

TOOLBOX

Statistical model rates billions of human mutations

BY JESSICA WRIGHT

16 APRIL 2014

A new statistical system ranks the potential harm done by each of the nearly 9 billion possible variants in the human genome, researchers reported in March in *Nature Genetics*¹.

The method — inspired by Nate Silver’s successful prediction of the 2012 presidential election — integrates 63 existing ratings into a single score for each variant.

Other available systems predict whether a certain variant probably affects the function of the protein or assess whether a variant is conserved: Variants that survive for millions of years are likely to be important.

Because these systems all measure such different things, however, no one rating has been able to rank all the types of variations seen in the genome. In particular, researchers were not able to directly compare variants in genes with those in regulatory regions.

This makes it difficult to interpret the massive amounts of data generated by next-generation sequencing studies, says lead researcher **Jay Shendure**, associate professor of genome sciences at the University of Washington. “Our ability to interpret genome sequence data hasn’t really kept up with our ability to sequence,” he says.

To aggregate these ratings into a single score, Shendure and his colleagues built a prediction model using 30 million variants, half of which represent harmful changes. The other half have remained stable during six million years of evolution and so are highly unlikely to be detrimental.

The researchers then looked at how well each of the 63 ratings predicts which group a specific variant in the model belongs to. They used this data to fuse these 63 ratings into a single combined measure, which they applied to each of 8.6 billion known variants.

As expected, mutations that disrupt a protein's sequence have the highest combined scores. The lowest scores are for variants found between genes. However, there are some surprises. For example, mutations that disrupt genes involved in scent have lower scores than similar mutations in other genes.

The researchers also looked at more than 1,600 variants found in children who have either autism or intellectual disability. Overall, those variants are more harmful than the ones seen in controls, but variants linked to intellectual disability are more severe than those associated with autism.

Other researchers have already used the scores in a study published last month in the *American Journal of Human Genetics*². They showed that autism-linked **variants inherited from a mother are more harmful** than those passed down from the father. The combined scores are **freely available online**.

References:

1: Kircher M. *et al. Nat. Genet.* **46**, 310-315 (2014) [PubMed](#)

2: Jacquemont S. *et al. Am. J. Hum. Genet.* **94**, 415-425 (2014) [PubMed](#)