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


Provided by University of Toronto