

SEQUENCING

High-resolution chromatin immunoprecipitation

Exonuclease digestion of immunoprecipitated chromatin sharpens protein binding peaks and improves signal.

Where do proteins bind in the genome? This question has occupied countless researchers hoping to unravel the mechanisms governing gene regulation. Most rely on chromatin immunoprecipitation (ChIP) to find where proteins physically associate with genomic DNA, but it is a famously noisy technique. Rhee and Pugh at Pennsylvania State University have come up with a solution that whittles away at protein binding sites to give a remarkably precise and clean signal.

The ChIP principle is simple. Protein-DNA interactions are stabilized by cross-linking, chromatin is sheared, and binding to a specific protein is selectively enriched by immunoprecipitation. DNA in the enriched fraction consists of overlapping fragments, yielding a distribution around any given binding site that can be detected by microarray hybridization or deep sequencing.

One limitation of ChIP is the necessity of shearing, which generates heterogeneously sized fragments that limit resolution. In addition, a substantial amount of unbound DNA is trapped in the precipitate and generates nonspecific signal. Statistical methods are needed to locate individual sites and to determine binding motifs, often with high uncertainty.

To get around this problem, Rhee and Pugh revisit the old concept of the nuclease protection assay. They add exonuclease to immunoprecipitated DNA while it is bound to resin, protecting it at protein-bound sites. The idea had not been applied to ChIP because of a conceptual block. “If you use nucleases to trim off all the excess, you end up chewing up almost all the DNA; just a little piece is protected and that’s not enough sequence to uniquely identify the fragment,” explains Pugh. “That’s where we all got stuck.”

The key to the technique, which they dub ChIP-exo, is the use of lambda exonuclease, which only trims DNA in the 5′-to-3′ direction on each strand and stops when it encounters protein. Everything downstream of the binding site and on the same

strand remains intact, so fragments are long enough to be identified. Each digested strand terminates on one side of the binding site, and examining both reveals the entire contact region. The result is that ChIP-exo sequencing gives extremely high resolution, with a standard deviation under one base for some proteins, almost two orders of magnitude below traditional ChIP-sequencing and hardly requiring statistics. “This reduces your search space almost down to nothing, to the binding element itself,” says Pugh.

The method has the added benefit of cutting noise. Exonuclease digestion removes all unprotected fragments including contaminating DNA. Noise is an obvious source of false positives, which may be as high as 30% in many standard ChIP samples. Pugh explains that an additional problem with noise is that it requires stringent filtering. “A vast majority of peaks are real bound locations that fall below the threshold, so you just throw them out,” he says. The authors uncovered 17,000 new sites for the human transcriptional repressor CTCF and low-occupancy sites for several yeast proteins.

Another implication is that enriched sequences do not need to be averaged over the genome to power statistical motif detection. “Proteins bind to degenerate sequences, making it very hard to pick out the individual motif in any one peak in standard ChIP,” notes Pugh. Their ChIP-exo results with yeast transcription factor Reb1 show that the same protein can bind different site variants based on genomic context (within or outside telomeres) to carry out unique functions.

Fine resolution enables another exciting application. Blocked sites can have a one-to-one relationship with the structure of the protein bound to DNA, allowing careful interpretation of the interaction. ChIP-exo adds steps to the ChIP protocol, but “if you’re looking for the entire enchilada,” quips Pugh, “then you need something with more resolution.”

Tal Nawy

RESEARCH PAPERS

Rhee, H.S. & Pugh, B.F. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408–1419 (2011).