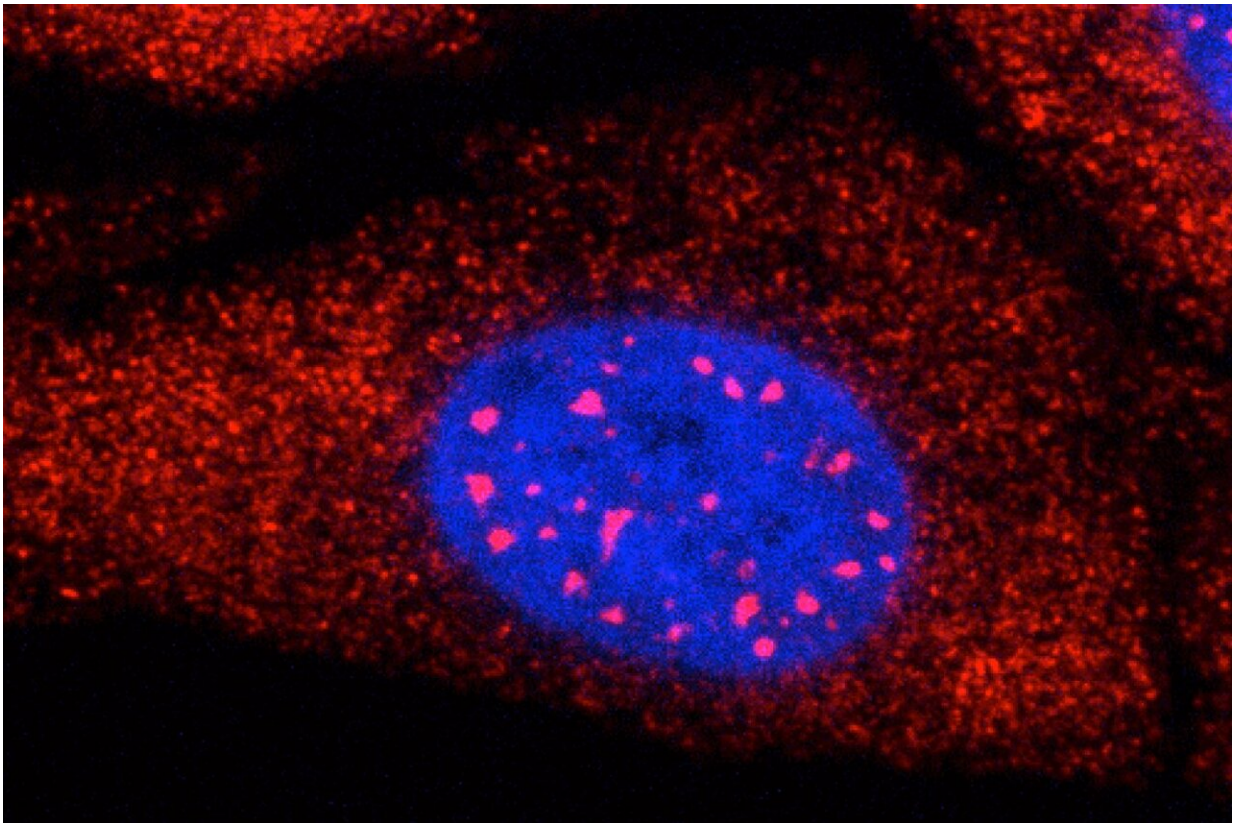


Protein production glitches in Huntington's disease revealed

February 19 2024, by Shafaq Zia



RNAs with expanded CAG repeats (red) form regions of high concentration in the nucleus (blue) of a cell. Credit: Rachel Anderson/ Whitehead Institute

DNA, often thought of as "the blueprint of life," contains instructions for building proteins that cells need to survive and function properly. But

DNA isn't perfect and errors can occur during replication. Sometimes, this can result in snippets of the DNA building blocks called nucleotides—G (guanine), A (adenine), T (thymine), C (cytosine)—getting repeated too many times in a row.

This can lead to a type of mutation, known as a nucleotide repeat expansions, that can alter the function and structure of vital proteins and can give rise to rare neurodegenerative conditions like Huntington's disease and amyotrophic lateral sclerosis (ALS).

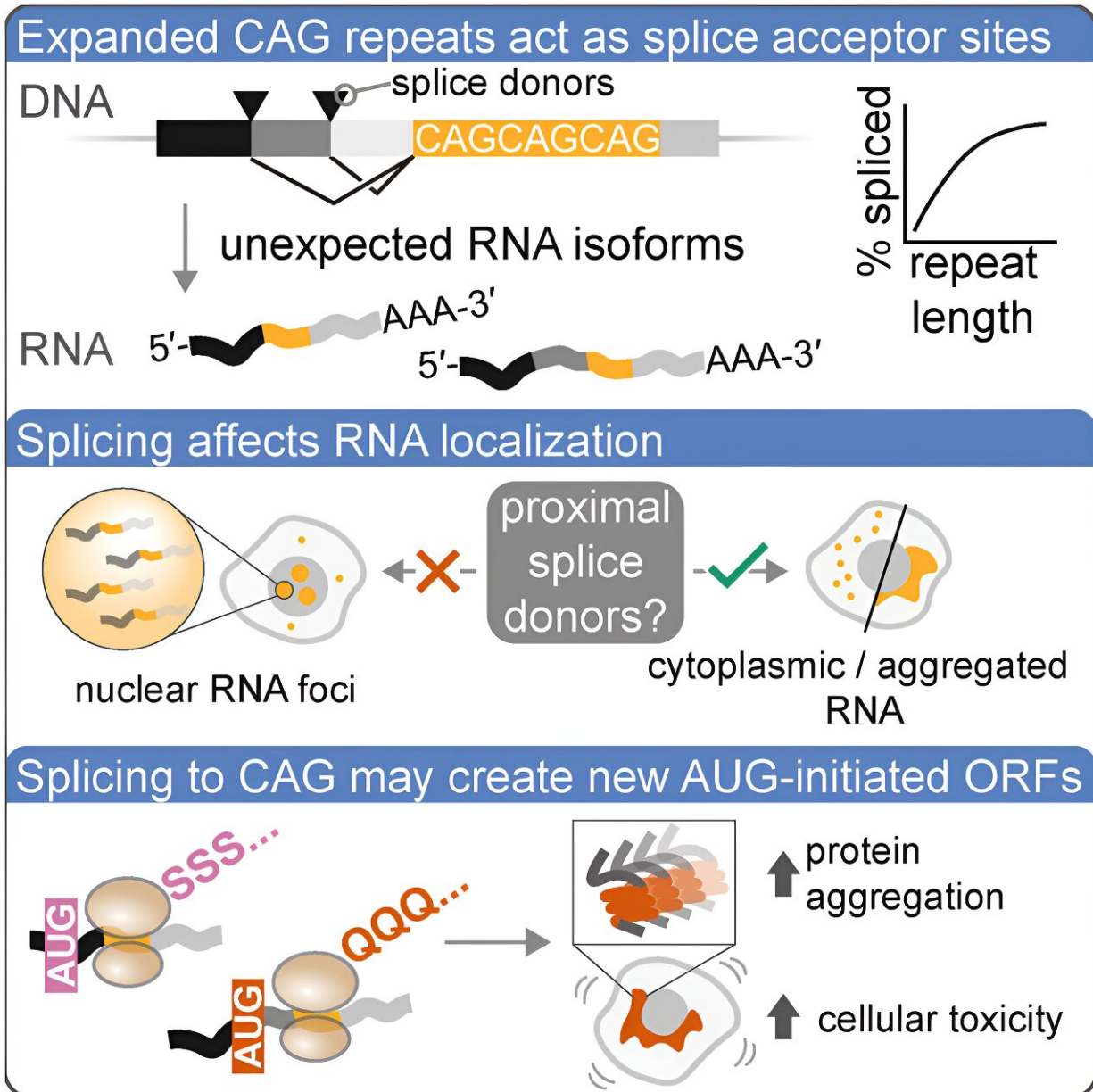
New research from Whitehead Institute Member Ankur Jain, graduate student Rachel Anderson, and their colleagues takes a closer look at how the repeat sequence involved in Huntington's disease—a CAG repeat—leads to the production of abnormal proteins that misfold and clump within cells, clogging up important cellular processes.

[Their findings](#), published in the journal *Molecular Cell* on January 30, reveal that the expanded CAG repeat can interfere with splicing. As depicted in the illustration below, this is the process where portions of RNA that do not encode proteins, also known as introns, are cut out. The remaining sections, called exons are then joined together to form the final messenger RNA that carries instructions for building a [protein](#).

According to the researchers, the expanded CAG repeat creates new markers, or splice acceptor sites, which leads to the cutting and pasting of genetic information to occur at different junctions than usual.

"The question of why the brains of patients with repeat expansion disorders have spurious proteins has confounded scientists for some time," says Jain, who is also an assistant professor of biology and a Thomas D. and Virginia W. Cabot Career Development Professor at the Massachusetts Institute of Technology. "Now, because we have an understanding of the molecular mechanism, we can try to target the

splicing pathway and diminish the production of these proteins."



Credit: *Molecular Cell* (2024). DOI: 10.1016/j.molcel.2024.01.006

Unfolding RNA hairpins

RNA is less stable than DNA, and common RNA analysis approaches rely on an enzyme called reverse transcriptase. Although usually in a cell, DNA is read into RNA, this enzyme reads RNA molecules into a complementary DNA strand (cDNA). This allows the researchers to closely analyze the RNA sequences without risking degradation of genetic information.

But reverse transcription of repeat-containing RNAs comes with its own challenges—these molecules tend to fold back on themselves, forming hairpin loops, and when these loops do not fully unwind during reverse transcription, researchers are left with gaps and errors in the cDNA.

In the new paper, Jain and Anderson used a different approach to sensitively reverse transcribe repeat-containing RNAs into cDNA. Specifically, the researchers worked with an enzyme called TGIRT (Thermostable Group II Intron Reverse Transcriptase) that stays active at high temperatures, allowing it to break open the hairpin structures and capture repeat-containing sequences at a higher fidelity.

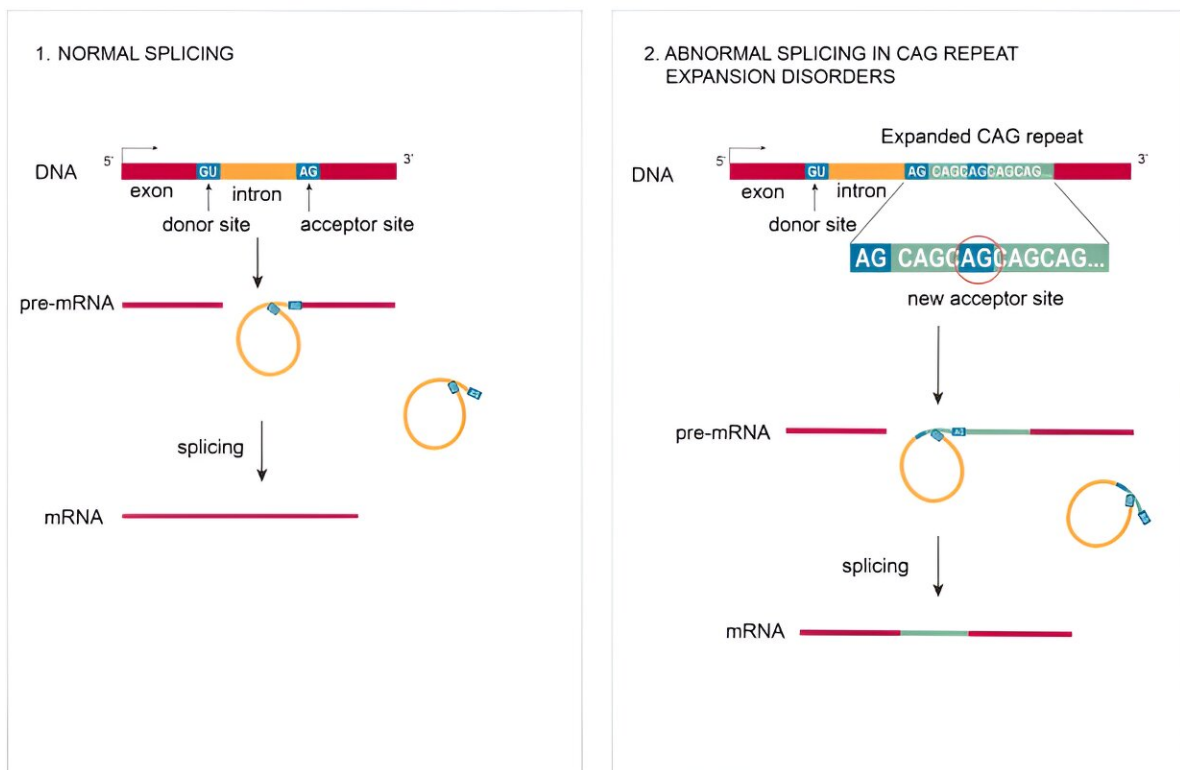
"When you heat up an egg, it turns yellow because the proteins in the egg are unfolding due to high temperature. We're exploiting the same thing but with RNA structures," says Anderson.

Then the researchers began mapping these repeats onto a reference genome, which serves as a guide for genetic information in a human, but they quickly ran into challenges. The "letters" that make up a human genome G-A-T-C combine in various sequences to form the strands of DNA in our cells.

This means, repeated patterns in the human genome are inevitable (repeat-based diseases only arise when a single sequence—like CAG—is repeated too many times in a row) and each pattern can occur at multiple locations in the genome. So pinpointing where the repeat-containing

RNA originated is like reconstructing a story from fragmented sentences without context.

"This is when we decided to approach the mapping of repeats differently," says Anderson. The researchers developed a new tool, called SATCfinder, that selects RNA sequences with at least three CAG repeats. These repeats are then computationally trimmed and the rest of the sequence is mapped onto the reference genome. The location, or mapping coordinates, of the pattern immediately before the CAG repeat is tracked, allowing the researchers to then figure out exactly where the repeats are supposed to go.



The expanded CAG repeat creates new markers, or splice acceptor sites, which leads to the cutting and pasting of genetic information to occur at different junctions than usual. This gives rise to abnormal RNA transcripts. Credit:

Jennifer Cook-Chrysos/ Whitehead Institute

Taking a closer look at splicing

[Previous work](#) from the Jain lab has shown that once repeat-containing RNAs leave the nucleus and reach the cell's cytoplasm, they form gel-like clumps.

Typically, in the cytoplasm, RNAs interact with cellular machinery that looks for a marker on RNA, called a start codon, to begin translating instructions for building proteins. Researchers have theorized that repeat-containing RNAs could be confusing the machinery, influencing it to translate instructions from different starting points. This process, called RAN translation, could then be leading to the creation of unneeded proteins that are not only prone to clumping together but also contribute to the gumming up of RNA in the cytoplasm.

But this explanation was not entirely satisfactory to Jain and Anderson, and they wanted to learn more about why repeat-containing RNAs lead to a haphazard translation of instructions in the first place. To investigate this, they created a set of sequences with the pattern "CAG" repeated 240 times consecutively. Just as they had expected, when these sequences reached the cytoplasm, they began to aggregate.

When the researchers performed RNA sequencing on these cells and analyzed the results using SATCfinder, they found their answer: CAG repeats in the RNA were often stitched to unexpected sequences, farther from the repeat in the DNA, with the intervening regions cut out. This meant that the presence of CAG repeats multiple times in a row led to the creation of new cutting and pasting sites at the edges of the repeat itself, creating abnormal RNA transcripts that then produce proteins that

misfold and clump together.

Now, researchers in the Jain lab are interested in further investigating how the expanded CAG repeat induces errors in splicing. They are also hoping to learn more about the extent to which these splicing errors contribute to the pathology of conditions like Huntington's disease.

"There's an array of mechanisms that come together and contribute to cell death in Huntington's. This is one piece of the puzzle that contributes to our molecular understanding of how these repeats distort cellular functions," says Jain.

More information: Rachel Anderson et al, CAG repeat expansions create splicing acceptor sites and produce aberrant repeat-containing RNAs, *Molecular Cell* (2024). [DOI: 10.1016/j.molcel.2024.01.006](https://doi.org/10.1016/j.molcel.2024.01.006)

Provided by Whitehead Institute for Biomedical Research

Citation: Protein production glitches in Huntington's disease revealed (2024, February 19)
retrieved 28 April 2024 from
<https://medicalxpress.com/news/2024-02-protein-production-glitches-huntington-disease.html>

| |
|--|
| <p>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.</p> |
|--|