

VIEWPOINT

For accurate results in autism, genetic databases need diversity

BY STORMY CHAMBERLAIN, LOUISA KALSNER

10 APRIL 2018

Clinicians increasingly are using results from advanced genetic testing technologies to identify a genetic cause for autism in people with the condition. Yet correct interpretation of genetic tests relies, in large part, on an accurate picture of the variation in the genomes of people who are not on the spectrum.

A lack of diversity constrains large genomic databases of the general population in the United States. Most databases do not adequately represent people from minority groups, such as black and Latino people. The frequency of specific genetic variants can vary by ethnic and racial group, so the underrepresentation of these groups affects research and clinical care.

We need to increase diversity in reference genetic databases to correctly interpret results from genetic tests, and connect specific variants to conditions such as autism.

Concern about the lack of diversity in autism studies is not new. Autism clearly affects **individuals of all races and ethnicities**, but disparities persist in access to diagnosis and treatment, and to research studies^{1,2}.

For instance, a review of 408 studies about how people with autism learn found that most of the participants were white³. So the findings may not apply to students from underrepresented groups.

A similar lack of diversity undermines genetic studies. Although **working with isolated populations** can help to rapidly identify rare, harmful mutations, researchers must consider whether those observed associations generalize to a broader population.

Some scientists have developed statistical methods to control for the limited diversity in their studies. But increasing that diversity would yield richer, more informative results.

Uniform data:

Starting in 2014, our team conducted advanced genetic tests in 100 individuals with autism. We also scored these individuals for autism traits and cognitive ability. We used two techniques to identify autism candidate genes: Clinicians typically use chromosomal arrays for testing individuals suspected of having autism; we also sequenced each person's **exome**, the small fraction of the genome that encodes proteins.

Studies have shown that combining these tools provides the best chance of finding a genetic cause for autism if there is one.

For comparison, we used the popular reference database **Exome Aggregation Consortium** (ExAC), a collection of exome sequences from more than 60,000 individuals from the general population.

We noticed an extraordinarily high number of rare, missense variants in the TSC2 gene in our autism population relative to the number in the ExAC database⁴. We found a similarly high frequency of these variants in the exome data housed in the **National Database for Autism Research**, a large government-sponsored repository.

These findings pointed to a role for these variants in autism, and we were excited. But soon, hints emerged to suggest the connection is spurious.

Spurious connection:

First, we found no difference in the rate of these rare variants between individuals with autism and their parents and unaffected siblings. We then began to second-guess our methods after seeing a study revealing that many black Americans were being misdiagnosed with a genetic form of heart disease⁵.

Doctors were flagging missense variants linked to heart disease on their genetic test reports because they are extremely rare in white Americans, and so are also rare in the reference genomic databases. However, the variants are relatively common in black Americans, including many who do not have heart disease — suggesting that they are in fact benign.

The heart disease story was eye-opening to us. Many of the so-called 'rare' variants that we were considering as autism-related were relatively common in at least one subpopulation in the ExAC database, suggesting they do not contribute to autism risk.

The heart disease study attributed misdiagnoses to a "paucity of diverse control data." We came to a similar conclusion. The ExAC database is still not as diverse as the overall U.S. population, let alone our racially mixed study cohort of individuals who visited Connecticut Children's Medical

Center in Hartford.

We recognize that increasing diversity in these databases is challenging and that they will never represent every population studied. But scientists must at least be aware that variants occur at different frequencies in subpopulations and strive to increase the diversity of the databases.

That awareness is essential for interpreting the significance of genes as they relate to a condition. It is also necessary for technicians, genetic counselors and physicians to put the results of genetic tests in proper context for individuals with autism and their families.

Stormy Chamberlain is associate professor of genetics and genome sciences at the University of Connecticut in Farmington. Louisa Kalsner is assistant professor of pediatrics and neurology at Connecticut Children's Medical Center in Hartford.

REFERENCES:

1. Durkin M.S. *et al. Am. J. Public Health* **107**, 1818-1826 (2017) [PubMed](#)
2. Tincani M. *et al. Research & Practice for Persons with Severe Disabilities* **34**, 81-90 (2009)
[Full text](#)
3. West E.A. *et al. J. Spec. Educ.* **50**, 151-163 (2016) [Full text](#)
4. Kalsner L. *et al. Mol. Genet. Genomic Med.* Epub ahead of print (2017) [PubMed](#)
5. Manrai A.K. *et al. N. Engl. J. Med.* **375**, 655-665 (2016) [PubMed](#)