

Researchers teach computer to interpret genome with video game-based technology

Sequencing of the human genome in 2003 was a monumental achievement. But it left us with more questions than answers: it gave scientists the three-billion-base-pair instruction manual for how a person is created, but not the knowledge of how to read it.

Now a research team led by Brendan Frey at the University of Toronto has created a sophisticated computer tool that uses machine learning — and hardware borrowed from the video-game industry — to peer into parts of the genome that were once “black boxes,” and to rank how likely variants in those regions are to give rise to diseases, including autism.

“We’ve increased by a factor of 10 how much of the genome we can analyze and understand,” says Frey, the Canada Research Chair in Biological Computation and a senior fellow of the Canadian Institute for Advanced Research.

The research, published online Thursday in the journal *Science*, is “a big deal,” says Jeremy Sanford, a professor at the University of California Santa Cruz who specializes in RNA biology. “This is a good step toward interpreting the less obvious features of the genome.”

In order to create the computer tool — recently dubbed “SPANR,” for SPlicing-based ANalysis of vaRiants” — the research team first acquired sophisticated graphics cards developed by video-game companies. Scientists have realized they are perfectly suited for deep learning, the type of high-level machine learning Frey’s lab wanted to undertake.

“We’ve taken these video-game cards that were causing teenagers to not do any work, and solved one of the hardest problems in science,” Frey jokes.

Teaching the computer how to read the genome is like teaching a child how to read words, Frey says. The child sees the word “cow” and a picture of a cow. Eventually, the child learns that those three letters in that order correspond with the picture of the animal. As the child learns to read, it recognizes the word “cow” in new contexts.

Frey’s team showed the computer system strings of DNA, and showed it how much protein those strings of DNA produce. By examining tens of thousands of such examples, the machine is eventually able to predict which proteins will be made for a given DNA sequence, including ones that differ between individuals. What the scientists were really interested in was the regulatory code, parts of genes that

provide the instructions for stitching proteins together, a process called splicing.

Read full, original article: [Computer taught how to read human genome](#)