

OPINION, Q&A

# Autism and the complete human genome: Q&A with Evan Eichler

BY CHARLES Q. CHOI

10 MAY 2022

It has been more than 20 years since scientists announced the completion of the Human Genome Project — even though the \$3 billion effort to sequence the 3 billion bases of human DNA was not, in fact, complete. Technological limitations meant that roughly 8 percent of the genome remained a mystery.

In April, the **Telomere-to-Telomere Consortium closed nearly all the gaps**, adding roughly 200 million bases of genetic information that codes for more than 1,900 genes.

This new treasure trove of data, **detailed in six papers in *Science***, stands to advance autism research, says **Evan Eichler**, professor of genome sciences at the University of Washington in Seattle.

*Spectrum* spoke with **Eichler**, who was part of the Human Genome Project and the Telomere-to-Telomere Consortium, about what secrets may emerge from once-murky regions of the genome.

*This interview has been edited for length and clarity.*

**Spectrum: How will having a more complete human genome affect autism research?**

**Evan Eichler:** Because our reference genome was incomplete, some gene sequences were not correctly mapped to their place in the genome. So when we would find a variant in an autism genome that was missing from the reference genome, we didn't always know where it was or which gene — or genes — it affected. This new telomere-to-telomere draft improves mapping across the board. The sequences we gather from people with autism are now more likely to be mapped to the right place.

One phenomenon often associated with autism is the deletion or duplication of DNA, known as

copy number variation (CNV). In our recent paper in *Science*, we analyzed the new **telomere-to-telomere data** and found that it was a better predictor of true copy number 9 times out of 10 when compared with the old reference. That means we're in a much better position to assess CNVs, which we know to contribute to autism, than with the old reference that was full of holes.

**S: How was this updated human genome sequence produced differently from previous ones?**

**EE:** We typically sequence autism genomes with short reads, which are just a few hundred bases long. When we do this, we're missing genetic variation that occur over long stretches of DNA, particularly structural variation such as large deletions, duplications and rearrangements of DNA. We have previously shown that about 75 percent of all structural variation in the human genome goes undetected if we rely only on short-read data. Roughly 10 percent of autism cases stem from known **structural variation in DNA**. If we can sequence the genomes of autism families with long reads, many thousands of bases long, we can explore the 75 percent of structural variation that was previously undetected and potentially find more genetic causes of autism.

**S: Are there potentially new autism-linked genes in the more complete genome?**

**EE:** About 500 genes have been mapped to complex regions of the new reference genome that were previously excluded from sequencing studies because our techniques didn't reliably map there. Among those genes are ones important for brain function, such as **SRGAP2C**. The number of copies of this gene influences where, how and when dendrites form during development, which influences the density and strength of synaptic connections. It's a gene incredibly important to brain function, whose duplicate copies we couldn't reliably detect with short reads.

Another gene, **ARHGAP11B** on chromosome 15, was previously found to be deleted in two people with autism and intellectual disability. It's known to increase neuronal stem cell division during development. That gene is typically not studied in autistic people because it was mapped to very repetitive regions of the genome that previous genome-sequencing techniques skipped over entirely.

**S: What could we learn about autism from the 'dark' regions of the human genome that do not code for proteins?**

**EE:** DNA that makes up the short arms of human chromosomes, called acrocentric DNA, may be important in autism. Those stretches were only sequenced in the last year or so of the Telomere-to-Telomere project.

There are gene families in acrocentric DNA that encode rDNA, which helps form the ribosomes that produce proteins in cells. We know autism is often linked with having too many or too few copies of genes; acrocentric DNA is another category of DNA we can now analyze for the same problem. If

we can compare the rDNA of people with autism with that of neurotypical individuals, any differences we see may help us understand the chances of developing autism.

**S: Now that we have a new reference genome, will some autism studies need to be repeated?**

**EE:** Yes, we'll need to run all autism genomes against this new, more complete reference genome. I'm particularly interested in looking for variations in genes on the X chromosome, which is linked with sex. There are significant sex differences in boys versus girls when it comes to autism, with boys four times more likely to be autistic than girls. Now, with the new reference genome, we can detect copy number variations and other genetic variation better than before, including on the X chromosome.

We would also like to look at unsolved cases of autism — those not linked to any known rare genetic variants and those without high polygenic risk scores, which reflect common genetic variants associated with the condition. These unsolved cases account for a very large fraction of kids with autism. Maybe in the dark regions of the genome, or in genes not characterized before, we can find answers — especially with long-read sequences of Mom, Dad and unaffected siblings to shed light on how these unsolved cases are genetically distinct or similar to their family members.

**S: What about methylated DNA — DNA with chemical tags called methyl groups on top. There is evidence that these epigenetic tags, which influence gene activity, play a role in autism.**

**EE:** With new long-read sequencing techniques such as nanopore sequencing, **we can distinguish methylated sequences** without having to amplify or convert the DNA beforehand, as was necessary with prior techniques. There might be differences in **epigenetic** modifications of DNA between people with and without autism that we missed before, which could help address some of the unsolved cases of autism.

**S: How feasible is it for researchers to conduct long-read sequencing, or for families to access it?**

**EE:** The major limitation is the cost. Sequencing and assembling a genome well with long reads costs about \$10,000, compared with about \$1,000 with short reads. Most insurance companies are not going to pay for long-read sequencing. There's also the issue of throughput. Since it started in 2016, **the SPARK project** has aimed to look at 50,000 families using exome sequences, which capture only the protein-coding regions of the genome. In that same time, we could look at just 50 families with long-read whole-genome sequencing. **[SPARK is funded by the Simons Foundation, Spectrum's parent organization.]**

But costs always come down with time. I think that long reads will replace short reads in 10 years. I

think every family deserves to have their genomes fully sequenced and characterized, to help them make decisions such as what the best care for their children should be. We just have to get the technology to a cost-effective point.

**S: The newly sequenced parts of the genome often contain highly repetitive DNA. Autism has been linked to the presence of such repetitive regions. What might these new regions tell us about autism?**

**EE:** The short answer is we don't know yet. But there is evidence that those regions are very relevant to autism. Two common genetic causes of autism include duplications on chromosome 15q, which account for about 1.5 percent of autism cases, and deletions on 16p11.2, which account for just under 1 percent of autism cases. We know that repetitive regions are hotspots for chromosomal damage that can lead to deletions and duplications, but they weren't precisely mapped. Now we can precisely map these regions of genomic instability and gain insights on how breaks occur there and potentially lead to autism.

**S: Does anything else come to mind with this new work?**

**EE:** One thing that's still a puzzle is that the same genetic variation strongly linked with autism can have very different outcomes in one kid versus another. Some children may be mildly affected, whereas others may be severely affected. I don't think the odyssey of this work ends with finding the primary genetic causes of autism. We need to understand the background in which these variations lie — the way they interact with other genetic variation — to understand their true outcome.

Another thing I have reflected on more recently is how most of the innovations we see with this new project were driven by scientists a full generation younger than me. I think that bodes well for the future, to have so many young people interested in solving these difficult problems. The future of human genetics research is in good hands.

**Cite this article: <https://doi.org/10.53053/GPXL5356>**