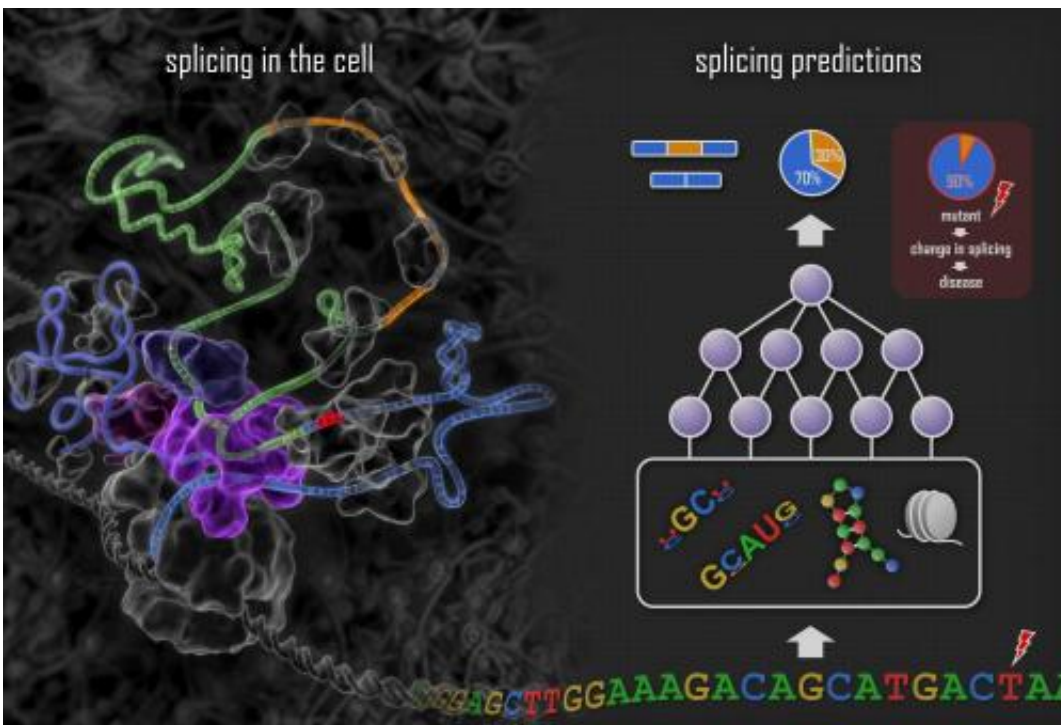


# 'Deep learning' reveals unexpected genetic roots of cancers, autism and other disorders

December 18 2014



Frey's team used computational deep learning techniques to train a system that mimics the process of splicing in the cell (left panel). Features such as motifs, RNA secondary structures and nucleosome positions are computationally determined from the DNA sequence (right panel), combinations of these features are combined to detect complex patterns, and these are combined again to predict how splicing will occur for the exon within the DNA sequence. The effect of a DNA mutation is assessed by applying the system to the sequence with and without the mutation and measuring the change in the computed splicing level. This procedure enables the identification of potentially harmful mutations from across a wide range of diseases. Frey's team applied it to spinal muscular atrophy, cancers and autism and in each case identified previously

unknown genetic determinants of the disease. Credit: Graham Johnson and Andrew Delong

In the decade since the genome was sequenced in 2003, scientists and doctors have struggled to answer an all-consuming question: Which DNA mutations cause disease?

A new computational technique developed at the University of Toronto may now be able to tell us.

A Canadian research team led by professor Brendan Frey has developed the first method for 'ranking' genetic [mutations](#) based on how living cells 'read' DNA, revealing how likely any given alteration is to cause disease. They used their method to discover unexpected genetic determinants of autism, hereditary cancers and spinal muscular atrophy, a leading genetic cause of infant mortality.

Their findings appear in today's issue of the leading journal *Science*.

Think of the [human genome](#) as a mysterious text, made up of three billion letters. "Over the past decade, a huge amount of effort has been invested into searching for mutations in the genome that cause disease, without a rational approach to understanding why they cause disease," says Frey, also a senior fellow at the Canadian Institute for Advanced Research. "This is because scientists didn't have the means to understand the text of the genome and how mutations in it can change the meaning of that text." Biologist Eric Lander of the Massachusetts Institute of Technology captured this puzzle in his famous quote: "Genome. Bought the book. Hard to read."

What was Frey's approach? We know that certain sections of the text,

called exons, describe the proteins that are the building blocks of all living cells. What wasn't appreciated until recently is that other sections, called introns, contain instructions for how to cut and paste exons together, determining which proteins will be produced. This 'splicing' process is a crucial step in the cell's process of converting DNA into proteins, and its disruption is known to contribute to many diseases.

Most research into the genetic roots of disease has focused on mutations within exons, but increasingly scientists are finding that diseases can't be explained by these mutations. Frey's team took a completely different approach, examining changes to text that provides instructions for splicing, most of which is in introns.

Frey's team used a new technology called '[deep learning](#)' to teach a computer system to scan a piece of DNA, read the genetic instructions that specify how to splice together sections that code for proteins, and determine which proteins will be produced.

Unlike other machine learning methods, deep learning can make sense of incredibly complex relationships, such as those found in living systems in biology and medicine. "The success of our project relied crucially on using the latest deep learning methods to analyze the most advanced experimental biology data," says Frey, whose team included members from University of Toronto's Faculty of Applied Science & Engineering, Faculty of Medicine and the Terrence Donnelly Centre for Cellular and Biomolecular Research, as well as Microsoft Research and the Cold Spring Harbor Laboratory. "My collaborators and our graduate students and postdoctoral fellows are world-leading experts in these areas."

Once they had taught their system how to read the text of the genome, Frey's team used it to search for mutations that cause splicing to go wrong. They found that their method correctly predicted 94 percent of

the genetic culprits behind well-studied diseases such as [spinal muscular atrophy](#) and colorectal cancer, but more importantly, made accurate predictions for mutations that had never been seen before.

They then launched a huge effort to tackle a condition with complex genetic underpinnings: autism spectrum disorder. "With autism there are only a few dozen genes definitely known to be involved and these account for a small proportion of individuals with this condition," says Frey.

In collaboration with Dr. Stephen Scherer, senior scientist and director of The Centre for Applied Genomics at SickKids and the University of Toronto McLaughlin Centre, Frey's team compared mutations discovered in the whole [genome](#) sequences of children with autism, but not in controls. Following the traditional approach of studying [protein](#)-coding regions, they found no differences. However, when they used their deep learning system to rank mutations according to how much they change splicing, surprising patterns appeared.

"When we ranked mutations using our method, striking patterns emerged, revealing 39 novel genes having a potential role in autism susceptibility," Frey says.

And autism is just the beginning—this mutation indexing method is ready to be applied to any number of diseases, and even non-disease traits that differ between individuals.

Dr. Juan Valcárcel Juárez, a researcher with the Center for Genomic Regulation in Barcelona, Spain, who was not involved in this research, says: "In a way it is like having a language translator: it allows you to understand another language, even if full command of that language will require that you also study the underlying grammar. The work provides important information for personalized medicine, clearly a key

component of future therapies."

**More information:** "The human splicing code reveals new insights into the genetic determinants of disease," by H.Y. Xiong et al. *Science*, [www.sciencemag.org/lookup/doi/10.1126/science.1254806](http://www.sciencemag.org/lookup/doi/10.1126/science.1254806)

Provided by University of Toronto

Citation: 'Deep learning' reveals unexpected genetic roots of cancers, autism and other disorders (2014, December 18) retrieved 30 April 2024 from <https://medicalxpress.com/news/2014-12-deep-reveals-unexpected-genetic-roots.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.