Genome-wide tiling arrays can be used to measure DNA or RNA hybridization at high-resolution across the genome. These are commonly used to identify protein-DNA binding sites (via chromatin immunoprecipitation [ChIP]), to map transcription start sites and expression levels, and to measure genomic copy number. However, methods for analyzing these data are still in their infancy.

We describe two procedures that effectively model ChIP-chip (protein-DNA binding localization) and transcript data respectively, collected on such tiling arrays. The first method is capable of deconvolving multiple, overlapping protein-binding locations in ChIP-chip data to a higher resolution than that of the tiling itself. It integrates a model for the enriched DNA fragments with a linear model of binding occurrences, and uses modern statistical regression with parameter selection to choose the most significant binding sites. The second method integrates a sequence-based RNA hybridization model with a regression tree-based segmentation model to describe simultaneously the effects of probe sequence on hybridization, and the “step-wise” signal of transcript binding across the genome. This method effectively reduces the small-scale noise in the data and thereby enables detection of transcripts and their start/stop sites with higher accuracy.

We have integrated both of these methods to perform a detailed analysis of transcriptional activity and regulatory dynamics in the archaeon Halobacterium. We show that they perform better at identifying TF binding sites and transcript locations and levels, than do previously described methods developed for these tasks.