

NEWS

Study links autism to genomic 'hotspots' of mutation

BY EMILY SINGER

21 JANUARY 2013

Cell; Michaelson et al.

Break point: Genomic hotspots, which are stricken with single-letter mutations, often lie in highly conserved regions of the genome.

Cell; Michaelson et al.

Break point: Genomic hotspots, which are stricken with single-letter mutations, often lie in highly conserved regions of the genome.

The genome appears to be littered with hotspots, areas that are prone to single-letter mutations, according to research published 21 December in *Cell*¹. In fact, the mutation rate varies highly across the genome, with the most mutation-prone regions being 100-fold more mutable than the least.

The findings, gleaned from sequencing the whole genomes of ten pairs of identical twins with autism and their parents, suggest that a number of genes linked to autism and other disorders lie within these regions.

The researchers focused specifically on *de novo*, or spontaneous, mutations that arise in the parents' eggs or sperm. They did so by looking for mutations present in both members of a pair of identical twins but not in their parents. **A number of studies** have linked ***de novo* mutations to autism**.

In 2007, lead investigator **Jonathan Sebat**, then a postdoctoral researcher in **Michael Wigler's** lab at Cold Spring Harbor Laboratory in New York, showed that people with autism are more likely to have **copy number variations** (CNVs), small deletions or duplications of DNA, than controls are².

CNVs lie within regions of the genome that are prone to structural changes, also called hotspots, because they have many repeated sequences.

"We wanted to take some of the lessons we learned from our CNV experience and apply them to the entire genome," says Sebat, now chief of the Beyster Center for Molecular Genomics of Neuropsychiatric Diseases at the University of California, San Diego.

The new findings suggest that single-letter mutations also fall into hotspots, though the reasons for that aren't clear.

"It's pretty complex, with a large number of different factors," says Sebat. "We are exploring this more and trying to understand which genes are most likely to break."

Tepid tempo:

The mutation rate varies widely across the genome, with regions ranging from hot to cold. "Most of the genome is tepid," Sebat says, with a mutation rate that varies by one- to three-fold.

About one percent of the genome is particularly vulnerable, with a mutation rate that is more than three times higher than the average. Some regions of the genome are even hotter, with a rate an order of magnitude higher than the average. A number of factors influence how mutation-prone a specific region is, including its sequence and how the DNA is packaged.

"Whether DNA is open and accessible or tightly wound up seems to be one of the strongest predictors of *de novo* mutations," says **Joshua Akey**, associate professor of genome sciences at the University of Washington in Seattle, who was not involved in the study.

Counter to intuition, many of the hotspots are within regions of the genome that remain stable over long periods of time. "The more functionally conserved a region is, the hotter [it is]," says Sebat.

Highly conserved regions typically include genes that are crucial to the function of the cell and may therefore be transcribed, or converted from DNA to RNA, more often than less important genes.

That may make them more susceptible to mutation. But it's also possible that a higher mutation rate provides some evolutionary advantage.

"It opens a whole Pandora's box of questions with respect to human evolution," says Sebat. "Is the hotness ancient or recent, and how has it influenced human evolution?"

The researchers also saw some instances in which a particularly dense cluster of mutations occurred in a narrow genomic region in a single child. These clusters occurred roughly once in each child.

"These are rather stunning observations," says Sebat. "We regard these as a slightly more hazardous type of mutation that has a higher chance of hitting a gene or multiple genes."

Regions prone to CNVs seem to be likely to also harbor single-letter mutations.

Previous sequencing studies of autism did not find the single-letter hotspots because they focused on the **exome**, or protein-coding region of the genome, rather than on the whole genome.

"There's been accumulating evidence that mutations are non-randomly distributed across the genome, but studying *de novo* mutations on a genome-wide scale has been technically impractical until recently," says Akey. "This is one of the most direct studies of *de novo* mutation to date."

Based on their findings, the researchers created a model that can predict with 90 percent accuracy how likely a certain gene is to carry specific mutations, such as those that signal a premature stop to the resulting protein.

The findings should help scientists home in on genomic regions that are important in disorders such as autism. However, given that these hotspots encompass a small proportion of the genome, most autism-linked mutations are likely to lie outside of them.

"Finding areas that are prone to *de novo* mutation is exciting, but we need a larger number of events to understand the mutational properties better," says **Evan Eichler**, professor of genome sciences at the University of Washington in Seattle, who was not involved in the study. "We'll probably need 100 genomes to get the fine scale."

Sebat's team is doing just that — sequencing an additional 110 families, 50 that have a child with autism and 60 control families.

"We are absolutely certain this will hold up with larger numbers," Sebat says. "The question is, what can we learn from larger numbers?"

Researchers plan to enter the genome sequences from this study into the **National Database for**

Autism Research, a **data repository** that is freely accessible to all autism researchers.

References:

1: Michaelson J.J. *et al. Cell* **151**, 1431-1442 (2012) **PubMed**

2: Sebat J. *et al. Science* **316**, 445-449 (2007) **PubMed**