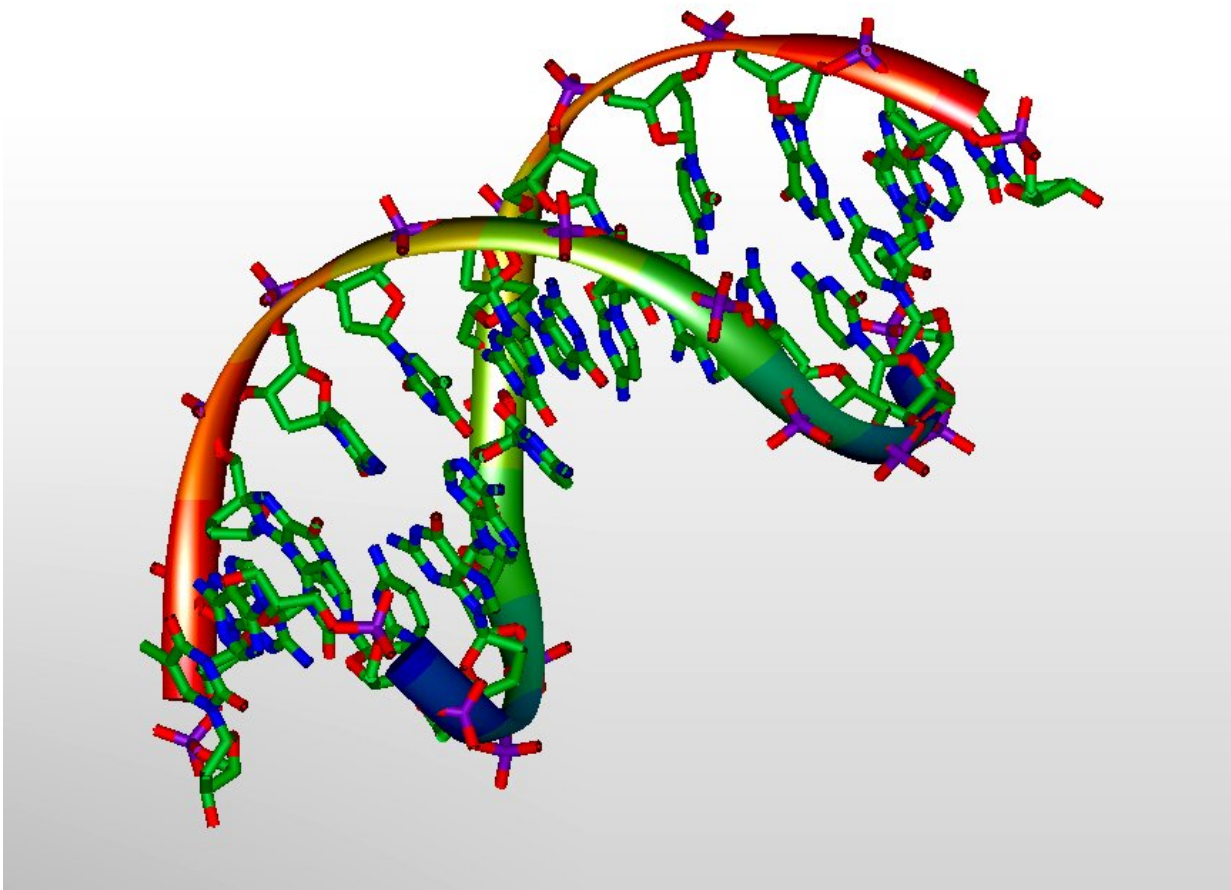# Researchers discover better way to identify DNA variants

July 12 2021

3D-model of DNA. Credit: Michael Ströck/Wikimedia/ GNU Free Documentation License

USC researchers have achieved a better way to identify elusive DNA variants responsible for genetic changes affecting cell functions and diseases.

Using computational biology tools, scientists at the university's Dornsife College of Letters, Arts and Sciences studied "variable-number tandem repeats" (VNTR) in DNA. VNTRs are stretches of DNA made of a short pattern of nucleotides repeated over and over, like a plaid pattern shirt. Though they comprise but 3% of the human genome, the repetitive DNA governs how some genes are encoded and levels of proteins are produced in a cell, and account for most of the structural variation.

Current methods do not accurately detect the variations in genes in some repetitive sequences. The new method by the USC scientists can detect variants among different populations of people and how it they affect gene expression, which helps to discover links between VNTR variation and traits or disease.

"This type of repetitive DNA has been called 'dark matter' of the human genome because it has been difficult to sequence and analyze how it varies," said Mark Chaisson, assistant professor of quantitative and computational biology and corresponding author of the study. "We showed that variation in dark matter can have a substantial effect on cellular processes, so future studies may use this approach to understand the genetic basis of disease and ways to improve our health."

The study was published in *Nature Communications* on July 12.

The study also says:

- Variants in genetic codes are responsible for Huntington's disease, Lou Gehrig's disease (ALS), schizophrenia, diabetes and attention-deficit disorder, as previous research has shown.

- While other tools, based on algorithms, have been developed to detect genetic variants, they provide incomplete information, especially for the VNTRs.
- The new software tool the USC scientists developed derives from a repeat-pangenome graph, a data structure that encodes population diversity and repetitions of VNTR locations on a chromosome to identify more gene sequences with better accuracy.

Provided by University of Southern California