

MICROARRAYS

A plan for when the ChIPs are down

A new microarray-based strategy offers a tool for the rapid and sensitive identification of binding sites throughout the genome for virtually any transcription factor.

Current estimates suggest that there are just over 200 DNA-binding transcription factors in yeast, and although much effort has gone into identifying the binding sites for these proteins, scientists are still far from a comprehensive genome-wide map of transcription factor binding sites for yeast or any other eukaryotic organism.

The combination of chromatin immunoprecipitation (ChIP) with DNA microarray studies—so-called ‘ChIP-chips’—has enabled researchers to identify a variety of specific transcription factor binding sites, but ChIP-chips are not without their limitations. Immunoprecipitation results can vary, and the process requires specific antibodies for each transcription factor. More importantly, each experiment provides only a snapshot of a factor’s DNA-binding state at a particular point in time or under specific conditions, limiting the ability to conduct generalized investigation of genome-wide binding.

Brigham and Women’s Hospital and Harvard Medical School investigator Martha Bulyk was interested in developing a more broadly applicable technique for binding site identification that surpassed some of the limitations of ChIP-chips. Her group, in collaboration with investigators at Yale and MIT, came up with an alternative strategy, protein binding microarrays (PBMs), which they demonstrate in a new study from *Nature Genetics* (Mukherjee *et al.*, 2004).

First, the transcription factor of interest is expressed as an epitope-tag fusion. This tag is used to purify the protein, which is then applied to a DNA microarray. In this particular study, the microarray contained an essentially comprehensive collection of intergenic sequences from the genome of *Saccharomyces cerevisiae*. Detection is achieved by fluorescently labeled antibodies targeted against the epitope tag. To normalize the resulting data, the relative amount of double-stranded

(ds) DNA is determined for each microarray spot on a duplicate chip with the dye SybrGreen; the comparison of fluorescence from each dye at each spot enables the identification of significant protein-DNA interactions (Fig. 1).

To test their PBM strategy, Bulyk’s team selected three yeast transcription factors whose binding had previously been characterized in ChIP-chip studies (Lieb *et al.*, 2001; Lee *et al.*, 2002): Abf1, Rap1 and Mig1. The group identified 189, 294 and 79 target sites for each protein, respectively, which were then subjected to computational analysis to determine the recognition motifs.

These sequences were compared against those identified in the ChIP-chip studies and in the TRANSFAC eukaryotic transcription factor database. Generally, the PBM motifs closely resembled the TRANSFAC and ChIP-chip sequences, and PBM also identified a considerable number of new putative binding targets. Gel shift experiments confirmed several of these sites and in at least one case reinforced the presence of a Rap1 site predicted by PBM but not by TRANSFAC.

Bulyk suggests that her group’s use of a particularly high standard for statistical significance may have somewhat restricted the number of sites identified. “[With] our cutoff—which was a very conservative one—we were seeing a false-positive rate of around 7–9%. If we’re a bit less conservative and pick a less strict cutoff, then we do see more sites coming up, but the false-positive rate increases a little bit. It’s really a matter of what you want to tune for—here we weren’t tuning so much for sensitivity as for specificity.”

Additional analysis reinforced the specificity of Bulyk’s approach. A comparison of genomic data from *S. cerevisiae* against four related yeast species, showed that sites identi-

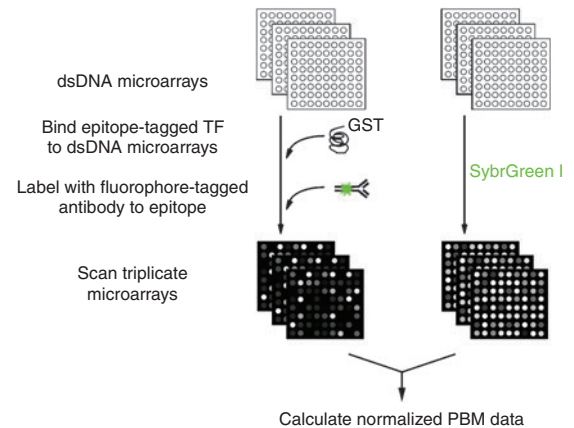


Figure 1 | An overview of the PBM process. Reprinted with permission from Nature Genetics.

fied by PBM in general were at least as likely to be closely conserved as sites identified by ChIP-chip. Furthermore, a number of the sites identified only by PBM were 100% conserved across all five yeast species, increasing the likelihood of their relevance.

Bulyk’s team is already planning broader studies in yeast, but is also looking to investigate higher eukaryotes. “We’re currently expanding this to look at human transcription factors,” she says, “[and] part of our lab is interested in predicting *cis*-regulatory modules... where you’re getting binding by a number of different transcription factors. So the better your information about what the binding specificities are, the more accurately, we think, you’ll be able to predict where the *cis*-regulatory modules are. And so that’s one of the ways we’re hoping to use this data.”

Michael Eisenstein

RESEARCH PAPERS

Mukherjee, S. *et al.* Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nat. Genet.* **36**, 1331–1339 (2004).

Lieb, J.D. *et al.* Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nat. Genet.* **28**, 327–334 (2001).

Lee, T. *et al.* Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799–804 (2002).