

Functional transcription factor binding motifs have a strong location bias towards the transcription start sites of human genes.

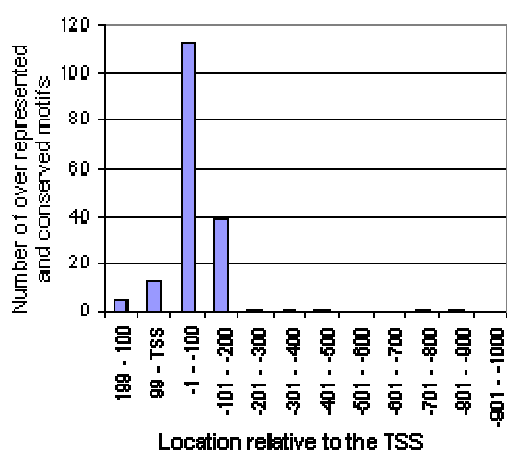
Yuval Tabach^{1,2}, Ran Brosh², Anat Reiner¹, Yossi Buganim², Or Zuk¹, Assif Yitzhaky¹, Varda Rotter² and Eytan Domany¹

¹Departments of Physics of Complex Systems and ²Molecular Cell Biology, The Weizmann Institute of Science, Rehovot 76100, Israel

Elucidating basic principles that underlie regulation of gene expression by transcription factors (TFs) is a central challenge of the post-genomic era. Transcription factors regulate expression by binding to specific DNA sequences; such a binding event is functional when it affects gene expression. Functionality of a binding site is reflected in conservation of the binding sequence during evolution, and is governed by several parameters, such as the binding strength, dictated mainly by the DNA sequence; distance of the binding site from the transcription start site (TSS) and DNA packing. Understanding how these parameters control functionality of different TFs in different biological contexts is an essential step towards identifying functional TF binding sites, a must for understanding regulation of transcription.

We introduce a novel method to search for functional TF binding motifs. The score of every putative binding sequence of a TF is compared to a threshold, whose value is optimized independently for every GO class, and for each distance window from the TSS of the target genes, taking into account also the location dependent nucleotide heterogeneity of the target genes' promoter. When used together with conservation between species, the method identifies with high probability functional binding sites that regulate groups of biologically related genes. We performed a high throughput analysis, searching the promoters (from 200bp downstream to 1000bp upstream the TSS), of more than 8000 genes from 134 functional Gene Ontology classes, for all known binding motifs, in human and mouse. We found that 86% of the identified functional binding motifs are located in the first 200bp upstream from the TSS (Figure 1). Our findings were put to several experimental tests.

By allowing a "flexible" threshold and combining our functional class specific search method with conservation between human and mouse, we are able to identify reliably functional TF binding motifs. This is an essential step towards constructing regulatory networks and elucidating the design principles that govern transcriptional regulation of expression. Surprisingly, the promoter region proximal to the TSS has quite similar importance for regulation of transcription in human/mouse as in bacteria and yeast.



Distribution, by location windows, of transcription factor binding motifs that are significantly over-represented in different functional Gene Ontology classes and conserved (in human and mouse).