

NEWS

# Geneticists join forces for autism sequencing push

BY ALLA KATSNELSON

16 JULY 2012

Strength in numbers: An international consortium plans to combine data from all of the large autism sequencing projects into a single bioinformatics framework.

**Strength in numbers:** An international consortium plans to combine data from all of the large autism sequencing projects into a single bioinformatics framework.

Researchers have launched an effort to yoke together disparate gene sequencing projects in the U.S., Canada and the U.K., aiming to double the number of known autism-related genes in the next three to four years.

The effort, spearheaded by the Autism Sequencing Consortium (ASC), a loose organization of more than 20 independent research groups, plans to pull together 20,000 **exome** sequences — the protein-coding parts of the genome — that have either already been completed or are under way at various institutions. It also aims to sequence another 10,000 exomes.

Mining these sequences should yield 50 to 100 novel autism risk genes, says **Joseph Buxbaum**, ASC's co-director and director of the Seaver Autism Center at Mount Sinai School of Medicine in

New York City.

A flurry of recent studies in autism genetics points to a model in which **many rare, spontaneous — or de novo — mutations** underlie the disease, and researchers increasingly believe they **will need tens of thousands of samples** to identify the genes involved. The only way to achieve that sample size, the researchers say, is through collaboration.

“Most people historically thought they could do this on their own,” says Buxbaum. “Nobody believes that anymore.”

## Sizable scheme:

The ASC formed in early 2010 to accelerate gene discovery in autism by sharing resources and data. It has several funders, including the National Institutes of Health (NIH), the research and advocacy organization **Autism Speaks**, and the Simons Foundation, SFARI.org’s parent organization.

One of its first successes is a trio of papers published in April in *Nature* that represent the **largest exome-sequencing effort** in so-called simplex families. Simplex families have only one child with autism, and unaffected siblings and parents.

Those studies together identified three new autism genes, confirmed three others, and found scores of candidate mutations to be examined.

Calculations from the *Nature* studies and a related paper from **Michael Wigler**’s lab at Cold Spring Harbor Laboratory in New York suggest that **the total number of genes** contributing to autism lies somewhere between 400 and 1,000.

“That’s the landscape,” says Buxbaum — not just for autism but probably for many childhood-onset developmental disorders. “There has to be very strong selection against those variants, so they have to be *de novo* or recent, and they have to be rare.”

The ASC plans to address this scale by combining data from all of the large autism sequencing projects into a single bioinformatics framework. The projects include the **UK10 project**, a U.K.-based initiative funded by the Wellcome Trust that aims to sequence 4,000 individuals from the general population and 6,000 with disorders such as autism; the **Simons Simplex Collection**, which is funded by the Simons Foundation and contains genetic samples from 2,700 families; and the **ARRA Autism Sequencing Collaboration**, a collection of 2,400 samples based at the Broad Institute in Cambridge, Massachusetts, and originally funded by the American Recovery and Reinvestment Act of 2009.

“We’re shooting [to identify] closer to 100 genes, says **Matthew State**, ASC co-chair and professor

of genetics and psychiatry at Yale School of Medicine. “That pool can really be a tipping point in understanding the nature of autism.”

The researchers hope to fund the project with a \$4.5 million grant from the NIH, for which they submitted a proposal last month. The money would allow for some sequencing, but would primarily go towards establishing a bioinformatics infrastructure at a central hub, located at Mount Sinai.

## Central resource:

One of the main goals — and challenges — of the project will be to develop the bioinformatics and statistical methods necessary for analyzing enormous datasets, and for working with sequences pooled from different sources.

The ASC has a ‘variant calling group’ for correctly identifying meaningful variants in the sequences, says Buxbaum. “[Their] job will be to either develop or compare existing tools, and have bake-offs to figure out which ones actually work,” he says. “We’re actually trying to make it open source in the end so that anybody can use it for anything.”

Another issue is the agreement to share data before they’re published — a driving principle of the group.

The principal investigators of the April *Nature* papers are all ASC members, so the teams were able to confirm their findings by comparing them to one another’s prior to publication. “The ability to share data led to the identification of genes that no single group would have identified alone,” says State.

ASC members have to agree that although individual groups can analyze and publish results from their own samples, data produced by the consortium will be published by the ASC as a whole. According to State, most ASC members are already participating in sample collections in which making data available to colleagues is the norm.

“Prospective data sharing is a bit fraught,” Buxbaum says, but people who didn’t think it was feasible when the ASC first launched are starting to come around. “The level of trust increases with time.”

The planned central data repository will not only simplify sharing by entering the data in a single format, it also appeases concerns of rogue usage by recording analyses, he notes.

“Undoubtedly pooling data on 30,000 exomes is going to give us information we couldn’t have gotten from individual studies,” notes **Jonathan Sebat**, associate professor of psychiatry and molecular medicine at the University of California, San Diego, who is not a member of the ASC. He cautions, however, that exome-sequencing research is still evolving, and the kind of innovation

needed to generate computational tools is more likely to come from fired-up postdoctoral fellows in individual labs than from a single centralized effort.

Even among ASC members, the optimism is tinged with caution. Putting together all the proposed samples may not be enough, says **Evan Eichler**, a member of the ASC and professor of genome sciences at the University of Washington in Seattle. To really characterize autism, researchers will need to sequence the exomes of individuals with particular variants.

Funding those sorts of efforts will be even more difficult, Eichler says. “[But] I think that’s actually where we need to move as a field.”