One Hand Clapping: Detection of condition-specific Transcription Factor Interactions from genome-wide gene activity Data

Sebastian Dümcke, Martin Seizl, Stefanie Etzold, Nicole Pirkl,
Dietmar Martin, Achim Tresch, Patrick Cramer

10.03.2011

Abstract

We present a simple and universal method that detects condition-specific interactions between transcription factors from a single genome-wide mRNA expression experiment. Based on a transcription factor-target gene annotation of reasonable quality, our method uses a simple linear regression model to assess from a single intervention experiment whether the common targets of two arbitrary transcription factors behave differently than expected from the genes that are targeted by only one of them. When applied to osmotic stress data in S.cerevisiae, One Hand Clapping produces consistent results across three types of expression measurements: Gene expression microarray data, RNA Polymerase II ChIP-on-chip binding data, and measurements of nascent mRNA. Of the predicted 8 novel, condition-specific interactions, we validated the interaction between Gcn4p and Arr1p experimentally.

1 Introduction

In classical genetic interaction studies such as quantitative epistasis analysis using epistatic miniarray profiles [1] or synthetic genetic arrays (SGAs [2]), an interaction between two gene products is predicted from the comparison of the phenotypic effects of the single gene knockouts with the effect of the double knockout. Three interventions (two single and one double gene knockout) are needed to detect a genetic interaction from a single phenotypic readout (cell growth or other fitness measure [3]). In this work we revert this approach and analyze the effects of a single, albeit global intervention on a high dimensional phenotype (genome-wide gene expression measurements) to predict interactions of transcription factor (TF) pairs. The requirement of only one intervention gave rise to the name of our method, "One Hand Clapping" (OHC).

The principle at the core of this method is that, given two TFs we expect the genes targeted by both regulators to behave significantly different than the genes targeted by either TF alone. The behaviour of the genes can be measured through mRNA expression or other means quantifying gene activity such as factor occupancies from ChIP-chip experiments or RNA-Seq measurements [4, 5]. This principle should hold especially under experimental interventions that leads to cooperative behaviour of the selected TF pair.

2 Results

2.1 OHC accurately predicts pairwise TF interactions

We first applied our method to mRNA expression data from the total fraction of Miller *et al.* [6] (data set D1, expression folds after 36 minutes of osmotic stress) using the filtered YEASTRACT database as TF-target annotation (see Methods section 4.1). The resulting interaction matrix is shown as heatmap (Supplemental Figure 2). The rows of the matrix were clustered and predictions made as described in the Methods section. We predict 59 pairwise interactions between TFs, 43 single TFs are left without interactions. Validation of the predictions is done through the BioGRID [7] database. It contains physical and genetic interactions from high and low throughput experiments in the literature for many yeast proteins. The sub graph of BioGRID corresponding to interactions between TFs present in the TF-target graph is shown in Supplemental Figure 1. We validate 13 interactions in this way (22.03% positive predictive value), which are highlighted in Figure 1A. A complete list of predicted and validated pairs is given as Supplemental Materials. Wilcoxon's test on the correlation distances between validated

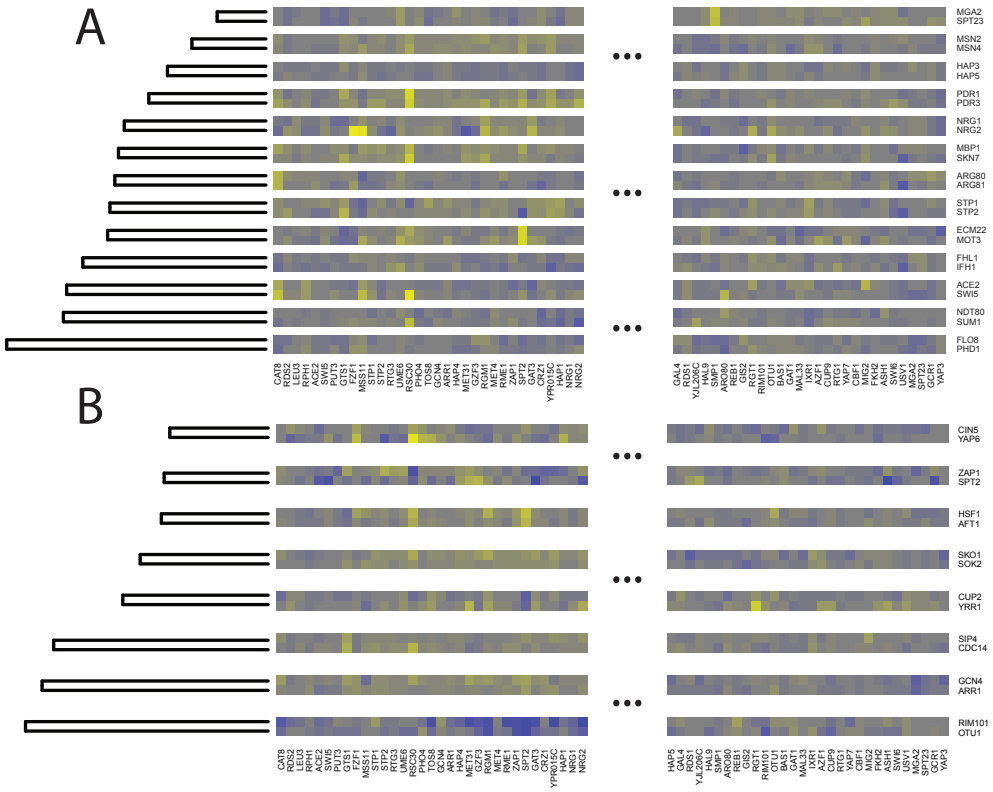
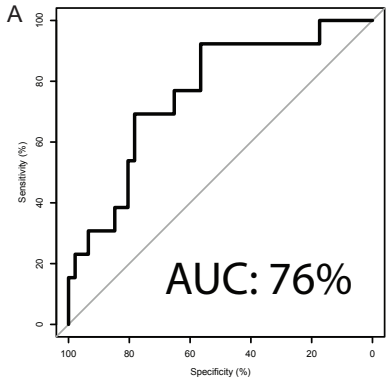


Figure 1: Excerpt from clustering dendrogram showing predicted interacting TF pairs along with the interaction profiles. Colors range from blue (strong negative interaction) to yellow (strong positive interaction term). A: 12 predictions validated by the BioGRID database B: 8 novel predictions across several data sets (see text).



B

predicted pairs

	D1	D2	D3	D4	total
D1	0	23	26	12	59
D2	23	0	23	10	60
D3	26	23	0	12	60
D4	12	10	12	0	57

validated pairs

	D1	D2	D3	D4	total
D1	0	8	12	7	13
D2	8	0	8	4	11
D3	12	8	0	7	14
D4	7	4	7	0	12

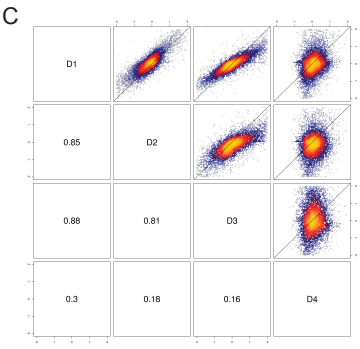


Figure 2: Validation of OHC predictions A: ROC curve using BioGRID interactions as gold standard B: Table of overlap between predicted and validated pairs for different data sets. 'total' refers to the number of pairs predicted resp. validated by the respective data set taken alone. D1: total mRNA fraction D2: labeled mRNA fraction D3: total mRNA fraction from Mitchell *et al.* D4: Pol2 ChIP-chip occupancy measurements C: pairwise comparison of expression values and occupancy values for all genes

and unvalidated predictions is significant at the 0.05 significance level (p-value 0.0036). This shows that interacting TF pairs are more closely related (considering our interaction measure and distance function) than unvalidated predictions. This is further investigated through a ROC plot (Figure 2A). The area under the curve (76%) shows a strong deviation from random predictions (diagonal).

2.2 OHC is stable and reproducible across independent experiments

To test the stability of our method we applied it to the mRNA expression data of the labeled fraction from the same osmotic stress experiment used previously (termed data set D2, see Methods). Both data sets are similar (Pearson's r=0.85, Figure 2C) and we expect similar results. On this data set we predict 60 pairwise interactions, 11 validated by the BioGRID database (18.33% prediction accuracy; predicted pairs: Nrg1p-Nrg2p Fhl1p-Ifh1p Stp1p-Stp2p Msn2p-Msn4p Mbp1p-Swi4p Ecm22p-Upc2p Cbf1p-Met28p Ndt80p-Sum1p Arg80p-Arg81p Hap3p-Hap5p and Mga2p-Spt23p). The validated interactions highly agree between both data sets, 8 pairs being validated by both runs (Figure 2B). The interactions Ace2p-Swi5p Ecm22p-Mot3p Pdr1p-Pdr3p Mbp1p-Skn7p and Flo8p-Phd1p found in the first data set are lost in the second, the interactions Mbp1p-Swi4p Ecm22p-Upc2p and Cbf1p-Met28p in the second are not present in the first data set. Comparison of all predicted interactions (Figure 2B) features an overlap of 23 pairwise interactions (38.66%).

Reproducibility was tested by running the method on another osmotic stress data set from [8](total mRNA measurement 30 minutes after addition of NaCl) termed D3 (Pearson's r=0.88, Figure 2C). The method predicts 60 pairwise interactions and 14 validated interactions (23.33%). The overlap with the previous two data sets is 26 and 23 pairs for data sets D1 and D2 respectively. Validated interactions agree more strongly; they overlap at 12 and 8 validated interactions for D1 and D2 respectively (Figure 2B). 8 validated interactions are found in all three data sets (Figure 1B shows the interaction profiles). It is interesting to notice that the data sets D3/D1 agree more closely than D3/D2 and D1/D2. This might be due to the fact that D1 and D3 measure the total mRNA at the extraction timepoint and thus include mRNAs transcribed before the onset of stress and not yet degraded, contrary to D2 which corresponds to the labeled mRNA fraction and thus contains only mRNAs transcribed after the onset of stress. Indeed D1/D3 have a higher correlation than D1/D2 and D3/D2 (Fig 2B).

2.3 Validation of novel predictions found by OHC

Eight predictions are found in all three data sets (D1, D2 and D3). They are Cin5p-Yap6p, Zap1p-Spt2p, Hasf1p-Aft1p, Sok1p-Sko2p, Cup2p-Yrr1p, Sip2p-Cdc14p, Gcn4p-Arr1p and Rim101p-Otu1p.

Considering the pair Cin5p-Yap6p we realized both TF had very similar binding motifs. Indeed a test for cooccurence of both motifs on all intergenic regions is highly significant (pvalue < 5.44e-12). This can be due to two causes: Either one factor binds the other prior to binding the DNA sequence and hence ChIP-chip experiments

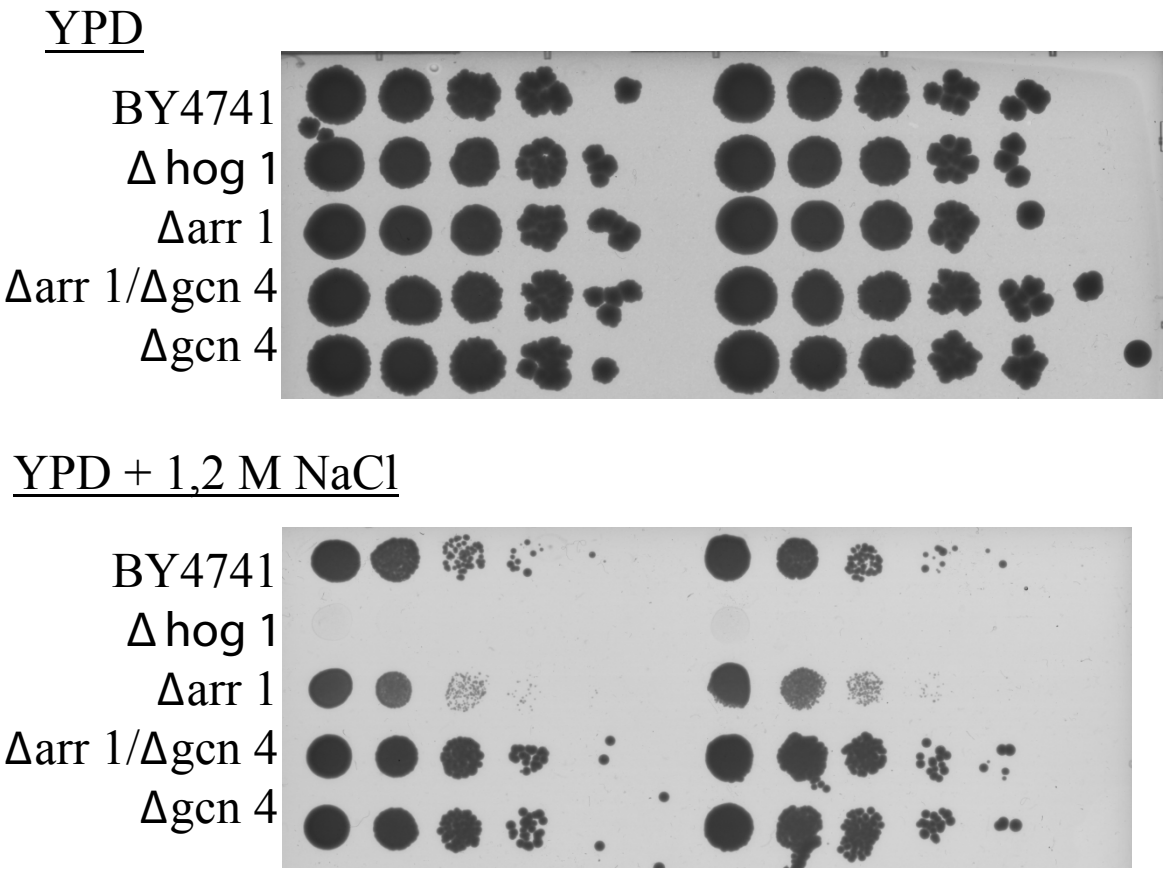


Figure 3: Growth assay on YPD and YPD containing 1.2M NaCl after 4 days at 30°C. Δhog1 is used as a negative control for salt stress. Growth phenotype is not affected in YPD medium. Wild type cells have decreased phenotype under osmotic stress. hog1 mutant is not viable under salt stress and arr1 mutant shows strong decrease in growth phenotype, while gcn4 mutant and double mutant do not. This shows the phenotypic rescue of the effect of the arr1 knockout in the double mutant.

would crosslink the same DNA stretches in both cases and leading to similar motifs or, alternatively both factors bind competitively to the same binding motif. The latter would directly validate this prediction.

Spot dilutions of the single and double knockouts of GCN4 and ARR1 on salt and YPD medium show a synthetic rescue of the detrimental effect of the ARR1 knockout in the double mutant (Figure 3). This effect is a strong indicator of an interaction between the gene products and validates our prediction.

2.4 OHC is applicable on a wide range of gene activity data

To show that the method also works on proxies of gene activity other than mRNA expression measurements, we used the RNA pol2 ChIP-chip data from [6] (termed D4). On this data set the method predicts 57 interactions, 12 of which can be validated (21.05% accuracy). Its performance is thus comparable to the performance on mRNA expression data. The predictions vary strongly as there are only 12, 10 and 12 predicted interactions shared with the data sets D1, D2 and D3 respectively (Figure 2B). This is due to a low correlation between the data sets varying between 0.16 and 0.3 (Pearson's correlation, see Figure 2C). Despite the low correlation, a core of 8 interaction is shared between all data sets (including 4 novel predictions) and shows that the method is robust enough to be adapted to all measures of gene activity.

3 Discussion

OHC finds *cis* and *trans* TF interactions We distinguish between two main types of combinatorial TF interactions: *cis* regulatory interactions and *trans* regulatory interactions [9]. *Cis* interactions are mediated by a specific TF binding site configuration at the *cis* regulatory region of a gene, possibly resulting in cooperative or competitive binding of TFs. Competitive binding occurs when two TFs share a common or overlapping binding motif. Cooperative binding of TFs occurs if two TFs are required to bind simultaneously to be functional, or if the binding of the second TF is enhanced by the binding of the first TF, which e.g. is the case for nucleosome-mediated cooperativity [10]. Nucleosome mediated cooperativity has gained much attention in the last years [11]. It can happen in several ways: the binding of one TF (often the one with stronger DNA binding affinity) evicts (part of) a nucleosome occupying the binding site of a second TF, whose binding to the promoter then becomes possible . Alternatively the first TF can recruit chromatin remodellers to the promoter which will evict nucleosomes blocking access to the DNA to other TFs .

Trans interactions are defined as direct protein-protein interaction of both TFs prior to DNA binding, either by forming a protein complex or by complex formation with other co-factors involved in polymerase recruitment and transcription initiation .

We investigated which type of interaction OHC detects. TFs pairs predicted by our method and validated by BioGRID include the following types of interaction: Ace2p-Swi5p [12] and Sum1p-Ndt80p [13] undergo competitive *cis* regulatory interactions, the first two TFs having identical binding sites, the latter two TFs overlapping binding sites. Mot3p-Ecm22p [14], Mbp1p-Skn7p [15], Arg80p-Arg81p [16], Hap3p-Hap5p [17] and Pdr1p-Pdr3p [18] are all examples of *trans* regulatory protein interactions forming prior to DNA binding. Ifh1p-Fhl1p are also *trans* acting but Fhl1p is already bound to the promoter of ribosomal protein genes and it is the phosphorylation of Ifh1p that enables the protein interaction [19].

We found a third category of predicted TF pairs, corresponding to homologous or functionally redundant proteins. Msn2p-Msn4p [20], Mga2p-Spt23p [21] and Stp1-Stp2p [22] belong to this category. They are found by our method due to their highly similar target gene sets. Nrg1p-Nrg2p is also found due to the target gene set overlap, as the target gene set for Nrg2p is a subset of the Nrg1p target set (as defined by YEASTRACT) although we did not find any evidence for a direct interaction. This is also the case of the pair Flo8p-Phd1p, both proteins are involved in parallel pathways regulating the FLO11 gene [23].

The TF annotation play an important role in OHC One cornerstone of the assumption that interacting TFs can be found by looking at commonly regulated target genes is the availability of a correct mapping of TFs to a set of target genes. Such a mapping is rarely available, especially for different growth conditions. This is a limitation of the method that will hopefully be alleviated with the advent of ChIP-seq data sets of TFs in many organism (this data begins to arrive to the ENCODE and modENCODE projects [24, 25, 26]). .

For the yeast *Saccharomyces cerevisiae* there are fortunately several options for a mapping between TF and target genes. An experimentally derived data set is provided by MacIsaac *et al.* [27] who reanalyzed ChIP-chip data for 128 TFs under standard conditions and a few selected stress conditions from [28]. Computationally derived TF-target graphs can be obtained using the method from [29] or by mapping TF motifs from databases ([30, 31]) to promoter regions in the genome. Finally TF-target relations mined from a manually curated literature repository can be found in the YEASTRACT database [32]. The latter is used in this paper as it is to our knowledge the most likely to contain associations from many different environmental conditions.

The predictions made by OHC are different from predictions based on target genes set alone. Indeed, Fisher's test for overlap does not correlate with the interaction scores from the linear model (Supplemental Figure 11). Also the method can predict interactions between TFs that have no overlap in target genes and thus no interaction score. This is possible because we predict interactions based on profile similarity which takes into account the interaction scores with all TFs. We found three interactions where this is the case: Kar4p-Stb1p, RDS1p-YJL206C and Cbf1p-Mig2p. We suspect these interaction to be physical interactions between the proteins

We have observed that TF pairs with small overlap get high interaction scores because they are more likely to behave uniformly and thus assigned a strong coefficient by the linear regression. Single TFs with large target sets have profiles that are very homogeneous with low interaction scores. They will rarely be similar to profiles of TFs with small single sets as these have often sparse profiles with few strong interaction scores. Efforts to adapt the distance measure used for the clustering did not improve the performance (data not shown).

The performance of OHC is strongly influenced by the assignment of target genes to TFs. When we run the method using the annotation from MacIsaacs *et al.* [27] containing 118 TFs, we predict 38 interactions, 6 validated by BioGRID (15.79% prediction accuracy). This performance difference is due to different annotated genes. While the annotation from MacIsaacs, based on Chip-chip data is of high quality, it does not suit our ap-

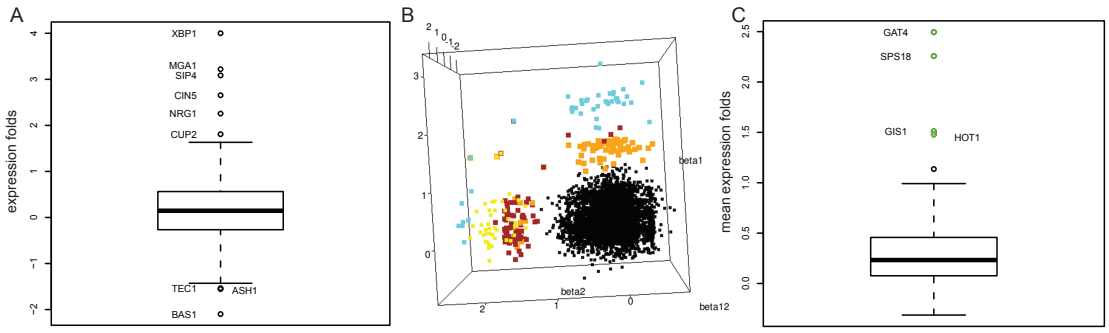


Figure 4: A: mRNA expression folds (data set D1) of genes coding for all TF from YEASTRACT. Strong differentially expressed genes are nor necessarily involved in the osmotic stress response pathway suggesting that TF activity is regulated post-transcriptionally. B: 3D plot of coefficients β_1, β_2 and β_{12} from the interaction model. β_1 is in the x-direction, increasing to the left, β_2 is in the y-direction increasing upwards and β_{12} in the z-direction. The coefficients of interactions involving Gis1p, Gat4p, Hot1p and Sps18p are highlighted in orange, brown, cyan and yellow respectively. There is no apparent correlation between the single effects and the interaction effect. C: mean expression folds in data set D1 of the target sets of each TF in YEASTRACT. The target sets of Gis1p, Gat4p, Hot1p and Sps18p (highlighted in green) show a strong (> log2(1.5) fold) differential expression.

plication. This might be due to the fact that it contains assignments made under standard experimental conditions while YEASTRACT contains many TF-target gene assignment under different stress and knockout conditions. It should be noted that YEASTRACT is a superset of MacIsaacs annotation and we hypothesize that, while the YEASTRACT graph probably contains many edges that are irrelevant under salt stress, this is accounted for and in a certain way beneficial to our model.

4 Materials and Methods

4.1 TF annotation

YEASTRACT is a manually curated repository of regulatory interactions between TF and genes. It can be visualized as a bipartite graph with edges between TF and genes representing a regulatory interaction (i.e TF A regulates gene B). This graph is filtered, removing TF with less than 10 annotated target genes. This leaves 165 TF with 167 annotated genes as median value across all TF. Size distribution of the annotated gene groups is shown as Supplemental Figure 12.

Figure 4A shows a box plot of expression folds of the TFs from YEASTRACT. Prominent differentially expressed TFs are explicitly shown (XBP1, MAG1, SIP4, CIN5, NRG1, CUP2, TEC1, ASH1 and BAS1). Most of these outliers are not directly involved in the salt stress or general stress response pathways, showing that TF activity is not regulated at the transcriptional level.

When looking at the coefficients β_1, β_2 and β_{12} from the regression model of all TF pairs in the YEASTRACT database there is no apparent structure (Figure 4B). Closer investigation finds that extreme values are due to pairwise interactions with a small set of four TFs (Hot1p, Sps18p, Gis1p, Gat4p colored points in Figure 4C). Indeed these TFs have target genes which are substantially differentially expressed, thus giving rise to a high β_{12} coefficient to every TF having a considerable overlap with one of those 4 TFs. The mean expression of all target genes is above that of all other TFs (Figure 4C). A Gene Ontology analysis revealed that they are stress responder genes involved in response to various stimuli and heat (Supplemental Table 1). We removed these four outlier TFs from the TF-target graph, leaving us with a final annotation containing 161 TFs.

4.2 Interaction model

The main idea of our method can be formalised thru a logistic linear regression. Given two TFs with target gene sets T_1 and T_2 as well as gene activity measurements e_g for all genes $g \in G$ we fit the following model:

$$e_g \sim \beta_0 + \beta_1 I(g \in T_1) + \beta_2 I(g \in T_2) + \beta_{12} I(g \in T_1 \cap T_2)$$

$I(g \in T_i)$ is a binary indicator function, taking value 1 if gene g belongs to the target set of TF T_i and 0 else. β_0 is the term for all genes that are not targets of either TF and represent the background in this model. β_1 and β_2 are the single terms for each TF while β_{12} is the interaction term. It should be noted that if the measurements e_g are differential mRNA expression measurements then $\beta_0, \beta_1, \beta_2$ and β_{12} can be interpreted as expression folds as well. The regression is performed for each TF pair separately as fitting many different TF pairs together would over fit the model.

After running the regression in an all-against-all fashion for a set of TFs T we are left with a symmetric interaction matrix $M^{|T|\times|T|}$ containing all interaction terms β_{12}(see Supplemental Figures 2, 4-6). We noticed that the interaction terms alone are not strong predictors of interaction (data not shown). We attribute this to the fact that the definition of the target sets T_1 and T_2 is imperfect as well as the possibility of higher order interactions between TFs.

4.3 Clustering

To strengthen the interaction measure we use a "guilt-by-association" principle. Instead of comparing single interaction values we compare interaction profiles (the rows of the interaction matrix M) of each TF by their correlation. More specifically we use $1 - r$ as distance measure, where r is Pearson's correlation coefficient.

One possible drawback of using Pearson's correlation is that it could potentially attribute a low distance to interaction profiles which have a similar variance yet a different amplitude in their effects. This is because of the scale invariance of this distance function. We do not observe this pitfall. Rather we find that the distance measure is dominated by few entries in each profile which agree well. We believe this to be a good sign as TFs probably interact with only a few partner under any given experimental condition, which is faithfully reproduced by this distance method.

We apply hierarchical average linkage clustering to the rows of M using the distance measure mentioned above. This is very similar to nearest neighbour (NN) clustering, which could have also been used, but is more stringent. In NN clustering each interaction profile is paired with the profile having the smallest distance. Since a NN can always be found this results in $|T|$ predicted TF pairs. To exclude some predictions an arbitrary threshold defining a maximum distance must be used.

The approach taken here differs from NN in that hierarchical clustering merges the closest TFs together into clusters in a pairwise fashion and will add other TFs that are closest to any of these 'core' cluster by adding a branch to the dendrogram under a same common ancestor.

Following the clustering procedure we select these 'core' cluster as predictions of the method. The algorithm to extract the 'core' cluster from the clustering dendrogram is provided as Supplemental Materials and uses a recursive descent down the dendrogram, collecting TF pairs sharing a common ancestor that is the ancestor only to this pair of TFs.

This is more stringent than NN as each TF is assigned once to its closest interaction partner, disregarding the absolute level of the correlation distance. Also this method predicts less interactions and allows TFs to be excluded from any interaction in the experimental condition studied without the need for an arbitrary distance cutoff. It should be noted, that using this procedure each TF can only partake in one interaction with another TF.

4.4 Gene activity data

In this paper we use several data sets as input to our method. First we use mRNA expression data from a timecourse experiment submitting yeast strain BY4741 to osmotic stress by adding 0.8 M NaCl (see [6] for more details). The paper uses a RNA labeling approach and provides microarray measurements for three fractions: labeled RNA, unlabeled RNA and total RNA. We use the total mRNA expression data (which corresponds to standard measurements without the labeling approach) after 36 minutes of osmotic stress. Throughout this paper we will always mean log expression folds (log quotient of expression under experimental condition against wild-type expression) when referring to expression data. We will also use the labeled fraction for validation purposes. The same publication also provides RNA Pol2 ChIP-chip experiments giving Pol2 occupancy data 24 minutes after addition of salt. We will use the mean occupancy on the gene (between transcription start site and polyadenylation site) as a proxy for gene activity.

To test the reproducibility of our method we will apply it to data from another publication from another independent laboratory: Mitchell *et al.* [8] studied anticipation to environmental changes in *E. coli* and *S. cerivisiae* including osmotic and oxidative stress. They use the same yeast strain and microarray platform (Affymetrix Yeast 2.0) as well as a similar protocol to Miller *et al.* which should render the measurements highly comparable. They measured mRNA expression (corresponding to the total fraction of Miller *et al.*) 30 minutes after addition of 0.8M

NaCl to the medium. Microarray data was downloaded as raw files from GEO (accession: GSE15936) and normalized using gcrma [33] (as implemented in R/Bioconductor) without quantile normalization (since we expect global effects of the perturbation on mRNA expression).

All expression values are median centered.

4.5 Yeast strains and growth assays

The *Saccharomyces cerevisiae* deletion strains *hog1*Δ*, arr1*Δ*, gcn4*Δ, as well as the wild type strain BY4741 were obtained from Open Biosystems (Huntsville, USA). The double deletion strain *arr1*Δ*/gcn4*Δ was generated by integrating a ClonNat cassette in the ARR1 locus of the *gcn4*Δ strain. Correct gene disruptions were verified by PCR. Spot dilutions were done to assess fitness and growth under salt stress. Equal amounts of freshly grown yeast cells in YPD were re-suspended in water, 10-fold dilutions were spotted on YPD plates and YPD plates with 1.2 M NaCl, and plates were incubated for 4 days at 30°C.

Supplemental Materials

Supplemental figures, tables and materials can be downloaded from:
`http://www.lmb.uni-muenchen.de/tresch/ohc/supp.pdf`

References

[1] Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, et al. (2005) Exploration of the Function and Organization of the Yeast Early Secretory Pathway through an Epistatic Miniarray Profile. Cell 123: 507–519.

[2] Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, et al. (2001) Systematic Genetic Analysis with Ordered Arrays of Yeast Deletion Mutants. Science 294: 2364–2368.

[3] Mani R, St, Hartman JL, Giaever G, Roth FP (2008) Defining genetic interaction. Proceedings of the National Academy of Sciences 105: 3461–3466.

[4] Bloom J, Khan Z, Kruglyak L, Singh M, Caudy A (2009) Measuring differential gene expression by short read sequencing: quantitative comparison to 2-channel gene expression microarrays. BMC Genomics 10: 221+.

[5] Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN (2010) RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics 26: 493–500.

[6] Miller C, Schwalb B, Maier K, Schulz D, Dümcke S, et al. (2011) Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Molecular Systems Biology 7.

[7] Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, et al. (2006) BioGRID: a general repository for interaction datasets. Nucleic Acids Res 34: D535–539.

[8] Mitchell A, Romano GH, Groisman B, Yona A, Dekel E, et al. (2009) Adaptive prediction of environmental changes by microorganisms. Nature 460: 220–224.

[9] Tirosh I, Reikhav S, Levy AA, Barkai N (2009) A Yeast Hybrid Provides Insight into the Evolution of Gene Expression Regulation. Science 324: 659–662.

[10] Mirny LA (2010) Nucleosome-mediated cooperativity between transcription factors. Proceedings of the National Academy of Sciences 107: 22534–22539.

[11] Raveh-Sadka T, Levo M, Segal E (2009) Incorporating nucleosomes into thermodynamic models of transcription regulation. Genome research 19: 1480–1496.

[12] Voth WP, Yu Y, Takahata S, Kretschmann KL, Lieb JD, et al. (2007) Forkhead proteins control the outcome of transcription factor binding by antiactivation. The EMBO Journal 26: 4324–4334.

[13] Pierce M, Benjamin KR, Montano SP, Georgiadis MM, Winter E, et al. (2003) Sum1 and Ndt80 Proteins Compete for Binding to Middle Sporulation Element Sequences That Control Meiotic Gene Expression. Mol Cell Biol 23: 4814–4825.

[14] Davies BSJ, Rine J (2006) A Role for Sterol Levels in Oxygen Sensing in Saccharomyces cerevisiae. Genetics 174: 191–201.

[15] Bouquin N, Johnson AL, Morgan BA, Johnston LH (1999) Association of the Cell Cycle Transcription Factor Mbp1 with the Skn7 Response Regulator in Budding Yeast. Mol Biol Cell 10: 3389–3400.

[16] Amar N, Messenguy F, El Bakkoury M, Dubois E (2000) ArgRII, a Component of the ArgR-Mcm1 Complex Involved in the Control of Arginine Metabolism in Saccharomyces cerevisiae, Is the Sensor of Arginine. Mol Cell Biol 20: 2087–2097.

[17] McNabb DS, Xing Y, Guarente L (1995) Cloning of yeast hap5: a novel subunit of a heterotrimeric complex required for ccaat binding. Genes & Development 9: 47-58.

[18] Mamnun YM, Pandjaitan R, Mahé Y, Delahodde A, Kuchler K (2002) The yeast zinc finger regulators Pdr1p and Pdr3p control pleiotropic drug resistance (PDR) as homo- and heterodimers in vivo. Molecular Microbiology 46: 1429–1440.

[19] Rudra D, Zhao Y, Warner JR (2005) Central role of Ifh1p-Fhl1p interaction in the synthesis of yeast ribosomal proteins. The EMBO Journal 24: 533–542.

[20] Martínez-Pastor MT, Marchler G, Schüller C, Marchler-Bauer A, Ruis H, et al. (1996) The Saccharomyces cerevisiae zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). The EMBO journal 15: 2227–2235.

[21] Zhang S, Skalsky Y, Garfinkel DJ (1999) MGA2 or SPT23 Is Required for Transcription of the ÎŽ9 Fatty Acid Desaturase Gene, OLE1, and Nuclear Membrane Integrity in Saccharomyces cerevisiae. Genetics 151: 473–483.

[22] Wielemans K, Jean C, Vissers S, André B (2010) Amino acid signaling in yeast: post-genome duplication divergence of the Stp1 and Stp2 transcription factors. The Journal of biological chemistry 285: 855–865.

[23] Pan X, Heitman J (2000) Sok2 Regulates Yeast Pseudohyphal Differentiation via a Transcription Factor Cascade That Regulates Cell-Cell Adhesion. Mol Cell Biol 20: 8364–8372.

[24] Consortium TEP (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 306: 636–640.

[25] Negre N, Brown CD, Ma L, Bristow CA, Miller SW, et al. (2011) A cis-regulatory map of the Drosophila genome. Nature 471: 527–531.

[26] Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, et al. (2010) Integrative Analysis of the Caenorhabditis elegans Genome by the modENCODE Project. Science 330: 1775–1787.

[27] MacIsaac K, Wang T, Gordon DB, Gifford D, Stormo G, et al. (2006) An improved map of conserved regulatory sites for saccharomyces cerevisiae. BMC Bioinformatics 7: 113+.

[28] Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, et al. (2004) Transcriptional regulatory code of a eukaryotic genome. Nature 431: 99–104.

[29] Beyer A, Workman C, Hollunder J, Radke D, Möller U, et al. (2006) Integrated assessment and prediction of transcription factor binding. PLoS computational biology 2: e70+.

[30] Bryne JC, Valen E, Tang MHE, Marstrand T, Winther O, et al. (2008) Jaspar, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucl Acids Res 36: D102–106.

[31] Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, et al. (2006) Transfac(r) and its module transcompel(r): transcriptional gene regulation in eukaryotes. Nucl Acids Res 34: D108–110.

[32] Teixeira MC, Monteiro P, Jain P, Tenreiro S, Fernandes AR, et al. (2006) The yeastract database: a tool for the analysis of transcription regulatory associations in saccharomyces cerevisiae. Nucl Acids Res 34: D446–451.

[33] Wu Z, Irizarry RA, Gentleman R, Martinez-Murillo F, Spencer F (2004) A model-based background adjustment for oligonucleotide expression arrays. Journal of the American Statistical Association 99: 909+.