# Supporting Material - Nonspecific transcription factor-DNA binding influences nucleosome occupancy in yeast

Ariel Afek[1#], Itamar Sela[1#], Noa Musa-Lempel[2], and David B. Lukatsky[1*],
[1]*Department of Chemistry, Ben-Gurion University of the Negev, Beer-Sheva 84105 Israel*
[2]*Department of Computer Science, Ben-Gurion University of the Negev, Beer-Sheva 84105 Israel*

[*] Corresponding author: lukatsky@bgu.ac.il
[#] These authors contributed equally to the work

## Materials and Methods

### Free energies of nonspecific TF-DNA binding genome-wide

We used the following procedure in order to assign the magnitude of the nonspecific TF-DNA binding free energy at each genomic location, $x_0$, within the interval $(-384;380)$ around the TSS for each of 5,014 yeast genes, distributed into four clusters. First, for a given gene, we position the middle of the sliding window of the width $L = 50$ bp at $x_0$. Second, for each TF (out of 256 randomly generated TFs) we compute the partition function, Eq. (2), and the free energy, $F = -k_B T \ln Z$, in a given sliding window. In order to get rid of the compositional bias in different genomic locations, we always compute the difference, $\Delta F = F - F_\infty$, where $F_\infty$ is computed for randomized sequence of the same width, $L$, in the same sliding window, and averaged over 50 random realizations, for a given TF. The randomization procedure permutes the positions of nucleotides, preserving the average nucleotide composition in the sliding window. Third, we repeat the latter procedure for each of 256 TFs, and average the free energies of all TFs at the genomic location, $x_0$, of this gene, $\langle \Delta F \rangle_{TF}$. Fourth, we move the sliding window to the next genomic location, $x_0 + \Delta x$, and repeat the entire procedure in order to obtain $\langle \Delta F \rangle_{TF}$ in this new location. In our analysis we computed $\langle \Delta F \rangle_{TF}$ in steps of 4 bp, $\Delta x = 4$. The contact energies of TF-DNA interactions are generated for each of 256 TFs, as described in the main text. We used $\sigma_\alpha = 2k_B T$ for the standard deviation of $P(K_\alpha)$, Eq. (1). The obtained free energy landscape is very weakly sensitive to the change of the sliding window width, $L$ (data not shown). We used, quite arbitrary, $M = 8$, for the length of each TF. Our conclusions are qualitatively robust for a wide range of $M$ (data not shown).

### Cumulative correlation functions genome-wide

We used the following procedure in order to compute the correlation function, $C_{\alpha\alpha}(x)$, Eq. (3), for each of four nucleotide types, $\alpha$. For each cluster of genes, aligned with respect to the TSS, we compute $C_{\alpha\alpha}(x)$ in a sliding window of width $L = 50$ bp, with the middle of the sliding window being positioned at $x_0$. In order to compute $\langle s^r_{\alpha\alpha}(x) \rangle$, we randomly permute the sequences in a given window, preserving the average nucleotide composition of each sequence. We used 50 random realizations in order to compute the average, $\langle s^r_{\alpha\alpha}(x) \rangle$. Next, we compute the cumulative correlation function, $\mathrm{H}_{\alpha\alpha}$, Eq. (4), at this location $x_0$, where the summation is performed up to a cutoff, $x_{\max} = 4$ bp. A different choice of the cutoff leads to a rescaling of the base level of $\mathrm{H}_{\alpha\alpha}$, without affecting the trend (data not shown). We then repeat the entire procedure for the next sliding window, $x_0 + \Delta x$, thus assigning the value of $\mathrm{H}_{\alpha\alpha}$ at each genomic location around the TSS, for a given cluster of genes. The obtained cumulative correlation functions, $\mathrm{H}_{\alpha\alpha}$, are very weakly sensitive to the change of the sliding window width, $L$ (data not shown).

### Transcriptional plasticity and TFBSs

In order to compute **Figure 6.A** and **B**, we used the classification of transcriptional plasticity from Refs. [1, 2]. In order to compute **Figure 6.C** and **D**, we used the TFBSs data from Ref. [3].

## References


1. Tirosh, I., N. Barkai. 2008. Two strategies for gene regulation by promoter nucleosomes. *Genome Res.* 18(7):1084-1091.
2. Ihmels J, Friedlander G, Bergmann S, Sarig O, Ziv Y, Barkai N. 2002. Revealing modular organization in the yeast transcriptional network. *Nat Genet*. 31(4):370-377.
3. MacIsaac KD, Wang T, Gordon DB, Gifford DK, Stormo GD, Fraenkel E. 2006. An improved map of conserved regulatory sites for Saccharomyces cerevisiae. *BMC Bioinformatics*. 7:113.

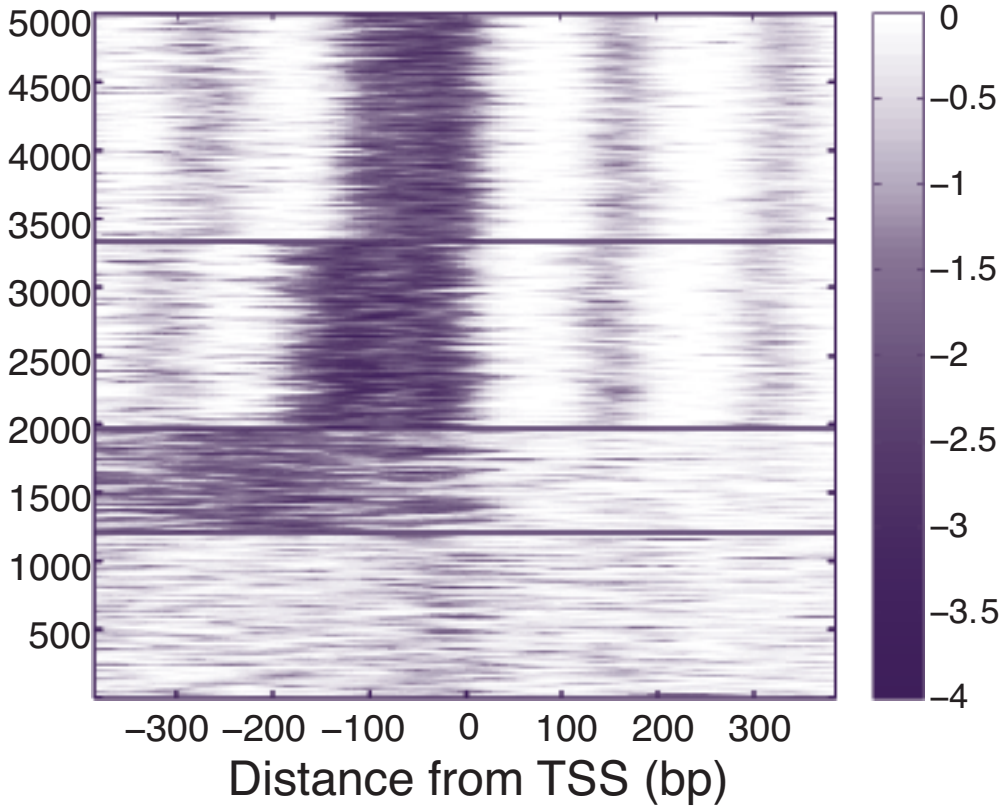


**Figure S1.** Graphical representation of the yeast nucleosome occupancy data, which we used for our analysis (see Figure 3 of Lee, W. *et al.* 2007. *Nat Genet*. 39(10):1235-1244). Average nucleosome occupancy, covering ~800 bp surrounding the TSS for 5,014 transcripts. The transcripts are clustered into four clusters according to the GO classification. Each cluster contains (from bottom to top): 1,211 stress response genes (cluster 0); 766 translation genes (cluster 1); 1,374 ribosome biogenesis and assembly genes (cluster 2); 1,663 organelle organization and biogenesis genes (cluster 3).