

Researchers use neural networks to study DNA

17 December 2020



Credit: Pixabay/CC0 Public Domain

HSE scientists have proposed a way to improve the accuracy of finding Z-DNA, or DNA regions that are twisted to the left instead of to the right. To do this, they used neural networks and a dataset of more than 30,000 experiments conducted by different laboratories around the world. Details of the study are published in *Scientific Reports*.

Over the 67 years that have passed since the discovery of the structure of DNA, scientists have found many structural variations of this molecule. Sometimes DNA structures do not at all resemble the usual double helix, which is called B-DNA: they can differ from B-DNA by the number of chains (from two to four), chain density and thickness, the way in which the nitrogenous bases are joined, and the direction of the twist of the helix.

One of the structures, Z-DNA, is composed of a double helix, twisted differently—to the left instead of to the right. It is known that regions of Z-DNA are found in the cells of various organisms (from bacteria to humans), arise under certain conditions (for example, in supercoiled DNA or high salt concentration), and can be combined with other DNA structures in one molecule. For example, if, for some reason, the B-DNA molecule is supercoiled to the extent that it complicates

transcription (synthesis of RNA based on DNA), some of its sections can twist in the opposite direction, thereby relieving unnecessary 'stress.' Scientists also suggest that Z-DNA can regulate transcription and increase the likelihood of mutations. Some research suggests that the formation of Z-DNA may be associated with certain diseases such as cancer, diabetes, and Alzheimer's. Recently, more and more studies have appeared that show the role of Z-DNA in the innate immune response—the reaction to viruses and other pathogens within the cell itself.

To learn more about the conditions of formation and the biological role of Z-DNA regions, it is necessary to have methods to find their location in the genome. The first genetic map with the markup of Z-DNA sites was compiled back in 1997, based on experimental data on the structural binding of consecutive nucleotides. In recent years, methods have emerged in which the location of regions other than B-DNA is predicted using computer algorithms. Advances in machine learning have made it possible to use another powerful tool for this task—[neural networks](#). Unlike most methods, neural networks can take into account many factors and do not require scientists to select in advance few most likely influential. But even for neural networks, the search for Z-DNA remains a difficult task, since there is not enough [experimental data](#): Z-DNA appears and disappears, and an experiment records only a small part of these regions. The researchers decided to test whether the accuracy of the neural networks increases with the addition of information from omics data, or information on how gene activity and protein synthesis in cells are regulated.

The scientists began by comparing how three types of neural networks—convolutional, recurrent, and a combination of the first two—can handle the task. A convolutional neural network is most often used for image processing, while a recurrent neural network is most often used to analyze texts. All three types

of neural networks have already been tested on problems related to the study of the genome. In total, the authors of the study trained and evaluated 151 models on the DNA dataset enlarged by omics data. One of the [recurrent neural networks](#), which the authors named DeepZ, yielded the best results, and they used it to predict novel Z-DNA regions in the human genome. Its accuracy significantly exceeds the accuracy of the existing algorithm, Z-Hunt.

With the help of DeepZ, the scientists mapped the entire sequence of the human genome, determining for each nucleotide the likelihood that it will end up inside a Z-DNA region. A sequence of several nucleotides for which the probability exceeded a certain threshold value was marked as a potential target site.

"The results of this study are important, because, with the help of neural networks, we were not only able to replicate the experiments, but also predict the potential sites of Z-DNA formation in the genome," said Maria Poptsova, leader of the study and Head of the Laboratory of Bioinformatics at the Faculty of Computer Science at HSE University. "The abundance of Z-DNA signals suggests that they are actively used to turn genes on and off. This is a faster signal than the genomic motifs. For example, the study by the group of scientists from Australia has shown that Z-DNA serves as a signal in training to suppress fear. Apparently, Z-DNA evolutionarily appeared in cases when a quick reaction to events was required. We plan to initiate joint projects with experimental groups to test the predictions."

The authors demonstrated a novel approach to predicting Z-DNA regions using omix data and deep learning methods. The neural network generated genome markup will help scientists conduct experiments to detect Z-DNA, the full spectrum of which is just beginning to emerge.

More information: Nazar Beknazarov et al, Deep learning approach for predicting functional Z-DNA regions using omics data, *Scientific Reports* (2020). [DOI: 10.1038/s41598-020-76203-1](https://doi.org/10.1038/s41598-020-76203-1)

Provided by National Research University Higher School of Economics

APA citation: Researchers use neural networks to study DNA (2020, December 17) retrieved 13 April 2021 from <https://phys.org/news/2020-12-neural-networks-dna.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.