Biology Research Software for Sequence Specificity of Nucleotide Binding Factors Based on Genomic Data

Technology #m07-094

“Lead Inventor: Harmen J. Bussemaker, Ph.D.

STV Reference: M06-084, M07-094

Understanding Transcription Factors in Gene Regulation Can Lead to Therapies: The use of high-throughput data in biology research is becoming increasingly ubiquitous. As data that describes gene expression (microarrays), DNA binding (ChIP-chip, CHIP-seq), SNP profiles, DNA sequence, etc. are more available, there is a need for computational models to analyze and convert these findings into useful biological insights. In particular, gene regulation has become a problem of high interest due to its potential impact on therapeutic strategy and design; e.g. a more comprehensive understanding of regulatory pathways and gene interactions within the cell can yield to novel targets for pharmacological intervention. However, gene regulation is a complex process, especially within a mammalian context, where combinatorial regulation, chromatin modification and expansive regulatory regions (enhancers, silencers, etc.) are all present. A recurrent area of focus within this context is to predict binding motifs of various transcription factors (TF) that physically bind to genomic regions and either activate or repress the transcriptional machinery that would express the target DNA. Understanding where TFs are binding would provide fundamental insight into which targets they regulate and how specific mutations affect their controlling ability.

Biology Research Software Deciphers Binding Affinity of Transcription Factor Proteins to DNA: MatrixREDUCE is an algorithm designed to decipher the sequence-specific binding affinity of transcription factor (TF) proteins to regions of the DNA. Using a novel statistical-mechanical model of the interaction between the TF and DNA (based on kinetics), MatrixREDUCE can accurately generate position-specific affinity matrices (PSAM) for TFs from genome-wide TF occupancy data, representing the change in binding affinity whenever a specific position within a reference binding sequence is mutated.

Applications:
• Prediction of accurate sequence-specific binding affinity for various TF protein
• Use of binding affinities to infer in vivo cellular regulatory interactions
• Predictive modeling of target gene expression given the expression level of a TF
• Predictive modeling of perturbations to specific TFs (e.g. drug-activity profiles) and their effect on expression level of other genes

Advantages:
• Algorithm is based on physical model of TF-DNA binding
• Close correlation of predicted PSAMs with experimental measurements of TF binding activity, as measured by Electrophoretic Mobility Shift Assay (EMSA) and reporter gene assays
• Uses the information for all probes in the dataset, eliminating the need to delineate "bound" and "unbound" sets a priori
• Does not require a background sequence model


Licensing Status: Available for Licensing and Sponsored Research Support

Inventors

Harmen J. Bussemaker