

Model unravels rules that govern how genes are switched on and off

December 4 2008

For years, scientists have struggled to decipher the genetic instruction book that details where and when the 20,000 genes in a human cell will be turned on or off. Different genes operate in each cell type at different times, and this careful orchestration is what ultimately distinguishes a brain cell from a liver or skin cell.

Now, scientists at Washington University School of Medicine in St. Louis report they have developed a model of gene expression in yeast that predicts with a high degree of accuracy whether a gene will be switched on or off. The study is now available in the advance online publication of *Nature*.

"A huge part of the magic in biology happens at the level of a cell deciding whether it is going to transcribe a gene or not," says senior author Barak Cohen, Ph.D., assistant professor of genetics. "We have found that just a few simple rules may underlie the complicated gene expression patterns that determine whether a particular gene will be expressed at high levels in one tissue and low levels in another tissue."

Since the discovery of DNA's double helical structure more than a half century ago, scientists have focused much of their attention on understanding the 2 percent of the genome that is made up of classic genes, which code for the production of proteins.

However, the instructions for turning these genes on or off are generally not in the genes themselves. Rather, they are buried in the 98 percent of

the genome that was once cast aside as little more than genetic "junk."

"In theory, we should be able to read those instructions," Cohen explains. "A cell can look at a piece of DNA and know where and when to express a particular gene. But the fundamental question we looked at starts with the premise that scientists can't do that at all."

Researchers have known for some time that the instructions for controlling gene expression lie in short DNA sequences, called promoters, embedded in long, rambling stretches of DNA at the front of most genes. Proteins known as transcription factors bind to promoters to either activate a particular gene or shut down its activity.

But in reality, the regulation of a gene is far more complex. Each promoter can simultaneously bind a number of transcription factors, some of which work to jump start a gene's activity and others to shut it down. Additionally, the binding sites in the promoter region tolerate DNA base substitutions, so the same transcription factor can bind with slightly different affinities, depending on the promoter's genetic sequence. Transcription factors then combine to create a net effect on gene activity that is far greater or less than expected, making it exceedingly difficult to quantify their influence on a gene.

The model Cohen and his colleagues developed boils down to a few simple rules the interactions of transcription factors with DNA and with each other. The model only takes into account how tightly transcription factor proteins bind to DNA in the promoter region and how tightly transcription factors bind to each other. These simple rules can explain most of the variation in gene expression between different promoters.

They then created 2,800 simple artificial promoters and asked if these rules were sufficient to understand the activity of these promoters. "Because if we can't boil it down and understand these complex

interactions on simple artificial promoters, then there's no hope for understanding real promoters," Cohen says.

The scientists constructed promoters that consisted of random combinations of three or four transcription factor binding sites, or building blocks, using a total of 18 different building blocks. They then recorded the DNA sequence of each promoter, along with its corresponding gene expression. By incorporating sophisticated mathematical equations and statistical analysis, they could eventually predict, given a particular promoter sequence, whether it would activate or suppress gene activity.

The team determined that 65 percent of the complex variation in gene expression from one cell to the next could be explained by the simple rules that focus on the binding affinity of transcription factors.

When the investigators tested their model on real promoters in the genome of yeast, they confirmed that it could accurately predict how the binding site for the transcription factor Mig1 dampens gene expression. The model identified all 40 genes already known to be regulated by Mig1. But additionally, by incorporating information from weak binding sites that other models have not taken into account, they also uncovered another eight genes not previously known to be influenced by Mig1.

"That our model can incorporate information from weak binding sites is really important because gene expression can be influenced by very subtle interactions in the promoter regions," Cohen says. "No other model has been able to account for these subtle interactions."

Some scientists have suggested that biochemical processes, including enzymatic reactions, are more important than binding affinity, but Cohen says his model disputes that assertion.

"Our model answers the question: How do cells read the instructions for gene expression," Cohen says. "They are read mostly by the simple binding of transcription factors to DNA. This binding either recruits the enzyme RNA polymerase, which begins the process of copying and transferring information stored in the genes, or blocks it."

While Cohen is still perfecting the model, he says it may eventually enable scientists to determine where and when all the genes in the human genome will be expressed just by looking at the genetic code in the promoter region.

In addition, the model may help researchers engineer artificial promoters that drive embryonic stem cells toward a particular fate or that will turn on a gene in only a particular cell type. If scientists know the code that turns on a particular gene, then they could conceivably design an artificial promoter as a potential treatment for disease, Cohen says.

He and his group are now designing more complicated synthetic promoters similar to those that occur naturally in yeast and higher organisms to determine whether they can continue to accurately predict variations in gene expression.

Paper: Gertz J, Siggia ED, Cohen BA. Analysis of Combinatorial c-Regulation in Synthetic and Genomic Promoters. *Nature*, online Nov. 23, 2008.

Source: Washington University School of Medicine

Citation: Model unravels rules that govern how genes are switched on and off (2008, December 4) retrieved 23 April 2024 from <https://medicalxpress.com/news/2008-12-unravels-genes.html>

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.