

SEMINAR

Timothy Hughes, Ph.D.

University of Toronto

Department of Medical Genetics & Microbiology

“Complexity, diversity, and conservation in the protein-DNA interactome”

Mapping the complete spectrum of protein-DNA interactions is paramount to understanding global gene regulation and to fully decoding the genome and interpreting its evolution. We are undertaking a brute force effort to determine the binding preferences of as many individual mouse transcription factors as possible, by cloning and purifying transcription factor DNA-binding domains and then determining their binding specificity using a microarray technique that surveys relative affinity to all possible 8-mer sequences. To date, we have binding profiles for proteins in 23 different structural classes, including the majority of SOX, ETS, ARID, IRF, AP-2, GCM, and homeodomain family members. These data reveal a surprisingly rich landscape of DNA sequence preferences, with many proteins exhibiting what appear to be multiple binding modes. Homeodomains provide a striking example of the biochemical repertoire that can be achieved with a single, simple domain structure: there are at least 65 distinct homeodomain DNA-binding activities in the mouse alone. Since the binding preferences correlate with conserved protein sequence features, the mouse data can be used to predict relative 8-mer binding sequences for homeodomains in species as distant as *Drosophila* and *C. elegans*, with the highest-scoring sequences corresponding to known *in vivo* binding sites. Our results suggest that variation in sequence recognition may be a factor in the functional diversity and evolutionary success of many transcription factor DNA-binding domain families, but also show that the sequence specificity of many animal regulatory factors has not changed substantially since the Cambrian era. Ongoing efforts are aimed at a more comprehensive effort, focusing on families such as zinc-fingers that have undergone the greatest expansions in mammals.

Date: Thursday, May 1st

Time: 2:10 - 3:00pm

Location: Howe Hall, Alliant Energy-
Lee Liu Auditorium

Presented by the L. H. Baker Center for Bioinformatics and Biological Statistics and IGERT
<http://www.bioinformatics.iastate.edu/seminars/index.html>