Supplemental Material for

"Distinguishing direct versus indirect transcription factor-DNA interactions"

Raluca Gordân, Alexander J. Hartemink, Martha L. Bulyk

I. Detailed methods

ChIP-chip data

We used the yeast ChIP-chip data from Harbison et al. (Harbison et al. 2004), who performed 352 experiments for 207 TFs under different environmental conditions: YPD (rich medium), Acid (acidic medium), Alpha (alpha factor pheromone treatment), BUT14 (butanol treatment for 14h), BUT90

(butanol treatment for 90 min), GAL (galactose medium), H202Hi (highly hyperoxic), H202Lo (mildly hyperoxic), HEAT (elevated temperature), Pi- (phosphate deprived medium), RAFF (raffinose medium), RAPA (nutrient deprived), SM (amino acid starvation), and THI- (vitamin deprived). We use the notation *TF_cond* to refer to the ChIP-chip experiment for transcription factor *TF* under environmental condition *cond*.

For each ChIP-chip experiment we ranked all the intergenic probes on the microarray according to their p-values, and defined the 'bound' sequences for this experiment to be those intergenic probes with a p-value < 0.001. Out of the 352 experiments, 237 contain at least 10 probes bound at a p-value < 0.001. We restricted our analysis to these 237 ChIP-chip data sets.

PBM-derived DNA binding motifs

We used PBM-derived DNA binding motifs for 139 TFs as reported by Zhu et al. (Zhu et al. 2009) and Badis et al. (Badis et al. 2008), who used universal PBMs (Berger et al. 2006) to determine high-resolution *in vitro* DNA binding specificity data for these TFs. We used position weight matrices (PWMs) as reported by each of these two studies (89 PWMs from the Zhu et al. study (Zhu et al. 2009) and 50 additional PWMs from the Badis et al. study (Badis et al. 2008)). To compare two motifs, we use the inter-motif distance described in (Harbison et al. 2004) and call the motifs 'similar' if the distance between them is < 0.25, and 'different' otherwise.

Nucleosome positioning data

We used *in vivo* nucleosome positioning information from Lee et al. (Lee et al. 2007) to compute, for each DNA site $\boldsymbol{S}$, the probability that the site is occupied by nucleosomes. Lee et al. used micrococcal nuclease digestion followed by microarray analysis to derive a high-resolution map of nucleosome occupancy across the whole yeast *Saccharomyces cerevisiae* genome. From this map we extracted, for every fourth position in the genome, the logarithm of the ratio between the signal intensity of nucleosomal DNA versus genomic DNA at that position, and then interpolated the data to obtain 1-bp resolution data. Next, we applied a logistic transformation to the log-ratio values to obtain, for each position in the genome, the probability of that position being occupied by a nucleosome.

As shown in Supplemental Figure 3A, we used the logistic transformation $g(x) = \lambda e^{\alpha x}/(1+\lambda e^{\alpha x})$, with α and λ chosen such that the following two criteria were met: 1) the distribution of nucleosome occupancy probabilities is bimodal, with the two modes close to the boundaries of the [0,1] interval

(similar to the distribution obtained by Yassour et al. (Yassour et al. 2008), and 2) approximately 80% of the positions in the genome are occupied by nucleosomes (as reported previously by Lee et al. (Lee et al. 2007)). Setting the parameter α equal to 8 ensures that the first criterion is met, while setting the parameter λ equal to 41.48 ensures that the fraction of the genome with nucleosome occupancy probability > 0.5 is ~80%.

In choosing the value for parameter α, we used as a guideline the posterior distribution of nucleosome occupancy probabilities of Yassour et al. (Yassour et al. 2008), which were computed using a complex hidden Markov model, but only for part of the yeast genome. We chose α=8 because it minimizes the Kullback-Leibler (KL) divergence between the distribution obtained by Yassour et al. (Yassour et al. 2008) and the distribution we obtained using our logistic transformation. To compute the KL divergence we discretized the [0,1] interval into 20 bins, each corresponding to a subinterval of size 0.05. The Kullback-Leibler divergence is a widely used measure of the distance between two probability distributions (Kullback 1959). Minimizing this measure ensures that the distribution we obtained is as similar as possible to the distribution computed by Yassour et al. (Yassour et al. 2008).

The parameter λ has the effect of shifting the logistic function to the left so that for all x > -0.47 (which is true for ~80% of the genome) we have g(x) > 0.5. Supplemental Figure 3B shows the raw nucleosome data and the probabilities of nucleosome occupancy over an arbitrary region of chromosome 4 (chromosomal coordinates 8,000 to 10,000).

Given a site $\boldsymbol{S}=S_1...S_W$ of width W and the probabilities of nucleosome occupancy for each position i in the site, we can compute the probability of site $\boldsymbol{S}$ being occupied by nucleosomes, or, alternatively, the probability of site $\boldsymbol{S}$ being free of nucleosomes:

$$P(S_1 \ldots S_W \text{ free}) = P(S_1 \text{ free}) \times P(S_2 \text{ free} \mid S_1 \text{ free}) \times \ldots \times P(S_W \text{ free} \mid S_{W-1} \text{ free}) \quad (1)$$

Each term $P(S_{i+1}$ free $\mid S_i$ free) can be written as:

$$P(S_{i+1} \text{ free} \mid S_i \text{ free}) = 1 - P(S_{i+1} \text{ occupied} \mid S_i \text{ free}) = 1 - \frac{1}{N} \times (S_{i+1} \text{ occupied}) \quad (2)$$

where N is set to 147, the average nucleosome width.

Scoring a DNA sequence according to a PWM

We scored DNA sequences using a model similar to GOMER (Granek and Clarke 2005). Other models such as MatrixReduce (Foat et al. 2006) or TRAP (Roider et al. 2007) may also be used to compute the

probability that a TF with a particular PWM binds a DNA sequence. However, both MatrixReduce and TRAP use parameters that need to be trained on the ChIP-chip data. Since we want to use the model to test how well certain motifs explain the ChIP-chip data, training those motifs on the data themselves would not be appropriate.

Let $\boldsymbol{T}$ denote a TF, and ϕ denote the PWM describing the DNA binding motif of $\boldsymbol{T}$: $\phi(b, j)$ = the probability of finding base b at location j within the binding site ($b \in$ {A, C, G, T} and $1 \le j \le W$, where W is the width of the motif). Let ϕ_0 denote the background model, a Markov model that describes general genomic sequences. Here, ϕ_0 is a 0^{th}-order Markov model trained on all intergenic sequences in yeast.

Given a DNA site $\boldsymbol{S} = S_1 S_2 \ldots S_W$, we can score it according to the PWM and background models, and use the ratio of the two scores as an approximation for the dissociation constant:

$$K_d(\boldsymbol{T},\boldsymbol{S}) = \prod_{j=1}^{W} \frac{\phi_0(S_j)}{\phi(S_j, j)}$$

Next, using the fact that $K_d(\boldsymbol{T},\boldsymbol{S}) = [\boldsymbol{T}]\cdot[\boldsymbol{S}] / [\boldsymbol{T}\cdot\boldsymbol{S}]$, we can write the probability that TF $\boldsymbol{T}$ binds the DNA site $\boldsymbol{S}$ as:

$$P(\boldsymbol{T} \text{ binds } \boldsymbol{S}) = \frac{[\boldsymbol{T}\cdot\boldsymbol{S}]}{[\boldsymbol{T}\cdot\boldsymbol{S}] + [\boldsymbol{S}]} = \frac{[\boldsymbol{T}]}{[\boldsymbol{T}] + K_d(\boldsymbol{T},\boldsymbol{S})} = 1 \Big/ \left(1 + \frac{1}{[\boldsymbol{T}]} \times \prod_{j=1}^{W} \frac{\phi_0(S_j)}{\phi(S_j, j)}\right) \tag{3}$$

where the concentration of free TF, $[\boldsymbol{T}]$, is set to the dissociation constant for the site with the optimal PWM score, as in the GOMER model (Granek and Clarke 2005). This implies that the optimal DNA site is occupied by the TF 50% of the time.

For a DNA sequence $\boldsymbol{X}$ longer than the motif width W, the probability that TF $\boldsymbol{T}$ binds $\boldsymbol{X}$ is computed as:

$$\begin{aligned} P(\boldsymbol{T} \text{ binds } \boldsymbol{X}) &= P(\boldsymbol{T} \text{ binds any } X_i \ldots X_{i+W-1})) \\ &= 1 - \prod_{i}^{n-W+1} \left(1 - P(\boldsymbol{T} \text{ binds } X_i \ldots X_{i+W-1})\right) \\ &= 1 - \prod_{i}^{n-W+1} \left(1 - 1 \Big/ \left(1 + \frac{1}{[T]} \times \prod_{j=i}^{i+W-1} \frac{\phi_0(X_j)}{\phi(X_j, j-i+1)}\right)\right) \end{aligned} \tag{4}$$

Incorporating nucleosome positioning information

So far we assumed that the probability that a TF binds a DNA site depends only on the specificity of the

factor for that particular site, which is a good assumption in the case of *in vitro* experiments. *In vivo*, however, many DNA regions are occupied by nucleosomes and thus are not accessible for binding by a TF. To take this into account, we first need to re-write equation (4) to include information about the accessibility of site ***S***:

$$\begin{aligned} P(\boldsymbol{T} \text{ binds } \boldsymbol{S}) &= P(\boldsymbol{T} \text{ binds } \boldsymbol{S} \mid \boldsymbol{S} \text{ free}) \times P(\boldsymbol{S} \text{ free}) + P(\boldsymbol{T} \text{ binds } \boldsymbol{S} \mid \boldsymbol{S} \text{ occupied}) \times P(\boldsymbol{S} \text{ occupied}) \\ &= P(\boldsymbol{T} \text{ binds } \boldsymbol{S} \mid \boldsymbol{S} \text{ free}) \times P(\boldsymbol{S} \text{ free}) \end{aligned} \quad (5)$$

where the second equality follows from the assumption that sites occupied by nucleosomes have zero probability of being accessed by TFs. Although a few TFs have been observed to bind nucleosomal DNA, our assumption is true for the vast majority of factors.

Taking into account nucleosome occupancy information, equation (4) becomes:

$$P(\boldsymbol{T} \text{ binds } \boldsymbol{X}) = 1 - \prod_{i}^{n-W+1} \left(1 - 1 / \left(1 + \frac{1}{[T]} \times \prod_{j=i}^{i+W-1} \frac{\phi_0(X_j)}{\phi(X_j, j-i+1)} \right) \times P(X_i \ldots X_{i+W-1} \text{ free}) \right) \quad (6)$$

where $P(X_i \ldots X_{i+W-1}$ free) is derived from the *in vivo* nucleosome positioning data of Lee et al. (Lee et al. 2007), as described above.

Given a DNA sequence, a PBM-derived motif, and the nucleosome occupancy information over that sequence, we use equation (6) to compute the probability that the TF binds that sequence, as shown in Figure 1A for the TF Gcn4 and the intergenic region iYER052C.

Analyzing data from a ChIP-chip experiment

Having defined the probability that a TF ***T*** binds a DNA sequence ***X***, we can use this probability to score every intergenic probe present on the microarrays used in (Harbison et al. 2004). For example, Figure 1B shows the probability of TF Gcn4 binding each yeast intergenic region. Next, for any particular ChIP-chip experiment we can define two sets of intergenic probes: the positive set (*i.e.*, the set of 'bound' probes), which contains all the probes with a p-value < 0.001, and the negative set (*i.e.*, the set of 'unbound' probes), which contains all the probes with a p-value > 0.5, as calculated by Harbison et al. (Harbison et al. 2004). We did not make any calls for the probes with intermediate p-values. Using these positive and negative sets from each ChIP-chip experiment, and the probabilities that TF ***T*** binds each of the probes, we compute the enrichment of the PBM-derived motif for TF ***T*** in the ChIP-chip data by an AUC value.

For each ChIP-chip experiment *TF_cond* we computed the AUC value of each of the 139 DNA

binding motifs derived from PBM data. The advantage of using PBM-derived motifs and not motifs derived from ChIP-chip data is that, since the PBM experiments were performed *in vitro*, the PBM-derived motifs reflect the direct DNA binding specificities of the TFs. We could also use additional motifs that have been reported in the literature, derived from small-scale experiments. However, such motifs are usually built from a small number of high affinity binding sites and may not correctly characterize medium or low affinity sites (Berger et al. 2006; Zhu et al. 2009).

Computing the statistical significance of AUC values

To assess whether the AUC value computed for a PBM-derived motif in a particular ChIP-chip data set is significant, we proceeded in three steps: 1) we randomly generated 1,000 motifs by permuting the nucleotides in each column of the initial motif; 2) for each random motif we computed its AUC value in the given ChIP-chip data set; and 3) we used the 1,000 AUC values to compute an empirical p-value for the AUC of the real motif. We chose this particular randomization procedure because it preserves the length and information content of the initial motif, both of which are important features in calculating an AUC value. We checked all the randomly generated motifs to make sure that our randomization procedure neither re-created the original motif nor any of the other motifs used in our analysis. Motif comparisons were performed using the inter-motif distance described by Harbison et al. (Harbison et al. 2004).

We consider an AUC value significant if it is at least 0.65 (*i.e.*, it explains the ChIP-chip data to some extent) and has an associated p-value ≤ 0.001 (*i.e.*, at most one of the 1,000 random motifs has an AUC value equal to or greater than the AUC value of the real motif).

II. Explaining the ChIP-chip data sets for which the motif of the profiled TF and additional motifs are significantly enriched

As mentioned in the main manuscript, in nine ChIP-chip experiments we noticed that the motifs of the significantly enriched TFs are not similar, although their AUC values are very close (Table 1, right column), suggesting that the enriched factors may be interacting or cooperating/competing in the profiled conditions. In seven of the nine cases we found independent experimental evidence reported in the literature for interaction or cooperation/competition between significantly enriched factors and the factors profiled in the ChIP-chip experiments

1. the significant enrichment of Mcm1 in the ChIP-chip experiments of Fkh2 profiled in hyperoxic conditions (Fkh2_H2O2Hi and Fkh2_H2O2Lo) can be explained by cooperation between the two factors: Fhk2 and Mcm1 cooperate to bind DNA at promoters of cell-cycle genes (Hollenhorst et al. 2001);
2. in the case of Sok2_BUT14, the significant enrichment of Phd1 is not surprising given that Sok2 negatively regulates pseudohyphal differentiation (Ward et al. 1995) and Phd1 is a transcriptional activator that enhances pseudohyphal growth (Gimeno and Fink 1994);
3. the enrichment of Tec1 in the Ste12_BUT90 data set is consistent with current knowledge about the physical interaction and cooperation between Ste12 and Tec1 in butanol-treated yeast cells (Chou et al. 2006);
4. the enrichment of Mcm1 in the Ste12_YPD data set can be explained by the physical interaction between these two factors (Bruhn and Sprague 1994);
5. in the case of Sum1_YPD we found experimental evidence reported in the literature that Sum1 and Ndt80 have overlapping, yet distinct sequence requirements for binding DNA and that they compete for binding to promoters containing the middle sporulation element (Pierce et al. 2003);
6. the fact that Mbp1 is enriched in the Swi4_YPD data set is not surprising given that Mbp1 and Swi4 have overlapping functions, both TFs being involved in cell cycle progression from G1 to S phase (Koch et al. 1993).

III. Comparison to the method of Liu et al. (2006)

Our method is similar to that presented by Liu et al. (Liu et al. 2006), in that both methods use nucleosome occupancy data in addition to DNA binding motifs to improve detection of *in vivo* TF-DNA interactions. However, there are major differences between the two methods. First, Liu et al. (Liu et al. 2006) analyze only one yeast TF, Leu3, and they choose this factor specifically because it binds DNA directly and it lacks known cofactors. By contrast, we applied our method to 139 TFs—all yeast TFs with a PBM-derived motif available (Badis et al. 2008; Zhu et al. 2009),—regardless of whether they bind DNA directly or indirectly. In fact, using our method we were able to detect indirect TF-DNA binding and *in vivo* TF-TF interactions. Second, Liu et al. (Liu et al. 2006) incorporated the nucleosome occupancy data by assuming an inhibitory effect of nucleosome occupancy (*i.e.*, by assuming that a twofold greater nucleosome occupancy inhibits TF-DNA binding by a fixed amount), and using a user-defined weight for this inhibitory effect. In their analyses, the authors either report results using different weights, or optimize the weight by choosing the value that shows the greatest improvement (in predicting TF binding) over not using nucleosome occupancy data. As the authors mention in their paper, the weight needs to be determined individually for each TF. Thus, it is not clear how one would compare the enrichment of motifs corresponding to different TFs. By contrast, our method does not require any parameters that are user-defined or that require training on the nucleosome data. The nucleosome dats is incorporated, in a principled way, in a probabilistic model of TF occupancy. The enrichment (AUC value) of binding motifs within ChIP data is computed in the same way for all TFs (the only difference being the TF binding motif), and the AUCs for different TFs are directly comparable.

IV. Comparison of results with versus without nucleosome occupancy

To compare the results of our method with versus without nuclesome occupancy data we empolyed two different strategies:

1. We divided the 237 ChIP-chip data sets into two categories: data sets in which a significantly enriched PBM-derived motif was found (Supplemental Table 4), and data sets in which no PBM-derived motif was significantly enriched (Supplemental Table 5). Next, for each data set we computed the maximum AUC over all 139 motifs, using our method both with and without nucleosome occupancy data. As mentioned in the main manuscript, in 60% of the ChIP-chip data sets in which a significantly enriched motif was found (Supplemental Table 4), the maximum AUC value is higher when nucleosome occupancy information is used as compared to when it is not used. For example, the AUC value for the Rap1 motif in the Rap1_YPD data set is 0.929 when using nucleosome data, and 0.895 when nucleosome occupancy data are not used. By contrast, in 71% of the data sets in which no motif was found to be significantly enriched (Supplemental Table 5), the maximum AUC value decreased when nucleosome occupancy data were used, which suggests that any observed motif enrichment may have been due to motif matches that are non-functional. In conclusion, using nucleosome occupancy data improves the results in at 60% of the data sets in which a significant motif was found and 71% of the data sets in which no significant motif was found.
2. Another way to compare the results of our method with versus without nucleosome data is to look at the number of ChIP-chip experiments for which the motif of the profiled TF was significantly enriched (*i.e.*, the data sets that are readily explained by direct DNA binding of the profiled factor). When using the nucleosome data, we found 61 such experiments, and for 7 of them the profiled TF did not appear significantly enriched when the analysis was done without the nucleosome data. In contrast, in only one ChIP-chip experiment we found the profiled TF significantly enriched *only* when nucleosome information was *not* used. This experiment was of Ste12 profiled after butanol treatment (Ste12_BUT14), an environmental condition were Ste12 has been shown to bind DNA indirectly, through Tec1 (Chou et al. 2006). Thus, even in this case including the nucleosome data gave better results.

A thorough analysis of the improvement obtained using nucleosome occupancy data would require complete knowledge about what TFs bind DNA directly in each ChIP-chip experiment, and additionally high-resolution nucleosome occupancy data in the same environmental conditions as the ChIP experiments, data that are not currently available.

V. Comparison of results with *in vivo* versus *in vitro* nucleosome occupancy

When nucleosome positioning is considered in the analysis, it would be preferable to use *in vivo* nucleosome positioning data rather than *in vitro* data or computational predictions of nucleosome occupancy, which may not accurately reflect *in vivo* nucleosome occupancy at functional TF binding sites. To show this, we repeated our analysis using *in vitro* nucleosome data (Kaplan et al. 2009). As expected, the overall results were slightly better than not using nucleosome data at all, but worse than the results using *in vivo* data. Furthermore, for a number of TFs the results were worse when using *in vitro* nucleosome data than no data at all. For example, in the case of Abf1, Rap1, and Reb1, factors that have been shown to remodel chromatin around their binding sites (Angermayr et al. 2003; Kaplan et al. 2009; Yarragudi et al. 2004), the AUC values are lower when using *in vitro* data (Abf1 AUC: 0.935; Rap1 AUC: 0.865; Reb1 AUCs: 0.840, 0.957, 0.916) than when not using nucleosome data (Abf1 AUC: 0.967; Rap1 AUC: 0.894; Reb1 AUCs: 0.852, 0.982, 0.952, respectively). Since the depletion around the binding sites of these TFs *in vivo* can be attributable to their own action, and not to the general properties of the DNA sequence, it is not surprising that for these TFs we get worse results using *in vitro* nucleosome data than no nucleosome data.

Furthermore, when looking at the number of ChIP-chip experiments for which the motif of the profiled TF was significantly enriched (*i.e.*, the data sets that are readily explained by direct DNA binding of the profiled factor), we found 61 such experiments when *in vivo* nucleosome data was used, compared to 54 when *in vitro* nucleosome data was used. In 6 of the 61 data sets the profiled TF did not appear significantly enriched when the analysis was done with *in vitro* nucleosome data. In contrast, in only one ChIP-chip experiment we found the profiled TF significantly enriched when *in vitro* nucleosome data was used, but not when *in vivo* nucleosome data was used. This experiment was of Ste12 profiled after butanol treatment (Ste12_BUT14), an environmental condition were Ste12 has been shown to bind DNA indirectly, through Tec1 (Chou et al. 2006). Thus, even in this case the *in vivo* nucleosome data gave better results.

VII. Analysis of enrichment and average nucleosome occupancy of Fhl1, Sfp1, and Rap1 8-mers in ChIP-chip data sets of the three factors

In this analysis, we used the 8-mer PBM enrichment scores (E-scores) directly to analyze the enrichment of high-affinity 8-mers (*i.e.*, 8-mers with E-scores above a range of thresholds, as described below) in various ChIP-chip data sets. Enrichment was measured by the p-value according to a hypergeometric distribution (Fisher's Exact test). Since the results of such an analysis are sensitive to the E-score threshold, we did not employ just one particular threshold, but instead we computed the p-value for 100 E-score thresholds equally spaced between -0.5 and 0.5. As shown in Supplemental Figure 4A, the pattern of enrichment of Rap1 8-mers in the Fhl1 ChIP-chip data sets is very similar to the Rap1 enrichment in Rap1_YPD, where Rap1 is known to bind DNA directly. Similarly, Supplemental Figure 4C shows that the Rap1 enrichment in Sfp1_SM follows the same trend as the Rap1 enrichment in Rap1_YPD, again suggesting that Rap1 is the factor that binds DNA directly in those ChIP-chip experiments.

We also analyzed the average probability of nucleosome occupancy (computed as described in above in the "Nucleosome positioning data" section) of Fhl1, Sfp1, and Rap1 8-mers with E-score above various thresholds. In Supplemental Figure 4B we show that: 1) high-affinity Rap1 8-mers are less occupied by nucleosomes (and thus more likely to be bound by TFs) as compared to high-affinity Fhl1 8-mers; and 2) the pattern of average nucleosome occupancy of Rap1 8-mers is very similar in Fhl1_RAPA, Fhl1_SM, Fhl1_YPD, and Rap1_YPD. Similar results were obtained for Sfp1 (see Supplemental Figure 4D).

Taken together, these results strengthen our predictions that Fhl1 and Sfp1 may bind DNA indirectly, in both cases through Rap1.

References

Angermayr, M., U. Oechsner, and W. Bandlow. 2003. Reb1p-dependent DNA bending effects nucleosome positioning and constitutive transcription at the yeast profilin promoter. *J Biol Chem* **278:** 17918-17926.

Badis, G., E.T. Chan, H. van Bakel, L. Pena-Castillo, D. Tillo, K. Tsui, C.D. Carlson, A.J. Gossett, M.J. Hasinoff, C.L. Warren, M. Gebbia, S. Talukder, A. Yang, S. Mnaimneh, D. Terterov, D. Coburn, A. Li Yeo, Z.X. Yeo, N.D. Clarke, J.D. Lieb, A.Z. Ansari, C. Nislow, and T.R. Hughes. 2008. A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. *Mol Cell* **32:** 878-887.

Berger, M.F., A.A. Philippakis, A.M. Qureshi, F.S. He, P.W. Estep, 3rd, and M.L. Bulyk. 2006. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol.* **24:** 1429-1435.

Bruhn, L. and G.F. Sprague, Jr. 1994. MCM1 point mutants deficient in expression of alpha-specific genes: residues important for interaction with alpha 1. *Mol Cell Biol* **14:** 2534-2544.

Chou, S., S. Lane, and H. Liu. 2006. Regulation of mating and filamentation genes by two distinct Ste12 complexes in Saccharomyces cerevisiae. *Mol Cell Biol* **26:** 4794-4805.

Foat, B.C., A.V. Morozov, and H.J. Bussemaker. 2006. Statistical mechanical modeling of genome-wide transcription factor occupancy data by MatrixREDUCE. *Bioinformatics* **22:** e141-149.

Gimeno, C.J. and G.R. Fink. 1994. Induction of pseudohyphal growth by overexpression of PHD1, a Saccharomyces cerevisiae gene related to transcriptional regulators of fungal development. *Mol Cell Biol* **14:** 2100-2112.

Granek, J.A. and N.D. Clarke. 2005. Explicit equilibrium modeling of transcription-factor binding and gene regulation. *Genome Biol* **6:** R87.

Harbison, C.T., D.B. Gordon, T.I. Lee, N.J. Rinaldi, K.D. Macisaac, T.W. Danford, N.M. Hannett, J.B. Tagne, D.B. Reynolds, J. Yoo, E.G. Jennings, J. Zeitlinger, D.K. Pokholok, M. Kellis, P.A. Rolfe, K.T. Takusagawa, E.S. Lander, D.K. Gifford, E. Fraenkel, and R.A. Young. 2004. Transcriptional regulatory code of a eukaryotic genome. *Nature* **431:** 99-104.

Hollenhorst, P.C., G. Pietz, and C.A. Fox. 2001. Mechanisms controlling differential promoter-occupancy by the yeast forkhead proteins Fkh1p and Fkh2p: implications for regulating the cell cycle and differentiation. *Genes Dev* **15:** 2445-2456.

Kaplan, N., I.K. Moore, Y. Fondufe-Mittendorf, A.J. Gossett, D. Tillo, Y. Field, E.M. LeProust, T.R. Hughes, J.D. Lieb, J. Widom, and E. Segal. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458:** 362-366.

Koch, C., T. Moll, M. Neuberg, H. Ahorn, and K. Nasmyth. 1993. A role for the transcription factors Mbp1 and Swi4 in progression from G1 to S phase. *Science* **261:** 1551-1557.

Kullback, S. 1959. *Information theory and statistics*. John Wiley and Sons, Inc., New York.

Lee, W., D. Tillo, N. Bray, R.H. Morse, R.W. Davis, T.R. Hughes, and C. Nislow. 2007. A high-resolution atlas of nucleosome occupancy in yeast. *Nat Genet* **39:** 1235-1244.

Pierce, M., K.R. Benjamin, S.P. Montano, M.M. Georgiadis, E. Winter, and A.K. Vershon. 2003. Sum1 and Ndt80 proteins compete for binding to middle sporulation element sequences that control meiotic gene expression. *Mol Cell Biol* **23:** 4814-4825.

Roider, H.G., A. Kanhere, T. Manke, and M. Vingron. 2007. Predicting transcription factor affinities to DNA from a biophysical model. *Bioinformatics* **23:** 134-141.

Ward, M.P., C.J. Gimeno, G.R. Fink, and S. Garrett. 1995. SOK2 may regulate cyclic AMP-dependent protein kinase-stimulated growth and pseudohyphal development by repressing transcription. *Mol Cell Biol* **15:** 6854-6863.

Yarragudi, A., T. Miyake, R. Li, and R.H. Morse. 2004. Comparison of ABF1 and RAP1 in chromatin opening and transactivator potentiation in the budding yeast Saccharomyces cerevisiae. *Mol Cell Biol* **24:** 9152-9164.

Yassour, M., T. Kaplan, A. Jaimovich, and N. Friedman. 2008. Nucleosome positioning from tiling microarray data. *Bioinformatics* **24:** i139-146.

Zhu, C., K. Byers, R. McCord, Z. Shi, M. Berger, D. Newburger, K. Saulrieta, Z. Smith, M. Shah, M. Radhakrishnan, A. Philippakis, Y. Hu, F. De Masi, M. Pacek, A. Rolfs, T. Murthy, J. LaBaer, and M. Bulyk. 2009. High-resolution DNA binding specificity analysis of yeast transcription factors. *Genome Research* **19:** 556-566.