialize ( Q ) distributions</td>
</tr>
<tr>
<td><strong>for</strong> ( i ) iterations ( \text{do} )</td>
<td><strong>for</strong> ( i ) iterations ( \text{do} )</td>
</tr>
<tr>
<td>( \text{(known factors)} ) update: ( Q(v_g), Q(\alpha_c) )</td>
<td>( \text{(genetic model)} ) update: ( Q(u_{g,n}), Q(b_{n,o}) )</td>
</tr>
<tr>
<td>( \text{(noise precision)} ) update: ( Q(\tau_g) )</td>
<td>( \text{(noise precision)} ) update: ( Q(\tau_g) )</td>
</tr>
<tr>
<td>( \text{(hidden factors)} ) update: ( Q(w_g), Q(\beta_k), Q(x_j) )</td>
<td>( \text{(known factors)} ) update: ( Q(v_g), Q(\alpha_c) )</td>
</tr>
<tr>
<td>( \text{(noise precision)} ) update: ( Q(\tau_g) )</td>
<td>( \text{(noise precision)} ) update: ( Q(\tau_g) )</td>
</tr>
<tr>
<td><strong>end for</strong></td>
<td><strong>end for</strong></td>
</tr>
<tr>
<td>( \text{(genetic model)} ) update: ( Q(u_{g,n}), Q(b_{n,o}) )</td>
<td>( \text{(hidden factors)} ) update: ( Q(w_g), Q(\beta_k), Q(x_j) )</td>
</tr>
<tr>
<td>( \text{(Out: Approximate posterior distribution of model parameters)} )</td>
<td>( \text{(noise precision)} ) update: ( Q(\tau_g) )</td>
</tr>
<tr>
<td></td>
<td>( \text{(Out: Approximate posterior distribution of model parameters)} )</td>
</tr>
</tbody>
</table>
4. Simulation dataset

We simulated 80 diploid individuals with 100 SNPs and 400 probe expression measurements. The simulated minor allele frequency was 0.4 for each SNP, and the allele configuration \( s_{n,j} \) of SNP \( n \) was encoded as \((1, 0), (1, 1), \) or \((1, 2)\), including a column for the mean. We independently simulated effects of known and hidden factors, as well as genetic associations, noise, and downstream effects. Noise level \( \psi_g \) of probe \( g \) was drawn from a normal distribution with mean 0 and inverse variance \( \tau_g \) drawn from \( \Gamma(3, 1) \), \( \psi_g \sim N(0, \tau_g^{-1}) \). We simulated associations between SNP genotypes and gene expression levels for 1% of the SNP-gene pairs. The genetic weight \( \theta_{g,n} \) for an association between probe \( g \) and SNP \( n \) was drawn from \( N(0, 4) \). 10 global factors affecting all gene expression levels were simulated. Individual factor levels \( x_{j,k} \) for factor \( k \) were drawn from \( N(0, 0.6) \). Weights \( w_{k,g} \) of factor \( k \) for probe \( g \) were drawn from \( N(0, \sigma^2_k) \), where \( \sigma^2_k \sim 0.8(\Gamma(2.5, 0.6))^2 \) for a heavy-tailed distribution. Three of the 10 simulated global factors were designated as known covariates \( f_{c,j} \). Further three probes that had a simulated SNP association were designated to have downstream effects on 30 other probes. The effect of probe \( g \) on probe \( h \) in individual \( j \) was simulated as additive factor of \( w'_{g,h}y_{g,j} \), where \( w'_{g,h} \sim N(8, 0.8) \) for strong downstream effects, and \( y_{g,j} \) is the expression level of probe \( g \) in individual \( j \).


Associating 10,000 probes against 34,143 linkage regions using standard permutation testing techniques is computationally infeasible. However, SNPs that have the same number of each genotype in a population, can be mapped onto each other via a permutation. We call such SNPs permutation equivalent, since if we performed a large enough number of permutations, they would share the same significance cutoff value. This can be used to overcome the computational cost barrier. We grouped SNPs into permutation equivalence classes, for instance there are 1579 unique ones in the CEU population. We precalculated 10,000 permutation LOD scores for each of the 10,000 probes against a member of each equivalence class and used the 0.01% tail of these scores as a significance cutoff value. When testing for significance of a region we used the cutoff values for the equivalence classes of the SNPs in that region. As in the cis-study, an association was called when the LOD score for at least one SNP in a linkage region exceeded the LOD-score cutoff for all SNPs in that region.

For a given region sharing permutations within an equivalence class is incorrect if the class contains multiple non-identical SNPs in that region. We empirically estimated a correction factor for independent permutations in any such region to 10 which leads to an FPR of 0.1%. With this FPR the number of expected false associations is still immense – we would expect 341,430 false positives. To obtain a more stringent FPR we used Bonferroni correction. Using this conservative approximation we estimated cutoff values for an FPR of \( 10^{-8} \) by increasing the LOD score cutoff values by 5.

APPENDIX B: SUPPLEMENTARY TABLES

<table>
<thead>
<tr>
<th>Chr.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>Total</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probes</td>
<td>1009</td>
<td>644</td>
<td>540</td>
<td>384</td>
<td>449</td>
<td>571</td>
<td>468</td>
<td>338</td>
<td>387</td>
<td>380</td>
<td>545</td>
<td>520</td>
<td>189</td>
<td>330</td>
<td>348</td>
<td>426</td>
<td>549</td>
<td>154</td>
<td>618</td>
<td>266</td>
<td>120</td>
<td>238</td>
<td>9861</td>
<td>-</td>
</tr>
<tr>
<td>Standard</td>
<td>25</td>
<td>23</td>
<td>12</td>
<td>24</td>
<td>14</td>
<td>28</td>
<td>19</td>
<td>12</td>
<td>12</td>
<td>2</td>
<td>16</td>
<td>20</td>
<td>24</td>
<td>7</td>
<td>18</td>
<td>15</td>
<td>21</td>
<td>37</td>
<td>10</td>
<td>29</td>
<td>7</td>
<td>7</td>
<td>17</td>
<td>394</td>
</tr>
<tr>
<td>fVBQTL</td>
<td>72</td>
<td>80</td>
<td>61</td>
<td>60</td>
<td>55</td>
<td>91</td>
<td>63</td>
<td>36</td>
<td>12</td>
<td>51</td>
<td>70</td>
<td>68</td>
<td>21</td>
<td>45</td>
<td>33</td>
<td>51</td>
<td>91</td>
<td>20</td>
<td>75</td>
<td>35</td>
<td>18</td>
<td>29</td>
<td>1156</td>
<td>0.08%</td>
</tr>
</tbody>
</table>

TABLE I: Number of probes with a cis association for individual chromosomes and effective overall False Discovery Rate for the CEU population (FPR= 0.1%) on raw expression data (Standard) and after accounting for hidden factors (fVBQTL).
### TABLE II: Number of probes with a *cis* association for individual chromosomes and effective overall False Discovery Rate for the YRI population (FPR= 0.1%) on raw expression data (Standard) and after accounting for hidden factors (fVBQTL).

<table>
<thead>
<tr>
<th>Chr.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>Total</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probes</td>
<td>1009</td>
<td>644</td>
<td>540</td>
<td>384</td>
<td>449</td>
<td>571</td>
<td>468</td>
<td>338</td>
<td>387</td>
<td>380</td>
<td>545</td>
<td>520</td>
<td>189</td>
<td>330</td>
<td>348</td>
<td>426</td>
<td>549</td>
<td>154</td>
<td>618</td>
<td>266</td>
<td>120</td>
<td>238</td>
<td>9816</td>
<td>-</td>
</tr>
<tr>
<td>Standard</td>
<td>34</td>
<td>28</td>
<td>25</td>
<td>15</td>
<td>31</td>
<td>29</td>
<td>9</td>
<td>23</td>
<td>17</td>
<td>24</td>
<td>39</td>
<td>12</td>
<td>29</td>
<td>17</td>
<td>10</td>
<td>22</td>
<td>921</td>
<td>1.8%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fVBQTL</td>
<td>107</td>
<td>121</td>
<td>93</td>
<td>67</td>
<td>134</td>
<td>82</td>
<td>67</td>
<td>99</td>
<td>22</td>
<td>38</td>
<td>65</td>
<td>62</td>
<td>73</td>
<td>126</td>
<td>35</td>
<td>102</td>
<td>47</td>
<td>28</td>
<td>48</td>
<td>1228</td>
<td>0.08%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE III: Number of probes with a *cis* association for individual chromosomes and effective overall False Discovery Rate for the CHB+JPT population (FPR= 0.1%) on raw expression data (Standard) and after accounting for hidden factors (fVBQTL).

<table>
<thead>
<tr>
<th>Chr.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>Total</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probes</td>
<td>1009</td>
<td>644</td>
<td>384</td>
<td>449</td>
<td>571</td>
<td>468</td>
<td>338</td>
<td>387</td>
<td>380</td>
<td>545</td>
<td>520</td>
<td>189</td>
<td>330</td>
<td>348</td>
<td>426</td>
<td>549</td>
<td>154</td>
<td>618</td>
<td>266</td>
<td>120</td>
<td>238</td>
<td>1203</td>
<td>0.8%</td>
<td></td>
</tr>
<tr>
<td>Standard</td>
<td>80</td>
<td>88</td>
<td>52</td>
<td>47</td>
<td>66</td>
<td>38</td>
<td>15</td>
<td>48</td>
<td>60</td>
<td>77</td>
<td>24</td>
<td>52</td>
<td>42</td>
<td>92</td>
<td>20</td>
<td>67</td>
<td>111</td>
<td>19</td>
<td>34</td>
<td>25</td>
<td>1203</td>
<td>0.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fVBQTL</td>
<td>181</td>
<td>210</td>
<td>131</td>
<td>135</td>
<td>221</td>
<td>158</td>
<td>115</td>
<td>39</td>
<td>130</td>
<td>181</td>
<td>168</td>
<td>67</td>
<td>126</td>
<td>35</td>
<td>102</td>
<td>47</td>
<td>28</td>
<td>48</td>
<td>6171</td>
<td>0.6%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE IV: Number of probes with a *cis* association for individual chromosomes and effective overall False Discovery Rate for the pooled population (FPR= 0.1%) on raw expression data (Standard) and after accounting for hidden factors (fVBQTL).

<table>
<thead>
<tr>
<th>Chr.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>Total</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probes</td>
<td>1009</td>
<td>644</td>
<td>540</td>
<td>384</td>
<td>449</td>
<td>571</td>
<td>468</td>
<td>338</td>
<td>387</td>
<td>380</td>
<td>545</td>
<td>520</td>
<td>189</td>
<td>330</td>
<td>348</td>
<td>426</td>
<td>549</td>
<td>154</td>
<td>618</td>
<td>266</td>
<td>120</td>
<td>238</td>
<td>9816</td>
<td>-</td>
</tr>
<tr>
<td>Standard</td>
<td>48</td>
<td>40</td>
<td>25</td>
<td>30</td>
<td>20</td>
<td>51</td>
<td>34</td>
<td>18</td>
<td>8</td>
<td>26</td>
<td>33</td>
<td>40</td>
<td>10</td>
<td>24</td>
<td>14</td>
<td>28</td>
<td>47</td>
<td>18</td>
<td>14</td>
<td>26</td>
<td>618</td>
<td>1.6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fVBQTL</td>
<td>107</td>
<td>121</td>
<td>93</td>
<td>62</td>
<td>79</td>
<td>47</td>
<td>66</td>
<td>31</td>
<td>16</td>
<td>62</td>
<td>78</td>
<td>70</td>
<td>20</td>
<td>48</td>
<td>53</td>
<td>88</td>
<td>25</td>
<td>76</td>
<td>35</td>
<td>17</td>
<td>40</td>
<td>1228</td>
<td>0.08%</td>
<td></td>
</tr>
</tbody>
</table>

### TABLE V: Magnitude and fraction of overlap between probes with a *standard* *cis* eQTL for different populations and methods. Total numbers for each population and method are given in parenthesis after the population. 991 probes had a standard eQTL in some population, and 160 in every population.

<table>
<thead>
<tr>
<th>Method</th>
<th>CEU (394)</th>
<th>YRI (521)</th>
<th>CHB+JPT (618)</th>
<th>Pooled (1203)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>394 (100%)</td>
<td>194 (49%)</td>
<td>265 (67%)</td>
<td>377 (95%)</td>
</tr>
<tr>
<td>YRI (521)</td>
<td>194 (37%)</td>
<td>521 (100%)</td>
<td>243 (46%)</td>
<td>462 (88%)</td>
</tr>
<tr>
<td>CHB+JPT (618)</td>
<td>265 (42%)</td>
<td>243 (39%)</td>
<td>618 (100%)</td>
<td>542 (87%)</td>
</tr>
<tr>
<td>Pooled (1203)</td>
<td>377 (31%)</td>
<td>462 (38%)</td>
<td>542 (45%)</td>
<td>1203 (100%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>CEU (1156)</th>
<th>YRI (1228)</th>
<th>CHB+JPT (1671)</th>
<th>Pooled (3003)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>371 (32%)</td>
<td>292 (25%)</td>
<td>418 (36%)</td>
<td>735 (63%)</td>
</tr>
<tr>
<td>YRI (1228)</td>
<td>277 (22%)</td>
<td>479 (39%)</td>
<td>380 (30%)</td>
<td>762 (62%)</td>
</tr>
<tr>
<td>CHB+JPT (1671)</td>
<td>328 (19%)</td>
<td>334 (19%)</td>
<td>585 (35%)</td>
<td>883 (52%)</td>
</tr>
<tr>
<td>Pooled (3003)</td>
<td>381 (12%)</td>
<td>504 (16%)</td>
<td>592 (19%)</td>
<td>1174 (39%)</td>
</tr>
</tbody>
</table>

### TABLE VI: Magnitude and fraction of overlap between probes with a *fVBQTL* *cis* eQTL for different populations and methods. Total numbers for each population and method are given in parenthesis after the population. 2416 probes had a fVBQTL eQTL in some population, and 506 probes in all populations.

<table>
<thead>
<tr>
<th>Method</th>
<th>CEU (1156)</th>
<th>YRI (1228)</th>
<th>CHB+JPT (1671)</th>
<th>Pooled (3003)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>371 (94%)</td>
<td>277 (70%)</td>
<td>328 (83%)</td>
<td>381 (96%)</td>
</tr>
<tr>
<td>YRI (1228)</td>
<td>292 (56%)</td>
<td>479 (91%)</td>
<td>334 (64%)</td>
<td>504 (96%)</td>
</tr>
<tr>
<td>CHB+JPT (1671)</td>
<td>418 (67%)</td>
<td>380 (61%)</td>
<td>585 (94%)</td>
<td>592 (95%)</td>
</tr>
<tr>
<td>Pooled (1203)</td>
<td>735 (61%)</td>
<td>762 (63%)</td>
<td>883 (73%)</td>
<td>1174 (97%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>CEU (1156)</th>
<th>YRI (1228)</th>
<th>CHB+JPT (1671)</th>
<th>Pooled (3003)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>1156 (100%)</td>
<td>613 (53%)</td>
<td>811 (70%)</td>
<td>1110 (96%)</td>
</tr>
<tr>
<td>YRI (1228)</td>
<td>613 (49%)</td>
<td>1228 (100%)</td>
<td>721 (58%)</td>
<td>1161 (94%)</td>
</tr>
<tr>
<td>CHB+JPT (1671)</td>
<td>811 (48%)</td>
<td>721 (43%)</td>
<td>1671 (100%)</td>
<td>1513 (90%)</td>
</tr>
<tr>
<td>Pooled (3003)</td>
<td>1110 (36%)</td>
<td>1161 (38%)</td>
<td>1513 (50%)</td>
<td>3003 (100%)</td>
</tr>
</tbody>
</table>
### TABLE VII: Count and percent overlap between probes in *trans* associations on different populations using standard method and after using fVBQTL.

<table>
<thead>
<tr>
<th>Standard Population</th>
<th>CEU (47)</th>
<th>YRI (78)</th>
<th>CHB+JPT (46)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEU (47)</td>
<td>47 (100%)</td>
<td>18 (38%)</td>
<td>22 (47%)</td>
</tr>
<tr>
<td>YRI (78)</td>
<td>18 (23%)</td>
<td>78 (100%)</td>
<td>18 (23%)</td>
</tr>
<tr>
<td>CHB+JPT (46)</td>
<td>22 (48%)</td>
<td>18 (39%)</td>
<td>46 (100%)</td>
</tr>
<tr>
<td>All populations</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 1 populations</td>
<td>32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any population</td>
<td>126</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>iVBQTL Population</th>
<th>CEU (72)</th>
<th>YRI (87)</th>
<th>CHB+JPT (76)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEU (72)</td>
<td>72 (100%)</td>
<td>26 (30%)</td>
<td>41 (31%)</td>
</tr>
<tr>
<td>YRI (87)</td>
<td>26 (30%)</td>
<td>87 (100%)</td>
<td>31 (36%)</td>
</tr>
<tr>
<td>CHB+JPT (76)</td>
<td>41 (54%)</td>
<td>31 (41%)</td>
<td>76 (100%)</td>
</tr>
<tr>
<td>All populations</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 1 populations</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any population</td>
<td>48</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE VIII: Pearson correlation coefficient between top 6 factors learned on the pooled HapMap data, and 4 indicator variables relating to the background of the individual. Correlations with absolute value above 0.6 are highlighted.

<table>
<thead>
<tr>
<th>Factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>0.12</td>
<td>-0.81</td>
<td>0.19</td>
<td>0.08</td>
<td>-0.00</td>
<td></td>
</tr>
<tr>
<td>CEU</td>
<td>0.68</td>
<td>-0.47</td>
<td>-0.21</td>
<td>-0.04</td>
<td>-0.27</td>
<td>0.04</td>
</tr>
<tr>
<td>CHB+JPT</td>
<td>-0.43</td>
<td>0.28</td>
<td>-0.24</td>
<td>-0.64</td>
<td>-0.08</td>
<td>0.03</td>
</tr>
<tr>
<td>YRI</td>
<td>-0.25</td>
<td>0.19</td>
<td>0.46</td>
<td>0.69</td>
<td>0.35</td>
<td>-0.08</td>
</tr>
</tbody>
</table>

### TABLE IX: Summary statistics for method performances on the yeast dataset presented in the main text. The parameters for different methods are varied by the number of allowed factors *K* (PCA, VBQTL) or by the significance cutoff *α* (PCAsig, SVA). Hidden factor summary is given by the number of factors found and the variance explained by the hidden factor effects. The number of probes with a *cis* and *trans* eQTL, as well as the sensitivity and specificity of recovering probes with a standard eQTL are given. FPR = 0.01, 1000 permutations.

<table>
<thead>
<tr>
<th>Method</th>
<th>K</th>
<th>α</th>
<th>Factors found</th>
<th>Variance explained</th>
<th>cis probes</th>
<th>cis sens.</th>
<th>cis spec.</th>
<th>trans probes</th>
<th>trans sens.</th>
<th>trans spec.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>702</td>
<td>1.00</td>
<td>1.00</td>
<td>1657</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>PCA 5</td>
<td>5</td>
<td>-</td>
<td>5</td>
<td>0.44</td>
<td>752</td>
<td>0.81</td>
<td>0.75</td>
<td>1041</td>
<td>0.47</td>
<td>0.75</td>
</tr>
<tr>
<td>PCA 15</td>
<td>15</td>
<td>-</td>
<td>30</td>
<td>0.78</td>
<td>630</td>
<td>0.57</td>
<td>0.63</td>
<td>197</td>
<td>0.08</td>
<td>0.65</td>
</tr>
<tr>
<td>PCA 60</td>
<td>60</td>
<td>-</td>
<td>60</td>
<td>0.91</td>
<td>237</td>
<td>0.23</td>
<td>0.69</td>
<td>38</td>
<td>0.02</td>
<td>0.76</td>
</tr>
<tr>
<td>PCAsig 0.01</td>
<td>6</td>
<td>-</td>
<td>6</td>
<td>0.48</td>
<td>735</td>
<td>0.77</td>
<td>0.73</td>
<td>639</td>
<td>0.28</td>
<td>0.73</td>
</tr>
<tr>
<td>PCAsig 0.10</td>
<td>6</td>
<td>-</td>
<td>6</td>
<td>0.48</td>
<td>735</td>
<td>0.77</td>
<td>0.73</td>
<td>639</td>
<td>0.28</td>
<td>0.73</td>
</tr>
<tr>
<td>PCAsig 0.30</td>
<td>6</td>
<td>-</td>
<td>6</td>
<td>0.48</td>
<td>735</td>
<td>0.77</td>
<td>0.73</td>
<td>639</td>
<td>0.28</td>
<td>0.73</td>
</tr>
<tr>
<td>SVA - 0.01</td>
<td>14</td>
<td>-</td>
<td>14</td>
<td>0.63</td>
<td>722</td>
<td>0.69</td>
<td>0.67</td>
<td>386</td>
<td>0.17</td>
<td>0.72</td>
</tr>
<tr>
<td>SVA - 0.10</td>
<td>14</td>
<td>-</td>
<td>14</td>
<td>0.63</td>
<td>722</td>
<td>0.69</td>
<td>0.67</td>
<td>386</td>
<td>0.17</td>
<td>0.72</td>
</tr>
<tr>
<td>SVA - 0.30</td>
<td>14</td>
<td>-</td>
<td>14</td>
<td>0.63</td>
<td>722</td>
<td>0.69</td>
<td>0.67</td>
<td>386</td>
<td>0.17</td>
<td>0.72</td>
</tr>
<tr>
<td>fVBQTL 5</td>
<td>5</td>
<td>-</td>
<td>5</td>
<td>0.33</td>
<td>855</td>
<td>0.84</td>
<td>0.69</td>
<td>866</td>
<td>0.39</td>
<td>0.75</td>
</tr>
<tr>
<td>fVBQTL 15</td>
<td>15</td>
<td>-</td>
<td>15</td>
<td>0.44</td>
<td>968</td>
<td>0.84</td>
<td>0.61</td>
<td>726</td>
<td>0.33</td>
<td>0.75</td>
</tr>
<tr>
<td>fVBQTL 30</td>
<td>30</td>
<td>-</td>
<td>30</td>
<td>0.49</td>
<td>1038</td>
<td>0.82</td>
<td>0.55</td>
<td>649</td>
<td>0.28</td>
<td>0.73</td>
</tr>
<tr>
<td>fVBQTL 60</td>
<td>60</td>
<td>-</td>
<td>60</td>
<td>0.50</td>
<td>1022</td>
<td>0.80</td>
<td>0.55</td>
<td>625</td>
<td>0.28</td>
<td>0.73</td>
</tr>
<tr>
<td>iVBQTL 5</td>
<td>5</td>
<td>-</td>
<td>5</td>
<td>-</td>
<td>829</td>
<td>0.85</td>
<td>0.72</td>
<td>958</td>
<td>0.44</td>
<td>0.77</td>
</tr>
<tr>
<td>iVBQTL 15</td>
<td>15</td>
<td>-</td>
<td>15</td>
<td>-</td>
<td>889</td>
<td>0.86</td>
<td>0.68</td>
<td>924</td>
<td>0.42</td>
<td>0.76</td>
</tr>
<tr>
<td>iVBQTL 30</td>
<td>30</td>
<td>-</td>
<td>30</td>
<td>-</td>
<td>906</td>
<td>0.86</td>
<td>0.66</td>
<td>948</td>
<td>0.43</td>
<td>0.75</td>
</tr>
<tr>
<td>iVBQTL 60</td>
<td>60</td>
<td>-</td>
<td>60</td>
<td>-</td>
<td>906</td>
<td>0.86</td>
<td>0.66</td>
<td>947</td>
<td>0.43</td>
<td>0.75</td>
</tr>
<tr>
<td>Method</td>
<td>K</td>
<td>α</td>
<td>Factors found</td>
<td>Variance explained</td>
<td>Probes</td>
<td>Sensitivity</td>
<td>Specificity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>----</td>
<td>-----</td>
<td>---------------</td>
<td>-------------------</td>
<td>---------</td>
<td>-------------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>1511</td>
<td>1.00</td>
<td>0.78</td>
<td>0.72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>5</td>
<td>-</td>
<td>5</td>
<td>1650</td>
<td>0.78</td>
<td>0.72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>15</td>
<td>-</td>
<td>15</td>
<td>1679</td>
<td>0.74</td>
<td>0.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>30</td>
<td>-</td>
<td>30</td>
<td>1816</td>
<td>0.71</td>
<td>0.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>60</td>
<td>-</td>
<td>60</td>
<td>921</td>
<td>0.45</td>
<td>0.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA.sig</td>
<td>-</td>
<td>0.01</td>
<td>9</td>
<td>1644</td>
<td>0.77</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA.sig</td>
<td>-</td>
<td>0.10</td>
<td>9</td>
<td>1644</td>
<td>0.77</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA.sig</td>
<td>-</td>
<td>0.30</td>
<td>9</td>
<td>1644</td>
<td>0.77</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVA</td>
<td>-</td>
<td>0.01</td>
<td>24</td>
<td>1816</td>
<td>0.73</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVA</td>
<td>-</td>
<td>0.10</td>
<td>24</td>
<td>1816</td>
<td>0.73</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVA</td>
<td>-</td>
<td>0.30</td>
<td>24</td>
<td>1816</td>
<td>0.73</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fVBQTL</td>
<td>5</td>
<td>-</td>
<td>5</td>
<td>2184</td>
<td>0.84</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fVBQTL</td>
<td>15</td>
<td>-</td>
<td>15</td>
<td>2303</td>
<td>0.80</td>
<td>0.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fVBQTL</td>
<td>30</td>
<td>-</td>
<td>30</td>
<td>2139</td>
<td>0.77</td>
<td>0.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fVBQTL</td>
<td>60</td>
<td>-</td>
<td>60</td>
<td>1811</td>
<td>0.70</td>
<td>0.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>iVBQTL</td>
<td>5</td>
<td>-</td>
<td>5</td>
<td>2209</td>
<td>0.85</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>iVBQTL</td>
<td>15</td>
<td>-</td>
<td>15</td>
<td>2482</td>
<td>0.84</td>
<td>0.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>iVBQTL</td>
<td>30</td>
<td>-</td>
<td>30</td>
<td>2226</td>
<td>0.80</td>
<td>0.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>iVBQTL</td>
<td>60</td>
<td>-</td>
<td>60</td>
<td>2225</td>
<td>0.80</td>
<td>0.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE X: Summary statistics for method performances on the mouse dataset presented in the main text. The parameters for different methods are varied by the number of allowed factors K (PCA, VBQTL) or by the significance cutoff \( \alpha \) (PCAsig, SVA). Hidden factor summary is given by the number of factors found and the variance explained by the hidden factor effects. The number of probes with an eQTL found as well as the sensitivity and specificity of recovering probes with a standard eQTLs are given. FPR = 0.01, 1000 permutations.

<table>
<thead>
<tr>
<th>Method</th>
<th>K</th>
<th>α</th>
<th>Factors found</th>
<th>Variance explained</th>
<th>cis probes</th>
<th>cis sens.</th>
<th>cis spec.</th>
<th>trans probes</th>
<th>trans sens.</th>
<th>trans spec.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>27</td>
<td>1.00</td>
<td>1.00</td>
<td>6</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>5</td>
<td>-</td>
<td>5</td>
<td>46</td>
<td>0.89</td>
<td>0.52</td>
<td>7</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>15</td>
<td>-</td>
<td>15</td>
<td>53</td>
<td>0.93</td>
<td>0.47</td>
<td>16</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>30</td>
<td>-</td>
<td>30</td>
<td>39</td>
<td>0.74</td>
<td>0.51</td>
<td>7</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>60</td>
<td>-</td>
<td>60</td>
<td>5</td>
<td>0.19</td>
<td>1.00</td>
<td>2</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>PCAsig</td>
<td>-</td>
<td>0.01</td>
<td>6</td>
<td>48</td>
<td>0.93</td>
<td>0.52</td>
<td>8</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>PCAsig</td>
<td>-</td>
<td>0.10</td>
<td>6</td>
<td>48</td>
<td>0.93</td>
<td>0.52</td>
<td>8</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>PCAsig</td>
<td>-</td>
<td>0.30</td>
<td>7</td>
<td>46</td>
<td>0.89</td>
<td>0.52</td>
<td>7</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>SVA</td>
<td>-</td>
<td>0.01</td>
<td>13</td>
<td>50</td>
<td>0.93</td>
<td>0.50</td>
<td>7</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>SVA</td>
<td>-</td>
<td>0.10</td>
<td>13</td>
<td>50</td>
<td>0.93</td>
<td>0.50</td>
<td>7</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>SVA</td>
<td>-</td>
<td>0.30</td>
<td>13</td>
<td>50</td>
<td>0.93</td>
<td>0.50</td>
<td>7</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>fVBQTL</td>
<td>5</td>
<td>-</td>
<td>5</td>
<td>46</td>
<td>0.93</td>
<td>0.44</td>
<td>9</td>
<td>0.17</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>fVBQTL</td>
<td>15</td>
<td>-</td>
<td>15</td>
<td>61</td>
<td>0.96</td>
<td>0.43</td>
<td>14</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>fVBQTL</td>
<td>30</td>
<td>-</td>
<td>30</td>
<td>68</td>
<td>0.96</td>
<td>0.38</td>
<td>9</td>
<td>0.17</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>fVBQTL</td>
<td>60</td>
<td>-</td>
<td>60</td>
<td>67</td>
<td>0.96</td>
<td>0.39</td>
<td>5</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>iVBQTL</td>
<td>5</td>
<td>-</td>
<td>5</td>
<td>45</td>
<td>0.93</td>
<td>0.56</td>
<td>9</td>
<td>0.17</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>iVBQTL</td>
<td>15</td>
<td>-</td>
<td>15</td>
<td>58</td>
<td>0.96</td>
<td>0.45</td>
<td>12</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>iVBQTL</td>
<td>30</td>
<td>-</td>
<td>30</td>
<td>59</td>
<td>0.96</td>
<td>0.44</td>
<td>13</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>iVBQTL</td>
<td>60</td>
<td>-</td>
<td>60</td>
<td>59</td>
<td>0.96</td>
<td>0.44</td>
<td>13</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

TABLE XI: Summary statistics for method performances on the human dataset presented in the main text. The parameters for different methods are varied by the number of allowed factors K (PCA, VBQTL) or by the significance cutoff \( \alpha \) (PCAsig, SVA). Hidden factor summary is given by the number of factors found and the variance explained by the hidden factor effects. The number of probes with a cis and trans eQTL found as well as the sensitivity and specificity of recovering probes with a standard eQTL are given. FPR = 0.01, 1000 permutations.