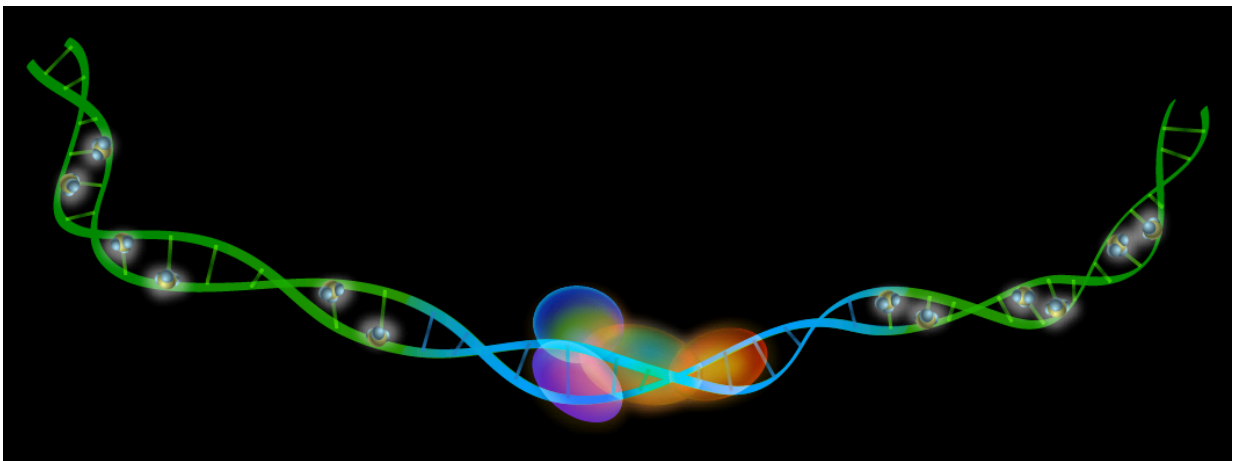


Team develops tool that maps functional areas of the genome to better understand disease

February 13 2017



A Salk team developed a tool that maps functional areas of the genome to better understand disease. Credit: Salk Institute

Most of us would be lost without Google maps or similar route-guidance technologies. And when those mapping tools include additional data about traffic or weather, we can navigate even more effectively. For scientists who navigate the mammalian genome to better understand genetic causes of disease, combining various types of data sets makes finding their way easier, too.

A team at the Salk Institute has developed a computational algorithm

that integrates two different data types to make locating key regions within the genome more precise and accurate than other tools. The method, detailed during the week of February 13, 2017, in *Proceedings of the National Academy of Sciences*, could help researchers conduct vastly more targeted searches for disease-causing genetic variants in the human genome, such as ones that promote cancer or cause metabolic disorders.

"Most of the variation between individuals is in noncoding regions of the genome," says senior author Joseph Ecker, a Howard Hughes Medical Institute investigator and director of Salk's Genomic Analysis Laboratory. "These regions don't code for proteins, but they still contain genetic variants that cause disease. We just haven't had very effective tools to locate these areas in a variety of tissues and cell types—until now."

Only about two percent of our DNA is made up of genes, which code for proteins that keep us healthy and functional. For many years, the other 98 percent was thought to be extraneous "junk." But, as science has developed ever more sophisticated tools to probe the genome, it has become clear that much of that so-called junk has vital regulatory roles. For example, sections of DNA called "enhancers" dictate where and when the gene information is read out.



From left Joseph Ecker and Yupeng He. Credit: Salk Institute

Increasingly, mutations or disruption in enhancers have been tied to major causes of human disease, but enhancers have been hard to locate within the genome. Clues about them can be found in certain types of experimental data, such as in the binding of proteins that regulate gene activity, chemical modifications of proteins (called histones) that DNA wraps around, or in the presence of chemical compounds called methyl groups in DNA that turn genes on or off (an epigenetic factor called DNA methylation). Typically, computational methods for finding enhancers have relied on histone modification data. But Ecker's new system, called REPTILE (for "regulatory-element prediction based on tissue-specific local epigenomic signatures"), combines histone

modification and methylation data to predict which regions of the genome contain enhancers. In the team's experiments, REPTILE proved more accurate at finding enhancers than algorithms that rely on histone modification alone.

"The novelty of this method is that it uses DNA methylation to really narrow down the candidate regulatory sequences suggested by histone modification data," says Yupeng He, a Salk graduate student and first author of the paper. "We were then able to test REPTILE'S predictions in the lab and validate them with experimental data, which gave us a high degree of confidence in the algorithm's ability to find enhancers."

The REPTILE algorithm operates in two general steps: training and prediction. For training, the Salk team taught REPTILE to recognize mammalian enhancers by feeding into the algorithm both the locations of known enhancers as well as genomic areas other than enhancers in the DNA. In the prediction step, the algorithm ran on nine mouse and five human cell lines and tissues whose enhancer regions were unknown and pinpointed the locations of potential enhancers. Finally, the team utilized data from laboratory experiments to test whether the predictions made by REPTILE in the prediction step corresponded to real regulatory regions. Because enhancers increase the activity of target genes, researchers can test the activity of DNA sequences by connecting them to a reporter gene and watching to see whether the supposed target gene ramps up. Using molecular tools, the team engineered mouse embryos so that enhancer activation would trigger the expression of linked reporters, which can be monitored by staining. So, if REPTILE predicted that a specific enhancer was linked to mouse forebrain development, the team was able to look for a staining pattern in the embryo's forebrain region. If they saw it, REPTILE's prediction was considered valid. The Salk team also tested REPTILE's predictions against four other commonly used enhancer-finding algorithms. Overall, REPTILE outperformed each one, finding enhancer regions with greater accuracy (getting closer

to them along the DNA strand) and fewer errors (misidentifications). In particular, REPTILE was more successful than the other systems at the invaluable task of finding enhancers in different tissue types than those it was trained on.

"The number of genetic variants in the genome is enormous," says Ecker. "So in terms of finding ones that cause disease, you really want to shine a spotlight on the regions you think are most important and identifying [enhancers](#) is a critical step in the process."

More information: Improved regulatory element prediction based on tissue-specific local epigenomic signatures, *PNAS*, www.pnas.org/cgi/doi/10.1073/pnas.1618353114

Provided by Salk Institute

Citation: Team develops tool that maps functional areas of the genome to better understand disease (2017, February 13) retrieved 11 May 2024 from <https://medicalxpress.com/news/2017-02-team-tool-functional-areas-genome.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>
--