

TOOLBOX

‘Dosage sensitivity map’ predicts active ingredients in copy number variants

BY PETER HESS

31 AUGUST 2022

A new map catalogs hundreds of rare copy number variants (CNVs) — duplications, deletions or insertions of large DNA segments — associated with 54 conditions, including autism. It also pinpoints individual genes within those CNVs that likely drive the various conditions’ traits, researchers explain in a new [study](#).

The reference could inform statistical models to predict which CNVs are more or less likely to be abundant among people with certain conditions, says study investigator [Ryan Collins](#), a graduate student in [Michael Talkowski](#)’s lab at Massachusetts General Hospital in Boston. “This is especially true for neurodevelopmental phenotypes like autism spectrum disorder, and others.”

Rare CNVs — those that show up in less than 1 percent of the population — can predispose people to a range of traits and neurodevelopmental conditions, [including autism](#) and some autism-linked syndromes. Scientists have suspected that the doubling or loss of certain genes within these CNVs leads to neurodevelopmental conditions, but they have historically struggled to identify these ‘driver genes,’ Collins says.

“We saw this problem: the need for better interpretation of these copy number variants,” Collins says. “We set out to build new, larger catalogs of these copy number variants in as many people as we could aggregate into a single research study.”

The analysis built upon existing CNV data from 17 different institutions and 950,278 people, about half of whom have at least one of 54 different physical or neurodevelopmental traits or conditions. The team systematically reviewed 200,000-nucleotide-long strips of each chromosome at a time, shifting focus by only 10,000 nucleotides from one strip to the next.

They identified a total of 558,113 rare CNVs of 100,000 or more nucleotides in size, a relatively “coarse resolution,” says senior investigator Talkowski, director of the Center for Genomic

Medicine at Massachusetts General Hospital.

Of those, 163 CNVs are significantly more common among people with health conditions than in controls, and 15 others have lower-confidence links, the team found. These 178 CNVs were significantly more likely than other CNVs to contain genes previously associated with any of 95 conditions.

All but 12 of the CNVs overlap with protein-coding genes. And more than half contain at least one driver gene, identified by an algorithm that combs through each variant for ‘dosage sensitive’ genes: those that are rarely mutated in the general population and have been previously linked to health or neurodevelopmental conditions. They identified 121 genes in total, including such top autism candidates as **MAGEL2**, **RAI1**, **SHANK3** and **UBE3A**.

A machine-learning model the team created scored the dosage sensitivity of 18,641 protein-coding genes throughout the genome, flagging 2,987 that are sensitive to deletions and 1,559 sensitive to duplications.

The paper describing the algorithm and ‘dosage sensitivity map’ appeared in August in *Cell*.

The procedures and algorithm the team employed can help other researchers analyze additional genetic datasets, says **David Ledbetter**, adjunct professor of psychiatry at the University of Florida in Gainesville, who was not involved in the work. A similar set of tools has been used to **predict the effect sizes** of CNVs on autism and IQ scores, and it will be interesting to compare the results of the two approaches, he adds.

This work can help clarify the relationship between a CNV’s size and its effect — two very different, often conflated, measures, says **Christa Lese Martin**, chief scientific officer at Geisinger Health System in Danville, Pennsylvania, who was not involved in the work. “You could have a [10 million nucleotide base-pair] deletion with no haploinsufficient genes in that region, versus a [100,000 base-pair] deletion with a gene that causes significant phenotypic effects.”

The map could also help researchers better understand the effects of variants in **noncoding genes**, which do not encode proteins, says **Sarah Elsea**, professor of molecular and human genetics at Baylor College of Medicine in Houston, Texas, who was not involved in the work. “Noncoding variation may have significant effects on gene expression but might be excluded or prioritized differently if the nearby gene was not known to be dosage-sensitive.”

The team is following up on this study by mapping CNVs at a higher resolution — 10,000 or even 1,000 nucleotides at a time, instead of 100,000, Collins says.

Cite this article: <https://doi.org/10.53053/WBFT2950>