Efficient mapping of genome-wide regulatory elements for biological insights

Shao-shan Carol Huang[1]
[1]*Joe Ecker Lab, Salk Institute*

We developed a high-throughput sequencing assay for rapid transcription factor binding site (TFBS) discovery, DNA affinity purification sequencing (DAP-seq), that uses *in vitro* prepared transcription factors (TFs) to capture native genomic DNA. We applied DAP- seq to 1,812 *Arabidopsis thaliana* TFs to resolve motifs for 529 factors and genome-wide enrichment maps for 349 factors. Cumulatively, the ~2.7 million experimentally- determined TFBSs captured the Arabidopsis cistrome and predicted thousands of TF target genes enriched for known and novel functions. Notably, DAP-seq target genes for many well-characterized hormone related TFs were enriched for Gene Ontology terms consistent with their known functions. Comparison of DAP-seq and ChIP-seq datasets showed that DAP-seq peaks predicted *in vivo* TF binding better than motif inference, potentially due to the ability of the assay to directly capture the impact of primary sequence and DNA methylation on binding affinities at individual TFBS. As a demonstration of the importance of genomic context, we showed that closely spaced motifs significantly affected TF binding by developing a model for cooperative auxin response factor (ARF) homodimer binding to complex motif repeats. Overall, DAP-seq enables rapid development of base-resolution cistrome atlases for a wide-array of applications for eukaryotic genomes.