SFB 680

Molecular Basis of Evolutionary Innovations

Molekulare Grundlagen evolutionärer Innovationen


# Rahul Siddharthan

**Institute of Mathematical Sciences, Chennai, India**


## *Predicting transcription factor binding sites: Going beyond position weight matrices*


Identifying transcription factor binding sites (TFBS) in silico is key in understanding gene regulation. TFBS are string patterns that exhibit some variability, commonly modelled as "position weight matrices" (PWMs). Though convenient, the PWM has significant limitations, in particular the assumed independence of positions within the binding motif; and predictions based on PWMs are usually not very specific to known functional sites. I describe a straightforward generalization of the PWM model, that considers frequencies of dinucleotides instead of individual nucleotides. Unlike previous efforts, this method considers all dinucleotides within an extended binding region, and does not make an attempt to determine a priori the significance of particular dinucleotide correlations. I describe how to use a "dinucleotide weight matrix" (DWM) to predict binding sites, dealing in particular with the complication that ist entries are not independent probabilities. Benchmarks show, for many factors, a dramatic improvement over PWMs in precision of predicting known targets. In most cases, significant further improvement arises by extending the commonly defined "core motifs" by about 10bp on either side. Though this flanking sequence shows no strong motif at the nucleotide level, the predictive power of the dinucleotide model suggests that the "signature" in DNA sequence of protein-binding affinity extends beyond the core protein-DNA contact region. In addition, I will briefly describe some ongoing work that incorporates various forms of prior information, and cooperation and competition between factors, in binding site prediction.


## June 4, 2010

## 2:00 p. m.

**Institute for Genetics, Lecture Room, ground floor**

**Host: Michael Lässig**

www.sfb680.uni-koeln.de