MatrixREDUCE: Software for quantifying sequence specificity of nucleotide-binding factors using genomic data

Gene expression regulation by a variety of proteins is a highly complex process and has great implications for the function of every biological system. High-throughput genomic data has become ubiquitous with the advent of cheap technology such as microarrays and DNA sequencing. Genomic information, however, requires effective algorithms to turn that data into useful and actionable information. This technology, MatrixREDUCE, is advanced software for the determination of gene sequence-specific binding affinities of transcription factors. Using genomic information, a computational model accurately predicts binding motifs of transcription factors without any previous empirical data, transforming common genomic data into highly useful and actionable transcription factor binding knowledge.

Advanced algorithm generates sequence-specific binding affinity of transcription factors for use in genomic manipulation

Regulation of gene expression by transcription factors is useful in gene therapy, drug discovery and biological research. It is critical, however, to be able to predict the affinity of a given transcription factor to a given nucleotide sequence, an affinity that results in increased or decreased expression of specific genes. MatrixREDUCE uses a statistical mechanics model based on kinetics to predict this affinity.

MatrixREDUCE has been shown to accurately generate position-specific affinity matrices (PSAM) for transcription factors from genome-wide transcription factor occupancy data, representing the change in binding affinity whenever a specific position within a reference binding sequence is mutated.

Lead Inventor:

Harmen Bussemaker, Ph.D.
Applications:

- Prediction of accurate sequence-specific binding affinity for various transcription factors
- Use of binding affinities to elucidate in vivo cellular regulatory functions
- Predictive modeling of target gene expression given known transcription factor levels
- Modeling of perturbations to specific transcription factors and their downstream effect on other genes

Advantages:

- Software algorithm is based on physical computational model of transcription factor-DNA binding
- Position-specific affinity matrices closely correlate to biological experimental data
- Does not require background sequence model
- Utilizes all available information without necessity to delineate bound and unbound sets
- Uses widely available high-throughput genomic data

Patent Information:

Patent Issued (US 8,219,323)
Tech Ventures Reference: IR M06-084, M07-094

Related Publications:


Inventors

Harmen J. Bussemaker