

# Understudied mutations have big impact on gene expression

April 12 2021, by Joshua Baxt

---



Credit: CC0 Public Domain

An international team of researchers led by computer scientists at the University of California San Diego have identified 163 variable number tandem repeats (VNTRs) that actively regulate gene expression. In a paper published in *Nature Communications* this week, the researchers provide new insights into this understudied mechanism, how it may drive disease and other traits and could ultimately impact patient care.

VNTRs are common genomic variations that appear as repeated DNA sequences longer than six base pairs.

"VNTRs have been difficult to identify through sequencing information, particularly short reads," said Vineet Bafna, a professor in the UC San Diego Department of Computer Science and Engineering and senior author on the paper. "We developed a fast computational method to identify these variations, which gave us a new lens to observe their potential impact on [gene expression](#) and disease."

While VNTRs are one of the most common genetic variations, detecting them has been both challenging and expensive. As a result, [genome-wide association studies](#) have mostly ignored VNTRs, excluding a potentially rich repository of disease-causing mutations.

Using the new method they developed, researchers found more than 10,000 VNTRs in a group of 652 people. Of those, 163 changed the way genes were expressed in 46 different tissue types. Nearly half had a particularly strong impact on Alzheimer's disease, familial cancers, diabetes and other conditions.

"We looked closely at these 163 VNTRs to understand their impact,"

said UC San Diego computer science Ph.D. student Mehrdad Bakhtiari. "One was affecting expression levels of a gene called AS3MT, which has been implicated in schizophrenia. The tandem repeat controls gene expression and the expression affects the disease."

## **A new neural-network based method**

The new computational method, called adVNTR-NN, rapidly finds VNTRs in short read genomic sequencing data. Short read technology breaks up DNA samples into relatively small pieces and uses sophisticated algorithms to piece together complete sequences. Popularized by Illumina, short reads are the most common, and least expensive, form of next-generation sequencing.

"Our computational method runs quite fast and it works in the most cost-effective sequencing technologies," said Bakhtiari, the first author on the paper. "So, it should be easy to scale and easy to use for people outside of research settings, for example, hospitals."

## **Next steps**

The authors have high confidence in these results, as 91% of the initial findings were validated in two other cohorts. However, there is still more work to do. The researchers want to better quantify the impact VNTRs have on specific [genes](#), as well as determining whether there is indeed a causal link between certain VNTRs and the diseases they have been linked to.

"We're learning more about the importance of complex and repetitive regions of the genome and the roles they play in human traits," said Melissa Gymrek, assistant professor in the Computer Science and Engineering department and the School of Medicine and coauthor on the

paper. "VNTRs have been almost impossible to look at in most available sequencing datasets on a large scale."

The team also wants to study specific conditions, such as autism spectrum disorder, to better understand patients' VNTR profiles and how those influence their conditions.

"These findings open a whole new window to study the role VNTRs play in gene expression," said Bafna. "Given that these sequences are physically larger than [single nucleotide polymorphisms](#) and are also quite prevalent, these efforts could eventually have a major impact on patient care."

**More information:** Mehrdad Bakhtiari et al. Variable number tandem repeats mediate the expression of proximal genes, *Nature Communications* (2021). [DOI: 10.1038/s41467-021-22206-z](https://doi.org/10.1038/s41467-021-22206-z)

Provided by University of California - San Diego

Citation: Understudied mutations have big impact on gene expression (2021, April 12) retrieved 28 April 2024 from <https://phys.org/news/2021-04-understudied-mutations-big-impact-gene.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.