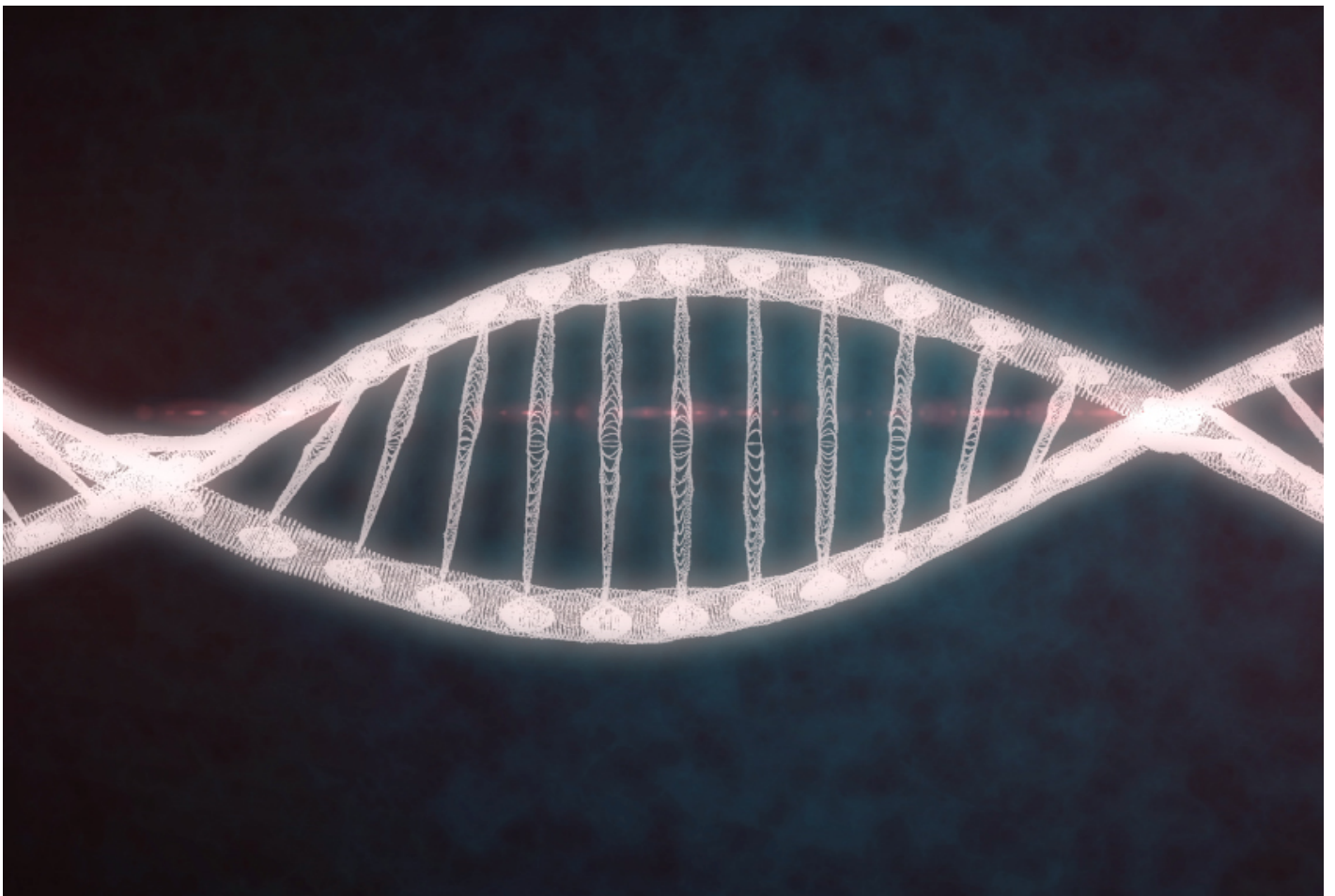


**TOOLBOX**

# Two tools speed the search for mutations in whole genomes

BY **MARISSA FESSENDEN**

14 SEPTEMBER 2018



Software to identify genetic variants, along with a new synthetic human genome, could help scientists discover mutations associated with conditions such as autism.

The software, called **Strelka2**, uses algorithms to compare the sequence of a reference genome with one from a person who has a condition. The tool detects small variants, including changes to a single DNA base as well as insertions and deletions of a few bases.

Sequencing data come in small chunks, or reads, that need to be assembled in the correct order. Many of these chunks are duplicates or can overlap. The new software assembles the pieces while checking for sequencing errors, as similar tools do. But it then uses a two-pronged approach that enables it to find variants faster and more accurately than others tools do.

For stretches with few repeated DNA segments, it quickly differentiates variants from sequencing errors. In sections with many repeated DNA segments, it slows down and uses a more complex calculation.

This approach works particularly well for finding rare variants in genomes, researchers reported 16 July in *Nature Methods*<sup>1</sup>.

### Known quantity:

Scientists evaluate the accuracy of tools such as Strelka2 by testing them on a genome with a known set of short variants. These so-called 'truth sets' may be computer generated or derived from a known human sample, often one called NA12878.

However, both sources present problems. Computer-generated genomes don't fully resemble results from tissues. And the NA12878 truth set is often used both to train tools that recognize variants and then to evaluate them. That overlap leads to biases.

Researchers have instead built a synthetic human genome with a known set of variants from the sequence of an unusual tissue source.

They used two human cell lines to build this synthetic genome. Both come from a rare noncancerous tumor that arises in the womb when a sperm fertilizes an egg that has lost all of its genetic material. The result is tissue that carries genes only from the sperm, and so is homozygous.

Homozygous genomes make it easier to differentiate errors from mutations, because any difference between the two copies of a gene is likely to be an error.

The researchers blended DNA from both cell lines to effectively create a genome of paired chromosomes. They sequenced this pair and identified the short variants. The resulting synthetic genome, called Syndip, is **available online** and described in the July issue of *Nature Methods*<sup>2</sup>.

### REFERENCES:

1. Kim S. *et al. Nat. Methods* **15**, 591-594 (2018) [PubMed](#)
2. Li H. *et al. Nat. Methods* **15**, 595-597 (2018) [PubMed](#)