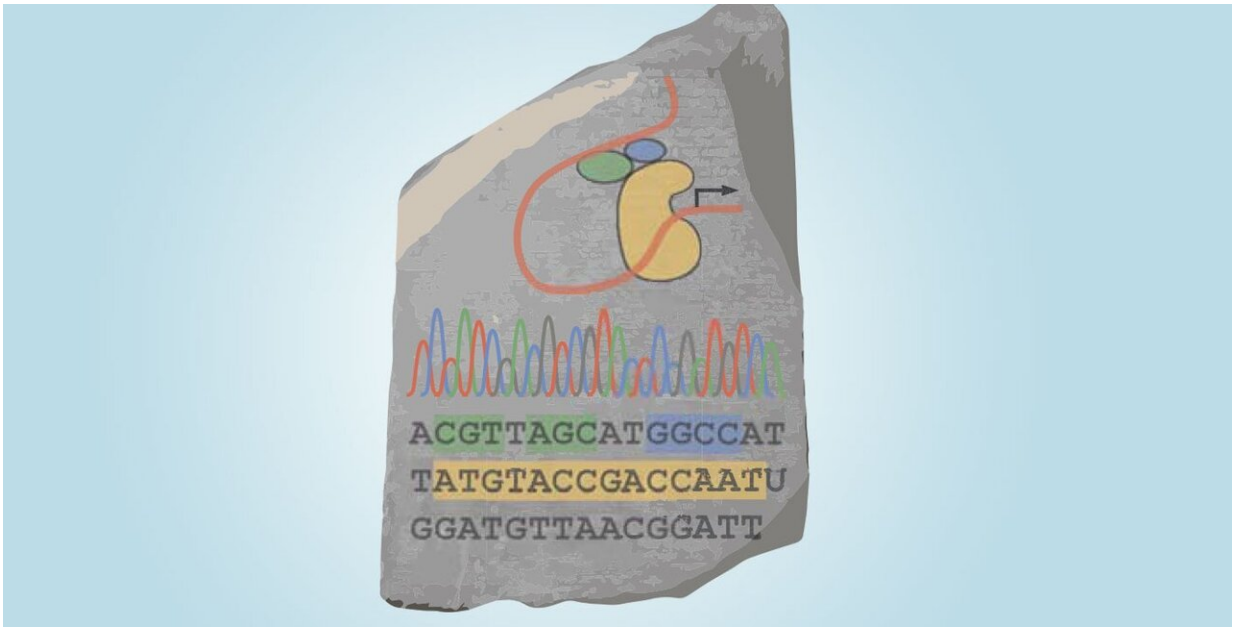
A 'genomic rosetta stone' for discovering the rules of gene regulation

October 26 2020, by Lori Dajose



Credit: R. Phillips

As early as 1975, biologists discovered that the protein-coding parts of the chimpanzee and human genomes are more than 99 percent identical. Yet, chimpanzees and humans are clearly different in significant ways. Why?

The answer lies in the fact that how DNA is used is as important as what it says. That is, the genes that make up a genome are not always being

used; they can be turned on or off or dialed up or down over time, and they interact with one another in complex ways. Some genes encode instructions for producing specific proteins and others encode information about regulating other genes.

Now, researchers in the laboratory of Rob Phillips, the Fred and Nancy Morris Professor of Biology and Biophysics, have developed a new tool for determining how various genes in the common bacterium Escherichia coli are regulated. Though E. coli has been used as a model organism in biology and bioengineering for decades, researchers understand the regulatory behavior of only about 35 percent of its genes. The new method from the Phillips laboratory sheds light on how nearly 100 previously uncharacterized genes are regulated and lays the foundation for studying many others.

A paper describing the new technique appears in the journal *eLife*.

Imagine you could read the alphabet and punctuation of some new language, but you could not understand what individual words meant or any of the rules of grammar. You could read a book and recognize each letter you read without having any comprehension of what a sentence or paragraph was saying. This is analogous to the challenge faced by biologists in the modern genomic era: Sequencing an organism's genome is now rapid and straightforward, but actually understanding how each gene is regulated is much more difficult. An understanding of gene regulation is key to understanding health and disease, and is important if we are to one day repurpose cells so they can do things that we have designed them to do.

"We've developed a general tool that researchers could use on nearly any microbial organism," says Phillips. "Our dream is that someone like Victoria Orphan [James Irvine Professor of Environmental Science and Geobiology] could go down to the ocean floor and come back with some

never-before-seen bacterium, and we could use our tool on it to determine not only the sequence of its genome but how it is regulated."

In the new method, researchers make systematic perturbations to the genome, and see what happens. Essentially, the equivalent of typographical errors are made in the genome, and the impact of those typos on cellular function is observed. For example, if you replace the letter "k" in the word "walk" with the letter "x" to make "walx," the intent of the original word is still fairly clear. This is not the case if you swap the letter "w" for a "t" to produce "talk." This suggests that the letter "w" carries important information about the meaning of the original word.

In the same way, making changes to a genome using the DNA alphabet allows researchers to figure out which letters are most important for the correct "meaning."

To validate their method, Phillips and colleagues first examined 20 particular E. coli genes that researchers already knew how to turn off and on. Their method correctly characterized these 20 genes. Next, the team moved on to 80 other, less-understood genes to understand how they work as well.

For now, the method has only been used on bacterial cells, but ultimately Phillips envisions being able to examine eukaryotic cells (such as human cells), which are more complex, with a modified version of the method.

"This was a decade-long project supported by the NIH Director's Pioneer Award, and required a sustained hard effort and funding," says Phillips. "This is the kind of project where there are no quick results."

The paper is titled "Deciphering the regulatory genome of Escherichia coli, one hundred promoters at a time."

More information: William T Ireland et al. Deciphering the regulatory genome of Escherichia coli, one hundred promoters at a time, *eLife* (2020). DOI: 10.7554/eLife.55308

Provided by California Institute of Technology