Transcription factor binding sites in space and time: ChEC-seq

November 15, 2015

L Koch

In ChEC-seq, a chromatin protein of interest is genetically fused to micrococcal nuclease (MNase) and expressed in cells. Living cells are permeabilized and calcium is added to activate the MNase, leading to cleavage of DNA in proximity to binding sites for the protein. Total DNA is then purified and sequenced.

Image provided by Dr. GE Zentner.

Transcription factors bind to distinct locations in the genome and promote the expression of genes by recruiting RNA polymerase. Traditionally, transcription factor binding sites have been identified by chromatin immunoprecipitation (ChIP), which uses an antibody against the transcription factor to pull out the associated DNA for sequencing. The DNA must be fragmented, either by digestion with micrococcal nuclease (MNase) or by sonication. Thus, in order to keep the transcription factor bound to the DNA before the immunoprecipitation step, protein is chemically cross-linked to DNA before the digestion, usually using formaldehyde. Formaldehyde preferentially generates protein-protein crosslinks and so can repeatedly capture fortuitous interactions, such that they appear biologically meaningful. Technically, ChIP can be challenging due to high background caused by the crosslinking, low quality antibodies, as well as poor solubility of the DNA binding protein. Finally, traditional ChIP methods create a snapshot of the binding that happened during the crosslinking incubation, which is extremely useful but does not provide kinetic information about protein binding.

In order to remedy these issues, Dr. Gabe Zentner, previously in the laboratory of Dr. Steven Henikoff (Basic Sciences) and now at his own lab at Indiana University, developed a method that could characterize transcription factor binding sites on a genome-wide scale without crosslinking or an antibody and could also provide kinetic information. Their method, termed ChEC-seq, was recently published in *Nature Communications*. In chromatin endogenous cleavage (ChEC), a DNA-binding protein of interest is fused to micrococcal nuclease (MNase), living cells are gently permeabilized, and then cleavage of the surrounding sequence, not protected by the protein, is induced by adding a high concentration of calcium. Next, total DNA is extracted and fragments are selected by size for sequencing. By using ChEC in combination with high-throughput sequencing (ChEC-seq), the authors were able to identify many more binding sites for the transcription factors they tested than had been reported by ChIP methods. Furthermore, because the cleavage of the surrounding sequence by the MNase depends on calcium addition, the authors were able to determine which sites were bound and the surround sequence cleaved quickly or slowly, terming these two classes of binding sites "fast" and "slow."

They wondered whether the differing binding kinetics of the two classes of sites reflected differences in the underlying sequence. Indeed, their analysis showed that sites in the fast class had strong matches to the known consensus motifs of the tested transcription factors. Conversely, the slow sites had lower motif scores. Importantly, the slow class of sites did not represent locations in the genome that were more open to transcription factor binding in general. In comparing datasets for each of the three major transcription factors tested, the authors found that the majority of the slow sites were unique to each transcription factor. This suggests that the slow sites are preferred by a given protein and do not reflect accessibility biases.

Recently, it has been suggested that the shape of the DNA can drive transcription factor binding in addition to sequence. The authors wondered if the sites in the slow binding class were enriched for certain shapes of DNA and whether that could account for their binding more than whether they matched the consensus sequence. They compared shape features of the DNA such as minor groove width, roll, propeller twist and helix twist between the fast and slow class of sites and found that the shape features were highly similar between the two. This result was consistent even when the authors ranked each site by motif match to account for similar sequences in each group. Finally, the authors compared the abilities of either a sequence or shape model to differentiate the two classes using L2-regularized linear regression. Their results strongly indicate that transcription factors recognize regular DNA shape patterns overall but that they only form long-lasting, stable interactions at sites with strong consensus motifs.

This work has illuminated key details about how transcription factors find their binding sites, which is an important part of understanding one of the first steps of normal and pathological biological processes. Many genes are turned on in response to specific cellular conditions, how do transcription factors find the right targets so that the cell can respond? This has been a difficult question to answer without a way to distinguish binding over time on a genome-wide level.

ChEC-seq has promise in distinguishing how other proteins interact with DNA over time. Indeed, "ChEC-seq is a very simple method and, because it doesn't rely on immunoprecipitation, gets around many of the issues associated with ChIP," said Dr. Zentner. "The major hurdle is generating an MNase fusion protein, but with the rapid adoption of CRISPR genome editing, I think it's likely that generating MNase fusions for ChEC-seq in non-yeast systems will become routine."