



**TITLE:** Integrating quantitative information from  
ChIP-chip Experiments into motif finding

**SPEAKER:** Sunduz Keles



**TIME:** 4:00 P.M.

**DATE:** Wednesday, February 14, 2007

**ROOM:** 140 BARDEEN

**ABSTRACT:**

Identifying binding locations of transcription factors within long segments of non-coding DNA is a challenging task. Recent ChIP-chip experiments utilizing tiling arrays are especially promising for this task since they provide high resolution genome-wide maps of the interactions between the transcription factors and DNA. Data from these experiments are invaluable for characterizing DNA recognition profiles (regulatory motifs) of transcription factors. A two step paradigm is commonly used for performing motif searches based on ChIP-chip data. First, candidate bound sequences that are in the order of 500-1000 base pairs are identified from ChIP-chip data. Then, motif searches are performed among these sequences. These two steps are typically carried out in a disconnected fashion in the sense that the quantitative nature of the ChIP-chip information is discarded in the second step. More specifically, all bound regions are assumed to equally likely have the motif(s) and the motifs are assumed to reside at any position of the bound regions with equal probability. We develop a flexible conditional two component mixture model (CTCM) that relaxes both of these common assumptions by adaptively incorporating ChIP-chip information. We discuss the challenges in fitting of these class of models. The performance of the proposed and existing methods are compared using simulated data and ChIP-chip data from recently available ENCODE studies. These studies indicate that CTCM efficiently utilizes the information available in the ChIP-chip experiments and has superior sensitivity and specificity especially when the motif of interest has low abundance among the ChIP-chip bound regions and/or has low information content.