

# High-resolution DNA binding specificity profiles of transcription factors and cis regulatory codes in DNA

**Martha L. Bulyk**

*Medicine, Pathology, and Health Sciences & Technology, Brigham & Women's Hospital and Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA*

**T**RANSSCRIPTION factors (TFs) regulate the expression of genes through sequence-specific interactions with DNA-binding sites. Using universal protein binding microarrays (PBMs), we determined high-resolution binding profiles for 89 known and predicted *S. cerevisiae* TFs, over more than 2.3 million gapped and ungapped 8-bp sequences ('k-mers'). We reported 50 new or significantly different direct DNA-binding site motifs for yeast DNA-binding proteins and motifs for eight proteins for which only a consensus sequence was previously known; in total, this corresponds to over a 50% increase in the number of yeast DNA-binding proteins with experimentally determined DNA-binding specificities. Among other novel regulators, we discovered proteins that bind the PAC (Polymerase A and C) motif (GATGAG) and regulate ribosomal RNA (rRNA) transcription and processing, core cellular processes that are constituent to ribosome biogenesis. In contrast to earlier data types, these comprehensive k-mer binding data permit us to consider the regulatory potential of genomic sequence at the individual word level. These k-mer data allowed us to re-annotate in vivo TF binding targets as direct or indirect and to examine TFs' potential effects on gene expression in ~1700 environmental and cellular conditions. These approaches could be adapted to identify TFs and cis regulatory elements in other organisms.