Supporting methods

Following the generative model (Equation (2), Methods) we use standard association and interaction statistics [1] to test for associations between known variables (genotype of SNP n, environment indicator, or mRNA expression level) and the inferred factor activations. For completeness, we first present the model and test statistics used for both associations and interactions, followed by the significance testing approach. The derivation is developed explicitly using the SNP genotype as the known variable and factor activation as the dependent variable; tests for the other covariates (or eQTL effects) are performed analogously.

Association and interaction test statistics

We perform independent tests for association between the activation $\mathbf{x}_k$ of individual factor k and genotype $\mathbf{s}_n$ of SNP n, fitting a liner model of the form

$$x_{k,j} = \mu_k + \underbrace{\beta_{k,n}s_{n,j}}_{\text{SNP effect}} + \epsilon_{k,j}, \tag{1}$$

(Methods), assuming Gaussian observation noise $\epsilon_{k,j} \sim \mathcal{N}(0, \sigma_{k,j}^2)$. For each pair of SNP n and factor k, we calculate the association log-odds (LOD) score

$$L_{k,n}^a = \log P(\mathbf{x}_k \,|\, \beta_{k,n}) - \log P(\mathbf{x}_k \,|\, \beta_{k,n} = 0) \tag{2}$$

as a test statistic. The weight in the foreground model $\beta_{k,n}$, the mean μ_k and the noise level $\sigma_{k,n}^2$ are fit by maximum likelihood independently for every evaluation.

Test statistics for the interaction terms are calculated analogously based on an independent interaction model. In short, we calculate the residuals of the factor analysis model and apply a standard interaction model between SNP n, factor k and gene g. This corresponds to the linear model

$$y_{g,j} = \mu_g + \overbrace{\underbrace{\theta_{g,n}s_{n,j}}_{\text{SNP effect}} + \underbrace{w_{g,k}x_{k,j}}_{\text{factor effect}}}^{\text{direct effects}} + \underbrace{\phi_{g,k,n}\left(s_{n,j}x_{k,j}\right)}_{\text{interaction term}} + \underbrace{\left[\sum_{l \neq k} w_{g,l}x_{l,j}\right]}_{\text{remaining factor effect}} + \psi_{g,j}, \tag{3}$$

where the expression level of gene probe g for individual j is described by fitted effects of the tested SNP $s_{n,j}$, learned factor activation $x_{k,j}$ and the interaction term $s_{n,j}x_{k,j}$ with the residuals explained by 0-meaned Gaussian noise $\psi_{g,j}$. The log-odds test statistic for the interaction between factor k and SNP n to influence gene g follows as

$$L_{g,k,n}^i = \log P(\mathbf{y}_g \,|\, \phi_{g,k,n}) - \log P(\mathbf{y}_g \,|\, \phi_{g,k,n} = 0). \tag{4}$$

The respective mean variable μ_g, weights $\theta_{g,n}$, $w_{g,k}$ (but not $w_{g,k'}$ where $k' \neq k$), and $\phi_{g,k,n}$, as well as noise variance $\psi_{g,j}$ are fitted independently using maximum likelihood for each factor, gene, SNP triplet. The contribution from all remaining factors is not refit to preserve the sparsity pattern learnt from the factor inference. To reduce the number of effective tests, we used the strongest interaction LOD score $\hat{L}_{g,n}^i = \max_k L_{g,k,n}^i$ across factors, thus performing tests for every SNP and gene pair. This approach is motivated by the assumption that at most a single factor is interacting with a given gene-SNP pair. The consistency of the strongest interacting factor is informative of the identifiability of the interaction effect (see below).

Significance testing using several random initialisations

For all our analysis of intermediate phenotypes, we generated factor inference results from $R = 20$ random initialisations of the model parameters to beter capture the variability in the model and avoid overfitting of inferred factor activations to local optima (See Statistical identifiability of factors below). Thus, we designed a significance testing scheme based on Q-values [2] that employs the full set of runs, taking the uncertainty in the factor posterior distributions into account. We present this approach for associations, testing interactions is analogous except the specifics of permutations highlighted in the text. In case of analyses where the multiple restarts are not used (i.e. all analyses not employing the inferred factors, such as trans eQTLs and genotype-environment interactions), we calculated Q-values from the single instance. In all cases, the null distribution of LOD scores was obtained by combining all calculated null statistics in the random restart.

Q-value calculation For every run $r = 1, \ldots, R$ of the factor anlysis model, we evaluated the test statistics of factor associations ($L_{k,n}^a$) for every pair of factor k and SNP s. This analysis was then repeated on 20 permuted datasets for each run with the genotypes shuffled with respect to the factor activations, while keeping individual segregants grown in two conditions paired. For interaction LOD scores, the factor activations and gene expression levels were not permuted with respect to each other. From this empirical null distribution of LOD scores in run r (across all SNPs and factors), we calculated Q-values $q_{n,r}^r$ (local FDR) for each candidate association [2] between SNP n and inferred posterior of factor k in this run.

Combining Q-values The Q-values from all runs were then combined into an overall Q-value $q_{k,n} = R^{-1} \sum_{r=1}^{R} q_{k,n}^r$, which was used to assess significance at a given FDR threshold.

From a probabilistic viewpoint, averaging Q-values over multiple restarts of the model can intuitively be thought of as integrating out the uncertainty from the factor inference. For example, for an association test assessing the significance of the weight $\beta_{k,n}$, we are truly interested in the probability of an association being absent (Bayesian Q-value, see for example [3]) given uncertain inference of the factor activation $P(\mathbf{x}_k \,|\, \mathbf{Y}, \boldsymbol{\pi})$. Conditioned on the observed data $\mathbf{Y}$ and prior $\boldsymbol{\pi}$ this probability follows as

$$P(\beta_{k,n} = 0 \,|\, \mathbf{Y}, \boldsymbol{\pi}, \mathbf{s}_n) = \int_{\mathbf{x}_k} P(\beta_{k,n} = 0 \,|\, \mathbf{x}_k, \mathbf{s}_n) P(\mathbf{x}_k \,|\, \mathbf{Y}, \boldsymbol{\pi}). \tag{5}$$

In general this integral is not analytically tractable. Assuming we have instead a number of R samples $\mathbf{x}_k^r$ from the factor posterior, the integral can be approximated by

$$\approx \frac{1}{R} \sum_{r=1}^{R} P(\beta_{k,n} = 0 \,|\, \mathbf{x}_k^r, \mathbf{s}_n) \tag{6}$$

in a Monte Carlo fashion. Finally, identifying the null probabilities as Bayesian Q-values we get

$$= \frac{1}{R} \sum_{r=1}^{R} q_{k,n}^r. \tag{7}$$

Note that the restarts from the factor analysis model are not exactly samples from its posterior but nevertheless characterise the posterior uncertainty sufficiently well (See also Simulation study below). Full MCMC sampling is computationally infeasible due to the size of the regulatory network; for a comparison of MCMC sampling and deterministic inference as employed here, see [5].
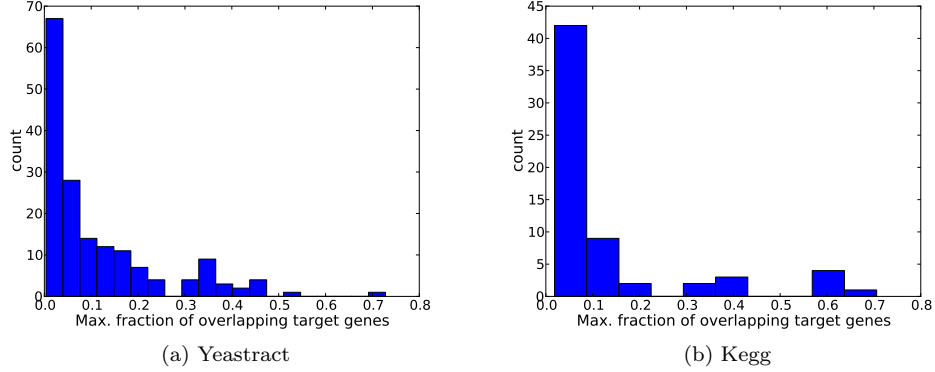
Figure 1: Histogram over the maximum fraction of overlapping target genes per factor for **(a)** Yeastract prior information and KEGG pathways respectively **(b)**. An overlap of 1.0 indicates that at least one factor shares identical target genes; an overlap of 0 indicates that no other factor has even a single target gene in common.

Consistency of the interacting factor On simulations (see below) and real data, most interactions have a single factor that consistently produces the highest LOD scores for almost all random initialisations of the model. This fraction can be used to quantify the degree to which an interaction can be pinpointed to a specific factor (See Simulation and Figure 3 in this text).

Statistical identifiability of factors

Factor analysis models have intrinsic symmetries with implications on the statistical identifiability of the activation of individual latent factors. The inner product of weights $\mathbf{W}$ and activations $\mathbf{X}$ is invariant under a general class of transformations of the form

$$\mathbf{Y} = \mathbf{W} \left(\mathbf{R} \mathbf{R}^T \right) \mathbf{X} = \left(\mathbf{W} \mathbf{R} \right) \left(\mathbf{R}^T \mathbf{X} \right) = \tilde{\mathbf{W}} \tilde{\mathbf{X}}. \tag{8}$$

Here, $\mathbf{R}$ is an arbitrary non-singular matrix, including rotations, rescaling and sign flips of factors. One approach to resolve these ambiguities are post-hoc transformations of the factor solution; for example the varimax rotation, which is frequently applied in PCA and ICA models (see [4] for discussion). An alternative is to address these ambiguities by introducing additional constraints into the model, explicitly restricting the space of valid factor solutions. In the sparse factor analysis used in this work, these contraints are introduces as prior sparsity pattern $P(z_{g,k} = 1) = \pi_{g,k}$. (See also discussion in [5]). Provided the prior sparsity profiles of the factors $\boldsymbol{\pi}_k$ are sufficiently distinct, this prior rules out unwanted symmetries including factor permutations. Figure 1 depicts the maximum fraction of overlapping target genes across factors for KEGG and Yeastract prior information as used for the analysis in the main text. These results show, that in particular the Yeastract data yields a high degree of factors with distinct prior connectivity profile. The relation of the prior orthogonality and the empirical reproducibility of the inferred factors is investigated on simulated data below (Simulation study).

Simulation study

We used a simulated dataset to check on the accuracy of the sparse factor analysis and the recovery of simulated associations and interactions.

Simulation procedure

We simulated $J = 80$ individuals with $N = 1200$ SNPs. The simulated minor allele frequency was 0.5 for each SNP, and the allele configuration $s_{n,j}$ of SNP n was encoded as $(1, 0)$ or $(1, 1)$, including a column for mean effects. Similar to the yeast dataset investigated in the main text, we simulated a single environmental factor. We considered $K = 50$ intermediate phenotypes that were meant to resemble transcript factors and a total of $G = 1000$ target genes. The annotated connectivity structure between factor and genes was subsampled from the full Yeastract [6] information, discarding factors that had fewer than 3 target genes to avoid unidentifiable factors (See Methods). Assuming an FPR 0.05 and FNR 0.05 we generated a prior connectivity matrix $\boldsymbol{\pi}$, providing uncertain information about the regulatory links. From this matrix we drew an actual factor-gene association matrix $\mathbf{Z}$ which was used in the dataset simulation.

Simulation of the factor activations and gene expression followed the causal direction of the true mechanisms. First, we simulated linear direct effects between individual SNPs and the environmental variable acting on the 50 factors that were meant to resemble transcription factors or pathway components (see Statistical model). The probability that a factor was genetically driven was set to 0.6, the probability of an environmental influence of the factor was 0.3 and with 0.1 probability the factor activation was set to be randomly Gaussian distributed. Next, conditioned on the SNP genotype and factor activations, we sampled independent gene expression values. Simulated effects from factors and SNPs on each of the 1000 gene expression values include direct effects from SNPs, factors and SNP-factor interaction effects as well as observation noise

$$y_{g,j} = \sum_{n=1}^{N} \left[\underbrace{\theta_{g,n} s_{n,j}}_{\text{SNP effect}} + \sum_{k=1}^{K} \left[z_{g,k} \underbrace{\gamma_{g,k} x_{k,j}}_{\text{factor effect}} + \underbrace{\phi_{g,k,n} \left(s_{n,j} x_{k,j} \right)}_{\text{interaction term}} \right] \right] + \underbrace{\epsilon_{g,,j}}_{\text{noise}}. \tag{9}$$

Note that the indicators $z_{g,k}$ ensure that any factor k has only downstream effects on its true target genes. The weights for direct factor effects were drawn to be Gaussian, $\gamma_{g,k} \sim \mathcal{N}(0, 1)$. Interaction weights $\phi_{g,k,n}$ and association weights to SNPs $\theta_{g,n}$ were drawn from a sparse Gaussian distribution. For 1% of all genes a direct effect was simulated and for 0.001% of all factor-SNP-gene triplets a non-zero interaction weight $\phi_{g,k,n}$ was simulated (each $\sim \mathcal{N}(0, 1)$). We drew the noise precision $\tau_{g,n,k}$ from a gamma distribution with $\tau_{g,n,k} \sim \Gamma(1, 0.1)$, and the individual noise levels $\epsilon_{g,n,k,j}$ were drawn from $\mathcal{N}(0, \frac{1}{\tau_{g,n,k}})$.

Model fitting

Given the simulated dataset, we followed the analysis procedure described in the main text (Methods). We run the sparse factor analysis model for 20 random restarts, each time inferring the transcript factor activations given the prior network structure ($\boldsymbol{\pi}$).

Evaluation of factor inference

Accuracy to recover regulatory links

The sparse factor analysis model was able to recover the true relational links between factors and genes with high accuracy (Table 1). For comparison, this table also includes the accuracy when solely using the imperfect prior knowledge $\boldsymbol{\pi}$. As the recovered factor activations and the connectivity structure are

Method	ACC	AUC
Prior	0.820 ± 0.002	0.722 ± 0.003
SparseFA	0.897 ± 0.010	0.877 ± 0.015

Table 1: Accuracy (ACC) and area under the operator curve (AUC) for recovering the true network structure on the simulated dataset using the prior network structure (Prior) and the sparse factor analysis model (SparseFA). Reported are mean accuracies and standard deviations estimated form 20 random restarts of the model or draws fromt the prior respectively.
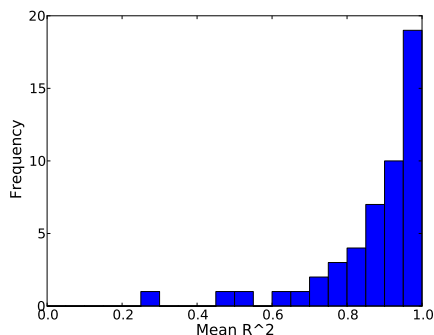

Figure 2: Histogram of average Pearson r^2 between the inferred factor activation in a single run and the simulation ground truth.

inherently linked, these results suggest that the sparse factor analysis model is also likely to recover accurate factor activations.

Factor identifiability

Next, we explicitly checked this hypothesis, investigating that the inferred factor activations $\mathbf{X}$ resemble the true, simulated activations. Despite the high accuracy in recovering the connectivity structure, this is interesting, as factor analysis-type models are generally prone to a large class of transformation invariances (See Statistical identifiability of factors). Twenty nine of 50 inferred factor posterior means had average Pearson's r^2 to the true simulated activation of greater than 0.9 (Figure 2). This shows that most factors activations are reproducibly inferred to be very similar to their true simulated state.

Accuracy to recover interactions

Next, we applied the interaction model (Association and interaction testing) to identify SNP-gene pairs for which there exists a significant interaction to at least one factor.

Using a Q-value cutoff of 0.01, we recovered 284 of the 600 simulated interactions and two false positives, corresponding to 47% sensitivity and 99.3% specificity, the latter in line with the chosen FDR cutoff. Thus, at this stringent cutoff, we are both sensitive and specific in recovering the interaction effects.

Furthermore, 268 of 284 SNP-gene pairs had strongest interaction LOD scores to the true interacting factor in the majority of the random restarts (Figure 3a). This shows that in most cases, we can accurately recover the true interacting factor. Fifteen of the remaining 16 interactions were with two factors that were not well identified (average r^2 with true activation 0.27 and 0.48, respectively), with another factor
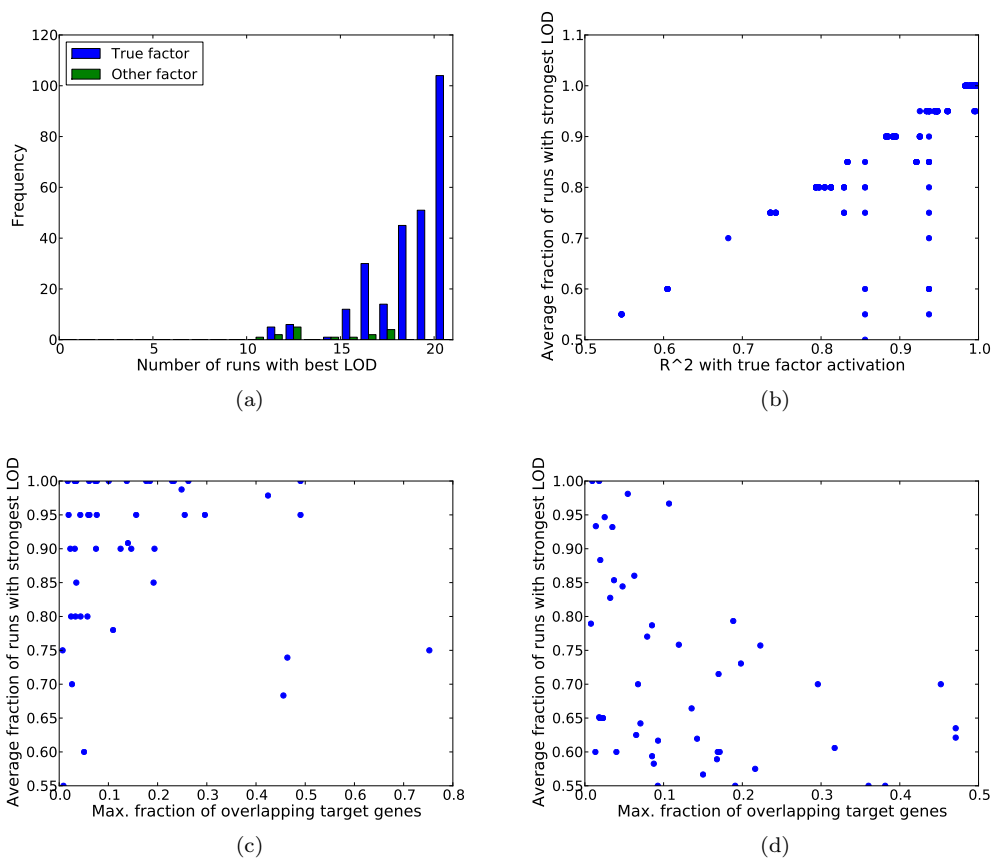
Figure 3: (a) Histogram of number of runs for each interacting SNP-gene pair where the true interacting factor (blue) or any one factor (green) has the strongest interaction LOD score. (b) Scatter plot of the fraction of times one factor has the strongest interaction LOD score and the average correlation of its inferred activation to the underlying truth. (c) Scatter plot of the fraction of times one factor has the strongest interaction LOD score and the uniqueness of its prior information. (d) Same as (c) for Yeastract prior used in the main study.

capturing their effect and reproducibly exhibiting the interaction signal (Figure 3a). However, in these cases, neither the true interacting factor, nor the factor strongest interacting showed the strongest LOD score in more than 17 runs.

In general, the more reproducible the factor, the more reproducibly its interaction effects are correctly the strongest ones across all factors (Pearson's r between average factor posterior correlation to true factor activation, and average fraction of random restarts for which the factor has the strongest LOD score = 0.93, Figure 3b). This effect is related to the uniqueness of the factor's prior information, with less unique factors showing less reproducible effects (Figure 3c).

Thus, care must be taken when interpreting interactions for which the interacting factor is not unambiguously defined, but it is possible to uncover the factor if its effect is well reproducible. The simulation data suggest that using a filter of 90% of runs yielding the strongest interaction LOD for one factor is in most cases sufficient to retain a high fraction of true positives.

In conclusion, even though the interaction effect LOD scores fluctuate between individual runs, we are able to find the interaction effects, and in most cases, also pinpoint the responsible factor. This shows that our testing approach robustly and elegantly incorporates the inherent variability in the model that is due to multiple posterior modes.

References

1. M. Lynch and B. Walsh. *Genetics and Analysis of Quantitative Traits.* Sinauer Associates, 1 edition, January 1998.

2. J.D. Storey and R. Tibshirani. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, 100(16):9440, 2003.

3. J.D. Storey. The positive false discovery rate: A Bayesian interpretation and the q-value. *Annals of Statistics*, 31(6):2013–2035, 2003.

4. P. Iosifina and W. Lorenz. Factor analysis for gene regulatory networks and transcription factor activity profiles. *BMC Bioinformatics*, 2007.

5. O. Stegle, K. Sharp, J. Winn, and M. Rattray. A comparison of inference in sparse factor analysis models. *Journal of Machine Learning Research*, 2009. in preparation.

6. M. C. Teixeira, P. Monteiro, P. Jain, S. Tenreiro, A. R. Fernandes, N. P. Mira, M. Alenquer, A. T. Freitas, A. L. Oliveira, and I. Sá-Correia. The YEASTRACT database: a tool for the analysis of transcription regulatory associations in Saccharomyces cerevisiae. *Nucleic Acids Research*, 34:D3–D5, 2006.