Technique finds gene regulatory sites without knowledge of regulators

November 19 2009

A new statistical technique developed by researchers at the University of Illinois allows scientists to scan a genome for specific gene-regulatory regions without requiring prior knowledge of the relevant transcription factors. The technique has been experimentally validated in both the mouse genome and the fruit fly genome.

"Our statistical approach is more widely applicable than other approaches, and is especially useful for investigating gene regulatory networks with relatively little prior characterization," said Saurabh Sinha, a professor of computer science and an affiliate of the university's Institute for Genomic Biology.

Sinha and colleagues at the U. of I., the University of Cambridge and the State University of New York described the technique and reported their findings in the Oct. 20 issue of the journal *Developmental Cell*. Sinha is a corresponding author of the paper.

Proper gene regulation depends upon certain combinations of [molecules](#) being attached to segments of DNA at the right place and at the right time. Such combinatorial regulation is arguably the most prominent theme in animal development, Sinha said.

The clusters of binding sites to which the necessary molecules attach are called
cis-regulatory modules. Finding the appropriate cis-regulatory modules is the first step toward unraveling the gene regulatory process.

Beginning with a short list of sample cis-regulatory module sequences (obtained through genetic assays) the computer algorithms search genome-wide for other sequences with similar-looking binding sites and similar functionality.

"Given a representative subset of cis-regulatory module sequences, we should be able to predict most of the other modules," Sinha said. "Ultimately, the task is to draw a complete picture of what molecules are sitting where, and what combination of molecules is controlling each gene."

The researchers performed extensive cross-validation tests representing hundreds of modules in the mouse genome and in the fruit fly [genome](). In addition, seven predictions were experimentally validated through genetic techniques, demonstrating the effectiveness of the approach for mammalian and insect genomes.

Sinha and colleagues are currently working on thermodynamics-based models for potential use in predicting cis-regulatory modules in different species.

"By describing the underlying chemistry in terms of molecules, DNA sequences and binding energies, it may be possible to use knowledge about gene regulation in one species to make predictions about [gene regulation]() in corresponding systems in very diverse species," Sinha said.

Source: University of Illinois at Urbana-Champaign ([news]() : [web]())