

NEWS

# Novel software spots scores of new autism candidates

BY JESSICA WRIGHT

19 SEPTEMBER 2016

Scientists have trained an algorithm to recognize the characteristics of autism genes. The program has pegged more than 2,000 new genes to the condition<sup>1</sup>.

The list, published 1 August in *Nature*, could help researchers trace the molecular pathways that underlie autism. It could also help them wade through the scores of new autism candidates to prioritize those warranting further research.

The new method is a departure from the conventional approach, in which researchers find likely autism genes by searching for mutations in people with autism. But most harmful autism mutations are exceedingly rare, so that method is likely to miss many autism genes. What's more, the molecular pathways involved in autism include genes that may never be found in mutated form.

The new method instead takes a new tack: It looks for genes that share features with known autism candidates.

"The estimate of how many genes are responsible for the disorder is vastly larger than what we actually know," says lead researcher **Olga Troyanskaya**, professor of computer science and integrative genomics at Princeton University.

The new method is an alternative to **statistical tools such as TADA**, which rank autism genes based solely on mutational status. It is also different from **an algorithm called DAWN**, which combines TADA data with gene-expression patterns to build networks.

These existing approaches are "highly curated," says **Stephan Sanders**, assistant professor of psychiatry at the University of California, San Francisco, who was not involved in the new study. The new algorithm, by contrast, "says let's put everything in the bucket and see what floats," he says.

However, that also means it is not always clear why the algorithm ranks certain genes higher than others, Sanders says. “It’s very difficult to know why a particular gene ends up in a certain place,” he says.

## Grading genes:

Last year, Troyanskaya’s group reported creating an algorithm to map tissue-specific interactions among all 25,825 genes in the human genome<sup>2</sup>. The network is based on gene expression, protein-binding patterns and other data from more than 14,000 publications. For example, two genes are likely to be connected if their proteins interact and they are expressed at the same time and place.

In their new work, the researchers trained their algorithm to recognize the ‘behavior’ of 594 known autism genes in the brain. They placed these genes into four categories based on their likely role in autism. The highest-ranked genes carry more weight in the algorithm than the low-ranked ones. The researchers then fed the computer 1,189 examples of genes that have never been linked to autism, but are involved in other conditions.

Using this input, the algorithm ranked each gene in the genome for the likelihood that it plays a role in autism. For autism-linked genes, the algorithm excludes their known ties to autism when assigning its score. The researchers found that genes with a score in the top 10 percent have high odds of contributing to the condition.

“For every gene in the genome, we now have a probability that it’s associated with autism. No one has ever been able to do this,” says Troyanskaya. “All the prior methods needed the gene to have prior genetic evidence.”

The top tier includes most genes in which spontaneous harmful mutations crop up in more than one person with autism. It also includes targets of **FMRP** and **CHD8**, two prominent autism candidates.

By contrast, spontaneous, harmful mutations found in the unaffected siblings of people with autism fan uniformly across categories.

## Top tier:

The researchers then looked at long stretches of DNA that are duplicated or deleted, called **copy number variations**, in autism. These stretches encompass multiple genes, and every researcher seems to have a different “favorite” for the link to autism, Troyanskaya says. Looking at the genes’ scores with the algorithm is an objective way to rate their involvement in autism, she says.

To look for key molecular pathways involved in autism, the researchers clustered genes that interact closely. They used databases to look for genes in each cluster that have the same

function. They then predicted a purpose for each cluster — sensory perception, for example.

Finding an autism-linked gene in a certain cluster would signal its particular function, Troyanskaya says.

The researchers also overlaid the top 10 percent of genes onto a map of gene expression in the brain. They found that autism-linked genes tend to be expressed in the brain during fetal development, consistent with previous evidence about **this period's importance in autism**.

The new method is useful because it permits gene discovery using existing data, says **Ivan lossifov**, assistant professor at Cold Spring Harbor Laboratory in New York, who was not involved in the new study. “This approach allows researchers to build on the amount of knowledge they already have.”

#### REFERENCES:

1. Krishnan A. *et al. Nat. Neurosci.* Epub ahead of print (2016) **PubMed**
2. Greene C.S. *et al. Nat. Genet.* **47**, 569-576 (2015) **PubMed**