6-9-2021

Inferring TF activities and activity regulators from gene expression data with constraints from TF perturbation data

Cynthia Z Ma

Michael R Brent

Bioinformatics, 37(9), 2021, 1234–1245
doi: 10.1093/bioinformatics/btaa947
Advance Access Publication Date: 22 December 2020
Original Paper

Gene expression

Inferring TF activities and activity regulators from gene expression data with constraints from TF perturbation data

Cynthia Z. Ma [1,2] **and Michael R. Brent** [1,2,3]*

[1]Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO 63110, USA, [2]Department of Computer Science and Engineering, Washington University, St. Louis, MO 63130, USA and [3]Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110, USA

*To whom correspondence should be addressed.

Associate Editor: Pier Luigi Martelli

Received on July 3, 2020; revised on September 26, 2020; editorial decision on October 21, 2020; accepted on October 27, 2020

Abstract

Motivation: The activity of a transcription factor (TF) in a sample of cells is the extent to which it is exerting its regulatory potential. Many methods of inferring TF activity from gene expression data have been described, but due to the lack of appropriate large-scale datasets, systematic and objective validation has not been possible until now.

Results: We systematically evaluate and optimize the approach to TF activity inference in which a gene expression matrix is factored into a condition-independent matrix of control strengths and a condition-dependent matrix of TF activity levels. We find that expression data in which the activities of individual TFs have been perturbed are both necessary and sufficient for obtaining good performance. To a considerable extent, control strengths inferred using expression data from one growth condition carry over to other conditions, so the control strength matrices derived here can be used by others. Finally, we apply these methods to gain insight into the upstream factors that regulate the activities of yeast TFs Gcr2, Gln3, Gcn4 and Msn2.

Availability and implementation: Evaluation code and data are available at https://doi.org/10.5281/zenodo.4050573.

Contact: brent@wustl.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

The activity level of a transcription factor (TF) in a given cell is the extent to which it is exerting its regulatory potential on its target genes. Cells process information, in part, by changing the activity levels of TFs, thereby changing the transcription rates of their target genes. Changes in TF activity can occur by several molecular mechanisms, including transcriptional and post-transcriptional regulation of the gene that encodes the TF, relocalization of the TF into or out of the nucleus, covalent modification of the TF such as phosphorylation, and non-covalent binding of the TF by other proteins. Because it has multiple molecular mechanisms, TF activity is difficult to measure directly. However, it may be possible to infer changes in TF activity level from changes in the expression levels of the TF's target genes (Alvarez *et al.*, 2016; Balwierz et al., 2014; Barenco et al., 2006; Boorsma et al., 2008; Boulesteix and Strimmer, 2005; Chen et al., 2013, 2017; Cheng et al., 2007; Cokus et al., 2006; Fröhlich, 2015; Fu et al., 2011; Gao et al., 2004; Garcia-Alonso et al., 2018; Gitter et al., 2013; Jiang et al., 2015; Khanin et al., 2007; Li et al., 2014; Liao et al., 2003; Nachman et al., 2004; Ocone and Sanguinetti, 2011; Sanguinetti et al., 2006; Schacht et al., 2014; Shi et al., 2009; Tchourine et al., 2018; Yu and Li, 2005; Zhu et al., 2013). If successful, this would provide insights into which TFs are involved in the transcriptional response to a given stimulus, such as a drug, extracellular signal or nutrient influx. In principle, TF activity inference could also be used to predict the transcriptomic effects of direct perturbations of TF activity levels, such as deletion or over-expression of a TF-encoding gene. Finally, inferred activity levels could be used to improve TF network mapping (Arrieta-Ortiz et al., 2015; Barenco et al., 2009; Bussemaker et al., 2001, 2017; Cokus *et al.*, 2006; Fu *et al.*, 2011; Gao *et al.*, 2004; Gitter *et al.*, 2013; Lam et al., 2016; Lee and Bussemaker, 2010; Nachman *et al.*, 2004; Shi *et al.*, 2009; Tchourine *et al.*, 2018; Wang et al., 2008; Yang et al., 2005; Yu and Li, 2005).

Most algorithms for inferring TF activity (TFA) from gene expression data fit the parameters of a mathematical model to the expression data (Balwierz *et al.*, 2014; Barenco *et al.*, 2009; Boulesteix and Strimmer, 2005; Bussemaker *et al.*, 2001; Chen *et al.*, 2017; Cokus *et al.*, 2006; Fröhlich, 2015; Fu *et al.*, 2011; Jiang *et al.*, 2015; Khanin *et al.*, 2007; Li *et al.*, 2014; Liao *et al.*, 2003; Nachman *et al.*, 2004; Sanguinetti *et al.*, 2006; Schacht *et al.*, 2014; Tchourine *et al.*, 2018; Yu and Li, 2005). These models include

parameters representing TFA levels, the values of which vary from one biological sample to another. Some models also include parameters that are constant across samples but vary as a function of the TF and target gene. These parameters reflect factors such as the affinity of the TF for sites in the promoter of each gene. In some approaches, these parameters, called *control strengths* (CSs), are obtained directly from TF binding data (Cokus *et al.*, 2006; Gao *et al.*, 2004; Jiang *et al.*, 2015; Li *et al.*, 2014) or by scanning models of TF binding specificity across promoters (Balwierz *et al.*, 2014; Bussemaker *et al.*, 2001; Conlon et al., 2003). In other approaches, they are treated as unknowns and obtained by fitting the model to gene expression data (Boulesteix and Strimmer, 2005; Fu *et al.*, 2011; Khanin *et al.*, 2007; Liao *et al.*, 2003; Nachman *et al.*, 2004; Sanguinetti *et al.*, 2006; Yu and Li, 2005). In this case, gene expression is typically modeled as depending linearly on the TFA levels and on the CSs (Boulesteix and Strimmer, 2005; Fu *et al.*, 2011; Liao *et al.*, 2003; Yu and Li, 2005), so the model as a whole is bilinear. More highly parameterized, non-linear models that more closely reflect the underlying biochemistry have also been tried when modeling a small number of TFs (Khanin *et al.*, 2007; Nachman *et al.*, 2004; Sanguinetti *et al.*, 2006). Because it has relatively few parameters and can be fit by a simple algorithm, we focus on the bilinear framework.

To infer the activity of a set of TFs from the expression of their target genes, an inference algorithm must know at least some of the targets of each TF. We refer to this input as a *TF network map* (Brent, 2016). TF network maps link each TF to the targets it has the potential to regulate directly, given the right conditions. These maps are qualitative, so they can be represented by binary adjacency matrices. Fitting the bilinear model yields a control strength for each edge of the input map. Multiplying these CSs by the TFA levels inferred for a sample of cells yields a sample-specific network map showing how strongly each TF is influencing the expression of each of its targets in that sample. In previous work featuring inferred CS values, qualitative network maps have mostly been constructed from binding location data obtained by chromatin immunoprecipitation (ChIP) (Arrieta-Ortiz *et al.*, 2015; Boscolo et al., 2005; Boulesteix and Strimmer, 2005; Chen *et al.*, 2017; Liao *et al.*, 2003; Nachman *et al.*, 2004; Ocone and Sanguinetti, 2011; Rogers et al., 2007; Sanguinetti *et al.*, 2006; Schacht *et al.*, 2014; Shi *et al.*, 2009; Tchourine *et al.*, 2018; Wang *et al.*, 2008; Yang *et al.*, 2005; Yu and Li, 2005). Garcia-Alonso *et al.* reported that a manually curated network map performed best for TFA inference in human (Garcia-Alonso et al., 2019), but curated networks include very few TFs and are not available for most organisms. Here, we analyze the effects on TFA inference accuracy when networks of comparable size are constructed from various high-throughput data sources.

In most previous studies, inferred TFA values were allowed to be positive or negative and their absolute value was interpreted as the magnitude of activity change relative to some reference sample. Thus, a smaller absolute TFA did not indicate less activity, but rather less change relative to the reference. As a result, the TFA levels did not distinguish between increasing and decreasing activity. Furthermore, the signs of the CS values had no meaning. Here, we propose, evaluate and optimize a version of the bilinear approach in which TFA values are constrained to be non-negative, so that zero represents no activity, equivalent to deletion of the gene encoding the TF. We also include parameters representing the expression of each gene when all its regulators have activity zero (*baselines*). The combination of baseline expression levels with the non-negativity constraint on TFA values differentiates our model from previously proposed models. Together, they make the parameters interpretable. Positive control strength indicates that the TF activates the target and negative control strength indicates that it represses the target. If a TF's activity is larger in one sample than in another, then the TF is more active in the former sample than in the latter. We make extensive use of gene expression data after direct perturbations of TF activities (Arrieta-Ortiz *et al.*, 2015; Tchourine *et al.*, 2018; Tran et al., 2005), constraining each control strength parameter to be positive (activating) or negative (repressing) based on the direction in which the target gene's mRNA level changes when the TF is perturbed. If the gene encoding a TF is deleted in a sample, the TF's activity is fixed at zero; if it is overexpressed, the TF's activity is constrained to be greater than its activity in unperturbed samples.

Evaluating the effects of various mathematical models, network mapping procedures and perturbation-derived constraints, requires accuracy metrics that are objective, quantitative and available for large numbers of TFs. This poses a challenge because TFA cannot be directly measured. As a result, most attempts to validate TFA inference algorithms have been small-scale and often qualitative, highlighting successes with just a few TFs. Some validation efforts have been based on inferring significant differential activity in a handful of samples subjected to stressors (Boorsma *et al.*, 2008) or small molecules known to affect a particular TF's activity (Azofeifa et al., 2018; Barenco *et al.*, 2006; Ocone and Sanguinetti, 2011). Others have been based on inferring activity patterns that appear to match the periodicity of cell cycles (Liao *et al.*, 2003; Nachman *et al.*, 2004; Sanguinetti *et al.*, 2006) or using changes in the nuclear localization of a GFP-tagged TF as a proxy for TFA (Boorsma *et al.*, 2008). Other evaluation efforts have been based on internal consistency measures (Berchtold et al., 2016), TF activity perturbations (Boorsma *et al.*, 2008; Garcia-Alonso *et al.*, 2019; Trescher and Leser, 2019) or identification of TFs important for proliferation of cancer cells (Alvarez *et al.*, 2016; Azofeifa *et al.*, 2018; Balwierz *et al.*, 2014; Barenco *et al.*, 2006; Chen *et al.*, 2013; Chen *et al.*, 2017; Cheng *et al.*, 2007; Fröhlich, 2015; Garcia-Alonso *et al.*, 2018; Jiang *et al.*, 2015; Li *et al.*, 2014; Ocone and Sanguinetti, 2011; Trescher and Leser, 2019; Tripodi et al., 2018; Zhu *et al.*, 2013). In this work, we take advantage of two independent, high-quality perturbation datasets in *Saccharomyces cerevisiae* to present multiple quantitative, large-scale validation metrics. By using one dataset for network construction and constraint generation, and the other for validation, we reveal which high-throughput data types are most valuable for TFA inference in the matrix factorization framework, identify best practices for achieving high accuracy, and show that given the right input data, TFA inference works reasonably well.

2 Results

2.1 Mathematical model of gene expression

We use a simple model in which the log expression level of a gene in a given sample is determined by its baseline expression level, when none of its regulators are active, plus the sum of the influences of all the TFs that regulate it. The influence of each TF is a product of the strength with which the TF regulates that gene (*control strength*) and the TF's activity in that sample:

$$\text{expression}_{i,\,k} = \text{baseline}_i + \sum\nolimits_{j \in \text{TFs}} (\text{controlStrength}_{i,\,j} \times \text{activity}_{j,\,k})$$

where $\text{expression}_{i,\,k}$ is the log expression level of gene i in sample k, baseline_i is the expression level of gene i absent any influence from TFs, $\text{controlStrength}_{i,\,j}$ is the condition-independent potential of TF j to activate or repress gene i, and $\text{activity}_{j,\,k}$ is the activity level of TF j in sample k. In matrix notation, $\mathbf{E} = \mathbf{CS} \bullet \mathbf{TFA}$, where $\mathbf{E}$ is a gene expression matrix (genes by samples), $\mathbf{CS}$ is a matrix of control strengths (genes by TFs) augmented to incorporate baselines, $\mathbf{TFA}$ is a matrix of TF activity levels (TFs by samples) and $\bullet$ indicates matrix multiplication (Fig. 1). Fitting the CS and TFA matrices to expression data is equivalent to factoring the expression matrix, under the constraints that CS signs are predetermined, TFA is non-negative and the activities of perturbed TFs are constrained according to the perturbation. The model is fit by minimizing the sum of squared errors based on this equation. This can be done by alternating linear regression (Li *et al.*, 2014; Liao *et al.*, 2003; Sanguinetti *et al.*, 2006; Tran *et al.*, 2005; Yu and Li, 2005) or with a more general non-linear optimization algorithm (Byrd *et al.*, 2006). See Section 3 for details.

2.2 Core evaluation metrics

Our approach to TFA evaluation relies on two independent expression datasets in which the activity of each TF has been

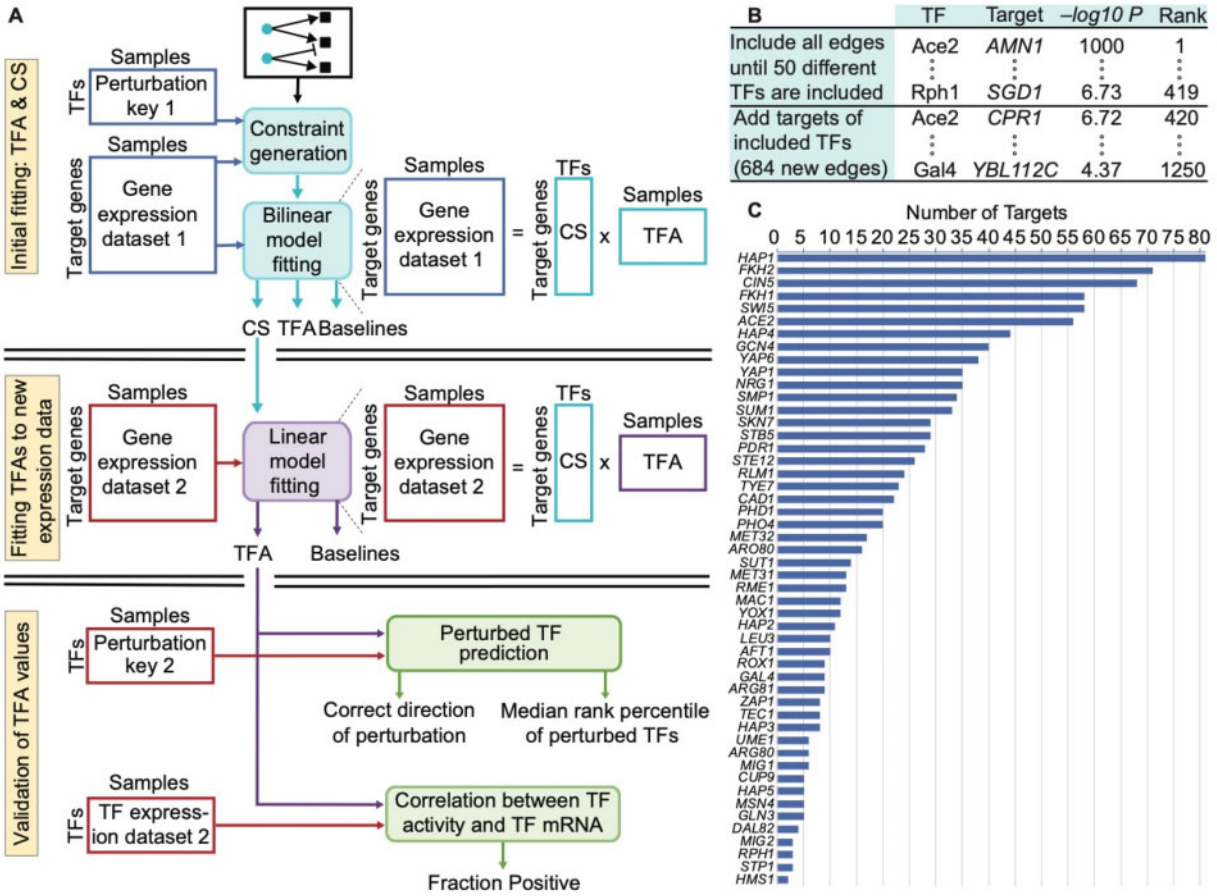

Fig. 1. Evaluation framework and ChIP-based network construction. (**A**) Overview of three-stage model fitting and TFA evaluation procedure. Gene expression levels and the perturbation key from dataset 1 are used only in the initial fitting. The CSs inferred in the initial fitting are fixed while the TFAs and baselines are refit to the target gene expression levels from dataset 2. The mRNA levels of the TFs and the perturbation key from dataset 2 are used only for evaluation. (**B**) Illustration of how edges were selected for the ChIP-based network. All edges were ranked according to their -log *P*-value for the TF binding in the promoter of the target. Edges were selected in rank order until there was at least one edge from 50 different TFs. Lower-ranked edges were then selected for those TFs until rank 1,250. After initial model construction, we removed any TFs with a single target and any set of TFs with identical targets, along with those targets. We then returned to the list and iteratively added edges that had previously been passed over until the network stabilized at 50 TFs. This yielded a network with 1,104 edges. (**C**) The number of targets for each of the 50 different TFs in the ChIP network

individually perturbed. The first dataset consists of expression profiles of strains in which a single TF was deleted from the genome (*TFKO* data) (Kemmeren et al., 2014). The second consists of expression profiles collected 15 min after overexpression of a single TF was induced using the ZEV system (*ZEV* data) (Hackett et al., 2020). These two datasets represent two very different growth conditions: the TFKO data profiles cells growing on synthetic complete medium in shake-flasks with no nutrient limitation, while the ZEV data profiles cells growing on minimal medium in phosphate-limited, continuous-flow chemostats. Two of our core evaluation metrics focus on whether TFA inference can determine which TF was perturbed in each sample and whether it was knocked out or overexpressed. We refer to this information as the *perturbation key*. First, we use both the perturbation key and the expression profiles from one dataset for network construction, constraint generation and fitting (Fig. 1A, top). Next, we hold the control strengths from the initial fit fixed and refit the TFA and baseline parameters to the expression profiles in a second perturbation dataset (Fig. 1A, middle). Crucially, the perturbation key for the second dataset is not used for either fit, so it can be used as independent evaluation data (Fig. 1A, bottom).

We use three core evaluation metrics for TFA accuracy. The *direction of perturbation* metric is the fraction of samples in which the direction of perturbation (deletion or overexpression) is inferred correctly, given the identity of the perturbed TF. For the second metric, all TFs in a perturbation sample are ranked, starting from the one whose inferred activity increased most, relative to an unperturbed sample (for overexpression) or decreased most (for deletion). We expect the perturbed TF to be highly ranked. The median rank percentile, across all perturbation samples, is the *median rank percentile* metric. The *positive correlation* metric is the fraction of TFs with positive correlations between their activity levels and their mRNA levels. To ensure that TFA-mRNA correlations were not driven by a few outliers, we constructed 1000 bootstrap samples of the gene expression profiles. For each bootstrap sample, we calculated the correlation of each TF's inferred TFA and its measured expression level. We then calculated the fraction of TFs with positive correlations in that bootstrap sample. The median fraction of positive correlations, across all bootstrap samples, is our third metric. (An analysis of correlation significance based on a model rather than bootstrapping yields the same conclusions; Supplementary Fig. S5.) We do not expect accurate TFA values to yield 100% for this metric, since post-transcriptional regulation is a key determinant of TFA levels, but we do expect the fraction of positive correlations to be substantially greater than half.

The metrics reported below are averages after training on each dataset, testing on the other and then switching the roles of the two datasets. *P*-values from these two train-test directions are combined by using Fisher's combined probability test (Fisher, 1954). The first fitting only used samples in which one of the network TFs was directly perturbed, while the second fitting also used samples in which other TFs were perturbed. For all analyses, we considered only the

179 TFs that were perturbed in both ZEV and TFKO datasets. For *median rank percentile*, inferred activity levels for each TF were log-transformed and then standardized across samples, converting them to *Z*-scores. This makes the activity levels of different TFs comparable.

2.3 Constructing a ChIP-based network

Suppose binding location data are available for many of the TFs in a given organism, along with gene expression profiles from many samples. For now, assume that the expression dataset does not contain direct TF perturbations or, if it does, that the perturbation key is not used. To generate a network map as input for TFA inference, all possible TF-target edges can be ranked according to the strength of evidence that the TF binds in the promoter of the gene. We did this for yeast, ranking edges according to their negative log *P*-value in a comprehensive ChIP-chip dataset (Harbison et al., 2004). This produced a single, global ranking of all edges involving all TFs (Fig. 1B). We then constructed a network map as described in Figure 1B. All networks are described further in Supplementary Methods and are provided as Supplementary Files.

2.4 Evaluating TFAs inferred from the ChIP network with correlation-based constraints

First, we evaluated the ChIP network without any constraints on the parameters (except non-negative TFA) and found that it did not perform better than chance on any metric (Supplementary Fig. S1). Next, we tried adding constraints on the signs of the control strength parameters, based on the intuition that a positive correlation between the mRNA levels of a TF and a target suggests activation while a negative correlation suggests repression. We constrained the sign of each control strength to match the sign of that correlation (even if the correlation was not significant) using the dataset reserved for the initial fitting, which improved performance. We refer to the ChIP-based network with correlation-based constraints as ChIP-CC (Supplementary Files S1–S4). Using ChIP-CC, the direction of activity change between a TF's perturbation sample and the unperturbed sample was predicted correctly for 66% of TFs ($P < 0.01$, binomial test; Fig. 2A, left), the median rank percentile was 76.5% ($P < 0.0001$, binomial test; Fig. 2A, middle) and 65% of TFs' activity levels were positively correlated with their mRNA levels (Fig. 2A, right; $P < 0.01$; see Materials and Methods for details).

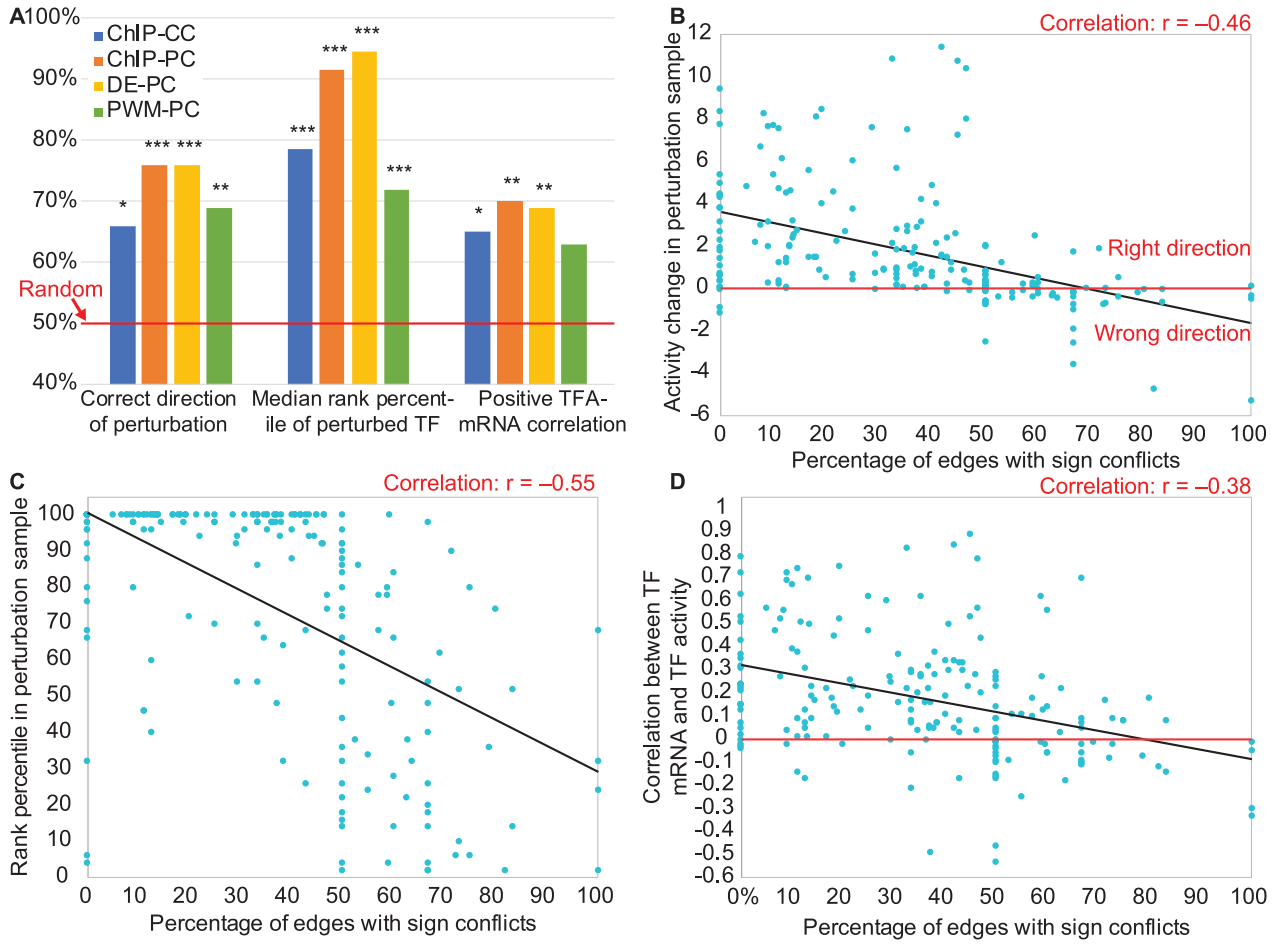

Fig. 2. Determinants of TFA inference accuracy. (**A**) Effects of network construction and constraint generation on TFA accuracy. Blue: ChIP network with correlation-based constraints. Orange: ChIP network with perturbation-based constraints. Yellow: Differential expression network with perturbation-based constraints. Green: Binding-specificity (PWM) network with perturbation-based constraints. Asterisks above the bars indicate magnitude of significance compared to a random model, with 1, 2 or 3 asterisks representing *P*-value thresholds of 0.01, 0.001 or 0.0001. (**B**) Vertical axis: The activity of each TF in the sample in which it was perturbed minus its activity in the unperturbed sample, oriented so that higher is better. TFs plotted below the horizontal axis have been inferred to change activity in the wrong direction. Horizontal axis: The fraction of each TF's targets for which the TFKO and ZEV datasets suggest conflicting CS signs. TFs with <50% conflict edges are almost all predicted in the correct direction, while most TFs with >50% conflict edges are not. (**C**) Vertical axis: Rank percentile of the perturbed TF's activity change in each perturbation sample (higher is better). Horizontal axis: Same as (B). TFs with a higher percentage of conflict edges tend to be ranked lower. (**D**) Vertical axis: median fraction of bootstrap samples in which a TF's mRNA level and its inferred activity level are positively correlated (see main text). TFs with a higher percentage of conflicting edges tend to have low or negative correlation. (B–D) Results from the 50-TF ChIP-PC and DE-PC networks, trained on each of the datasets and tested on the other, have been combined, but each individual set of 50 points showed similar, highly significant correlations

2.5 Adding constraints obtained from TF perturbation data

When expression data after direct TF perturbation are available, the perturbation key can be used to constrain both the CS signs and the activity of the TF perturbed in each sample, as described above. We evaluated the effect of using the perturbation key from the first dataset to constrain the control strengths and activity levels during the first fit (Fig. 1A, top). The perturbation key for the second dataset is only used for evaluation (Fig. 1A, bottom), not during the second fit (Fig. 1A, middle). We call the ChIP network with perturbation-constraints ChIP-PC (Supplementary Files S7–S10). Performance on all three metrics increased, relative to using correlation-based constraints (Fig. 2A, blue and orange bars).

2.6 Generating network and constraints from TF perturbation data, without binding data

If expression data from direct TF perturbations is available, it is possible to build a network from the perturbation data rather than ChIP data. The same network building procedure is used, but instead of TF-target interactions being ranked by the strength of ChIP evidence, they are ranked by the absolute value of the log fold change of the target when the TF is perturbed (DE-PC, Supplementary Files S13–S16). The performance of this network with perturbation-based constraints was similar to that of ChIP-PC, with the biggest change being an increase in median rank percentile from 92% to 96% (Fig. 2A, orange and yellow bars). We conclude that differential expression data from direct TF perturbations are necessary and sufficient for accurate TFA inference performance—binding location data are not necessary.

2.7 Using binding specificity models in network generation

A popular source of data for building gene regulatory networks is models of TF binding specificity, typically represented as position weight matrices (PWMs) (Boorsma *et al.*, 2008; Boscolo *et al.*, 2005; Bussemaker *et al.*, 2001; Cheng *et al.*, 2007; Garcia-Alonso *et al.*, 2018; Lee and Bussemaker, 2010). To test this approach, we ranked all possible TF-target interactions by the maximum, across all positions in the target gene's promoter, of the negative log *P*-value for presence of the TF's motif, as defined in the ScerTF database (Spivak and Stormo, 2012) and scored by FIMO (Grant et al., 2011) (see Supplementary Material). We then built a network from this ranking just as we did with the ChIP-chip data and the differential expression data (Fig. 1B). Using this network, we optimized with perturbation-based constraints (PWM-PC, Supplementary Files S19–S22). This network performed significantly worse than both ChIP-PC and DE-PC networks (Fig. 2A, green bars).

2.8 Increasing the number of TFs

To infer activity for more than 50 TFs, the input network maps can be extended by considering more than just the 1,250 top ranked edges. To quantify the loss in performance from using lower ranked edges, we used blocks of 2,000 edges of decreasing rank to make independent networks of 50 TFs (Supplementary Methods). Both ChIP-PC and DE-PC lost accuracy steadily as lower ranked edges were used. DE-PC lost accuracy more slowly in the first two metrics (Fig. 3A and B) while both networks lost ground at the same rate for TFA-mRNA correlation (Fig. 3C). Thus, DE-PC appears to be the better choice for building larger networks with more TFs.

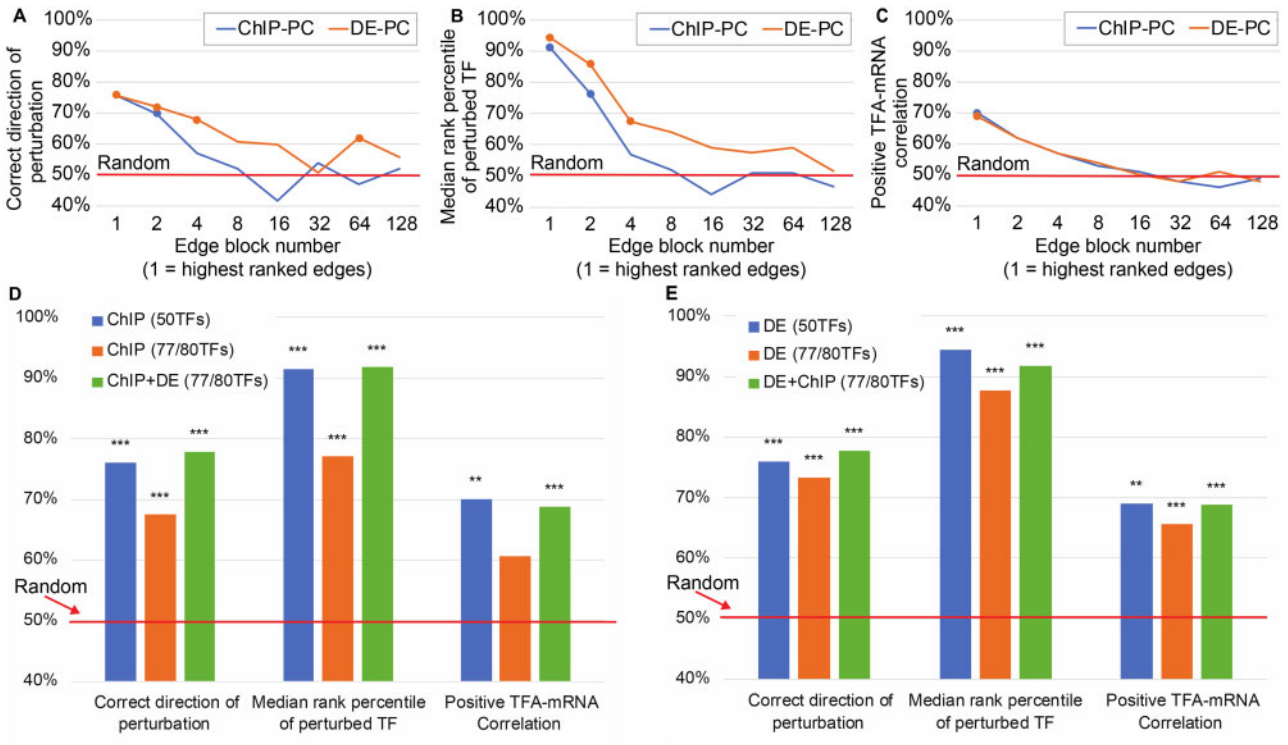

Fig. 3. Effects of increasing the number of network TFs on accuracy. (**A–C**) Accuracy metrics for networks constructed from the ChIP or DE edge lists by taking successively lower ranked edges. Edges were divided into blocks of 2,000 and blocks are plotted in an exponential series. For example, Block 1 is edges ranked 1–2,000 and Block 4 is edges ranked 6,001–8,000. Points are plotted for results that are significantly better than random ($P < 0.001$). (A) Percent of TFs whose direction of perturbation is predicted correctly. (B) Median rank percentile of the perturbed TF. (C) Percent of TFs with a positive TF-mRNA correlation. In A and B, the ChIP-PC performance starts out similar to DE-PC, but it drops faster, to no better than random in any measure by Block 4. (**D**) Comparison of two ways of increasing the number of TFs in the network—going further down the list of ChIP edges or using 50-TF ChIP and DE networks and averaging standardized TFAs of TFs that are in both networks. Consistent with A–C, performance degrades when lower ranked edges are included in the ChIP network. Inferring TFAs separately and averaging them, by contrast, yields performance on a larger network that is as good as performance on the smaller, 50-TF networks. (E) Same as D, but blue and orange bars are for DE networks

If both ChIP and TF perturbation data are available, another option is to infer activities separately for 50-TF ChIP-PC and DE-PC networks and combine the results, averaging the standardized activities for TFs that are in both networks. This provides activities for a larger number of TFs with accuracy that is better than including lower-ranked edges from a single data source (Fig. 3D and E, orange versus green bars).

2.9 Effect of optimizing control strengths on TFA accuracy

To determine how much optimizing control strengths contributes to the accuracy of inferred TFAs, we compared TFAs obtained by optimization of both TFA and CS matrices to TFAs obtained by using fixed control strengths of +1 for activation or -1 for repression (*signed binary* CSs). Signed binary CSs were also used in (Arrieta-Ortiz *et al.*, 2015; Chen *et al.*, 2017; Gitter *et al.*, 2013; Tchourine *et al.*, 2018). The optimized CSs performed slightly better on some metrics and some networks, but there was little difference overall (Fig. 4A and B). To investigate the potential value of CS optimization further, we considered two additional metrics.

Known regulators of TF activity The TFKO dataset contains 1,484 samples from strains in which a gene was deleted, including many known regulators of TF activity. Our next evaluation task was to determine the target TFs in samples where a known regulator of TF activity is perturbed. To evaluate performance on this task, we compiled a map of known TF activity regulators and their targets (Supplementary File S31). For each TF, an activity regulator was assigned to it if there was published literature that proposed a direct interaction that affects TF activity by a specific mechanism, such as phosphorylation, nuclear localization or complex formation. Network maps, constraints and initial fitting used the ZEV dataset. The resulting CS matrix was then used to fit TFAs and baselines to the 194 TFKO samples in which a known TFA regulator was perturbed. In each perturbation sample, all TFs were ranked by the absolute difference between their standardized log activities in the perturbed and unperturbed samples. The absolute value was used because the literature is not always clear on the direction of regulation. We then plotted the fraction of literature-supported targets that were ranked above a given percentile in the sample in which their TFA regulator was perturbed (Fig. 4C). The optimized CS matrices (solid lines) identify more known TFA regulators than the signed binary matrices (dashed lines), especially at rank percentiles above 85%. The ChIP-PC network (blue) also outperforms the DE-PC (orange) by this metric.

ZEV time course data Although we have focused on the ZEV data from 15 min after TF induction, they are part of time courses with samples taken 2.5, 5, 10, 15, 20, 30, 45, 60 and 90 min after induction (some timepoints are not available for some TFs). The expression data show that each TF is rapidly induced to a very high

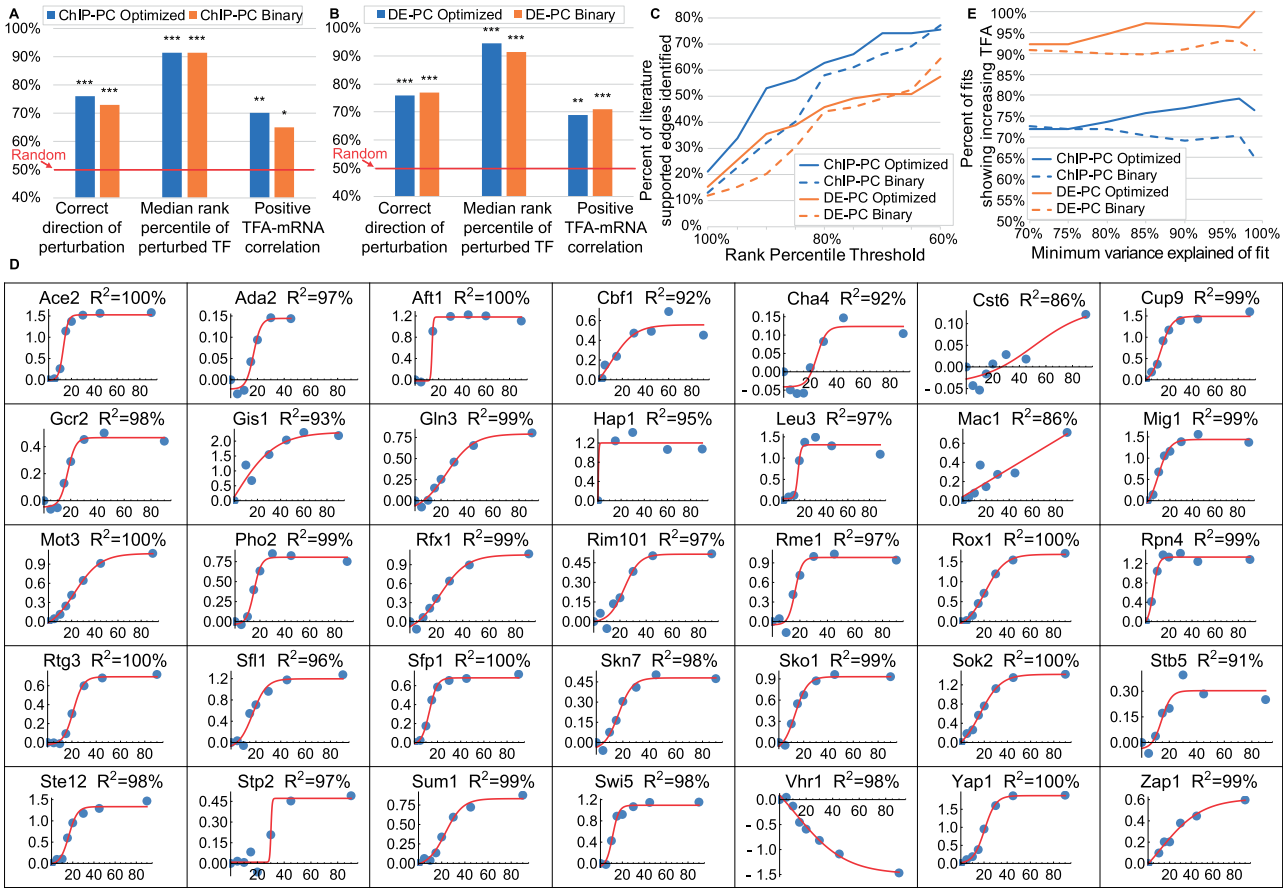

Fig. 4. Impact of using a CS matrix optimized on a different dataset versus using a signed binary CS matrix. (**A**) ChIP-PC network. (**B**) DE-PC network. (**C**) Percent of literature-supported edges between TFA regulators and TFs identified, as a function of minimum rank percentile for identification. Solid lines: CS matrix optimized on the ZEV dataset and used to infer TFAs in the samples in which a TF regulator was deleted. Dashed lines: Signed binary CS matrix. For TFs whose change in standardized log activity from WT ranks above 85th percentile, more literature supported edges are identified by using optimized CS matrices than by using signed binary matrices. (**D**) Sigmoidal fits to log2 fold change of TFAs inferred for the ZEV time course data, using the DE-PC network and a CS matrix optimized on the TFKO dataset, relative to the 0 min timepoint. Only fits with variance explained above 85% are shown. In all but one of the 35 fits, TF activity is correctly inferred to be increasing (97%). Only Vhr1 activity is inferred to change in the wrong direction, probably because 9 of its 11 targets have sign conflict (80%, see Fig. 2B–D). (**E**) After fitting sigmoidal curves as in D and imposing various thresholds on the variance explained by the fit, the percentage of fits that correctly show increasing activity. The DE-PC network (orange lines) performs better than the ChIP-PC network (blue lines). For each network, using a CS matrix optimized on the TFKO data (solid lines) generally shows better performance than using a signed binary CS matrix (dashed lines), and this effect increases as the variance explained by the sigmoidal fits increases

level and remains highly expressed throughout the 90 min. Therefore, we decided to infer TFAs for the entire time course, using a CS matrix derived from the TFKO data. For each time point, we computed a log fold change of the induced TF's inferred activity, relative to its activity at time 0. We then fit a 4-parameter, sigmoidal, saturating curve to the time series (Section 3; Fig. 4D). To eliminate curves that fit poorly, we tried several different thresholds on the variance explained by the sigmoidal fit. For each threshold, we calculated the fraction of fits that showed increasing activity throughout the time series (expected behavior) rather than decreasing activity. Overall, the TFAs with optimized CS matrices (solid lines) performed better than those with signed binary CS matrices (dashed lines) and this effect got stronger for curves with better fits (Fig. 4E). Furthermore, the DE-PC network (orange lines) performed substantially better than the ChIP-PC network (blue lines).

2.10 Condition-independence of control strengths

Control strengths are intended to be condition-independent, quantitative measures of each TF's potential to regulate each target gene. We have seen that good performance on TFA inference tasks can be obtained by using control strengths optimized on a different dataset (Figs 1A, 2A, 3 and 4). Another indication that inferred CSs are transferable between datasets is that using the CS values inferred from one dataset for TFA inference in a second dataset increases the variance explained by 5%, relative to using the signed binary CSs (Supplementary Fig. S2).

Another way to test the condition-independence of control strengths is to calculate the correlation between CSs inferred from two datasets collected in different growth conditions. To maximize the number of TFs and their target genes we could use, we first created a large network from the union of edges from ChIP-PC and the two DE-PC networks, one derived from the TFKO data and the other from the ZEV data. After filtering out edges with conflicting sign constraints (see Section 3) and dropping two TFs that were left with only a single target, this new network (Union-PC, Supplementary File S25) contains 94 TFs, 1,416 target genes and 2,731 edges. We optimized both TFA and CS matrices using both the TFKO and the ZEV datasets together (see Section 3). As an initial validation of the resulting CS matrix, we used it to infer TFAs in a new dataset consisting of 69 double-deletion strains (Sameith et al., 2015) (See Supplementary Material). The results for predicting the direction of perturbation (86.4% correct) and identifying the perturbed TF (rank percentile 95.7%) were even better than the results for smaller networks (compare to Fig. 2A). The percentage of TFs that showed positive correlation between their mRNA levels and their inferred TFAs, 66%, is slightly below the results for the smaller networks, but still much better than chance. Readers can use the optimized CS matrix (Supplementary File S26) for TFA inference in other datasets.

To calculate the correlation of CS values inferred in two growth conditions, we re-optimized Union-PC using only the TFKO data (synthetic complete medium with 2% glucose and no nutrient limitation) and then only the ZEV data (minimal medium with 2% glucose in phosphate-limited chemostats). Focusing on the 76 TFs with at least 5 targets, we constructed 1000 bootstrap samples of target genes for each TF and for each sample, calculated the correlation between target genes' CS inferred from one dataset and their CS inferred from the other dataset. We then scored each TF by its median correlation across bootstrap samples. The median score (across TFs) was +0.32 and 61 of 76 TFs had positive scores (80%). (Supplementary Fig. S6 shows these correlations and their significance thresholds calculated with a traditional model-based approach.) This shows that, while there may be some over-fitting of CSs to a particular growth condition, there is also a substantial amount of condition independence.

2.11 Evaluating control strengths directly

As another evaluation, we asked whether the control strengths inferred for the targets of a TF would correspond in any way to the strength with which the TF binds to the promoters of those targets in genome-wide binding location data. To do this, we turned to binding data obtained by the transposon calling cards method (Mayhew and Mitra, 2016; Wang et al., 2011). In this method, a TF is linked to a transposase, which deposits a transposon in the genome near where the TF is bound. The number of transposons in a gene's promoter is an approximate measure of the amount of time the TF spends bound to that promoter. We predicted that promoter occupancy would be positively correlated with the inferred control strength, in most cases. Importantly, we considered only the genes that were targets of a TF in the input network and therefore had inferred control strengths. The input network itself does not contain quantitative binding strength information. We optimized both TFA and CS using the Union-PC network and both the TFKO and ZEV data together. Using 1,000 bootstrap samplings of target genes for inferring TFAs, the median correlation between inferred CS and measured transposons was calculated for each TF. Of the 11 TFs with calling cards data and at least 5 target genes, 9 (82%) have positive median correlation between measured binding events and inferred CS values, with a median correlation of 0.31. A histogram of correlations and their model-based *P*-values are shown in Supplementary Figure S7.

2.12 Analyzing time courses after glucose influx

Next, we used the CS matrix for Union-PC (Supplementary File S26) to infer baselines and TFAs for three expression time courses after yeast cells were provided with glucose. In the first dataset, yeast cells growing in galactose-limited chemostats were provided glucose to a final concentration of 0.02% (w/v) or 0.2% (Ronen and Botstein, 2006). In the second, batch cultures depleted glucose over a 24 h growth period before being transferred to fresh media with 2% glucose (Apweiler et al., 2012). We plotted the inferred activity of each TF as a function of time, fit both a 4-parameter sigmoid curve (as in Fig. 4) and a 6-parameter impulse curve (Chechik and Koller, 2009), chose one of the two by the Bayes Information Criterion (Schwarz, 1978) and filtered out poor fits (R^2<80%) (see Supplementary Material). For four well-studied TFs, Gcr2, Gln3, Gcn4 and Msn2, the inferred TFA levels and fits for all three time courses are shown as points and lines inside turquoise circles in Figure 5. The shapes of the activity curves in response to glucose made sense: Gcr2, an activator of glycolytic genes, increased in activity upon glucose addition; Gln3, Gcn4 and Msn2, activators of genes needed during nutrient deprivation, generally decreased activity upon glucose addition (Gln3 showed a very small increase in one time course). Gcr2 activity also makes sense in terms of the glucose concentrations added, returning to baseline quickly at the lowest concentration, more slowly at the intermediate concentration (note the last gold data point, which is not reflected in the fitted curve), and not at all at the highest concentration.

To gain a deeper understanding of the network structure, the control strengths, and how they led to the TFA patterns shown in Figure 5, we carried out enrichment analysis on the network targets of the four TFs using Gene Ontology biological process annotations and Kyoto Encyclopedia of Genes and Genomes metabolic pathway annotations (Liao et al., 2019). We first discarded targets that, after optimization, had absolute CS$\leq 10^{-4}$, then analyzed the remaining activated and repressed targets of each TF separately, and removed redundant terms (Bodenhofer et al., 2011). All significantly enriched annotations that apply to four or more target genes are listed above the black squares in Figure 5. Each TF is connected to a target square by an arrow if the corresponding annotation was enriched among its activated targets and a T-head if the annotation was enriched among its repressed targets. Gcn4 and Msn2 had no repressed targets and the few repressed targets of Gcr2 had no enriched annotations that applied to four or more targets. The annotations and directions of regulation of target sets made sense in terms of the known function of each TF (Broach, 2012; Conrad et al., 2014; De Virgilio, 2012; Ljungdahl and Daignan-Fornier, 2012; Rodkaer and Faergeman, 2014).

For each gene set in Figure 5, we calculated the median log fold change in each time course (not shown) and fit sigmoidal or impulse curves to them (shown in black boxes). In general, the shapes of the

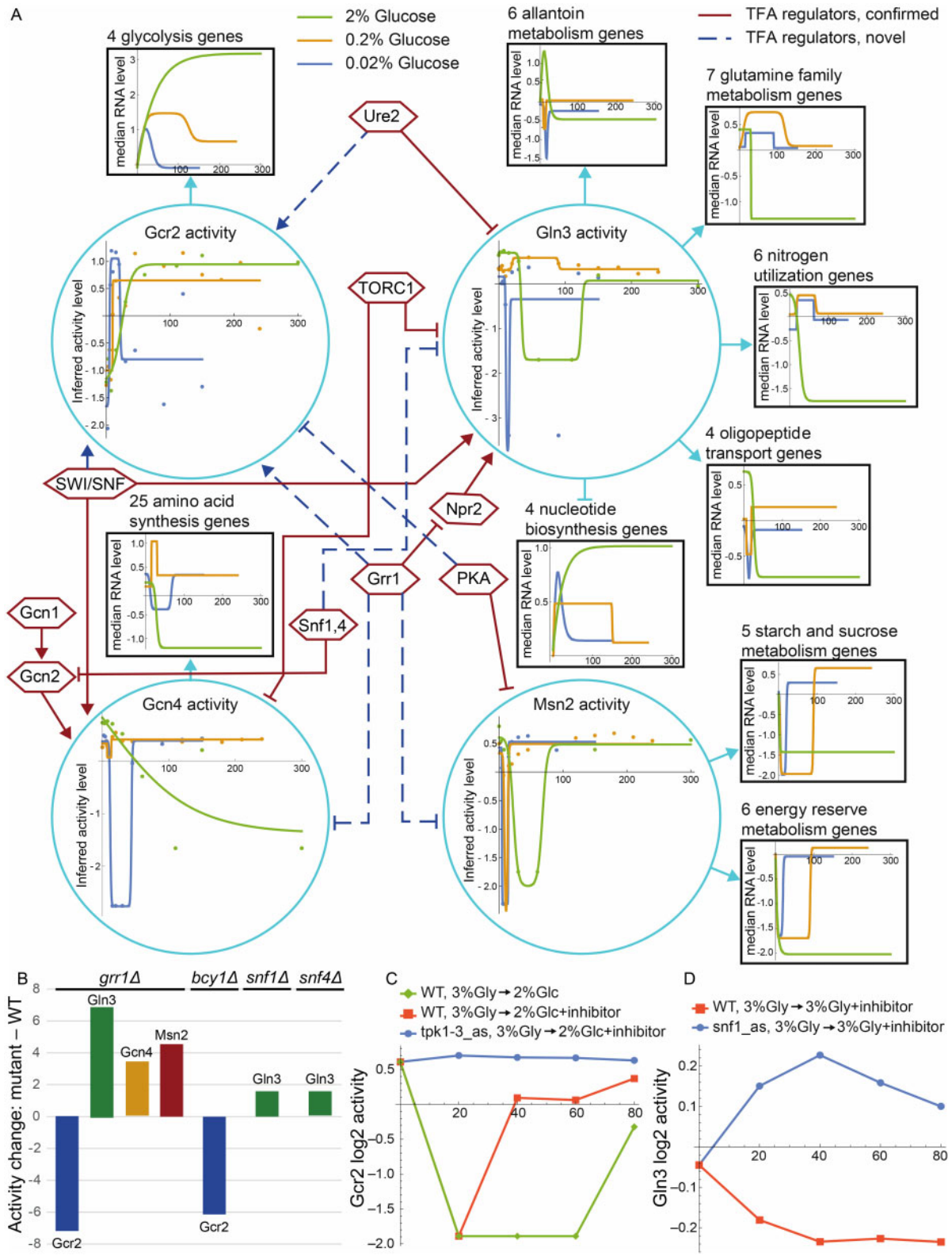

Fig. 5. Gcr2, Gln3, Gcn4 and Msn2, their activity regulators, activity changes in response to different glucose concentrations, target gene sets and target set expression patterns. (**A**) Turquoise circles: Changes in inferred TF activity after addition of 2% glucose to post-diauxic-shift shake-flasks with synthetic complete medium (green) or addition of 0.2% (gold) or 0.02% (blue) glucose to cultures grown in galactose-limited chemostats with minimal medium. Points are log2 of inferred activity level and lines are impulse or sigmoidal fits to the points, chosen by the Bayes Information Criterion. Black boxes: Sets of target genes that are regulated in the same direction and are annotated to a Gene Ontology or KEGG term enriched among targets of the TF that regulates them. Arrowheads indicate activation and T-heads repression. Colored lines are impulse or sigmoidal fits to the median log2 fold-change of the annotated genes at each time point, relative to time 0. Hexagons: TF activity regulators inferred from analysis of two datasets as described in the text. Solid maroon lines indicate clear literature support while dashed blue lines indicate hypothesized novel edges. (**B**) Change in activity of TFs in response to deletion of *GRR1*, *BCY1* (inhibitory subunit of PKA), *SNF1* or *SNF4* (activating subunit of Snf1 complex). (**C**) Gcr2 activity after addition of 2% glucose to cells growing on 3% glycerol. In wild-type cells, glucose initially reduces Gcr2 activity (green, orange). (This response is different from Gcr2's response to glucose under the conditions of Fig. 5A.) Addition of the Tpk1-3 inhibitor with glucose to analog-sensitive cells (blue) eliminates that response, suggesting that PKA represses Gcr2 activity. This is consistent with the observation that deletion of *BCY1* reduces Gcr2 activity in 2% glucose (B). (**D**) Gln3 activity is slightly elevated when inhibitor is added to cells growing in 3% glycerol and expressing an analog-sensitive Snf1 (blue), relative to WT cells (orange), suggesting that the Snf1,4 complex represses Gln3 activity. This is consistent with the observation that deletion of either Snf1 or Snf4 increases Gln3 activity in 2% glucose (B)

inferred activity curves for each TF made sense in light of the expression patterns of their target groups. For example, the inferred activity patterns of Gcr2 mirror those of the glycolysis genes it activates. Note that curves are expected to be inverted for repressed targets and that some of the lines occlude others at early time points, hiding early dips. Although the mRNA levels of target genes show a clear relationship to inferred activity levels, they do not always mirror each other perfectly. For example, the amino acid synthesis genes spike upward after addition of 0.2% glucose, whereas the inferred activity of Gcn4 dips and returns, presumably responding to influential target genes whose individual expression levels do not follow the pattern of the median levels shown.

Finally, we attempted to identify key regulators driving the TFA patterns observed in Figure 5, by inferring TFAs from two more datasets, using the same CS matrix. The first consists of expression profiles of strains deleted for 567 different regulatory factors, including many known TF activity regulators (Kemmeren *et al.*, 2014). The second includes time courses after addition of rapamycin (an inhibitor of TORC1) or inhibitors of Snf1, Tpk1-3 or Sch9 (Zaman et al., 2009). The inferred TFAs for these datasets are provided as Supplementary Files S29 and S30. By examining the effects of these perturbations on the inferred activities of Gcr2, Gln3, Gcn4 and Msn2 and comparing them to published literature, we were able to confirm many previously described regulatory interactions (Fig. 5A, solid maroon lines) and to identify a few potentially new ones (Fig. 5A, dashed blue lines). The TF regulatory edges in Figure 5A were derived by manual curation of the changes in inferred TF activity when TF activity regulators are perturbed in these datasets. The evidence supporting most of the hypothesized novel TF regulatory interactions is shown in Figure 5B and C, but space limitations precluded showing the evidence for the SWI/SNF or Ure2 activating Gcr2. A description of the TFA regulatory interactions shown in Figure 5 and of the literature supporting those that were previously known can be found in the online Supplementary Material. To summarize key highlights, we have identified several likely regulators of Gcr2 activity, about which little was previously known, and discovered that Grr1, previously known for its role in glucose repression, is probably a positive regulator of glycolysis (via Gcr2) and a negative regulator of stress-induced TFs. These novel observations, which were mined from gene expression data via TF activity inference, constitute a rich trove of hypotheses for future experimental investigation. The repression of Gcn4 and Msn2 activity by Grr1 is lent plausibility by the fact that the mRNA levels of *GCN4 and MSN2* increase significantly when Grr1 is deleted (the *GCR2* mRNA level does not change). One possible mechanism for regulation of Gcn4, Gcr2 and Msn2 by Grr1, a change in the nuclear concentration of the TF, could be tested by perturbing Grr1 in strains in which one of these TFs is linked to a fluorescent protein and comparing images of the perturbed and unperturbed samples. A complementary approach would be to carry out calling cards experiments on each TF in unperturbed cells and in cells in which Grr1 has been perturbed. This could detect changes in promoter occupancy by the putatively regulated TFs.

3 Materials and methods

Model fitting In initial fits (Fig. 1A, top), models are fit to gene expression data by least-squares linear regression, alternating between TFA and CS matrices (which include baselines), starting from 20 random initializations of the CS matrix. If a TF does not regulate a gene in the TF network map, the corresponding CS is held at zero; otherwise, it is constrained to be either positive (activating) or negative (repressing). If a TF is deleted in a sample, its activity is held at zero; if it is overexpressed, its activity is constrained to be greater than its activity in unperturbed samples. Except for deletion samples, all activities are constrained to be ≥ 0.0001. When learning a CS matrix, the mean activity of each TF, across all samples, is constrained to be one, since scaling a TF's activities and control strengths by inverse factors does not affect the predictions. After each iteration, the non-baseline control strengths are held constant while the TFAs and baselines are fit to the second dataset by alternating linear regression without constraining the mean activity of a TF (Fig. 1A, middle). Optimization against the first dataset is halted when R^2 in the second dataset peaks or after 100 iterations, whichever comes first. This halting criterion does not use the perturbation key of the second dataset, which determines the gold standard for evaluation. It does not use the expression levels of the TFs, either, as they are removed from the input expression profiles. Supplementary Figures S3 and S4 replicate the results of Figure 2A without using the expression profiles from the second dataset to determine the stopping criteria. Comparable results were obtained by using Knitro (Byrd et al., 2006), a general non-linear solver, to optimize all parameters at once. Code for model fitting is available at https://doi.org/10.5281/zenodo.4050573.

Ranking TFs To calculate the rank percentile of a perturbed TF, the log2 of the activity values for each TF are standardized across all samples. In TF knockout samples, the standardized log2 activity values of each TF are sorted from lowest (most negative) to highest (most positive); in TF overexpression samples, they are sorted from highest to lowest. The rank percentile is 100 - (rank - 1) / numTFs. The same procedure is used for identifying targets of TFA regulators (Fig. 4C), except that the sorted values are absolute difference of standardized log2 activity of each TF in each perturbed sample from the non-perturbed sample. Code for this evaluation is available at https://doi.org/10.5281/zenodo.4050573.

Correlation between TFA and TF-mRNA For each dataset, the second fittings (Fig. 1A, middle) included an unperturbed sample and 179 samples in which a TF was perturbed. For 1000 bootstraps of the 180 samples, the fraction of TFs whose mRNA and TFA values were positively correlated was calculated, and the median fraction across bootstrap samples was reported. TFA values were not logged or standardized before calculating correlation. Code for this evaluation is available at https://doi.org/10.5281/zenodo.4050573.

Union-PC Combined, the ChIP-PC networks with sign constraints from either TFKO or ZEV and the DE-PC networks with edges and sign constraints from either TFKO or ZEV contain 3,133 unique edges between 96 TFs and 1,592 target genes. Edges with conflicting sign constraints from different networks were dropped and TFs with only one target gene were removed, along with their target gene, leaving 2,731 edges between 94 TFs and 1,416 target genes. When optimizing this network on both datasets together, the perturbation samples for each network TF in each dataset were included, along with an unperturbed sample. CS parameters were shared, but different baseline values were used for each dataset to compensate for potential constant shifts in gene expression measurements.

4 Discussion

The ability to accurately infer changes in TF activity from changes in gene expression profiles provides a vital tool in the systems-biology toolbox. It enables us to look inside the cell, seeing not only the output of the circuits that control gene expression, but also their internal state. Observing the internal states of regulatory circuits is the key to understanding how these circuits control the cell's transcriptional program in response to internal and external signals. We demonstrated how this tool can be used to gain new insights into the regulation of TF activity, such as the probable activation of Gcr2 and repression of Gcn4, Gln3 and Msn2 by the Grr1 ubiquitin ligase in the presence of glucose (Fig. 5). In future work, we hope to automate this process of identifying TFA regulators and develop genome-scale benchmarks with which to evaluate it.

Previous studies introduced various versions of the matrix factorization approach (Boulesteix and Strimmer, 2005; Liao *et al.*, 2003; Sanguinetti *et al.*, 2006; Yu and Li, 2005). Here, we presented an objective, genome-scale evaluation of this approach, using multiple measures of accuracy and multiple independent datasets. We found that, when inferring TFA and CS matrices, it is essential to have data in which TFs are directly perturbed and constraints derived from such data. We also showed that a CS matrix derived from the TFKO and ZEV data can be used to successfully analyze other expression datasets that do not contain direct perturbations of TF

activity. In two of our core metrics, matrix factorization with constraints yielded TFA values that cover more than half the distance from random (50%) to ceiling (100%; Fig. 2A). However, this probably underestimates the true accuracy of the method, since even perfect TFA inference would not necessarily yield 100% on these metrics. For example, perturbation of a TF's activity could lead to an equal or bigger change in the activity of a downstream TF, pushing the true rank percentile of the perturbed TF below 100%. Similarly, the true percentage of TFs whose activity is positively correlated with their mRNA level is almost certainly less than 100%, due to post-transcriptional regulation.

Building the input network from binding specificity models (e.g. PWMs) is a popular approach that is condition-independent in principle, but it did not do well in our evaluation (Fig. 2A). Network structures derived from either perturbation-response data (DE) or currently available binding location data (ChIP) perform about equally when using the highest-scoring edges, but as more edges are added, the DE network is more robust (Fig. 3). For the ChIP-PC network, we observed the best performance when using only the top 1,250 edges, even though the recommended *P*-value threshold of 0.001 includes at least three times more. Using edges ranked 2,001 through 3,250 (Fig. 3A–C, Block 2) resulted in decreased performance on all metrics. The extended ChIP-PC networks of 77–80 TFs, which considered the top 2,000 edges rather than the top 1,250, also resulted in decreased performance on all the metrics (Fig. 3D). Since perturbation-response data are needed for sign constraints in any case, generating binding location data in addition may not be worth the effort and expense required. In other words, perturbation-response data is both necessary and sufficient for good performance.

Another recent paper (Trescher and Leser, 2019) evaluated TFA inference algorithms by using expression data after TF knockdowns in human cell lines and *E. coli*, similar to our second metric. It reported that the perturbed TF was rarely among those with the greatest inferred activity changes and that there was very little agreement among the algorithms tested. One factor that likely contributes to the difference in findings is that the algorithms they tested did not use sign constraints on control strengths and could not distinguish between increasing and decreasing activity. In the absence of sign constraints, we also saw poor performance. Another difference is that the input networks they used were largely based on manual curation rather than automated processing of high-throughput data, so they lacked confidence scores, making it impossible to select only the most confident edges. We saw performance degrade when less confident edges were included in the network. Finally, their evaluation was carried out using perturbations of only a handful of TFs for evaluation, making the findings vulnerable to sampling error.

We were surprised to find that, by our three basic metrics, optimizing control strengths is not necessary for achieving good performance—signed binary control strengths, taken directly from the input network and sign constraints, do almost as well (Fig. 4A and B). Optimized CS matrices result in somewhat better performance on two other metrics—detecting literature-supported TFA regulators (Fig. 4C) and detecting the trend in TF activity from a time course (Fig. 4D and E). Furthermore, we found evidence that the optimized control strengths correlated positively with those learned by optimizing on data from different growth conditions. They also correlated positively with the strength of binding to target promoters in independent binding data. Nonetheless, these correlations were far from perfect, so the limited impact of optimizing CS matrices on TFA accuracy may reflect the fact that the control strengths in our testing framework are optimized on one dataset while TFAs are optimized and evaluated on another (Fig. 1). In many real applications, control strengths and activities would be optimized on the same dataset. Improvements to the mathematical model could also increase the importance of CS optimization (see below). For now, however, using signed binary control strengths may be a reasonable choice for some organisms, especially when a limited amount of gene expression data is available. When control strengths are not optimized, the overall optimization changes from non-linear to linear, making it much faster and simpler.

Any approach to TFA inference relies on having a reasonably large and accurate set of targets for each TF. This can be challenging for TFs that are not very active in any of the conditions in which the data used to build the network were obtained. For example, Gal4 had only three targets in the Union-PC network. One of those, *GAL10*, is an established target with a known role in galactose metabolism, but CS optimization reduces the link between Gal4 and *GAL10*, emphasizing instead the link between Gal4 and dubious ORF YDR544C. This may be due to the fact that none of the samples to which we fit the activity levels were grown with galactose, so the expression of *GAL10* does not vary much. As a result, there is little need to explain *GAL10* expression as resulting from changes in Gal4 activity. A possible approach to this problem would be to discard inferences about TFs that have two or fewer significant targets after CS optimization.

Analysis of the error patterns on our three core benchmarks showed that poor performance was highly correlated with the fraction of a TF's targets that exhibited opposite signs in the TFKO and ZEV datasets (Fig. 2B–D). Apparent sign conflict can occur when the true sign of regulation is consistent, but the effect of the perturbation on the target gene is so weak that random measurement noise leads to a sign error. This can be remedied by using multiple perturbation datasets to determine sign and discarding edges with sign conflicts, as we did when constructing the Union-PC network. However, conflicts can also occur because some TFs are repressors in some conditions and activators in others. For example, Rgt1 represses *HXT1* in low glucose but activates it in high glucose (Polish et al., 2005). This points to a limitation of any model that constrains TFAs to be non-negative and CSs to be one sign or the other. A possible solution would be to release the non-negativity constraint on TFAs when there is sufficient evidence of a true sign change that applies to most of its targets.

The matrix factorization approach has several limitations. First, predicted gene expression does not saturate as TF activity gets large, whereas in reality each gene has maximum and minimum expression levels and the binding sites for each TF eventually become fully occupied. Second, the model assumes that each TF-target relationship is either activating or repressing in all conditions. Third, TF-TF interactions, such as competitive or cooperative binding, are not accounted for. Fourth, this approach does not model condition-dependent epigenetic effects on the susceptibility of each gene to regulation, for example by making the promoter more or less accessible to TFs. Fifth, as parameters are optimized during the fitting process, the variance explained is not a good predictor of a model's value for accurate TFA inference. Finally, when control strengths learned on one dataset are used to model a different dataset, the variance explained in the second dataset is much lower than in the first. This suggests a degree of over fitting that might be remedied by parameter shrinkage. However, improving the variance explained in cross-validation is not guaranteed to improve the accuracy of the TFA parameters learned.

We found that TFA inference in yeast works reasonably well when best practices are followed, but there is still room for improvement. We anticipate improvements coming from better network maps. One likely source of better maps is new, more accurate methods for measuring TF binding locations (Bergenholm et al., 2018; Holland et al., 2019; Kang et al., 2020; Mayhew and Mitra, 2016; Shively et al., 2019). The input network could also be improved by obtaining TF perturbation data from cells grown in new conditions. More improvement in TFA inference could come with better sign constraints and the possibility of allowing negative TF activity when the data strongly justify it. Mathematical models that more closely reflect the underlying biochemistry could also lead to better results, although such models come with new challenges. As new approaches are tried, the benchmarks presented here can be used to determine whether they robustly improve the accuracy of TFA inference, across multiple datasets and network maps.

Acknowledgements

The authors thank Dhoha Abid for her help with data analysis and Artelys Knitro for their help in testing their non-linear optimizer.

Funding

This work was supported by the National Institutes of Health [AI087794 and GM129126 to M.B.].

Conflict of Interest: none declared.

References

Alvarez,M.J. *et al.* (2016) Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.*, **48**, 838–847.

Apweiler,E. *et al.* (2012) Yeast glucose pathways converge on the transcriptional regulation of trehalose biosynthesis. *BMC Genomics*, **13**, 239.

Arrieta-Ortiz,M.L. *et al.* (2015) An experimentally supported model of the *Bacillus subtilis* global transcriptional regulatory network. *Mol. Syst. Biol.*, **11**, 839.

Azofeifa,J.G. *et al.* (2018) Enhancer RNA profiling predicts transcription factor activity. *Genome Res.*, **28**, 334–344.

Balwierz,P.J. *et al.* (2014) ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. *Genome Res.*, **24**, 869–884.

Barenco,M. *et al.* (2009) rHVDM: an R package to predict the activity and targets of a transcription factor. *Bioinformatics*, **25**, 419–420.

Barenco,M. *et al.* (2006) Ranked prediction of p53 targets using hidden variable dynamic modeling. *Genome Biol.*, **7**, R25.

Berchtold,E. *et al.* (2016) Evaluating transcription factor activity changes by scoring unexplained target genes in expression data. *PLoS One*, **11**, e0164513.

Bergenholm,D. *et al.* (2018) Reconstruction of a global transcriptional regulatory network for control of lipid metabolism in yeast by using chromatin immunoprecipitation with lambda exonuclease digestion. *mSystems*, **3**.

Bodenhofer,U. *et al.* (2011) APCluster: an R package for affinity propagation clustering. *Bioinformatics*, **27**, 2463–2464.

Boorsma,A. *et al.* (2008) Inferring condition-specific modulation of transcription factor activity in yeast through regulon-based analysis of genomewide expression. *PLoS One*, **3**, e3112.

Boscolo,R. *et al.* (2005) A generalized framework for network component analysis. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **2**, 289–301.

Boulesteix,A.L. and Strimmer,K. (2005) Predicting transcription factor activities from combined analysis of microarray and ChIP data: a partial least squares approach. *Theor. Biol. Med. Model.*, **2**, 23.

Brent,M.R. (2016) Past roadblocks and new opportunities in transcription factor network mapping. *Trends Genet.*, **32**, 736–750.

Broach,J.R. (2012) Nutritional control of growth and development in yeast. *Genetics*, **192**, 73–105.

Bussemaker,H.J. *et al.* (2017) Network-based approaches that exploit inferred transcription factor activity to analyze the impact of genetic variation on gene expression. *Curr. Opin. Syst. Biol.*, **2**, 98–102.

Bussemaker,H.J. *et al.* (2001) Regulatory element detection using correlation with expression. *Nat. Genet.*, **27**, 167–171.

Byrd,R. *et al.* (2006) *Knitro: An Integrated Package for Nonlinear Optimization*. Springer, Boston, MA, pp. 35–59.

Chechik,G. and Koller,D. (2009) Timing of gene expression responses to environmental changes. *J. Comput. Biol.*, **16**, 279–290.

Chen,J. *et al.* (2013) Genome-wide signatures of transcription factor activity: connecting transcription factors, disease, and small molecules. *PLoS Comput. Biol.*, **9**, e1003198.

Chen,Y. *et al.* (2017) Systems-epigenomics inference of transcription factor activity implicates aryl-hydrocarbon-receptor inactivation as a key event in lung cancer development. *Genome Biol.*, **18**, 236.

Cheng,C. *et al.* (2007) Inferring activity changes of transcription factors by binding association with sorted expression profiles. *BMC Bioinformatics*, **8**, 452.

Cokus,S. *et al.* (2006) Modelling the network of cell cycle transcription factors in the yeast *Saccharomyces cerevisiae*. *BMC Bioinformatics*, **7**, 381.

Conlon,E.M. *et al.* (2003) Integrating regulatory motif discovery and genome-wide expression analysis. *Proc. Natl. Acad. Sci. USA*, **100**, 3339–3344.

Conrad,M. *et al.* (2014) Nutrient sensing and signaling in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.*, **38**, 254–299.

De Virgilio,C. (2012) The essence of yeast quiescence. *FEMS Microbiol. Rev.*, **36**, 306–339.

Fisher,R.A. (1954) *Statistical Methods for Research Workers*.Oliver and Boyd, Edinburgh.

Fröhlich,H. (2015) biRte: Bayesian inference of context-specific regulator activities and transcriptional networks. *Bioinformatics*, **31**, 3290–3298.

Fu,Y. *et al.* (2011) Reconstructing genome-wide regulatory network of E. coli using transcriptome data and predicted transcription factor activities. *BMC Bioinformatics*, **12**, 233.

Gao,F. *et al.* (2004) Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data. *BMC Bioinformatics*, **5**, 31.

Garcia-Alonso,L. *et al.* (2019) Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.*, **29**, 1363–1375.

Garcia-Alonso,L. *et al.* (2018) Transcription factor activities enhance markers of drug sensitivity in cancer. *Cancer Res.*, **78**, 769–780.

Gitter,A. *et al.* (2013) Linking the signaling cascades and dynamic regulatory networks controlling stress responses. *Genome Res.*, **23**, 365–376.

Grant,C.E. *et al.* (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017–1018.

Hackett,S.R. *et al.* (2020) Learning causal networks using inducible transcription factors and transcriptome-wide time series. *Mol. Syst. Biol.*, **16**, e9174.

Harbison,C.T. *et al.* (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature*, **431**, 99–104.

Holland,P. *et al.* (2019) Predictive models of eukaryotic transcriptional regulation reveals changes in transcription factor roles and promoter usage between metabolic conditions. *Nucleic Acids Res.*, **47**, 4986–5000.

Jiang,P. *et al.* (2015) Inference of transcriptional regulation in cancers. *Proc. Natl. Acad. Sci. USA*, **112**, 7731–7736.

Kang,Y. *et al.* (2020) Dual threshold optimization and network inference reveal convergent evidence from TF binding locations and TF perturbation responses. *Genome Res.*, **30**, 459–471.

Kemmeren,P. *et al.* (2014) Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell*, **157**, 740–752.

Khanin,R. *et al.* (2007) Statistical reconstruction of transcription factor activity using Michaelis–Menten kinetics. *Biometrics*, **63**, 816–823.

Lam,K.Y. *et al.* (2016) Fused regression for multi-source gene regulatory network inference. *PLoS Comput. Biol.*, **12**, e1005157.

Lee,E. and Bussemaker,H.J. (2010) Identifying the genetic determinants of transcription factor activity. *Mol. Syst. Biol.*, **6**, 412.

Li,Y. *et al.* (2014) Regression analysis of combined gene expression regulation in acute myeloid leukemia. *PLoS Comput. Biol.*, **10**, e1003908.

Liao,J.C. *et al.* (2003) Network component analysis: reconstruction of regulatory signals in biological systems. *Proc. Natl. Acad. Sci. USA*, **100**, 15522–15527.

Liao,Y. *et al.* (2019) WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.*, **47**, W199–W205.

Ljungdahl,P.O. and Daignan-Fornier,B. (2012) Regulation of amino acid, nucleotide, and phosphate metabolism in Saccharomyces cerevisiae. *Genetics*, **190**, 885–929.

Mayhew,D. and Mitra,R.D. (2016) Transposon calling cards. *Cold Spring Harb. Protoc.*, **2016**, pdb.top077776.

Nachman,I. *et al.* (2004) Inferring quantitative models of regulatory networks from expression data. *Bioinformatics*, **20**, i248–256.

Ocone,A. and Sanguinetti,G. (2011) Reconstructing transcription factor activities in hierarchical transcription network motifs. *Bioinformatics*, **27**, 2873–2879.

Polish,J.A. *et al.* (2005) How the Rgt1 transcription factor of *Saccharomyces cerevisiae* is regulated by glucose. *Genetics*, **169**, 583–594.

Rodkaer,S.V. and Faergeman,N.J. (2014) Glucose- and nitrogen sensing and regulatory mechanisms in *Saccharomyces cerevisiae*. *FEMS Yeast Res.*, **14**, 683–696.

Rogers,S. *et al.* (2007) Bayesian model-based inference of transcription factor activity. *BMC Bioinformatics*, **8**, S2.

Ronen,M. and Botstein,D. (2006) Transcriptional response of steady-state yeast cultures to transient perturbations in carbon source. *Proc. Natl. Acad. Sci. USA*, **103**, 389–394.

Sameith,K. *et al.* (2015) A high-resolution gene expression atlas of epistasis between gene-specific transcription factors exposes potential mechanisms for genetic interactions. *BMC Biol.*, **13**, 112.

Sanguinetti,G. *et al.* (2006) Probabilistic inference of transcription factor concentrations and gene-specific regulatory activities. *Bioinformatics*, **22**, 2775–2781.

Schacht,T. *et al.* (2014) Estimating the activity of transcription factors by the effect on their target genes. *Bioinformatics*, **30**, i401–407.

Schwarz,G. (1978) Estimating the dimension of a model. *Ann. Stat.*, **6**, 461–464.

Shi,Y. *et al.* (2009) A combined expression-interaction model for inferring the temporal activity of transcription factors. *J. Comput. Biol.*, **16**, 1035–1049.

Shively,C.A. *et al.* (2019) Homotypic cooperativity and collective binding are determinants of bHLH specificity and function. *Proc. Natl. Acad. Sci. USA*, **116**, 16143–16152.

Spivak,A.T. and Stormo,G.D. (2012) ScerTF: a comprehensive database of benchmarked position weight matrices for *Saccharomyces* species. *Nucleic Acids Res.*, **40**, D162–168.

Tchourine,K. *et al.* (2018) Condition-specific modeling of biophysical parameters advances inference of regulatory networks. *Cell Rep.*, **23**, 376–388.

Tran,L.M. *et al.* (2005) gNCA: a framework for determining transcription factor activity based on transcriptome: identifiability and numerical implementation. *Metab. Eng.*, **7**, 128–141.

Trescher,S. and Leser,U. (2019) Estimation of transcription factor activity in knockdown studies. *Sci. Rep.*, **9**, 9593.

Tripodi,I.J. *et al.* (2018) Detecting differential transcription factor activity from ATAC-Seq data. *Molecules*, **23**, 1136.

Wang,C. *et al.* (2008) Motif-directed network component analysis for regulatory network inference. *BMC Bioinformatics*, **9**, S21.

Wang,H. *et al.* (2011) Calling Cards enable multiplexed identification of the genomic targets of DNA-binding proteins. *Genome Res.*, **21**, 748–755.

Yang,Y.L. *et al.* (2005) Inferring yeast cell cycle regulators and interactions using transcription factor activities. *BMC Genomics*, **6**, 90.

Yu,T. and Li,K.C. (2005) Inference of transcriptional regulatory network by two-stage constrained space factor analysis. *Bioinformatics*, **21**, 4033–4038.

Zaman,S. *et al.* (2009) Glucose regulates transcription in yeast through a network of signaling pathways. *Mol. Syst. Biol.*, **5**, 245.

Zhu,M. *et al.* (2013) REACTIN: regulatory activity inference of transcription factors underlying human diseases with application to breast cancer. *BMC Genomics*, **14**, 504.