

TOOLBOX

Whole genomes may power search for autism risk factors

BY JESSICA WRIGHT

23 AUGUST 2017



Scientists this week **announced the release** of nearly 7,000 whole-genome sequences from a collection of families that each have one child with autism and unaffected family members.

The release brings the total to nearly 9,000 sequences from families in the Simons Simplex Collection (SSC), including more than 2,300 from children with autism. Some of the sequences are still in transit from the sequencing facility. (The collection and the majority of the sequencing efforts are funded by the Simons Foundation, *Spectrum's* parent organization.)

Another project, called **MSSNG**, has also amassed 5,205 whole-genome sequences, including 2,626 from children with autism, with another large release scheduled for October. The set is expected to eventually include 10,000 sequences from families with autism, only some of which have a single child with autism.

The SSC is specifically designed to reveal mutations that appear spontaneously, or *de novo*, in children with autism. Studies of the **exomes** — the portions of the genome that code for protein — have yielded **dozens of new candidate genes** for the condition.

A decade ago, when the study launched, sequencing a single genome was a costly and ambitious prospect. Advances in technology, along with dropping costs, have since made sequencing genomes more practical.

Having access to whole-genome sequences allows researchers to probe the little-explored regions between genes. It also provides more precise information about genes, **revealing flips, swaps** and other structural variations.

“The genomic resource is critical for assessing other forms of genetic variation that cannot be accessed by exomes,” says **Evan Eichler**, professor of genome sciences at the University of Washington in Seattle. Eichler plans to analyze the new sequences.

The new release adds to sequences from 553 families in the SSC made available a year ago. Researchers have used those sequences to try to find ways to assess which **variants in noncoding regions** of the genome are harmful.

However, such assessments require **thousands of genome sequences**, as well as enough computing power to analyze them. The addition of the nearly 7,000 sequences, from roughly 1,800 families, significantly powers up the search for these variants but may still not be enough.

“If as a field we want to stimulate as much analysis of this data as possible to make significant discoveries faster, we will need to seek researchers in other fields such as computer science and statistics,” says **Lucia Peixoto**, assistant professor of biomedical sciences at Washington State University in Pullman.

Researchers **can apply** to gain access to the sequence data to study autism or a related condition.