

An Ensemble Model of Competitive Multi-factor Binding of the Genome

Todd Wasson¹, Alexander J. Hartemink²

1 Introduction

Hundreds of different factors adorn the eukaryotic genome, binding to it in large number. These DNA binding factors (DBFs) include nucleosomes, transcription factors (TFs), and other proteins and protein complexes, such as the origin recognition complex (ORC). DBFs compete with one another for binding along the genome, yet current models of genome binding do not consider different types of DBFs together simultaneously. Additionally, binding is a stochastic process that results in a continuum of binding probabilities at any position along the genome, but current models tend to consider positions as being either binding sites or not.

2 Results

We present a model that allows a multitude of DBFs, each at different concentrations, to compete with one another for binding sites along the genome. The result is an ‘occupancy profile’, a probabilistic description of the DNA occupancy of each factor at each position. We implement our model efficiently as the software package COMPETE. We demonstrate how modeling nucleosome binding alters TF binding, and vice versa, and illustrate how factor concentration influences binding occupancy. Binding cooperativity between nearby TFs arises implicitly in our model via mutual competition with nucleosomes. Our method applies not only to TFs, but also recapitulates known occupancy profiles of a well-studied replication origin with and without ORC binding. Importantly, the sequence preferences our model takes as input are derived from *in vitro* experiments. This ensures that the calculated occupancy profiles are the result of the forces of competition represented explicitly in our model and the inherent sequence affinities of the constituent DBFs.

In Figure 1, we present a specific example in yeast showing how the binding of TFs can displace nucleosomes from the genome. Depicted is the region surrounding the *GTR1* promoter on Chromosome XIII. In the top panel is the occupancy profile in which nucleosomes compete for positions along the genome without any other DBF competitors (as in the models of [4] and [2]). Nucleosome positions at the left are somewhat ambiguous, but become more stably positioned to the right. Five to six reasonably stably positioned nucleosomes can be seen, starting near the start codon of *YML122C*. In contrast, the bottom panel reveals how nucleosomes are positioned once the binding of Pho4, Pho2, and Nrg1 is taken into account. The binding of Pho4, Pho2, and Nrg1 to sites in the *GTR1* promoter displaces the two nucleosomes that would otherwise reside there. This binding also helps to stably position a nucleosome near the 3’ end of the coding region of *YML122C*, and hems in a single nucleosome near the 5’ end. The strong binding of the TFs form boundaries that prevent nucleosomes from translating upstream or downstream; instead, overall nucleosome occupancy in the region drops dramatically, illustrating the importance of modeling direct competition between nucleosomes and TFs.

¹Program in Computational Biology and Bioinformatics, Institute for Genome Sciences & Policy, Duke University, Durham, NC 27708-0090, USA. E-mail: todd.wasson@duke.edu

²Department of Computer Science, Box 90129, Duke University, Durham, NC 27708-0129, USA. E-mail: amink@cs.duke.edu

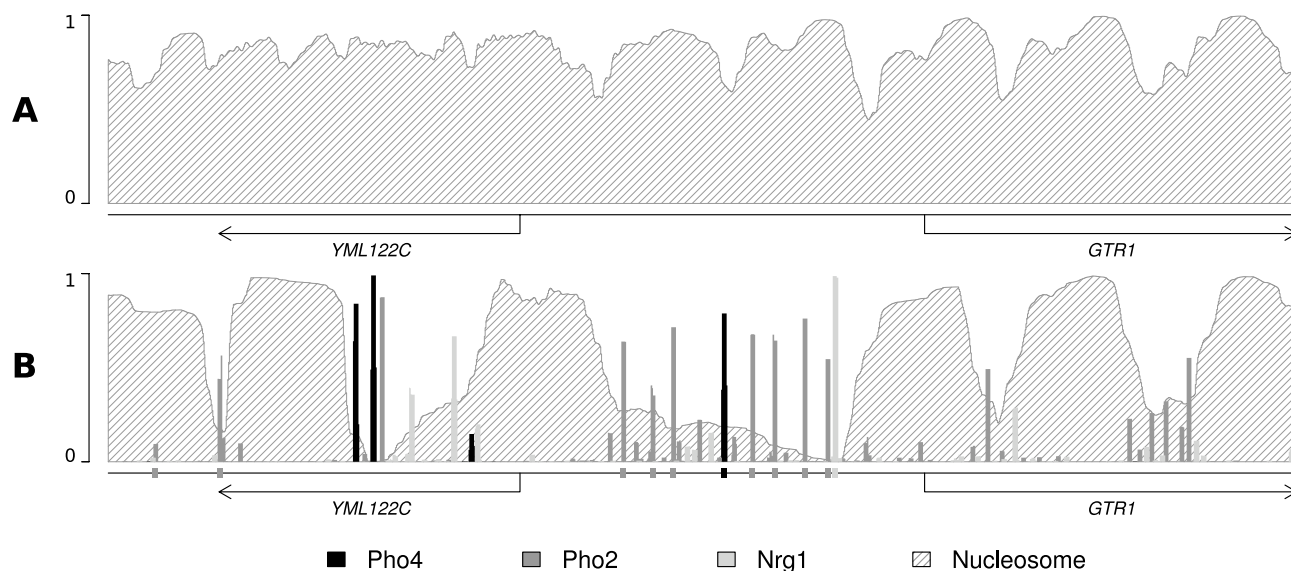


Figure 1: Transcription factor binding can displace nucleosomes. Shown are DBF occupancy profiles. X-axis is genome position, y-axis is probability of occupancy. Plots of each shade of gray indicate the probability of the respective DBF occupying each position. The black line underneath each plot represents the genome; oriented arrows thereon demarcate coding regions of genes according to SGD [1]; shaded boxes thereon denote strong transcription factor binding site matches according to [3]. The promoter region of *GTR1* is known to contain high-affinity binding sites of Pho4 and Nrg1, as well as several of Pho2. (A) An occupancy profile including only nucleosomes depicts five to six reasonably well-positioned nucleosomes beginning atop the start codon of *YML122C*. (B) Addition of Pho4, Pho2, and Nrg1 to the model results in the displacement of several of the previously bound nucleosomes, and the reduction of binding frequency of those remaining, to various extents. Additionally, a well-positioned nucleosome is established near the 3' end of *YML122C*. The strong TF binding in this region precludes nucleosome repositioning upstream or downstream along the sequence; the strongly-bound TFs displace some nucleosomes and hem in the remaining others, resulting in their well-defined positions.

References

- [1] J. M. Cherry, C. Adler, C. Ball, S. A. Chervitz, S. S. Dwight, E. T. Hester, Y. Jia, G. Juvik, T. Roe, M. Schroeder, S. Weng, and D. Botstein. 1998. SGD: Saccharomyces Genome Database. *Nucleic Acids Res.* 26:73–79.
- [2] N. Kaplan, 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. 2008. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* Online Publication.
- [3] K.D. MacIsaac, T. Wang, D.B. Gordon, D.K. Gifford, G.D. Stormo, and E. Fraenkel. 2006. An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinformatics* 7:113.
- [4] E. Segal, Y. Fondufe-Mittendorf, L. Chen, A. Thåström, Y. Field, I.K. Moore, J.P. Wang, and J. Widom. 2006. A genomic code for nucleosome positioning. *Nature* 442:772–778.