

VIEWPOINT

Focus on function may help unravel autism's complex genetics

BY ARAVINDA CHAKRAVARTI, TYCHELE TURNER

26 JUNE 2012

Hidden links: To identify rare mutations in complex disorders, researchers may need to screen the genomes of thousands of individuals.

In the past few years, technologies that enable the sequencing of whole genomes have made possible numerous advances in genetics. However, understanding the molecular basis of human diseases that have multiple underlying genetic mutations, and unraveling their physiology, remain challenging tasks. Nowhere is this more apparent than in autism for which, despite its high heritability and striking increase in prevalence, the genetic architecture remains foggy. The solution may be to focus on the underlying genetic functions that are disrupted and to identify common affected pathways in the hopes of revealing a mutation 'type' that is common in autism.

A major reason for our ignorance of autism genetics — or the genetics of any multifactorial disease, for that matter — is that our perspective on genetic diseases stems almost entirely from our successful experiences with single-gene disorders¹. Classical Mendelian disorders are invariably rare, but they are the mainstay of human genetics, with inherited symptoms that either are life-threatening or affect one's ability to have children. Natural selection weeds out damaging mutations, which arise spontaneously, or *de novo*, in each new generation.

The mutations and the resulting disease are both rare, usually found in less than 0.1 percent of the population. There are striking exceptions, however: The sickle cell mutation is damaging when present on both gene copies but can be advantageous as a single copy, thereby becoming common — meaning that it is present in greater than five percent of the population.

We do not sufficiently appreciate that the ease of genetic analysis in Mendelian disorders arises from the specific properties of their mutations. Each mutation is necessary and sufficient for the phenotype, or physical traits (thus leading to single-gene inheritance), very rare (owing to a balance between mutation and selection) and absent even in large groups of controls. As a result, these mutations are readily identified in an individual's genetic sequence as highly deleterious function-busters: mutations in sequences that stitch different regions of a gene together, those that prematurely terminate the protein or those that change the protein code at evolutionarily conserved spots. This also appears to be true for mutations in **MECP2**, **FMR1**, **PTEN** and **SHANK3**², which have been identified in a minority of people with severe autism. The frustrating difficulty for understanding the majority of multifactorial autism cases is that none of these three features appear to be common.

Fascinating puzzle:

Geneticists have latched onto investigating autism precisely because it is a fascinating genetic puzzle, which when solved will deeply contribute to understanding its molecular basis and lead the way to directed therapies. At a basic level, our task is to identify the specific genes or alleles — alternative forms of a gene on a single chromosome — involved, the nature of harmful alleles, the cumulative effect of these alleles in individuals with the disorder, and how they relate to autism symptoms.

Screening the genomes of many people with autism, whether by using genetic sequences arrayed on diagnostic chips or by sequencing the coding region or even the entire genome, has generated the primary data for autism gene discovery. But our track record so far is poor^{3,4,5}. Studies have identified many autism gene candidates, but frankly, few of their findings are replicable or have the strength of genetic evidence to provide new insight or to help us pursue new therapies⁶. This has led to calls for examining **larger and larger numbers** of genomes of people with the disorder, which may indeed be necessary. But we also need to examine and understand why autism mutations themselves have been elusive.

How does one identify autism mutations spread out across many genes, with effects that are neither necessary nor sufficient, with frequencies either rare or common, and that are also present in controls?

The geneticist's answer to this question is 'enrichment analysis': showing a statistically significant excess of a variant in individuals with the disorder than in controls, independent of its overall frequency.

There are two realities that compromise this approach: First, most genomic variants are rare and require large sample sizes to confirm their role. Second, capturing causal variants by capturing all variants also requires large sample sizes to ensure that the enrichment observed is not a false positive. Not many autism gene candidates would pass statistical stringency tests, making it crucial to increase study sizes.

Scanning the entire genome for alleles identifies both the minority of alleles that are linked to autism and the majority of millions of autism-unrelated variants. The major challenge for genetics research is whether we can sort the autism-causal from the autism-irrelevant variants using biological clues.

In the future, when every base in the human genome has been accurately and comprehensively annotated with respect to function, this will be possible. But annotation of genomes is far from complete and depends on large-scale functional analyses at multiple levels, which are still in their infancy⁷.

Defining features:

Given the state of the field, even recognizing all Mendelian-disease alleles has been challenging, and the causative mutation in about ten percent of cases remains unidentified. Clearly, we are not perfect in recognizing pathogenic, or harmful, coding alleles that can compromise function in multifactorial disorders; doing so for the noncoding pathogenic alleles is very much harder⁸. How do we proceed with this challenge?

Multifactorial-disease alleles must have some defining genetic features, although they may be different from Mendelian ones. The way forward, rather than the brute-force approach of looking at enrichment in tens of thousands to hundreds of thousands of individuals⁹, is to understand what these defining features are.

I suspect that in-depth studies into a few such genes may be revealing of underlying biology and defining genetic features. We need to answer a number of questions: Are the mutations mostly in gene coding or noncoding regions? Are they rare or common? If coding, then which biological processes do they disrupt? If noncoding, which regulatory changes are most often involved? Are they at the DNA, RNA or protein level, and how does alteration of gene dosage affect the relevant biological processes? If multiple genes are involved, then how do they relate to one another? Why do not all genes in a pathway harbor autism variants? And finally, does disrupting the same genes, by mutation or environmental perturbation, in an experimental model produce the same phenotype¹⁰?

We have proceeded for 20 years believing that we can identify disease genes by their genomic location alone, without considering the influence of biochemistry, cell biology or physiology¹¹. Perhaps it is time to borrow from these fields to obtain a deeper knowledge of the functions

encoded in our genomes and how they are compromised in multifactorial human disease.

In the end, the genetic features of autism might be 'simple' as well, but elucidating this simplicity will require great effort. For example, one compelling hypothesis for autism susceptibility is that, just as for Mendelian disorders, an amalgam of harmful alleles at many loci are, as a whole, necessary and sufficient, rare, and absent in controls.

Environmental factors or the influence of development may also act on these genotypes to modulate their responses, as they do not on the genes for Mendelian traits. However, these agents act on specific genes and not the whole genome. If this is true, then using this understanding to identify autism genes will be groundbreaking.

Aravinda Chakravarti is the director of the Center for Complex Disease Genomics at the Johns Hopkins University School of Medicine. Tychele Turner is a graduate student in his laboratory.

References:

- 1: **Antonarakis S.E.** et al. *Nat. Rev. Genet.* **11**, 380-384 (2010) [PubMed](#)
- 2: **Abrahams B.S. and D.H. Geschwind.** *Nat. Rev. Genet.* **9**, 341-355 (2008) [PubMed](#)
- 3: **Arking D.E.** et al. *Am. J. Hum. Genet.* **82**, 160-164 (2008) [PubMed](#)
- 4: **Weiss L.A.** et al. *Nature* **461**, 802-808 (2009) [PubMed](#)
- 5: **Sanders S.J.** et al. *Nature* **485**, 237-241 (2012) [PubMed](#)
- 6: **Pinto D.** et al. *Nature* **466**, 368-372 (2010) [PubMed](#)
- 7: **Birney E.** et al. *Nature* **447**, 799-816 (2007) [PubMed](#)
- 8: **Chakravarti A. and A. Kapoor** *Science* **335**, 930-931 (2012) [PubMed](#)
- 9: **Kryukov G.V.** et al. *Proc. Natl. Acad. Sci. USA* **106**, 3871-3876 (2009) [PubMed](#)
- 10: **Sparrow D.B.** et al. *Cell* **149**, 295-306 (2012) [PubMed](#)
- 11: **Collins F.S.** *Nat. Genet.* **1**, 3-6 (1992) [PubMed](#)