

Algorithm created by deep learning finds potential therapeutic targets throughout genome

August 6 2020



Credit: CC0 Public Domain

A team of researchers from New Jersey Institute of Technology (NJIT) and Children's Hospital of Philadelphia (CHOP) have developed an algorithm through machine learning that helps predict sites of DNA

methylation—a process that can change the activity of DNA without changing its overall structure—and could identify disease-causing mechanisms that would otherwise be missed by conventional screening methods.

The paper was published online this week by the journal *Nature Machine Intelligence*.

DNA methylation is involved in many key cellular processes and an important component in gene expression. Likewise, errors in methylation can be linked to a variety of human diseases. While genomic sequencing tools are effective at pinpointing polymorphisms that may cause a disease, those same methods are unable to capture the effects of methylation because the individual genes still look the same. Specifically, there has been considerable effort to study DNA methylation on N6-adenine (6mA) in [eukaryotic cells](#), which include human cells, but while [genomic data](#) is available, the role of methylation in these cells remains elusive.

"Previously, methods that had been developed to identify these methylation sites in the genome were very conservative and could only look at certain nucleotide lengths at a given time, so a large number of methylation sites were missed," said Hakon Hakonarson, MD, Ph.D., Director of the Center for Applied Genomics (CAG) at CHOP and one of the senior co-authors of the study. "We needed to develop a better way of identifying and predicting methylation sites with a tool that could identify these motifs throughout the genome that may have a robust functional impact and are potentially disease causing."

In order to address this issue plaguing the [research community](#), CAG and its partners at NJIT turned to [deep learning](#). Zhi Wei, Ph.D., a professor of computer science at NJIT and a senior co-author of the study, worked with Hakonarson and his team to develop a deep learning algorithm that

could predict where these sites of methylation happened, which would then help researchers determine the effect they might have on certain nearby genes.

Wei calls his software Deep6mA. To predict where these methylation sites might be found, Wei led the development of a neural network, which is a [machine learning](#) model that attempts to learn in similar ways to a brain. Neural networks have been utilized in cellular research before, but this is its first application to study DNA methylation sites on natural multicellular organisms.

Wei cited four advantages of the new method: automation of the sequence feature representation of different levels of detail; integration of a broad spectrum of methylation sequences flanking genes of interest; enabling of the potential visualization of inherent sequence motifs for interpretation; and facilitation of model development and prediction in large-scale genomic data.

The study team applied this algorithm to three different types of representative organisms: *A. thaliana*, *D. melanogaster*, and *E. coli*, the first two being eukaryotic. Deep6mA was able to identify 6mA methylation sites down to the resolution of a single nucleotide, or basic unit of DNA. Even in this initial confirmation study, the researchers were able to visualize regulatory patterns that they had been unable to observe using previously existing methods.

"One limitation is that our proposed prediction is purely based on sequence information," Wei said in his discussion statement of the study. "Whether a candidate is a 6mA [site](#) or not will also depend on many other factors. Methylation, including 6mA, is a dynamic process, which will change with cellular context. In the future, we would like to take other factors into consideration [such as] [gene expression](#). We hope to predict 6mA across cellular context by integrating other data."

"We already know that a number of genes have a disease-causing mechanism brought about by methylation, and while this study was not done in [human cells](#), the eukaryotic cell models were very comparable," Hakonarson said. "Genomic scientists looking to translate their findings into clinical applications would find this tool very useful, and the level of precision could eventually lead to the discovery of specific cells or targets that are candidates for therapeutic intervention."

More information: Fei Tan et al, Elucidation of DNA methylation on N6-adenine with deep learning, *Nature Machine Intelligence* (2020).
[DOI: 10.1038/s42256-020-0211-4](https://doi.org/10.1038/s42256-020-0211-4)

Provided by Children's Hospital of Philadelphia

Citation: Algorithm created by deep learning finds potential therapeutic targets throughout genome (2020, August 6) retrieved 3 May 2024 from <https://phys.org/news/2020-08-algorithm-deep-potential-therapeutic-genome.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.