Citations

ChIPing report

Although chromatin immunoprecipitation (ChIP) is a key technique for epigenomic studies, it is limited by the antibodies it depends on. If a protein is not known to be involved in epigenetic regulation, its significance could be missed entirely. Alternatively, antibodies may be unavailable or cross-react with other targets. This can be a particular problem for antibodies directed against posttranslationally modified histones, for example singly versus doubly acetylated forms. An unbiased detection method that would sidestep these issues is introduced in a publication from Byrum et al. in *Cell Reports*. The ChIP alternative is called ChAP-MS, for chromatin affinity purification with mass spectrometry. Working in yeast, Byrum et al. engineered a LexA DNA binding site directly upstream of the transcription start site for the GAL1 gene. This same strain was designed to constitutively express a LexA-protein A fusion. The LexA portion directs the fusion to the GAL1 locus, and the protein A component enables capture by IgG-coated beads. The ChAP-MS procedure involves formaldehyde crosslinking, cell lysis, sonication, affinity purification, gel electrophoresis of eluants, trypsinization, and then high-resolution MS. Although the affinity purification can be done under stringent conditions (1 M NaCl and 1 M urea), contamination by non-specifically

associated proteins remains a preeminent concern. To minimize this noise, the authors used an isotopic labeling scheme: experimental samples were grown in isotopically light medium, while a control strain (protein fusion but no LexA DNA binding site) was grown in heavy medium. If the two samples are mixed, specifically interacting proteins should be 100% light while contaminants would be a 50/50 mix of light and heavy. Using this strategy, Byrum et al. were able to isolate proteins associated with a ~1,000-base pair fragment of chromatin. Since the GAL1 locus is already well characterized, the authors could confirm that many of the proteins they isolated jibed with the previous literature. However, ChAP-MS also picked up new factors and histone posttranslational modifications. Importantly, the method simultaneously characterizes multiple factors associating with a given region instead of the one-at-a-time testing characteristic of ChIP, which means this yeast-based method should yield a wealth of unbiased information about epigenetic control of chromatin function.

Byrum et al. 2012. ChAP-MS: A method for identification of proteins and histone posttranslational modifications at a single genomic locus. *Cell Rep* 2:198-205.

Double or nothing

The capabilities of next-generation DNA sequencing technologies have grown so rapidly that it seems harsh to focus on its limitations. However, sequencing errors persist. When a sequence variant is very rare, distinguishing a true mutation from an artefact is tricky: the background error rate, estimated to be 1%, is sufficient to complicate deep sequencing studies in metagenomics, paleogenomics, and clinical testing. Previous studies have reported workflows that reduce the level of technical error twentyfold. However, this is still many times higher than the actual mutation frequency in normal cells. Schmitt et al., writing in *Proceedings of the National Academy of Sciences*, describe a strategy that employs the informational redundancy of double-stranded DNA in order to detect and discard sequencing artefacts. The method uses a modified library preparation phase. Duplex Tags, which contain a 12-base randomized sequence, are incorporated into the standard sequencing adapters that are then ligated to the sheared target DNA. Each fragment to be sequenced has two tags, and reads that share the same two tags in the same order are derived from the same original DNA strand. For a read to be retained, the sequence must be shared by a set threshold of members of this tag family. This step eliminates misreads because of amplification mistakes or random errors

The two-step method for distinguishing true mutants from technical errors in massively parallel sequencing. Source: *PNAS*

in sequencing calls. In the second error-correcting step, the consensus sequences from the two families sharing the same two tags are compared. Only those variants that show up in both (that is, in the consensus read from each original DNA strand) are accepted. Using this methodology, the authors calculate a technical error would arise just once for every 10^9 bases sequenced. The power of this low error rate is evident in a test involving human mitochondrial DNA.

Standard sequencing suggested a mutation frequency on the order of 10^{-3} . Sequences that passed the family consensus sequence quality control step gave an estimate approximately tenfold lower. However, checking the reads from each strand of the original double-stranded sequence against each other gave a mutation frequency of 3.5×10^{-5} , a value lining up with population genetics studies. Although it is based on a simple premise, this duplex sequencing methodology provides a detailed, validated workflow that dramatically enhances the sensitivity of deep sequencing, unlocking its full potential for genomics, forensics, and medicine.

Schmitt et al. Detection of ultra-rare mutations by next-generation sequencing. *Proc Natl Acad Sci USA*. [Epub ahead of print, August 1, 2012; doi:10.1073/pnas.1208715109].

Selected and written by Nijse Dorman, Ph.D.

BioTechniques 53:207 (October 2012)
doi 10.2144/000113931

To purchase reprints of this article, contact:
biotechniques@fosterprinting.com