

NEWS

# The final frontier: Autism geneticists take on the noncoding genome

BY LAURA DATTARO

12 JULY 2022

*Listen to this story:*

<https://www.spectrumnews.org/wp-content/uploads/2022/07/audio-9c40c0fb-805d-4c26-a656-d4aed1f923bb-encodings.mp3>

Like most geneticists, Ryan Doan learned in school that the vast majority of the genome is useless — so-called “junk DNA” that doesn’t code for proteins. But in 2014, while doing postdoctoral research, Doan started to rethink that belief. He was bothered by the fact that autism genetics research, which has largely focused on the coding genome, hasn’t made the progress many had hoped for — especially in providing autistic people with genetic information that informs potential treatments.

“We’re not finding as much as we would’ve thought,” says **Doan**, assistant professor of pediatrics at Boston Children’s Hospital in Massachusetts. “The next best place is trying to branch out into the noncoding regions.”

Now, Doan’s autism research is primarily focused on the largely unexplored 99 percent of the genome that lies beyond the protein-coding exome. According to his **unpublished work**, at least 3 percent of autistic people have noncoding mutations that contribute to their condition.

“The future seems bright, but the noncoding space will be difficult for quite a while.” Ivan Iossifov

Aided by new databases and cheaper whole-genome sequencing, many autism genetics researchers are, like Doan, taking tentative steps into the wide-open noncoding space. Their results so far are mixed, and the challenges remain large. Whole-genome sequencing still costs two or three times as much as exome sequencing, which limits sample size, and the effects of

noncoding mutations are likely to be more subtle than those of their coding counterparts, Doan and other scientists say. But many say they hope that probing noncoding DNA will unearth genetic causes of autism for more people and reveal new details about the condition's biology.

"The future seems bright, but the noncoding space will be difficult for quite a while," says **Ivan lossifov**, associate professor of genetics at Cold Spring Harbor Laboratory in New York. For now, everyone is simply taking baby steps, he says — "very expensive baby steps."

Researchers had no way to navigate the noncoding genome's 3 billion base pairs until the launch of the Encyclopedia of DNA Elements (ENCODE) in 2003. A little more than a decade later, its spinoff, **psychENCODE**, started to map gene regulatory elements within that vast uncharted space in the human brain and other tissues — work that is still underway.

Those maps made it possible for researchers to begin devising targeted strategies to explore the links, if any, between noncoding mutations and autism. It might be tempting to search the entire noncoding space to ensure that important mutations connected to autism aren't missed — especially given how little is known about the DNA there. But starting with stretches of DNA with known functions, such as the promoters and enhancers that help regulate a gene's expression, stands to increase the likelihood that any discovered mutations will be meaningful.

"Some people are very agnostic to location," says **Santhosh Girirajan**, associate professor of genomics at Pennsylvania State University. "And some are looking at some star in some galaxy somewhere."

Promoters — the focus of Doan's study — are located next to the genes they regulate. Enhancers, which may be farther away, carry more mutations in autistic people than in their non-autistic siblings, according to a 2021 **analysis**. In autistic people, genes linked to autism also tend to have an overabundance of transposons — sections of noncoding DNA that can "jump" randomly around the genome and disrupt other genes — another **study** found.

lossifov is surveying yet another source of noncoding DNA: stretches located within genes called introns. About 6 percent of autistic people have an intron mutation that likely contributes to their condition, according to his 2021 **analysis** of these sections in nearly 2,000 autistic children and their non-autistic siblings. To bolster the finding, his team is studying gene expression levels, reasoning that if a gene with an intron mutation has atypical expression in autistic people, it's likely that mutation is involved in the condition.

Early results look "promising," lossifov says. "Expression abnormalities in a gene are rare enough that they can be used as this very useful filter for pointing at de novo noncoding mutations which might be contributory."

For researchers who are exploring the entire noncoding space, machine learning is proving to be a

useful tool. A 2018 **analysis** of whole genomes from nearly 2,000 families with one autistic and one non-autistic child, for example, initially turned up no relevant noncoding mutations compared with controls. But using a machine-learning tool that identifies multiple types of noncoding variants revealed an excess of mutations in promoter regions among the autistic participants.

Similarly, only a neural network trained on functional genomics data could spot differences between autistic children and their non-autistic siblings across some 200,000 noncoding variants in another 2021 **study**. More noncoding mutations occurred near autism-linked genes in children with autism than in those without. Overall, though, noncoding mutations occurred equally closely to the nearest gene in both autistic and non-autistic people, “highlighting the challenge of identifying these causal mutations,” the investigators wrote.

Noncoding and coding mutations may contribute to autism in similar proportions: they are found in about 4.3 and 5.4 percent of autistic children, respectively, according to a 2019 **analysis** that used machine learning to estimate an individual mutation’s likelihood of contributing to the condition.

Yet a third strategy involves looking at the whole noncoding space but limiting the analysis to a cohort that’s more likely to have rare mutations. A February **study** of 22 families with high rates of inter-family marriage, for example, found likely disease-causing variants in promoters and enhancers for five autism-linked genes. The team is now using CRISPR to study the variants’ functions in cells, as well as repeating the work in a new **cohort of African children with autism**.

“Eventually, all of this information in aggregate will be able to tell us about the molecular mechanisms underlying autism,” says lead investigator **Maria Chahrour**, assistant professor of genetics and neuroscience at the University of Texas Southwestern Medical Center in Dallas.

Even when a noncoding mutation contributes to autism, its individual effect is small, the results so far suggest. That means noncoding mutations probably aren’t acting on their own to cause autism, Girirajan says. Rather, several may act together or in tandem with a coding mutation.

How noncoding mutations affect the genome might also be far more subtle and difficult to nail down than for coding mutations. A given mutation may matter in only one cell type or at a specific point in development, for example. Parsing this kind of complexity, while enormously challenging, could help to explain autism’s heterogeneity, Girirajan says. Autism subtypes might reflect not just mutations in a specific gene, but how a gene’s expression varies across time.

“Eventually, all of this information in aggregate will be able to tell us about the molecular mechanisms underlying autism.” Maria Chahrour

“It’s so complex. We are living in a naive land where everything is genes,” Girirajan says. “What we are not thinking about is gene regulation at different stages of development and tissues. Oh gosh.”

To move forward, Girirajan and others say, the field needs to build up whole-genome databases in a big way: At present, autism researchers have access to the exomes of around 50,000 autistic people, and even that has been barely enough to find results in the much simpler coding space, Doan says.

For the noncoding space, “you cut your samples 5-fold, but increase complexity 50-fold,” he says. “You have a huge power problem and that’s just something we have to deal with for a while.”

Geneticists also need to refine the maps that autism researchers are using to find their way. The ENCODE project, for one, is **working to release** data on the time periods and cell types in which promoters, enhancers and other regulatory elements influence genes.

Still, results from other fields are encouraging: Other neuropsychiatric conditions are now linked to many mutations in the noncoding region. Of 22 regions implicated in schizophrenia in **one large study**, for example, 13 are located in noncoding regions within or between genes.

“In autism, this is still behind,” lossifov says, but adds that it is only a matter of time before similar findings emerge. “There’s no doubt.”

**Cite this article:** <https://doi.org/10.53053/WHLV1876>