

# Algorithm reveals complex protein dynamics behind gene expression

October 25 2013

---

(Phys.org) —In yet another coup for a research concept known as "big data," researchers at the Stanford University School of Medicine have developed a computerized algorithm to understand the complex and rapid choreography of hundreds of proteins that interact in mindboggling combinations to govern how genes are flipped on and off within a cell.

To do so, they coupled findings from 238 DNA-[protein](#)-binding experiments performed by the [ENCODE project](#)—a massive, multiyear international effort to identify the functional elements of the human genome—with a laboratory-based technique to identify binding patterns among the proteins themselves.

The analysis is sensitive enough to have identified many previously unsuspected, multipartner trysts. It can also be performed quickly and repeatedly to track how a cell responds to environmental changes or crucial developmental signals.

"At a very basic level, we are learning who likes to work with whom to regulate around 20,000 human [genes](#)," said Michael Snyder, PhD, professor and chair of genetics at Stanford. "If you had to look through all possible interactions pair-wise, it would be ridiculously impossible. Here we can look at thousands of combinations in an unbiased manner and pull out important and powerful information. It gives us an unprecedented level of understanding."

Snyder is the senior author of a paper describing the research published

Oct. 24 in *Cell*. The lead authors are postdoctoral scholars Dan Xie, PhD, Alan Boyle, PhD, and Linfeng Wu, PhD.

Proteins control gene expression by either binding to specific regions of DNA, or by interacting with other DNA-bound proteins to modulate their function. Previously, researchers could only analyze two to three proteins and DNA sequences at a time, and were unable to see the true complexities of the interactions among proteins and DNA that occur in living [cells](#).

The challenge resembled trying to figure out interactions in a crowded mosh pit by studying a few waltzing couples in an otherwise empty ballroom, and it has severely limited what could be learned about the dynamics of gene expression.

The ENCODE, for the Encyclopedia of DNA Elements, project was a five-year collaboration of more than 440 scientists in 32 labs around the world to reveal the complex interplay among regulatory regions, proteins and RNA molecules that governs when and how genes are expressed. The project has been generating a treasure trove of data for researchers to analyze for the last eight years.

In this study, the researchers combined data from genomics (a field devoted to the study of genes) and proteomics (which focuses on proteins and their interactions). They studied 128 proteins, called trans-acting factors, which are known to regulate gene expression by binding to [regulatory regions](#) within the genome. Some of the regions control the expression of nearby genes; others affect the expression of genes great distances away.

The researchers used 238 data sets generated by the ENCODE project to study the specific DNA sequences bound by each of the 128 trans-acting factors. But these factors aren't monogamous; they bind many different

sequences in a variety of protein-DNA combinations. Xie, Boyle and Snyder designed a machine-learning algorithm to analyze all the data and identify which trans-acting factors tend to be seen together and which DNA sequences they prefer.

Wu then performed immunoprecipitation experiments, which use antibodies to identify protein interactions in the cell nucleus. In this way, they were able to tell which proteins interacted directly with one another, and which were seen together because their preferred DNA binding sites were adjoining.

"Before our work, only the combination of two or three regulatory proteins were studied, which oversimplified how gene regulators collaborate to find their targets," Xie said. "With our method we are able to study the combination of more than 100 regulators and see a much more complex structure of collaboration. For example, it had been believed that a key regulator of cell proliferation called FOS typically only works with JUN protein family members. We show, in addition to JUN, FOS has different partners under different circumstances. In fact, we found almost all the canonical combinations of two or three trans-acting factors have many more partners than we previously thought."

To broaden their analysis, the researchers included data from other sources that explored protein-binding patterns in five cell types. They found that patterns of co-localization among proteins, in which several proteins are found clustered closely on the DNA to govern [gene expression](#), vary according to cell type and the conditions under which the cells are grown. They also found that many of these clusters can be explained through interactions among proteins, and that not every protein bound to DNA directly.

"We'd like to understand how these [interactions](#) work together to make different [cell types](#) and how they gain their unique identities in

development," Snyder said. "Furthermore, diseased cells will have a very different type of wiring diagram. We hope to understand how these cells go astray."

Provided by Stanford University Medical Center

Citation: Algorithm reveals complex protein dynamics behind gene expression (2013, October 25) retrieved 2 May 2024 from <https://phys.org/news/2013-10-algorithm-reveals-complex-protein-dynamics.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--