

 GENE EXPRESSION

Complex interactions

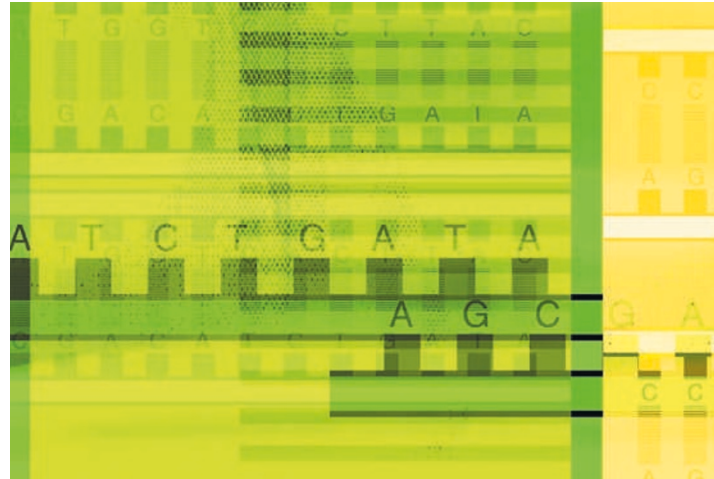
Transcription factors (TFs) and their DNA-binding sites are crucial components of gene regulatory networks that control myriad cellular processes. However, the DNA-binding specificities of only a few TFs have been sufficiently characterized to enable the prediction of the sequences that they can and cannot bind to. A new study in *Science* reports the DNA-binding specificities of 104 TFs and reveals unexpected complexity and diversity in DNA recognition.

Using protein-binding microarrays (PBMs) with all possible DNA sequence variants of a given length, Bulyk and colleagues determined the DNA-binding specificities of 104 known and predicted mouse TFs, which represent 22 structural classes of DNA-binding domains found in metazoa. For each TF, an algorithm was used to identify the single octamer that has the highest PBM enrichment score (the highest binding affinity). Next, they systematically tested the relative preference of each nucleotide variant at

each position in the binding sequence and obtained a rank-ordered list of the binding preferences of each TF for every k -mer (in which k is “the number of informative nucleotide positions in the binding site”).

Virtually every TF has a unique binding sequence preference and even proteins that have high amino acid sequence identity have distinct DNA-binding profiles. Notably, almost half of the TFs analysed recognize multiple distinct sequence motifs, and these ‘secondary motifs’ are bound by most TFs almost as well as the primary motifs. These secondary motifs are occupied by specific TFs *in vivo*.

This previously unknown widespread existence of secondary motifs has important implications for understanding how proteins interact with their DNA-binding sites. Indeed, further analysis of one category of secondary motifs suggests that some TFs recognize their DNA-binding sites through many completely different interaction modes.



“ Virtually every TF has a unique binding sequence preference... ”

Although it remains to be tested whether the binding of the same TF to distinct sequence motifs leads to different physiological responses, the authors argue that “this complexity in DNA recognition may be important in gene regulation and in evolution of transcriptional networks”.

Francesca Cesari

ORIGINAL RESEARCH PAPER Badis, G. et al. Diversity and complexity in DNA recognition by transcription factors. *Science* 14 May 2009 (doi:10.1126/science.1162327)