doi:10.1038/nature07521

LETTERS

Analysis of combinatorial *cis*-regulation in synthetic and genomic promoters

Jason Gertz[1], Eric D. Siggia[2] & Barak A. Cohen[1]

Transcription factor binding sites are being discovered at a rapid pace[1,2]. It is now necessary to turn attention towards understanding how these sites work in combination to influence gene expression. Quantitative models that accurately predict gene expression from promoter sequence[3–5] will be a crucial part of solving this problem. Here we present such a model, based on the analysis of synthetic promoter libraries in yeast (*Saccharomyces cerevisiae*). Thermodynamic models based only on the equilibrium binding of transcription factors to DNA and to each other captured a large fraction of the variation in expression in every library. Thermodynamic analysis of these libraries uncovered several phenomena in our system, including cooperativity and the effects of weak binding sites. When applied to the *S. cerevisiae* genome, a model of repression by Mig1 (which was trained on synthetic promoters) predicts a number of Mig1-regulated genes that lack significant Mig1-binding sites in their promoters. The success of the thermodynamic approach suggests that the information encoded by combinations of *cis*-regulatory sites is interpreted primarily through simple protein–DNA and protein–protein interactions, with complicated biochemical reactions—such as nucleosome modifications—being downstream events. Quantitative analyses of synthetic promoter libraries will be an important tool in unravelling the rules underlying combinatorial *cis*-regulation.

Thermodynamic models of gene regulation have shown promising results in eukaryotic systems[6,7] when applied to small gene sets. Owing to limitations in studying genomic promoters, the number of observations in these studies is small compared to the number of molecular events that are modelled, and over-fitting is therefore a serious concern. An approach that circumvents this limitation is to model the expression of synthetic promoters[8–10]. As conceivably any promoter sequence can be created and analysed, a large portion of possible regulatory element combinations can be evaluated.

We constructed synthetic promoter libraries consisting of random combinations of three to four transcription factor binding sites (TFBS), or building blocks (Table 1 and Supplementary Information). In total, we analysed 2,807 promoters among 7 libraries using 18 different building blocks. All promoters were placed upstream of a medium strength basal promoter driving yellow fluorescent protein (YFP; Supplementary Fig. 1) and integrated into the yeast genome at the *TRP1* locus. The level of gene expression directed by each synthetic promoter was quantified by flow cytometry of 25,000 individual cells per promoter (Fig. 1a, b).

Figure 1c shows the expression levels of 429 synthetic promoters from the L1 library (see Supplementary Tables 1–7 for expression and sequence of all promoters). Basal promoter only controls (Fig. 1c, shown in red) were used to estimate the technical variance of our expression measurements, which is 1.3% of the total variance of the L1 library; the average technical variance for all libraries is 0.8% of the total variance. The biological replicate variance, which refers to the gene expression differences between independent transformants that have the same synthetic promoter by chance, is 35% of the total variance in the L1 library and 17% on average. Therefore, a perfect model relating promoter sequence to our expression data would explain 65% of the variance in expression driven by the different promoters in the L1 library.

We constructed a thermodynamic model of the relationship between promoter sequence and expression. The purpose of the model was to provide a formal mathematical framework for predicting the activity of novel combinations of *cis*-regulatory sites, and to gain insight into the mechanisms that generate diverse expression levels from different arrangements of the same *cis*-regulatory sites. We used a model first proposed in ref. 11, and later modified in ref. 12. The main assumption of this model is that gene regulation is controlled completely by the equilibrium binding of proteins to DNA and to each other. Enzymatic events, such as chromatin modifications and polymerase phosphorylation, are not taken into account. The model consists of parameters that describe the changes in free energy of particular DNA–protein and protein–protein interactions that can occur on the promoters. These parameters are used to calculate the probability of RNA polymerase (RNAP) being bound to each promoter in the library (see Supplementary Information). We then assume that the probability of RNAP being bound to a given promoter is directly proportional to the intensity of YFP fluorescence measured for that promoter.

In every library, thermodynamic models explained 44–59% of the variance in expression (Table 1), which is more than double the amount of variance explained by the best models of genome-wide expression data[4,5]. The thermodynamic model for the L1 library

Table 1 | Summary of synthetic promoter libraries

Library	Building blocks	Number of promoters	Number of parameters fitted	Fraction of variance explained (R^2)
L1	Mig1, Gcr1, Spacer	429	5	0.49
L1-test	Same as L1	83	0	0.44
L1-weak	Same as L1 plus a weak Mig1 site	266	1	0.44
L2	Mig1, Reb1, Rap1, Gcr1 (different from L1)	471	4	0.59
L3	Adr1, Hap2/3/4/5, CSRE, Rgt1	596	6	0.47
L4	Cbf1, Gcn4, Met31/32, Nrg1	381	10	0.54
L5	Msn2/4, Smp1, Xpb1	581	4	0.57

[1]Center for Genome Sciences, Department of Genetics, Washington University in Saint Louis School of Medicine, 4444 Forest Park Avenue, St Louis, Missouri 63108, USA. [2]Center for Studies in Physics and Biology, The Rockefeller University, New York, New York 10021, USA.

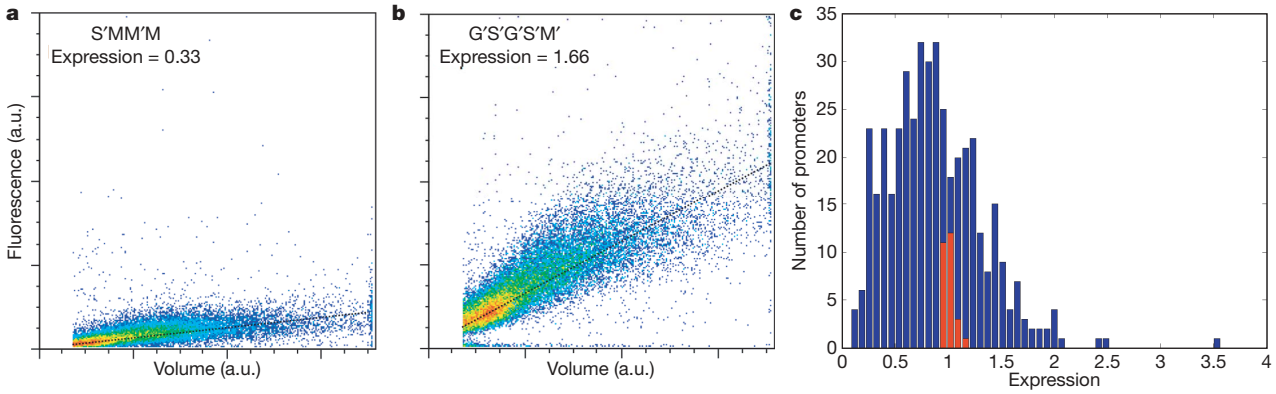

Figure 1 | Gene expression measurements. **a**, **b**, Graphs of cell volume versus fluorescence for 25,000 individual cells containing the promoters S′MM′M (**a**) and G′S′G′S′M′ (**b**). Here S = Spacer, G = Gcr1 site, M = Mig1 site and prime indicates a site in the reverse orientation. **c**, Histogram of expression values for all L1 library members. Expression values were computed as the average fluorescence/volume ratio for 25,000 individual cells, and then normalized to plate controls. Control promoters with no library insert are shown in red. a.u., arbitrary units.

captured 49% of the variance in expression (Supplementary Fig. 2; 75% of the available variance). The overall success of the thermodynamic approach indicates that expression driven by combinations of binding sites can be generally and accurately modelled by simply considering protein–DNA and protein–protein binding events.

To determine the predictive power of our model for the L1 library, we constructed the L1-test library, which consists of novel combinations of the L1 building blocks. With the same parameter values from the L1 library, the model still captures 44% of the variance in expression, implying that the model is not over-fitted. This lack of overfitting is not surprising, considering that each model contains about 6 parameters fitted to an average of over 400 promoters. The Mig1 parameter values found in the L1 library were held constant among thermodynamic models for three other libraries (L1-test, L1-weak and L2) that all exhibited high predictive power (see Supplementary Table 8 for all parameter values). Our model for the L1 library predicts that the 'Spacer' building block, which we designed to contain no known or predicted regulatory sequence elements, can recruit RNAP to promoters. As about half of the DNA-binding proteins in yeast do not yet have an associated *cis*-regulatory motif[1], it is likely that the Spacer site is actually an unidentified *cis*-regulatory element. The ability of the model to incorporate an unknown sequence element and accurately predict its behaviour points to a strength of the approach.

Analysis of the model for the L1 library suggests that Mig1 binds cooperatively to the synthetic promoters. Because nothing in the previous literature suggested cooperativity between Mig1 monomers, we decided to analyse Mig1 cooperativity independent of the model. We fitted a Hill equation relating percentage repression to the number of Mig1 sites, with the assumption that 100% repression occurs with five Mig1-binding sites. We found that a Hill coefficient of 3.4 ± 0.25 and $K = 1.8$ (where K is the number of Mig1 sites that causes half maximal repression) gives the best fit, suggesting cooperativity. Figure 2a shows that the observed data fit well to the Hill equation and that without cooperativity the fit is substantially worse. These results are consistent with the thermodynamic model and suggest that Mig1 acts cooperatively to repress transcription in our system, which led us to examine the influence of low affinity TFBS on expression.

Low affinity, or weak, TFBS are known to play important roles in prokaryotic promoters[13] and have been postulated to be important in eukaryotic gene regulation[14]. However, their quantitative effect on gene expression is difficult to determine. To study the effects of weak TFBS, we constructed a library (L1-weak) incorporating a building block matching a Mig1-binding site that was shown to have low affinity for Mig1 *in vitro*[15]. The sequence of this weak site scores below any reasonable cut-off in a genome scan for Mig1 sites based on a weight matrix derived from known Mig1 sites[16,17]. In our system, the low affinity Mig1 site behaved as a weaker repressor than the strong site (Fig. 2b). However, when there are strong Mig1 sites present in a promoter, the weak sites behave as strong sites. When comparing promoters with the same building block content except for the

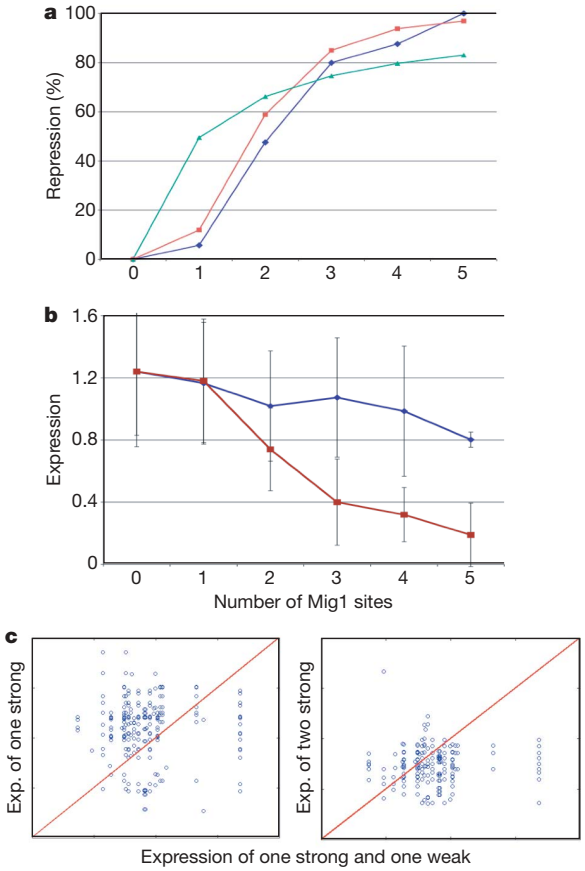

Figure 2 | Mig1-binding sites act cooperatively, and a weak Mig1 site represses weakly. **a**, Hill equation with a Hill coefficient of 3.4 and $K = 1.8$ (red) fits the observed data (blue) well, compared to a Hill equation with a Hill coefficient of 1 (green). **b**, Plot of average expression versus the number of weak sites (blue) without strong sites, and versus the number of strong sites (red) without weak sites. Error bars, ± 1 s.d. **c**, Plots of expression for pairs of promoters that are almost identical except that either one strong Mig1 site or two strong Mig1 sites replace one strong and one weak Mig1 site. A blue circle represents one promoter pair and the red line represents equal expression.

number of Mig1 sites, promoters with one weak and one strong Mig1 site exhibit lower expression compared to promoters with one strong Mig1 site (Fig. 2c; $P < 10^{-8}$, sign test, $n = 211$) and the same expression as promoters with two strong Mig1 sites (Fig. 2c; $P > 10^{-2}$, sign test, $n = 177$). This behaviour suggests that strong and weak Mig1 sites interact cooperatively to repress transcription in our system. This interaction produces complex patterns of expression in the L1-weak library.

The thermodynamic model of transcriptional regulation accurately captures many of the complexities of expression in the L1-weak library by adding only one adjustable parameter to the L1 library model parameters—namely, the relative affinity of Mig1 for the weak site. The optimal value of the new parameter corresponds to a 6.7-fold lower relative affinity for Mig1 than the stronger Mig1 site. This value is in good agreement (and within a 95% confidence interval) with an independent computational analysis of a position specific weight matrix for Mig1, which predicted a 9.0-fold lower affinity of the weak site for Mig1 (refs 16, 18). The similarity of the R^2 of this model with that of the L1 library model demonstrates that we are capturing the additional complexities caused by the effect of weak Mig1 sites on expression.

We next examined the possibility that weak sites contribute to Mig1 repression of genomic promoters. Weak Mig1-binding sites are over-represented in *S. cerevisiae* promoters compared to shuffled *S. cerevisiae* promoters ($P < 10^{-3}$, simulation, $n = 1{,}000$; Supplementary Fig. 3). Weak sites are found in 24% of all promoters, while 39% of promoters containing a significant match to a Mig1 weight matrix also contain a weak site ($P < 10^{-12}$, hypergeometric test), indicating that strong and weak sites tend to co-occur. Of 33 genes that are known to be regulated by Mig1 (refs 19, 20), and whose promoters contain a significant match to a Mig1 weight matrix, 20 also contain a weaker Mig1 site in their promoters compared to 8 genes expected by chance. According to our model of gene regulation, promoters with one strong and one weak site are more sensitive to changes in Mig1 concentrations than are promoters with either two strong or two weak sites; promoters with one strong and one weak site therefore may be best suited to respond to changes in available carbon sources (Supplementary Fig. 4). These results suggest that combinations of strong and weak Mig1-binding sites are commonly found together in genomic promoters, and may provide a sensitive strategy for glucose repression.

We sought to determine if the properties of Mig1 repression found in the synthetic promoter libraries were informative when studying genomic promoters. In the *S. cerevisiae* genome, 359 promoters have a significant match to a Mig1 weight matrix and 33 of these promoters correspond to one of 136 documented Mig1-regulated genes. To compare these results directly to our model, we applied the thermodynamic model of Mig1 repression to genomic promoters (see Methods). Out of the top 359 promoters ranked by the thermodynamic model for the strength of Mig1 repression, 41 correspond to one of the 136 documented Mig1-regulated genes. Using the regulatory rules encoded in our thermodynamic model, we explain eight (24%) more known Mig1-regulated genes (*HXT9*, *HXT12*, *HXT13*, *GSY1*, *SOR1*, *ICS2*, *YIL172C*, *YOL153C*) than by simply looking for promoters with a significant match to a Mig1 weight matrix. For example, the *SOR1* promoter does not harbour a significant match to a Mig1 site but contains a number of weak sites that cluster together (Fig. 3a). As cooperativity between Mig1 sites is an important part of our quantitative model, we correctly predicted that *SOR1* is Mig1 regulated and also identified the likely binding sites of Mig1 in this promoter.

Using the thermodynamic model, we also predicted a number of Mig1-regulated genes that were not previously known to be Mig1 targets (Supplementary Table 9). *MIG2*, a paralogue of *MIG1* that represses and binds the same site as Mig1 (ref. 15), was predicted by the model to be auto-regulated on the basis of its promoter sequence (Fig. 3b). To validate this prediction, we measured *MIG2* promoter activity (see Methods) in strains deleted for both *MIG1* and *MIG2*. *MIG2* promoter activity increased significantly in the *mig1Δ mig2Δ* strain as compared to wild type ($P < 10^{-3}$, *t*-test, $n = 24$), showing that *MIG2* is auto-regulated by Mig1/Mig2 (Fig. 3c). The prediction from the model was that *MIG2* expression would increase 1.8-fold in a *mig1Δ mig2Δ* strain, and we observed a 1.5-fold change. The regulation of *MIG2* by Mig1/Mig2 represents a previously unreported negative feedback loop in the glucose repression network that was identified on the basis of our analysis of synthetic promoters.

Using a simple system, we have succeeded in constructing an accurate model of the relationship between promoter sequence and gene expression. In part this was because we sampled a much larger fraction of promoter space using our library than we could by sampling genomic promoters. Thus, we were able to fit models containing a small number of parameters to data containing large numbers of observations. We found that a completely thermodynamic model based on the equilibrium binding of the transcription factors and RNAP to each other, and to their *cis*-regulatory sites, was a reasonable way to capture the relationship between promoter sequence and gene expression in our system for all of the libraries examined. This does not imply that kinetic processes, such as histone or RNAP modification, are unimportant in gene regulation; however, it does suggest that the information encoded in a promoter is decoded primarily by the sequence-specific binding of transcription factors. Our results support the idea that the complexity and variation in gene regulation could stem from very simple rules describing the binding of proteins to DNA and to each other[12,13,21].

METHODS SUMMARY

To create the building blocks that make up the synthetic promoters, oligonucleotide pairs (each with a 5′ phosphate) were annealed by being boiled and then slowly cooled to room temperature (see Supplementary Information for building block sequences). 15 μl of 50 μM double-stranded building blocks were then ligated with 200 U of T4 DNA ligase (New England Biolabs) for 2 hours at 16 °C. The ligation products were then purified using a Microcon YM-100 column (Millipore) to reduce the number of short promoters. 15 ng of purified ligation

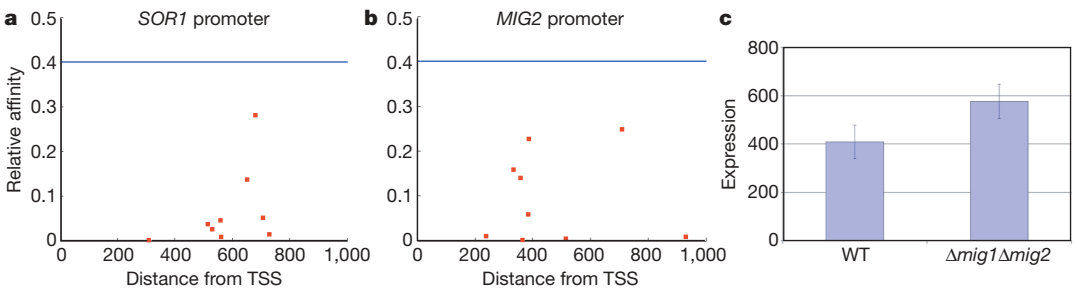

Figure 3 | Thermodynamic model explains Mig1 repression in the genome. **a**, **b**, Mig1-binding sites in the promoters of *SOR1* (**a**) and *MIG2* (**b**). The affinity of Mig1 for the site based on a position weight matrix score relative to the strong site is plotted versus the location upstream of the translation start site (TSS). The horizontal line represents the significance threshold for the weight matrix and each square represents a Mig1 site. **c**, *MIG2* promoter activity in a wild-type (WT) strain and a *mig1Δ mig2Δ* strain. Error bars, ±1 s.d.

product were then ligated into the BamH1 site of the integrating reporter plasmid pJG102 (20 ng) and transformed into *Escherichia coli*. Transformants were scraped into Luria broth plus carbenicillin, grown overnight and then DNA was extracted using the GenElute HP Plasmid Maxiprep kit (Sigma). 130 μg of library DNA was digested with BglI, BamH1, Sal1 and EcoR1 (200 U each) and transformed into yeast as described[22]. Colonies growing on medium lacking uracil were picked into 96-well plates and Trp^- colonies were then identified by replica plating onto medium lacking tryptophan. We observed that some building blocks were represented slightly more than others, even though they were added at equal molar concentrations. The relative abundance of each building block in each library scaled similarly to the melting temperature of the building block.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 July; accepted 1 October 2008.
Published online 23 November 2008.

1. Harbison, C. T. *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **431,** 99–104 (2004).
2. Hu, Z., Killion, P. J. & Iyer, V. R. Genetic reconstruction of a functional transcriptional regulatory network. *Nature Genet.* **39,** 683–687 (2007).
3. Beer, M. A. & Tavazoie, S. Predicting gene expression from sequence. *Cell* **117,** 185–198 (2004).
4. Bussemaker, H. J., Li, H. & Siggia, E. D. Regulatory element detection using correlation with expression. *Nature Genet.* **27,** 167–171 (2001).
5. Das, D., Banerjee, N. & Zhang, M. Q. Interacting models of cooperative gene regulation. *Proc. Natl Acad. Sci. USA* **101,** 16234–16239 (2004).
6. Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U. & Gaul, U. Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature* **451,** 535–540 (2008).
7. Zinzen, R. P., Senger, K., Levine, M. & Papatsenko, D. Computational models for neurogenic gene expression in the *Drosophila* embryo. *Curr. Biol.* **16,** 1358–1365 (2006).
8. Murphy, K. F., Balazsi, G. & Collins, J. J. Combinatorial promoter design for engineering noisy gene expression. *Proc. Natl Acad. Sci. USA* **104,** 12726–12731 (2007).
9. Ligr, M., Siddharthan, R., Cross, F. R. & Siggia, E. D. Gene expression from random libraries of yeast promoters. *Genetics* **172,** 2113–2122 (2006).
10. Cox, R. S. III, Surette, M. G. & Elowitz, M. B. Programming gene expression with combinatorial promoters. *Mol. Syst. Biol.* **3,** 145 (2007).
11. Shea, M. A. & Ackers, G. K. The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation. *J. Mol. Biol.* **181,** 211–230 (1985).
12. Buchler, N. E., Gerland, U. & Hwa, T. On schemes of combinatorial transcription logic. *Proc. Natl Acad. Sci. USA* **100,** 5136–5141 (2003).
13. Ptashne, M. & Gann, A. *Genes and Signals* (Cold Spring Harbor Laboratory Press, 2002).
14. Tanay, A. Extensive low-affinity transcriptional interactions in the yeast genome. *Genome Res.* **16,** 962–972 (2006).
15. Lutfiyya, L. L. *et al.* Characterization of three related glucose repressors and genes they regulate in *Saccharomyces cerevisiae*. *Genetics* **150,** 1377–1391 (1998).
16. Hertz, G. Z. & Stormo, G. D. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* **15,** 563–577 (1999).
17. Matys, V. *et al.* TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* **31,** 374–378 (2003).
18. Nehlin, J. O. & Ronne, H. Yeast MIG1 repressor is related to the mammalian early growth response and Wilms' tumour finger proteins. *EMBO J.* **9,** 2891–2898 (1990).
19. Monteiro, P. T. *et al.* YEASTRACT-DISCOVERER: new tools to improve the analysis of transcriptional regulatory associations in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **36,** D132–D136 (2008).
20. Teixeira, M. C. *et al.* The YEASTRACT database: a tool for the analysis of transcription regulatory associations in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **34,** D446–D451 (2006).
21. Ptashne, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386,** 569–577 (1997).
22. Gietz, R. D. & Schiestl, R. H. Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols* **2,** 38–41 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Mitra, G. Stormo, M. Johnston, K. Varley, S. Doniger and members of the Cohen laboratory for discussions and suggestions, and J. Sabina for technical help with LacZ experiments. B.A.C. and J.G. were supported by the NIH (R01 GM078222) and E.D.S. was supported by the NSF (DMR0129848). J.G. was also supported by an NSF Graduate Research Fellowship (DGE-0202737).

Author Contributions J.G. performed all experiments and analyses. B.A.C. and J.G. designed the experiments and wrote the paper. E.D.S. conceived the idea of applying the thermodynamic model to the synthetic promoter libraries, and contributed to its development.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to B.A.C. (cohen@genetics.wustl.edu).

METHODS

Strains and plasmids. The strain harbouring the synthetic promoter library was derived from the haploid strain BY4742 (*MAT**a** his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) as described[23]. The library of promoters was constructed in plasmid pJG102. pJG102 was created by amplifying the *TSA1* promoter from genomic DNA using the following primers: forward, 5′-CGCGGATCCGGCTCGGGTTGGCAAAGTCGGC-3′; reverse, 5′-CGCGGATCCACTAGTGTGTATGTATGTGTATGTAGTTGTGG-3′. The amplification product was digested with BamH1 and ligated into the BglII and BamH1 sites of pJG100 (GenBank accession number FJ175377) to create pJG102. To measure *MIG2* promoter activity we used the plasmid pBM4346 described in ref. 24 and the strains BY4743 and YM6682 described in refs 23 and 24 respectively, which were provided by M. Johnston (Washington Univ.).

Expression analysis. Cultures were grown to log phase in 2 ml 96-well plates in 500 μl of synthetic complete media with 2% glucose and lacking uracil. The fluorescence intensities and electronic volumes of 25,000 events from each well were measured on a Beckman Coulter Cell Lab Quanta SC with a multi-plate loader. For each well, the mean of fluorescence divided by electronic volume for 25,000 events was taken as the expression value for that well. On each plate the expression value of the four no-insert controls were averaged to calculate a plate effect to account for changes in laser intensity or growth differences. Each expression value on the plate was then divided by the plate effect.

Sequencing. Synthetic promoters were sequenced directly from library members. First colony-PCR was used to amplify the synthetic promoter using the following primers: forward, 5′-GAGTTGATGAATCTCGGTGTG-3′; reverse, 5′-GAACTGGCAATTTACCAGTAGTAC-3′. The PCR reactions were treated with ExoI and shrimp alkaline phosphatase. The PCR products were then sequenced using the forward primer with Big Dye mix v3.1 from Applied Biosystems. Promoter sequences were called only if sequence matched both ends of the PCR product and the number of building blocks found by WU-BLAST[25] matched the expected number of building blocks based on the length of the read.

Thermodynamic model. For details about the thermodynamic framework and parameter fitting, see Supplementary Information.

Genome analysis of thermodynamic model. We used a thermodynamic model of Mig1 repression on genomic promoters. The promoter sequences used were obtained from ref. 26. First each sequence was scanned using Patser[27] with an experimentally derived Mig1 position weight matrix[28], with a cut-off of 2, which is well below the significance cut-off of 7.11. The relative affinity of the weak and strong Mig1 sites predicted by the model was consistent with the difference in weight matrix scores between the weak site and the strong site. We therefore calculated the natural log of relative affinities (q in Supplementary Information) for each site found by Patser by subtracting 8.01 (the score of the strong site) from the score calculated by Patser. Cooperativity between Mig1 sites in the model was only between Mig1 proteins bound next to each other without any interfering factors. As it is impossible to know what other transcription factors are bound at every promoter, we made the simplifying assumption that Mig1 proteins can only interact with each other when the sites are within 40 base pairs. The thermodynamic model only involving Mig1 sites was then run using the Mig1 parameters described in the manuscript.

***MIG2* expression analysis.** Strains carrying the plasmid containing the *MIG2* promoter upstream of LacZ were grown to mid-log phase in synthetic complete media with 2% glucose and lacking leucine. The cells were disrupted with YPER (Pierce) and then β-galactosidase was monitored using Novagen's Betafluor β-Galactosidase Assay Kit measured on a BioTek Synergy HT. The velocity of the reaction was divided by the OD_{600} of each sample to calculate expression. Six replicates were performed for each strain.

23. Brachmann, C. B. *et al.* Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14,** 115–132 (1998).
24. Kaniak, A., Xue, Z., Macool, D., Kim, J. H. & Johnston, M. Regulatory network connecting two glucose signal transduction pathways in *Saccharomyces cerevisiae*. *Eukaryot. Cell* **3,** 221–231 (2004).
25. Gish, W. WU BLAST. ⟨http://blast.wustl.edu⟩ (1995-, 2008).
26. Cliften, P. *et al.* Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301,** 71–76 (2003).
27. Hertz, G. Z. & Stormo, G. D. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* **15,** 563–577 (1999).
28. Nehlin, J. O. & Ronne, H. Yeast MIG1 repressor is related to the mammalian early growth response and Wilms' tumour finger proteins. *EMBO J.* **9,** 2891–2898 (1990).