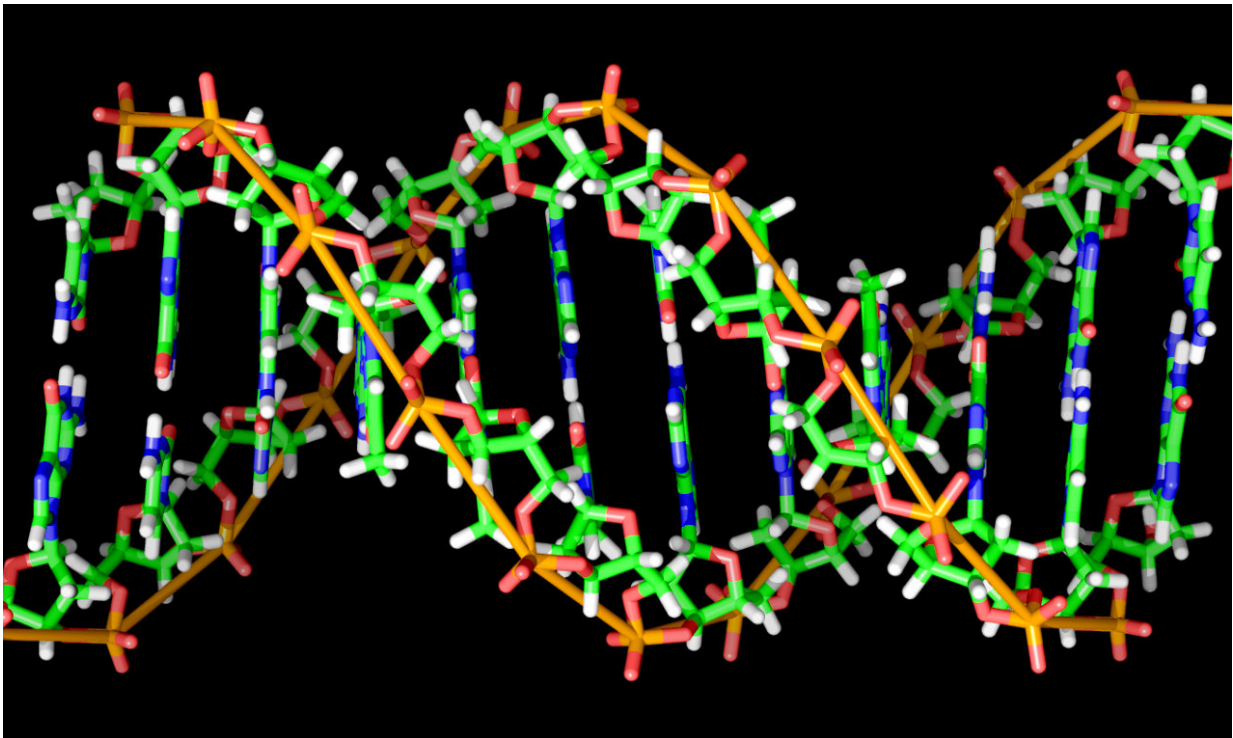


DNA barcodes that reliably work: A game-changer for biomedical research

June 21 2018



This illustration shows the most common structure of DNA found in a cell, called B-DNA. Credit: Richard Wheeler (Zephyris). Used under the Creative Commons Attribution-ShareAlike 3.0 license.

In the same way that barcodes on your groceries help stores know what's in your cart, DNA barcodes help biologists attach genetic labels to biological molecules to do their own tracking during research, including

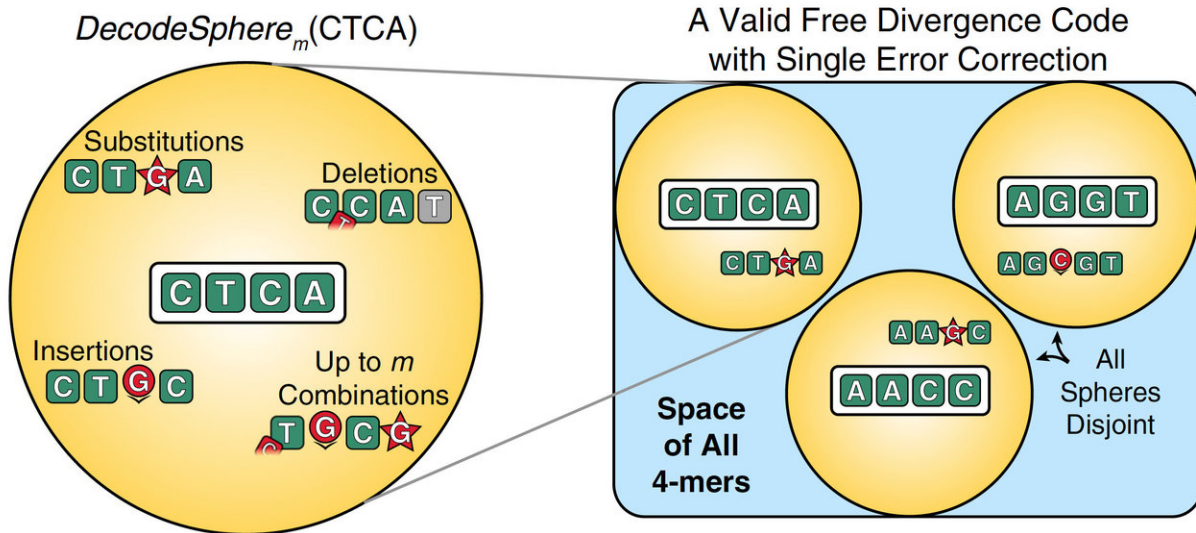
of how a cancerous tumor evolves, how organs develop or which drug candidates actually work. Unfortunately with current methods, many DNA barcodes have a reliability problem much worse than your corner grocer's. They contain errors about 10 percent of the time, making interpreting data tricky and limiting the kinds of experiments that can be reliably done.

Now researchers at The University of Texas at Austin have developed a new [method](#) for correcting the errors that creep into DNA barcodes, yielding far more accurate results and paving the way for more ambitious medical research in the future.

The team—led by postdoctoral researcher John Hawkins, professor Bill Press and assistant professor Ilya Finkelstein—demonstrated that their new method lowers the error rate in barcodes from 10 percent to 0.5 percent, while working extremely rapidly. They describe their method, called FREE (filled/truncated right end edit) barcodes, today in the journal *Proceedings of the National Academy of Sciences*.

The researchers have applied for a patent and are making the method freely available for academic and noncommercial use.

With DNA barcodes, scientists can study how a cancerous tumor evolves, not just as a whole, but as a large collection of individual cells that evolve differently to reveal which cells are vulnerable to therapeutics and which aren't. Scientists interested in growing replacement organs for injured or sick people can use DNA barcodes to better understand how organs naturally develop. And researchers looking to screen millions of potential drugs to find one that binds to a certain molecule, and thus has the potential to treat a disease, can use DNA barcodes to find the proverbial needle in a haystack.



A key step in the new method for correcting errors in DNA barcodes involves creating "decode spheres" around each potential barcode and then solving a "sphere packing" problem to maximize the number of usable barcodes. Credit: University of Texas at Austin

"DNA barcodes are a part of a great deal of cutting-edge research in medicine and drug development, and to be able to improve the accuracy and efficiency of so many of these is very exciting," said Hawkins. "And maybe even more exciting is that now with these better barcodes, this allows us to have larger, more ambitious experiments that weren't possible before."

A DNA [barcode](#) contains a short string of letters that equates to a unique code, using the four letters found in DNA: A, C, G and T. These barcodes are stuck onto molecules, such as cellular proteins or drug candidates, as a way of keeping track of where they all go, sometimes by the millions, and how they interact with other molecules. About one-tenth of the time, however, errors occur—such as one letter being

replaced by the wrong letter, an extra letter being inserted, or a letter being deleted—potentially skewing the results of critical biomedical research.

One of the keys to this new error-correction method is to select just the right barcodes from the beginning. This method involves choosing a string of letters for each barcode such that even if a small error creeps in—say, a G is substituted for a C—it will still be more like the intended barcode than any other. The method requires throwing out many possible strings of letters, but the researchers minimized this loss by borrowing an approach from computer science called sphere packing.

"My contribution has been designing a way to find those barcodes such that even if there is an error in it, you know which original barcode it came from," Hawkins said.

Alternative error-correcting methods for DNA barcodes, such as what are known as Levenshtein codes, require throwing away up to 100 times as many barcodes as with the FREE method, and they are up to 1,000 times slower to decode the results. As a result, whereas existing technology made projects with hundreds of millions of barcodes nearly impossible, the new technology allows for rapid, accurate results.

More information: John A. Hawkins et al. Indel-correcting DNA barcodes for high-throughput sequencing, *Proceedings of the National Academy of Sciences* (2018). [DOI: 10.1073/pnas.1802640115](https://doi.org/10.1073/pnas.1802640115)

Provided by University of Texas at Austin

Citation: DNA barcodes that reliably work: A game-changer for biomedical research (2018, June 21) retrieved 25 April 2024 from <https://phys.org/news/2018-06-dna-barcodes-reliably-game->

changer-biomedical.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.