



RESEARCH2REALITY

Shining a light on research & innovation.

Lifting Our Genomic Knowledge Into the Cloud

An interview with University of Toronto researchers:

Stephen Scherer, Geneticist

Brett Trost, Computational Biologist

Stephen Scherer

The genome is massive in a way. There's 3 billion chemical letters in each of our DNA, and they come in pairs. But to actually sequence them and decode what that means, you need to sequence it about 50 times or so. When we did the first 5,200 genomes in our autism project here, on this floor, at the time in about 2015 it was the largest data transfer that Google had received in its history. We've now doubled that and we're doing roughly 10,000 genomes a year here.

Brett Trost

Currently we have about 11,500 whole genome sequences from individuals with autism and their families, and each of these data sets is a massive amount of data by itself. And then when you have 11,500 of them, you need a lot of computing power and a lot of computing storage in order to be able to process this data.

Stephen Scherer

We have these imaging devices that can figure out the order and identity of the nucleotides in short stretches, but then we need to string that together

TODAY'S RESEARCH. TOMORROW'S REALITY.



into the three billion string of letters. That has to go down to the high performance computing facility here. But if you want to do thousands at a time, and then compare them to tens of thousands of other genomes, we have to do that in the cloud. And then to really make sense of it we have to use machine learning approaches.

Brett Trost

Computer science is becoming more and more important for being able to make meaningful inferences from biological data. The human genome is so complex that you can't really just look at simple patterns in the data and make sense of it. But with machine learning methods, we can actually give the computer lots and lots of information, that one human can't necessarily make sense of himself or herself, but the computer can learn the patterns and make sense of it. And we can use these methods to try to predict, for example, whether genes might be associated with disorders like autism.

Stephen Scherer

Our team here is founded around human geneticists, but in fact 30 percent of our staff are computational biologists. Multidisciplinary teams in this realm of big science are critical.