Appendix A

Variational inference in PEER

Supplementary Methods

Implementation of non-Bayesian models

Standard expression QTL model

To ensure a common ground when comparing different methods, we used a well established linear regression approach introduced by Lander and Botstein (1989) 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

$$L_{n,g} = \log \left\{ \prod_j \frac{P(y_{g,j} \mid s_{n,j}, \boldsymbol{\theta}_1)}{P(y_{g,j} \mid \boldsymbol{\theta}_0)} \right\} = \log \left\{ \prod_j \frac{\mathcal{N}(y_{g,j}; u_{n,j} s_{n,j} + \mu_{g,1}, \sigma^2_{g,1})}{\mathcal{N}(y_{g,j}; \mu_{g,0}, \sigma^2_{g,0})} \right\} \tag{A.1}$$

which assess how well a particular gene expression level is modelled when the observed genetic state $s_{n,j}$ is taken into account, compared to how well it is model-led by a background model ignoring the genetic effect. The probe expression levels $y_{g,j}$ can either be the raw measurements, residuals after subtracting the estimated effect of hidden and known factors, or ranks for a non-parametric statistic.

Significance of an association was evaluated in three different ways:

1. **2-tailed t test on expression values** uses the Student's t distribution with $N-2$ degrees of freedom to assess the significance of the statistic $t = (N-2)^{0.5}\rho(1-\rho^2)^{-0.5}$ based on the correlation coefficient $\rho^2 = 1 - \exp(-2L_{n,g}N^{-1})$ between the genotype and the expression levels. We called an association significant if $|t|$ was greater than the $\frac{10^{-3}}{2S}$ tail of the t_{N-2} distribution, which corresponds to a 10^{-3} Bonferroni-corrected per-gene false positive rate when performing tests for S SNPs.

2. **Rank correlation** uses the same test, but on the ranks of expression values.

3. **Permutation testing** (Lynch and Walsh, 1998) repeats the analysis in Equation (A.1) 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 δ tail of the null distribution for a given false positive rate (FPR) δ. 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 δ tail (Stranger et al., 2007).

The posterior of the switch variable for the probabilistic genetic model is not used for the final tests to put all methods on equal footing.

PEER framework

VBQTL and the alternative compared methods are implemented within the PEER (Probabilistic Estimation of Expression Residuals) framework. Here, we give a full self-contained treatment of the framework and the implemented inference algorithms.

Likelihood models

The likelihood model of PEER for observed expression levels $\mathbf{Y}$ is

$$P(\mathbf{Y} \mid \mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(M)}, \boldsymbol{\tau}) = \mathcal{N}(\mathbf{Y} \mid \mathbf{Y}^{(1)} + \cdots + \mathbf{Y}^{(M)}, \boldsymbol{\Sigma}), \tag{A.2}$$

where $\boldsymbol{\Sigma} = \mathrm{diag}\{\frac{1}{\tau_g}\}$ is the diagonal matrix constructed from noise precisions $\{\tau_g\}$ and $\{\mathbf{Y}^{(m)}\}$ are the contributions of expression variability for each of M models. The noise model is per gene, similar to a factor analysis model, where gamma priors are put on the noise precisions,

$$P(\tau_g) = \Gamma(\tau_g \mid a_\tau, b_\tau). \tag{A.3}$$

In experiments we used vague gamma prior parameters, $a_\tau = 1, b_\tau = 100$. Each of the M models itself depends on parameters $\boldsymbol{\theta}^{(m)}$ and possibly other data $\mathcal{D}^{(m)}$

$$P(\mathbf{Y}^{(m)} \mid \boldsymbol{\theta}^{(m)}, \mathcal{D}^{(m)}). \tag{A.4}$$

Genotype effect model. The expression level $y_{g,j}^{(1)}$ of the gth gene probe in the jth individual is explained by linear effects of genotypes of N SNPs $\mathbf{s}_j = (s_{1,j}, \ldots, s_{N,j})$:

$$P(y_{g,j}^{(1)} \mid \mathbf{s}_j, \mathbf{b}_g, \mathbf{u}_g, \tau_g) = \mathcal{N}(y_{g,j}^{(1)} \mid \sum_{n=1}^{N} b_{n,g} \cdot (u_{n,g} s_{n,j}), \frac{1}{\tau_g}) \tag{A.5}$$

$$P(b_{n,g}) = \mathrm{Bernoulli}(b_{n,g} \mid p_{\mathrm{ass}}) \tag{A.6}$$

$$P(u_{n,g}) = \mathcal{N}(u_{n,g} \mid 0, 1). \tag{A.7}$$

The weight $\mathbf{u}_g = (u_{1,g}, \ldots, u_{N,g})$ indicates the magnitude of the effect, and the binary variables $\mathbf{b}_g = (b_{1,g}, \ldots, b_{N,g})$ determine whether it is significant (true) or not (false), taking the Bernoulli prior on the switch variable $P(b_{n,g}) = \mathrm{Bernoulli}(b_{n,g} \mid p_{\mathrm{ass}})$ into account. 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 (Section Standard expression QTL model) 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 $P(b_{n,g} = 1 \,|\, y_{g,j}, s_{j,n}) = \sigma(\text{LOD score})$ where $\sigma()$ is the sigmoid function $\sigma(x) = 1/(1 + e^{-x})$.

2) Known factor model. The effect of the measured C covariates in the jth individual, $\mathbf{f}_j = (f_{1,j}, \ldots, f_{C,j})$, where the weights of their effect on a gene g is $\mathbf{v}_g = (v_{g,1}, \ldots, v_{g,C})$ is modelled as:

$$P(y^{(2)}_{g,j} \,|\, \mathbf{f}_j, \mathbf{v}_g, \tau_g) = \mathcal{N}(y^{(2)}_{g,j} \,|\, \sum_{c=1}^{C} v_{g,c}\, f_{c,j}, \frac{1}{\tau_g}) \tag{A.8}$$

$$P(v_{g,c} \,|\, \alpha_c) = \mathcal{N}(v_{g,c} \,|\, 0, \frac{1}{\alpha_c}) \tag{A.9}$$

$$P(\alpha_c) = \Gamma(\alpha_c \,|\, a_\alpha, b_\alpha). \tag{A.10}$$

The gamma prior on the inverse covariances for each factor introduces automatic relevance detection (ARD) Mackay (1995); Neal (1996), driving the weights of unused factors to 0 and thereby switching them off. This is explained in more detail below.

3) Hidden factor model. Analogously to known factors, expression variability is modelled by linear effects from K hidden factors $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_K\}$:

$$P(y^{(3)}_{g,j} \,|\, \mathbf{x}_j, \mathbf{w}_g, \tau_g) = \mathcal{N}(y^{(3)}_{g,j} \,|\, \sum_{k=1}^{K} w_{g,k}\, x_{k,j}, \frac{1}{\tau_g}) \tag{A.11}$$

$$P(w_k, \beta_k) = \prod_{g=1}^{G} \mathcal{N}(w_{g,k} \,|\, 0, \frac{1}{\beta_k}) \tag{A.12}$$

$$P(x_{k,j}) = \mathcal{N}(x_{k,j} \,|\, 0, 1) \tag{A.13}$$

$$P(\beta_k) = \Gamma(\beta_k \,|\, a_\beta, b_\beta). \tag{A.14}$$

The factor activations $\mathbf{X}$ 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 $a_\alpha = 10^{-7}G$ and $b_\alpha = 10^{-1}G$, where G is the total number of gene probes. Similar prior settings were used for the weights of the known factors $\mathbf{v}_c$. We put a standard normal prior on the hidden factors, $x_{k,j} \sim \mathcal{N}(x_{k,j} \,|\, 0, 1)$.

Variational inference

As outlined in Methods we use variational Bayesian inference Jordan et al. (1999) for parameter learning in the framework. The basic principle of variational methods is to approximate the exact joint posterior distribution over all parameters by a factorised Q-distribution. Individual factors of the Q-distribution are refined by minimisation 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 factorisations and update rules for the general framework and the individual models.

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

$$P(\{\mathbf{Y}^{(m)}\}_{m=1}^{M}, \{\boldsymbol{\theta}^{(m)}\}_{m=1}^{M}, \,|\, \mathcal{D}) \tag{A.15}$$

by a factorised approximation over parameters for individual models

$$Q(\boldsymbol{\Theta}) = \prod_{m=1}^{M} Q(\boldsymbol{\theta}^{(m)}) Q(\mathbf{Y}^{(m)}). \tag{A.16}$$

Here we defined the abbreviation $\mathcal{D} = \{\mathbf{Y}, \{\mathcal{D}^{(m)}\}_{m=1}^{M}\}$, summarising all observed data; expression levels $\mathbf{Y}$ as well as model-specific data $\{\mathcal{D}^{(m)}\}_{m=1}^{M}$. Note that as the expression contributions $\mathbf{Y}^{(m)}$ are not observed they also resemble parameters

that need to be inferred. Strictly speaking these are not treated as random variables of the model, but Gaussian messages that comprise the first and second moments of the expression variability contribution of a respective model. The distributions of parameters $\boldsymbol{\theta}^{(m)}$ for individual models are in turn factorised. The set $\boldsymbol{\Theta} = \{\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(M)}\}$ denotes the set of all parameters from all models.

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 minimisation of the KL-divergence between P and Q with respect to Q_i which leads to

$$\tilde{Q}(\Theta_i) \propto \exp\{\langle \log P(\mathcal{D}, \boldsymbol{\Theta}) \rangle_{Q(\Theta_j), i \neq j}\}. \tag{A.17}$$

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

$$P(\mathbf{Y} \mid \boldsymbol{\Theta}) = \mathcal{N}(\mathbf{Y} \mid \mathbf{Y}^{(1)} + \cdots + \mathbf{Y}^{(M)}, \boldsymbol{\Sigma}) \prod_{m=1}^{M} P(\mathbf{Y}^{(m)} \mid \boldsymbol{\theta}^{(m)}, \mathcal{D}^{(m)}) \tag{A.18}$$

this allows generic update rules for all model parameters to be derived. Substituting in Equation (A.16) for each $Q(\cdot)$, we obtain the following approximate distributions:

(Approximate distributions)

$$Q(\boldsymbol{\tau}) = \prod_{g=1}^{G} \Gamma(\tau_g \mid \tilde{a}_{\tau_g}, \tilde{b}_{\tau_g}) \tag{A.19}$$

$$Q(\mathbf{Y}^{(m)}) = \prod_{g=1}^{G} \prod_{j=1}^{J} \mathcal{N}(y_{g,j}^{(m)} \mid \tilde{m}_{Y_{g,j}^{(m)}}, \frac{1}{\tilde{\tau}_{Y_{g,j}^{(m)}}}), \tag{A.20}$$

and similar factorisations for each of the models (given below). The parameter

update equations for the framework parameters follow as:

(Update rules)

$$\tilde{a}_{\tau_g} = a_\tau + \frac{1}{2}\sum_{j=1}^{J}\left\langle\left(y_{g,j} - \sum_{m=1}^{M} y_{g,j}^{(m)}\right)^2\right\rangle \tag{A.21}$$

$$\tilde{b}_{\tau_g} = b_\tau + \frac{J}{2}. \tag{A.22}$$

Genotype effect model The update equations for the models introduced in the main text (Inference) follow similarly. For the models, we give the approximate factorisations employed, and the resulting update equations that are derived in identical manner to the treatment above.

(Approximate distributions)

$$Q(\mathbf{B}) = \prod_{n=1}^{N}\prod_{g=1}^{G}\text{Bernoulli}(b_{n,g} \,|\, \tilde{p}_{b_{n,g}}) \tag{A.23}$$

$$Q(\mathbf{U}) = \prod_{n=1}^{N}\prod_{g=1}^{G}\mathcal{N}(\mathbf{u}_{n,g} \,|\, \tilde{\mathbf{m}}_{\mathbf{u}_{n,g}}, \tilde{\mathbf{\Sigma}}_{\mathbf{u}_{n,g}}) \tag{A.24}$$

(Update rules)

$$\tilde{\Sigma}_{\mathbf{u}_{n,g}} = \mathbf{I} + \langle \tau_g \rangle \left\langle b_{n,g}^2 \right\rangle \sum_{j=1}^{J} \mathbf{s}_{n,j}^{\mathrm{T}} \mathbf{s}_{n,j} \tag{A.25}$$

$$\tilde{\mathbf{m}}_{\mathbf{u}_{n,g}} = \tilde{\Sigma}_{\mathbf{u}_{n,g}}^{-1} \left(\langle \tau_g \rangle \langle b_{n,g} \rangle \sum_{j=1}^{J} \mathbf{s}_{n,j} \left\langle z_{g,j}^{(1)^{\setminus n}} \right\rangle \right) \tag{A.26}$$

$$\tilde{m}_{y_{g,j}^{(1)}} = \sum_{n=1}^{N} \langle b_{n,g} \rangle \langle \mathbf{u}_{n,g} \rangle \mathbf{s}_{n,j} \tag{A.27}$$

$$\tilde{\tau}_{y_{g,j}^{(1)}} = \left[\sum_{n=1}^{N} \left\langle b_{n,g}^2 \right\rangle \left\langle \mathbf{u}_{n,g}^2 \right\rangle \mathbf{s}_{n,j}^2 \right], \tag{A.28}$$

where we define

$$\left\langle z_{g,j}^{(1)^{\setminus n}} \right\rangle = z_{g,j}^{(1)} - \sum_{m \neq n} \langle b_{m,g} \rangle \langle \mathbf{u}_{m,g} \rangle \mathbf{s}_{m,j} \tag{A.29}$$

and the residual expression dataset for the mth model

$$z_{g,j}^{(m)} = y_{g,j} - \sum_{l \neq m}^{M} y_{g,j}^{(l)}. \tag{A.30}$$

$$\tag{A.31}$$

The approximate posterior over the indicator variables can be obtained from

$$\begin{aligned} \tilde{p}_{b_{n,g}} &\propto p_{ass} \cdot \exp \left\{ -\frac{1}{2} \sum_{j=1}^{J} \left\langle \left(z_{g,j}^{(1)^{\setminus n}} - b_{n,g} \mathbf{u}_{n,g} \mathbf{s}_{n,j} \right)^2 \right\rangle \right\} \\ (1 - \tilde{p}_{b_{n,g}}) &\propto (1 - p_{ass}) \cdot \exp \left\{ -\frac{1}{2} \sum_{j=1}^{J} \left\langle \left(z_{g,j}^{(1)^{\setminus n}} \right)^2 \right\rangle \right\}, \end{aligned} \tag{A.32}$$

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

(A.33)

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

(A.34)

(Approximate distributions)

$$Q(\mathbf{X}) = \prod_{j=1}^{J} \mathcal{N}(\mathbf{x}_j \,|\, \tilde{\mathbf{m}}_{\mathbf{x}_j}, \tilde{\Sigma}_{\mathbf{x}_j}) \tag{A.35}$$

$$Q(\mathbf{W}) = \prod_{g=1}^{G} \mathcal{N}(\mathbf{w}_g \,|\, \tilde{\mathbf{m}}_{\mathbf{w}_g}, \tilde{\Sigma}_{\mathbf{w}_g}) \tag{A.36}$$

$$Q(\boldsymbol{\beta}) = \prod_{k=1}^{K} \Gamma(\beta_k \,|\, \tilde{a}_{\beta_k}, \tilde{b}_{\beta_k}) \tag{A.37}$$

(Update rules)

$$\tilde{\Sigma}_{\mathbf{x}_j} = \Sigma_{\mathbf{x}_j} + \left\langle \mathbf{W}^T \mathrm{diag}\left(\boldsymbol{\tau}\right) \mathbf{W} \right\rangle \tag{A.38}$$

$$\tilde{\mathbf{m}}_{\mathbf{x}_j} = \tilde{\Sigma}_{\mathbf{x}_j}^{-1} \left\langle \mathbf{W}^{\mathrm{T}} \right\rangle \mathrm{diag}\left\langle \boldsymbol{\tau} \right\rangle \left(\left\langle \mathbf{z}_j^{(3)} \right\rangle \right) \tag{A.39}$$

$$\tilde{\Sigma}_{\mathbf{w}_g} = \text{diag}\,\langle \boldsymbol{\beta} \rangle + \langle \tau_g \rangle \sum_{j=1}^{J} \left\langle \mathbf{x}_j \mathbf{x}_j^{\mathrm{T}} \right\rangle \tag{A.40}$$

$$\tilde{\mathbf{m}}_{\mathbf{w}_g} = \tilde{\Sigma}_{\mathbf{w}_g}^{-1} \left(\langle \tau_g \rangle \sum_{j=1}^{J} \langle \mathbf{x}_j \rangle \left(\left\langle \mathbf{z}_j^{(3)} \right\rangle \right) \right) \tag{A.41}$$

$$\tilde{m}_{y_{g,j}^{(3)}} = \sum_{k=1}^{K} \langle w_{g,k} \rangle \langle x_{j,k} \rangle \tag{A.42}$$

$$\tilde{\tau}_{y_{g,j}^{(3)}} = \left[\sum_{n=1}^{N} \left\langle b_{n,g}^2 \right\rangle \left\langle \mathbf{u}_{n,g}^2 \right\rangle \mathbf{s}_{n,j}^2 \right] \tag{A.43}$$

$$\tag{A.44}$$

Initialisation. The initial states of hidden factor model weights $Q(\mathbf{w}_g)$ and levels $Q(\mathbf{x}_j)$ are determined from a PCA solution, and the weights for known factors $Q(\mathbf{v}_g)$ are initialised to the maximum likelihood estimate. The parameters for remaining Q distributions for all models are deterministically initialised to corresponding prior means. A random initialisation is possible as well, however, additional computation time is required for multiple restarts, and the inference becomes non-deterministic. We have not explored the implications of this alternative here as the maximum likelihood initialisation performs robustly well in practise.

Bottleneck approximation. The genetic association model accounts for additive association signals from all considered SNPs. The corresponding variational updates of the indicator variables in Equation (A.32) can be unstable in practise. In particular, if multiple correlated SNPs are in association to a single gene, variational learning is prone to being trapped in local optima, attributing the effect to only one of them. Hence, the inferred state of the indicator variables $\mathbf{B}$ depends on the order in which these updates are carried out. To obtain meaningful results, the update sequence needs to be randomised and typically large numbers of restarts are required. This procedure implies prohibitive computational cost, particularly for large datasets. To avoid this additional computation, these updates are instead implemented greedily. For each gene g only a single non-zero entry in the indicator matrix is permitted, corresponding to

the SNP with the greatest evidence for an association. This leads to a sparse association matrix $\mathbf{B}$.

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 $Q(\boldsymbol{\theta}_i)$ distributions. In experiments, we used 3 iterations of the full model. Within each full iteration, the genetic model was iterated 3, known factor model 30 and hidden factor model 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 Standard expression QTL model) 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. 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.

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 ψ_g of probe g was drawn from a normal distribution with mean 0 and inverse variance τ_g drawn from $\Gamma(3, 1)$, $\psi_g \sim \mathcal{N}(0, \tau_g^{-1})$. 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 $\mathcal{N}(0, 4)$. A total of 10 global factors affecting all gene expression levels were simulated. Individual factor levels $x_{j,k}$ for factor k were drawn from $\mathcal{N}(0, 0.6)$. Weights $w_{k,g}$ of factor k for probe g were drawn from $N(0, \sigma_k^2)$, where $\sigma_k^2 \sim 0.8(\Gamma(2.5, 0.6))^2$ for a heavy-tailed weight 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 $w'_{g,h} \sim \mathcal{N}(8, 0.8)$ for strong downstream effects, and $y_{g,j}$ is the expression level of probe g in individual j.

Appendix B

Supplementary Tables

Chr.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y	**Total**	**FDR**
Probes	1009	644	540	384	449	571	468	338	387	380	545	520	189	330	348	426	549	154	618	266	120	238	328	15	9816	-
CEU																										
Standard	23	21	12	24	14	26	18	12	3	17	21	24	5	16	15	21	35	9	29	8	8	14	7	0	382	2.57 %
fVBQTL	61	69	53	57	45	83	44	36	12	48	61	68	16	41	32	55	82	20	69	29	17	30	23	0	1051	0.93 %
YRI																										
Standard	37	32	23	19	21	42	27	17	9	27	31	30	9	24	16	24	38	12	30	18	8	26	9	0	529	1.86 %
fVBQTL	79	94	75	48	56	91	66	38	17	58	79	65	26	48	48	59	94	22	77	40	19	43	27	0	1269	0.77 %
ASI																										
Standard	36	37	19	28	19	48	30	15	9	24	33	36	10	19	12	24	43	16	42	16	10	19	9	0	554	1.77 %
fVBQTL	91	105	88	55	58	111	73	55	19	59	87	78	31	56	52	61	109	30	96	43	22	37	28	0	1444	0.68 %
pooled																										
Standard	68	77	56	48	42	79	52	32	14	46	48	66	21	39	34	43	82	21	71	31	19	37	19	0	1045	0.94 %
fVBQTL	159	191	158	115	120	202	138	101	36	120	168	159	54	104	96	113	181	51	170	78	33	85	60	4	2696	0.36 %

Table B.1: Number of probes with a *cis* association for individual chromosomes and per-probe false discovery rate for the considered populations (per-probe FPR= 0.100%, Bonferroni corrected for testing multiple SNPs per probe, 2-tailed t test) on raw expression data (Standard) and after accounting for hidden factors (fVBQTL).

		Standard eQTLs			
		CEU (382)	YRI (529)	CHB+JPT (554)	Pooled (1045)
Standard	CEU (382)	382 (100%)	194 (50%)	236 (61%)	356 (93%)
	YRI (529)	194 (36%)	529 (100%)	228 (43%)	409 (77%)
	CHB+JPT (554)	236 (42%)	228 (41%)	554 (100%)	490 (88%)
	Pooled (1045)	356 (34%)	409 (39%)	490 (46%)	1045 (100%)
fVBQTL	CEU (1051)	365 (34%)	282 (26%)	358 (34%)	662 (62%)
	YRI (1269)	276 (21%)	510 (40%)	356 (28%)	675 (53%)
	CHB+JPT (1444)	305 (21%)	322 (22%)	531 (36%)	788 (54%)
	Pooled (2696)	370 (13%)	486 (18%)	527 (19%)	1028 (38%)
		fVBQTL eQTLs			
		CEU (1051)	YRI (1269)	CHB+JPT (1444)	Pooled (2696)
Standard	CEU (382)	365 (95%)	276 (72%)	305 (79%)	370 (96%)
	YRI (529)	282 (53%)	510 (96%)	322 (60%)	486 (91%)
	CHB+JPT (554)	358 (64%)	356 (64%)	531 (95%)	527 (95%)
	Pooled (1045)	662 (63%)	675 (64%)	788 (75%)	1028 (98%)
fVBQTL	CEU (1051)	1051 (100%)	591 (56%)	717 (68%)	1007 (95%)
	YRI (1269)	591 (46%)	1269 (100%)	697 (54%)	1120 (88%)
	CHB+JPT (1444)	717 (49%)	697 (48%)	1444 (100%)	1350 (93%)
	Pooled (2696)	1007 (37%)	1120 (41%)	1350 (50%)	2696 (100%)

Table B.2: Magnitude and fraction of overlap between probes with a **Standard** of **fVBQTL***cis* eQTL respectively, for different populations and methods. Total numbers for each population and method are given in parenthesis after the population. 955 probes had a standard eQTL in some population, and 148 in every population. 2236 probes had a fVBQTL eQTL in some population, and 477 in every population.

Population	1. eQTLs	2. fVBQTLs	3. Pooled eQTLs	2. & 3.	2. - 1.	3. - 1.	(2. - 1.) &(3. - 1.)
CEU	382	1051	871	485	686	582	204
YRI	529	1269	796	476	759	507	188
CHB+JPT	554	1444	709	501	913	378	170

Table B.3: Overlap of VBQTLs in one population (2.) with standard eQTLs found when pooling the other two populations (3.). Overlaps are given both for all QTLs (2. & 3.) and only for additional ones (2. - 1. & 3. - 1.) compared to standard eQTLs in the population. Per-probe eQTL FPR=0.1%, Bonferroni corrected for testing multiple SNPs per probe, 2-tailed t test.

Standard

Population	CEU (47)	YRI (78)	CHB+JPT (46)
CEU (47)	47 (100%)	18 (38%)	22 (47%)
YRI (78)	18 (23%)	78 (100%)	18 (23%)
CHB+JPT (46)	22 (48%)	18 (39%)	46 (100%)
All populations	13		
> 1 populations	32		
Any population	126		

fVBQTL

Population	CEU (72)	YRI (87)	CHB+JPT (76)
CEU (72)	72 (100%)	26 (36%)	41 (57%)
YRI (87)	26 (30%)	87 (100%)	31 (36%)
CHB+JPT (76)	41 (54%)	31 (41%)	76 (100%)
All populations	25		
> 1 populations	48		
Any population	162		

Table B.4: Count and percent overlap between probes in *trans* associations on different populations using standard method and after using fVBQTL.

Factor	1	2	3	4	5	6
Gender	0.12	0.16	**-0.81**	0.19	0.08	-0.00
CEU	**0.68**	-0.47	-0.21	-0.04	-0.27	0.04
CHB+JPT	-0.43	0.28	-0.24	**-0.64**	-0.08	0.03
YRI	-0.25	0.19	0.46	**0.69**	0.35	-0.08

Table B.5: 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.

Method	K	α	Factors found	Variance explained	*cis* probes	*cis* spec.	*cis* sens.	*trans* probes	*trans* spec.	*trans* sens.
Standard	–	–	0	0.00	20	1.00	1.00	0	0.00	0.00
PCA	5	–	5	0.52	35	0.54	0.95	0	0.00	0.00
PCA	15	–	15	0.70	38	0.45	0.85	0	0.00	0.00
PCA	30	–	30	0.82	29	0.45	0.65	0	0.00	0.00
PCA	60	–	60	0.94	4	0.75	0.15	0	0.00	0.00
PCAsig	–	0.01	7	0.56	37	0.51	0.95	0	0.00	0.00
PCAsig	–	0.1	7	0.56	37	0.51	0.95	0	0.00	0.00
PCAsig	–	0.3	7	0.56	37	0.51	0.95	0	0.00	0.00
SVA	–	0.01	12	0.65	38	0.50	0.95	0	0.00	0.00
SVA	–	0.1	12	0.65	38	0.50	0.95	0	0.00	0.00
SVA	–	0.3	12	0.65	38	0.50	0.95	0	0.00	0.00
fVBQTL	5	–	5	0.52	34	0.59	1.00	0	0.00	0.00
fVBQTL	15	–	15	0.69	51	0.39	1.00	0	0.00	0.00
fVBQTL	30	–	30	0.70	55	0.36	1.00	0	0.00	0.00
fVBQTL	60	–	60	0.70	55	0.36	1.00	0	0.00	0.00
iVBQTL	5	–	5	0.52	34	0.59	1.00	0	0.00	0.00
iVBQTL	15	–	15	0.69	51	0.39	1.00	0	0.00	0.00
iVBQTL	30	–	30	0.70	54	0.37	1.00	0	0.00	0.00
iVBQTL	60	–	60	0.70	54	0.37	1.00	0	0.00	0.00

Table B.6: Summary statistics for method performances on the human chromosome 19 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. Per-probe eQTL FPR = 0.001, Bonferroni corrected for testing multiple SNPs per probe, 2-tailed t test.

Method	K	α	Factors found	Variance explained	*cis* probes	*cis* spec.	*cis* sens.	*trans* probes	*trans* spec.	*trans* sens.
Standard	–	–	0	0.00	445	1.00	1.00	746	1.00	1.00
PCA	5	–	5	0.28	478	0.77	0.82	501	0.79	0.53
PCA	15	–	15	0.53	481	0.64	0.69	132	0.77	0.14
PCA	30	–	30	0.70	392	0.60	0.53	57	0.75	0.06
PCA	60	–	60	0.86	105	0.66	0.16	5	1.00	0.01
PCAsig	–	0.01	7	0.34	468	0.72	0.76	229	0.80	0.25
PCAsig	–	0.1	7	0.34	468	0.72	0.76	229	0.80	0.25
PCAsig	–	0.3	7	0.34	468	0.72	0.76	229	0.80	0.25
SVA	–	0.01	14	0.52	482	0.65	0.71	144	0.78	0.15
SVA	–	0.1	14	0.52	482	0.65	0.71	144	0.78	0.15
SVA	–	0.3	14	0.52	482	0.65	0.71	144	0.78	0.15
fVBQTL	5	–	5	0.34	547	0.72	0.89	409	0.81	0.45
fVBQTL	15	–	15	0.55	668	0.59	0.88	364	0.80	0.39
fVBQTL	30	–	30	0.62	719	0.54	0.87	349	0.79	0.37
fVBQTL	60	–	60	0.62	722	0.54	0.87	348	0.78	0.37
iVBQTL	5	–	5	0.32	616	0.68	0.95	650	0.76	0.66
iVBQTL	15	–	15	0.50	785	0.54	0.96	694	0.73	0.68
iVBQTL	30	–	30	0.57	821	0.52	0.95	746	0.71	0.71
iVBQTL	60	–	60	0.57	825	0.51	0.95	739	0.71	0.70

Table B.7: 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. Per-probe eQTL FPR = 0.001, Bonferroni corrected for testing multiple SNPs per probe, 2-tailed t test.

Method	K	α	Factors found	Variance explained	*cis* probes	*cis* spec.	*cis* sens.	*trans* probes	*trans* spec.	*trans* sens.
Standard	–	–	0	0.00	560	1.00	1.00	369	1.00	1.00
PCA	5	–	5	0.25	639	0.84	0.96	418	0.76	0.86
PCA	15	–	15	0.48	614	0.82	0.90	409	0.72	0.80
PCA	30	–	30	0.74	708	0.70	0.88	488	0.59	0.78
PCA	60	–	60	0.91	354	0.82	0.52	178	0.76	0.37
PCAsig	–	0.01	12	0.39	601	0.84	0.91	376	0.76	0.77
PCAsig	–	0.1	13	0.41	589	0.85	0.90	371	0.75	0.76
PCAsig	–	0.3	13	0.41	589	0.85	0.90	371	0.75	0.76
SVA	–	0.01	24	0.67	687	0.74	0.91	501	0.58	0.79
SVA	–	0.1	24	0.67	687	0.74	0.91	501	0.58	0.79
SVA	–	0.3	24	0.67	687	0.74	0.91	501	0.58	0.79
fVBQTL	5	–	5	0.32	876	0.63	0.98	590	0.56	0.90
fVBQTL	15	–	15	0.51	1028	0.54	0.99	716	0.46	0.89
fVBQTL	30	–	30	0.67	973	0.56	0.98	657	0.49	0.88
fVBQTL	60	–	60	0.70	932	0.59	0.98	626	0.51	0.87
iVBQTL	5	–	5	0.32	895	0.62	0.99	613	0.55	0.91
iVBQTL	15	–	15	0.51	1036	0.53	0.99	723	0.46	0.90
iVBQTL	30	–	30	0.55	1056	0.52	0.99	729	0.46	0.90
iVBQTL	60	–	60	0.55	1049	0.53	0.99	728	0.45	0.90

Table B.8: 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 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. Per-probe eQTL FPR = 0.001, Bonferroni corrected for testing multiple SNPs per probe, 2-tailed t test.

Factor	Q-value	mean(LOD_s)	Covariate
Oaf1p	5.54E-03	42.9 (r^2=0.30)	Probe
Pdr3p	2.09E-02	14.6	SNP XV 132423
Rtg3p	3.01E-02	21.4	SNP XIV 449639
Reb1p	3.70E-02	41.5	Env
Reb1p	0.00E+00	78.1 (r^2=0.51)	Probe
Thi2p	0.00E+00	52.2	SNP VI 5852
Kar4p	0.00E+00	45.7	SNP V 183958
Hcm1p	0.00E+00	38.9 (r^2=0.29)	Probe
Rpn4p	2.25E-02	56.1	Env
Rpn4p	2.44E-02	35.4 (r^2=0.24)	Probe
Pdc2p	1.84E-02	16.4	SNP XII 611967
Gis1p	4.18E-02	11.9	SNP XV 193911
Ino2p	1.48E-02	11.9	SNP II 603790
Upc2p	2.90E-02	11.7	SNP I 55215
Adr1p	4.98E-02	41.7	Env
Met32p	1.90E-02	15.8	SNP IX 277908
Met32p	1.04E-03	23.4 (r^2=0.19)	Probe
Sum1p	0.00E+00	115.2	SNP XV 838599
Stp1p	1.36E-02	23.6 (r^2=0.19)	Probe
Gcn4p	2.28E-02	66.7	Env
Gcn4p	3.00E-02	72.4 (r^2=0.42)	Probe
Swi4p	6.09E-03	39.7	Env
Spt2p	8.70E-05	34.1	SNP XV 10337
Gat1p	2.44E-02	23.5 (r^2=0.19)	Probe
Hac1p	4.56E-02	20.5	Env
Cdc14p	0.00E+00	42.3	SNP X 307178
Pho4p	2.90E-02	15.5	SNP XIII 28694
Mig1p	5.77E-04	151.3	Env
Mig1p	3.30E-02	51.1 (r^2=0.35)	Probe
Aft1p	3.83E-02	10.9	SNP XV 180210
Hsf1p	2.60E-02	64.3	Env
Hsf1p	3.79E-04	31.1 (r^2=0.24)	Probe
Tos8p	5.79E-03	60.0	Env
Tos8p	1.92E-02	14.7 (r^2=0.12)	Probe
Gts1p	7.33E-03	43.1	SNP V 17399
Yap3p	1.53E-03	21.6	SNP VII 73452
Opi1p	3.24E-02	22.5	SNP V 15817
Stp2p	1.63E-02	70.4	Env
Stp2p	3.41E-02	61.7 (r^2=0.39)	Probe
Rsc30p	1.00E-03	29.7	SNP VIII 221933

Factor	Q-value	mean(LOD_s)	Covariate
Rsc30p	4.97E-02	60.7 (r^2=0.41)	Probe
Ste12p	2.22E-02	156.7	Env
Ste12p	4.31E-05	85.1 (r^2=0.51)	Probe
Zap1p	3.67E-02	35.1	Env
Gzf3p	3.63E-02	110.2	SNP III 210748
YJL206C	8.80E-04	46.0	SNP VIII 92978
Cbf1p	3.70E-02	34.7	Env
Put3p	1.47E-02	10.3	Env
Put3p	2.26E-02	7.0 (r^2=0.06)	Probe
Phd1p	2.51E-02	12.9	SNP XIII 46084
Phd1p	6.45E-04	24.5 (r^2=0.19)	Probe
Hap4p	4.84E-02	79.0 (r^2=0.41)	Probe
Abf1p	0.00E+00	52.4	Env
Bas1p	3.46E-02	72.9	SNP IV 289639
Rfx1p	4.78E-02	29.7	Env
Ifh1p	4.61E-02	15.7	Env
Hap1p	0.00E+00	38.7	SNP XII 607076
Hap1p	0.00E+00	96.4 (r^2=0.59)	Probe
Pdr8p	5.93E-03	14.2	SNP XII 27765
Sfp1p	0.00E+00	104.6	Env
Yap1p	0.00E+00	225.2	Env
Yap1p	0.00E+00	84.9 (r^2=0.52)	Probe
Yox1p	0.00E+00	93.6	Env
War1p	8.89E-03	36.5	SNP III 301446
Msn2p	3.35E-02	21.0	SNP XV 154309
Mcm1p	8.37E-03	76.7	Env
Mcm1p	3.28E-02	21.5 (r^2=0.17)	Probe
Fkh2p	4.90E-02	17.7	Env
Fkh2p	4.42E-02	10.5 (r^2=0.09)	Probe
Met4p	2.21E-04	79.0	Env
Met4p	4.77E-02	32.9 (r^2=0.24)	Probe
Sko1p	1.76E-02	36.3	SNP XV 180222
Gcr2p	6.25E-04	22.7	SNP XIV 486861
Gcr2p	4.36E-02	8.2 (r^2=0.07)	Probe
Gis2p	3.79E-02	12.6	SNP XIV 582954
Cin5p	2.35E-02	45.6	Env
Hms1p	3.21E-02	27.3	Env
Sfl1p	0.00E+00	39.1	SNP I 186488
Pip2p	4.34E-02	35.4 (r^2=0.25)	Probe
Usv1p	9.62E-04	41.3	SNP XI 98330

Factor	Q-value	mean(LOD_s)	Covariate
Rox1p	4.72E-02	35.5	SNP XIV 449639
Fhl1p	3.76E-02	31.7 (r^2=0.25)	Probe
Arr1p	3.50E-02	111.9	Env

Table B.9: Properties of inferred yeastract factor activations. Q-value and average LOD score of association with SNPs (with best locus) or environment indicator is given for associations with combined Q-value < 0.050

Factor	Q-value	mean(LOD_s)	Covariate
Glycolysis / Gluconeogenesis (00010)	4.63E-02	19.9	SNP XIV 486861
Nitrogen metabolism (00910)	0.00E+00	119.9	SNP XII 433955
Lysine biosynthesis (00300)	4.00E-05	25.6	SNP II 479166
Tryptophan metabolism (00380)	0.00E+00	29.2	SNP XV 779974
Arginine and proline metabolism (00330)	0.00E+00	46.7	SNP XV 59733
Aminoacyl-tRNA biosynthesis (00970)	4.50E-02	21.7	SNP XIV 486861
Metabolic pathways (01100)	0.00E+00	393.2	Env
Fatty acid metabolism (00071)	7.66E-03	67.1	SNP I 55329

Table B.10: Properties of inferred kegg factor activations. Q-value and average LOD score of association with SNPs (with best locus) or environment indicator is given for associations with combined Q-value < 0.050

Factor	Q-value	mean(LOD_s)	Covariate
Factor 1	0.00E+00	289.5	Env
Factor 4	0.00E+00	19.9	SNP XV 89211
Factor 5	0.00E+00	61.4	SNP XIV 449639
Factor 7	1.96E-02	10.9	SNP XV 446514
Factor 8	2.34E-04	16.2	SNP XII 681096
Factor 9	1.11E-03	15.6	SNP XII 659357
Factor 10	1.32E-03	15.1	SNP XII 672779
Factor 11	0.00E+00	19.2	SNP XII 634225
Factor 12	0.00E+00	17.7	SNP II 506661
Factor 14	6.23E-03	12.6	SNP XI 180221
Factor 15	1.99E-03	14.2	SNP III 76127
Factor 16	1.65E-02	11.2	SNP XIII 404546
Factor 17	2.54E-02	10.1	SNP XV 838599
Factor 18	3.12E-02	9.8	SNP XIII 216022
Factor 19	3.15E-02	9.7	SNP XV 619862
Factor 20	0.00E+00	21.3	SNP II 506661
Factor 21	2.25E-03	13.8	SNP XV 842027
Factor 22	0.00E+00	24.1	SNP V 395442
Factor 23	2.36E-03	14.1	SNP XIII 78655
Factor 24	0.00E+00	18.5	SNP III 75021
Factor 25	1.08E-02	11.5	SNP XV 496730
Factor 26	9.58E-03	11.6	SNP IX 195965
Factor 27	1.98E-02	10.9	SNP II 486640
Factor 28	3.32E-02	9.7	SNP XVI 454307

Table B.11: Properties of inferred freeform factor activations. Q-value and average LOD score of association with SNPs (with best locus) or environment indicator is given for associations with combined Q-value < 0.050

Locus	Factor	Q-value	mean(LOD_s)
III 79091	War1p	4.53E-02	26.3
III 79091	Thi2p	9.65E-03	12.3
III 79091	Gzf3p	3.63E-02	110.2
IV 106892	Bas1p	3.46E-02	72.9
IV 106892	Gzf3p	3.90E-02	9.8
IV 106892	Yap3p	1.73E-03	35.2
V 6335	Gts1p	7.33E-03	43.1
V 6335	Opi1p	3.24E-02	22.5
V 6335	Kar4p	0.00E+00	45.7
V 420595	Rsc30p	3.60E-02	10.5
V 420595	Kar4p	0.00E+00	40.5
V 420595	Hap1p	1.89E-02	10.4
V 420595	Sfl1p	3.78E-04	36.4
VII 55458	Gts1p	1.79E-02	13.0
VII 55458	Yap3p	1.53E-03	21.6
VII 449898	Gzf3p	4.24E-02	12.2
VII 449898	Pdr8p	4.52E-02	14.7
XII 611810	Hap1p	0.00E+00	38.7
XII 611810	Pdc2p	1.84E-02	16.4
XII 611810	Pdr8p	2.21E-02	13.2
XIII 46084	Pho4p	2.90E-02	15.5
XIII 46084	Phd1p	2.51E-02	12.9
XIII 46084	Ino2p	4.69E-02	11.3
XIV 449639	Rox1p	4.72E-02	35.5
XIV 449639	Gcr2p	6.25E-04	22.7
XIV 449639	Rtg3p	3.01E-02	21.4
XIV 449639	Gis2p	3.79E-02	12.6
XV 174364	Pdr3p	2.09E-02	14.6
XV 174364	Sko1p	1.76E-02	36.3
XV 174364	Spt2p	8.70E-05	34.1
XV 174364	Aft1p	3.83E-02	10.9
XV 174364	Gis1p	4.18E-02	11.9
XV 174364	Msn2p	3.35E-02	21.0
XV 380725	Gis1p	4.79E-02	9.5
XV 380725	Sum1p	6.18E-03	13.2
XVI 932310	Rsc30p	4.52E-02	14.2
XVI 932310	Sfl1p	2.76E-02	14.8

Table B.12: Associations to loci with more than one yeastract factor association. Q-value and average LOD score are given for all factors associated to each locus.

Factor	Q-value	mean(LOD_s)	Covariate
Locus	Factor	Q-value	mean(LOD_s)
XIV 486861	Aminoacyl-tRNA biosynthesis (00970)	4.50E-02	21.7
XIV 486861	Glycolysis / Gluconeogenesis (00010)	4.63E-02	19.9

Table B.13: Associations to loci with more than one kegg factor association. Q-value and average LOD score are given for all factors associated to each locus.

Locus	Factor	Q-value	mean(LOD_s)
II 486640	Factor 5	1.73E-02	11.1
II 486640	Factor 7	2.00E-02	10.8
II 486640	Factor 8	2.11E-02	10.6
II 486640	Factor 12	0.00E+00	17.7
II 486640	Factor 20	0.00E+00	21.3
II 486640	Factor 27	1.98E-02	10.9
II 697894	Factor 20	3.11E-02	9.8
II 697894	Factor 12	2.01E-03	14.3
III 91287	Factor 8	3.40E-02	9.6
III 91287	Factor 15	1.99E-03	14.2
III 91287	Factor 16	3.51E-02	9.4
III 91287	Factor 17	4.78E-02	8.8
III 91287	Factor 24	0.00E+00	18.5
III 91287	Factor 28	3.49E-02	9.4
V 350744	Factor 14	4.96E-02	8.7
V 350744	Factor 22	0.00E+00	24.1
IX 195965	Factor 25	4.18E-02	9.0
IX 195965	Factor 26	9.58E-03	11.6
IX 195965	Factor 4	3.05E-02	9.8
XII 635380	Factor 4	4.21E-02	9.0
XII 635380	Factor 8	2.34E-04	16.2
XII 635380	Factor 9	1.11E-03	15.6
XII 635380	Factor 10	1.32E-03	15.1
XII 635380	Factor 11	0.00E+00	19.2
XII 635380	Factor 12	1.50E-03	14.9
XII 635380	Factor 23	2.53E-02	10.0
XIII 28622	Factor 18	3.12E-02	9.8
XIII 28622	Factor 23	2.36E-03	14.1
XIII 28622	Factor 7	2.56E-02	10.1
XIV 418269	Factor 5	0.00E+00	61.4
XIV 418269	Factor 30	3.37E-02	9.6
XIV 418269	Factor 8	1.67E-03	14.7
XV 96633	Factor 18	4.94E-02	8.7
XV 96633	Factor 4	0.00E+00	19.9
XV 96633	Factor 5	2.38E-02	10.3
XV 96633	Factor 24	9.55E-03	11.6
XV 838599	Factor 17	2.54E-02	10.1
XV 838599	Factor 21	2.25E-03	13.8

Table B.14: Associations to loci with more than one freeform factor association. Q-value and average LOD score are given for all factors associated to each locus.

Locus	Chr	Pos.	1. Probes with *trans* associations	2. Probes with downstream factor associations	3. (2.) with stronger factor association	1.&2	$\frac{1.\&2.}{1.}$	$\frac{1.\&3.}{1.}$	$\frac{1.\&3.}{1.\&2}$
AMN1	2	555575	51	73	73	3	0.06	0.06	1.00
HAP1	12	644082	66	53	53	31	0.47	0.47	1.00
PHO84	13	46084	31	454	454	11	0.35	0.35	1.00
MKT1	14	449639	218	514	508	21	0.10	0.07	0.71
IRA2	15	174364	271	1443	1438	164	0.61	0.59	0.97

Table B.15: *trans* eQTL peaks with at least 50 associations. For each peak, the number of significant associations to probe expression levels (1.), number of associations for Yeastract factor activations significantly associated with the peak (2.), number of genes more strongly associated with the factor than the peak locus genotype (3.) are given, together with the number and fraction of *trans* eQTLs explained by the factors, fraction of *trans* eQTLs more strongly associated with the factor, and fraction of *trans* eQTLs associated with a factor that are more strongly associated with the factor.

Sample	Generation	Replica	Type	Ploidy	Condition	Timepoint	Coverage
WA-NA_Initial_R1_F6_T0	6	1	Pool	Haploid	Permissive	0	23.8
WA-NA_Initial_R2_F6_T0	6	2	Pool	Haploid	Permissive	0	13.1
WA-NA_Heat_R1_F6_T4	6	1	Pool	Haploid	Heat 40C	2	19.3
WA-NA_Heat_R2_F6_T4	6	2	Pool	Haploid	Heat 40C	2	25.7
WA-NA_Initial_R1_F6_S1	6	1	Segregant	Haploid	Permissive	0	20.3
WA-NA_Initial_R2_F6_S1	6	2	Segregant	Haploid	Permissive	0	27.4
WA-NA_Mock_R1_F12_T4	12	1	Pool	Haploid	Permissive	2	115.4
WA-NA_Heat_R1_F12_T4	12	1	Pool	Haploid	Heat 40C	2	129.3
WA-NA_Mock_R2_F12_T4	12	2	Pool	Haploid	Permissive	2	105.7
WA-NA_Initial_R2_F12_T0	12	2	Pool	Haploid	Permissive	0	107.3
WA-NA_Heat_R2_F12_T2	12	2	Pool	Haploid	Heat 40C	1	54.8
WA-NA_Heat_R2_F12_T4	12	2	Pool	Haploid	Heat 40C	2	83.7
WA-NA_Heat_R2_F12_T6	12	2	Pool	Haploid	Heat 40C	3	65.9
WA-NA_Diploid-heat_R2_F12_T6	12	2	Pool	Diploid	Heat 40C	3	32.6
WA-NA_Diploid-heat_R1_F12_T4	12	1	Pool	Diploid	Heat 40C	2	88.6
WA-NA_Paraquat_R1_F12_T4	12	1	Pool	Haploid	Paraquat	2	150

Table B.16: Average sequencing coverage at segregating sites for different intercross generations, ploidies, conditions, and selection timepoints.

Chromosome	Location	Combined change (R1 + R2)
1	11998	0.38
1	207560	0.29
2	472111	0.33
4	1444248	-0.26
4	373030	0.3
4	430662	0.35
4	474894	0.39
4	572931	0.53
4	700611	-0.35
7	1081499	-0.59
8	261643	0.28
9	77497	0.27
10	420908	0.27
10	450702	0.26
10	492479	0.26
10	613016	0.45
12	388635	-0.38
12	491120	-0.28
12	967942	-0.35
14	49576	0.3
15	184627	0.39
15	580877	-0.28

Table B.17: Regions selected for during intercross rounds between F6 and F12 generations.

Chromosome	Location	Combined allele frequency change (R1 + R2)
1	119382	0.31
2	472031	-0.52
2	517350	-0.68
4	1313885	0.42
4	454021	-0.31
4	496586	-0.3
7	131690	0.3
7	859960	0.83
9	292345	-0.32
10	234117	-0.39
10	420908	-0.42
10	679911	-0.28
12	140165	0.38
12	730764	-0.28
13	743221	-0.27
13	893719	-0.56
14	480623	0.46
15	1032447	-0.76
15	179760	-1.27

Table B.18: Heat QTLs detected with artificial selection. All loci with total allele frequency change of at least 0.3, and at least 0.1 in both replicas are given.

Gene	F12 T4	F12 T0	Change T0 - T4
Q0045	0.4	36.5	-36.1
Q0250	2	37	-35
Q0255	1.7	36.1	-34.4
Q0060	0.3	31.9	-31.6
Q0115	1.2	32	-30.8
Q0275	1.8	32.3	-30.5
Q0105	3.3	32.8	-29.5
Q0050	0.2	27.7	-27.5
Q0120	1.6	28.3	-26.7
Q0070	0.2	26	-25.8
Q0085	1.9	26.1	-24.2
Q0065	0.2	22	-21.8
Q0182	0.7	18.3	-17.6
Q0032	0.9	12.3	-11.4
Q0142	0.3	11.3	-11
YLR162W	44.5	55.4	-10.9
Q0140	3.3	13.2	-9.9
Q0130	2.7	11.4	-8.7
Q0144	2.2	10.8	-8.6
Q0143	0.7	7.9	-7.2
Q0080	0.1	6.2	-6.1
YDR366C	11.5	17.6	-6.1
Q0110	0.7	6	-5.3
Q0010	13.7	18	-4.3
Q0092	0	3.5	-3.5
Q0017	0.1	2.5	-2.4
YEL074W	4.1	5.9	-1.8
YIR044C	1.1	2.9	-1.8
YIL174W	0.7	1.9	-1.2
YJL225C	2.1	3.3	-1.2
YNL337W	1.6	2.8	-1.2
YOL166C	1.6	2.8	-1.2
YHR216W	3.4	4.4	-1
YLR465C	2.6	0.9	1.7
YDR340W	8.3	3.9	4.4

Table B.19: Genes changing in copy number upon selection.