

TOOLBOX

Machine learning flags ‘mosaic’ mutations that may contribute to autism

BY TARA SANTORA

19 FEBRUARY 2020

A new technique detects rare genetic mutations that occur in a subset of a person’s cells¹. Pinpointing these ‘mosaic’ mutations is the first step in studying their contribution to conditions such as autism.

About 8 percent of the mutations that arise spontaneously in people with autism **follow a mosaic pattern**. The earlier these mutations occur after conception, the more cells they affect. Because mosaic mutations occur in only a fraction of cells, though, studies often **overlook them**. Traditional genetic tests do not always catch them and sometimes misclassify them.

The new technique, called MosaicForecast, uses whole-genome sequencing and machine learning to identify mosaic mutations that appear with varying frequency throughout the body. Researchers tested the method on postmortem brain tissue from 60 autistic people and 15 controls.

People typically have two copies of any stretch of DNA, one from each parent. If they have three copies, one of them is likely to be mosaic. The researchers applied this logic to identify likely mosaics. They first identified candidate mosaic mutations from sequencing data. They then analyzed 25 percent of these variants for more than 30 genetic features — such as relative position in the DNA sequence — to determine the true mosaics.

In the final step, they trained a machine-learning algorithm on the data. They then applied the algorithm to the remaining 75 percent of the mutations to parse the mosaic mutations in that group.

The method correctly flagged up to 90 percent of mosaic mutations affecting a single DNA base, and up to 80 percent of those that are DNA insertions or deletions. It identifies slightly fewer mosaic mutations than other methods do, but it is less likely to mislabel inherited mutations as being mosaic.

The researchers have used the method to identify **mosaic mutations relevant to autism**; their unpublished results are posted on the preprint server bioRxiv.

REFERENCES:

1. Dou Y. *et al. Nat. Biotechnol.* Epub ahead of print (2020) **PubMed**