

DeepBind predicts where proteins bind, uncovering disease-causing mutations

July 27 2015



A depiction of the double helical structure of DNA. Its four coding units (A, T, C, G) are color-coded in pink, orange, purple and yellow. Credit: NHGRI

A new tool called DeepBind uses deep learning to analyze how proteins bind to DNA and RNA, allowing it to detect mutations that could disrupt cellular processes and cause disease.

CIFAR Senior Fellow Brendan Frey (University of Toronto), supervising lead authors Babak Alipanahi and Andrew Delong, developed the method using deep learning—a machine learning technique pioneered by CIFAR fellows in the Neural Computation & Adaptive Perception program and now used by companies such as Google and Facebook.

Hundreds of thousands of proteins in human cells attach themselves to DNA and RNA and regulate gene expression. "They are master controllers of the cell," Alipanahi says. But many of these proteins are picky about which sequences they will bind to.

DeepBind can analyze noisy experimental data to determine the DNA and RNA sequences to which a set of proteins will bind. Then, it can look at a new sequence and compute a score saying how likely it is that any of these proteins would bind to it. Given a sequence with a mutation, the tool can analyze whether binding changes. Mutations that add or delete protein binding sites can alter gene expression patterns and lead to disease.

Frey, a senior fellow in both the program in Genetic Networks and the program in Neural Computation & Adaptive Perception, says, "DeepBind is one example of a series of technologies that we have developed by combining our expertise in deep learning and genome biology. I believe that these technologies will revolutionize health care and precision medicine. They are game changers."

Frey's group has been developing disruptive machine learning techniques to understand the genome for over a decade. Just last year, technology leaders from companies like Facebook, Google and Deep Mind agreed

that the next "big thing" for deep learning is its application in health care and genomic medicine. Based on their successes over the past decade, Frey's group couldn't agree more.

DeepBind's first analyses of human genetic data, described in *Nature Biotechnology*, has already provided new information about disruptions to protein binding in [mutations](#) tied to cancers, haemophilia and familial hypercholesterolemia—a hereditary condition associated with very high levels of cholesterol. The system also revealed that a genetic mutation leading to abnormal development of the brain's cerebral cortex may be more complex than previously thought. Specifically, a mutation that is known to delete one binding site was found to actually delete two binding sites at the same time.

DeepBind follows the breakthrough development of the "human splicing code" by Frey's group in collaboration with CIFAR senior fellows Stephen Scherer (The Hospital for Sick Children and the University of Toronto) and Timothy Hughes (University of Toronto). Their system identified new genes that could be implicated in autism spectrum disorders, hereditary non-polyposis colon cancer and spinal muscular atrophy by identifying errors in gene splicing. Since many proteins regulate splicing, the researchers intend to use their insights from DeepBind to improve the human splicing code. The new tool substantially expands the types of mutations they can analyze to promoters and enhancers.

Frey says that the concepts behind DeepBind were highly influenced by CIFAR collaborators who are at the forefront of developing deep learning technologies. "CIFAR played a crucial role in establishing the research network that led to our breakthroughs in [deep learning](#) and genomic medicine," Frey says.

More information: Predicting the sequence specificities of DNA- and

RNA-binding proteins by deep learning, *Nature Biotechnology*, [DOI: 10.1038/nbt.3300](https://doi.org/10.1038/nbt.3300)

Provided by Canadian Institute for Advanced Research

Citation: DeepBind predicts where proteins bind, uncovering disease-causing mutations (2015, July 27) retrieved 6 May 2024 from <https://phys.org/news/2015-07-deepbind-proteins-uncovering-disease-causing-mutations.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.