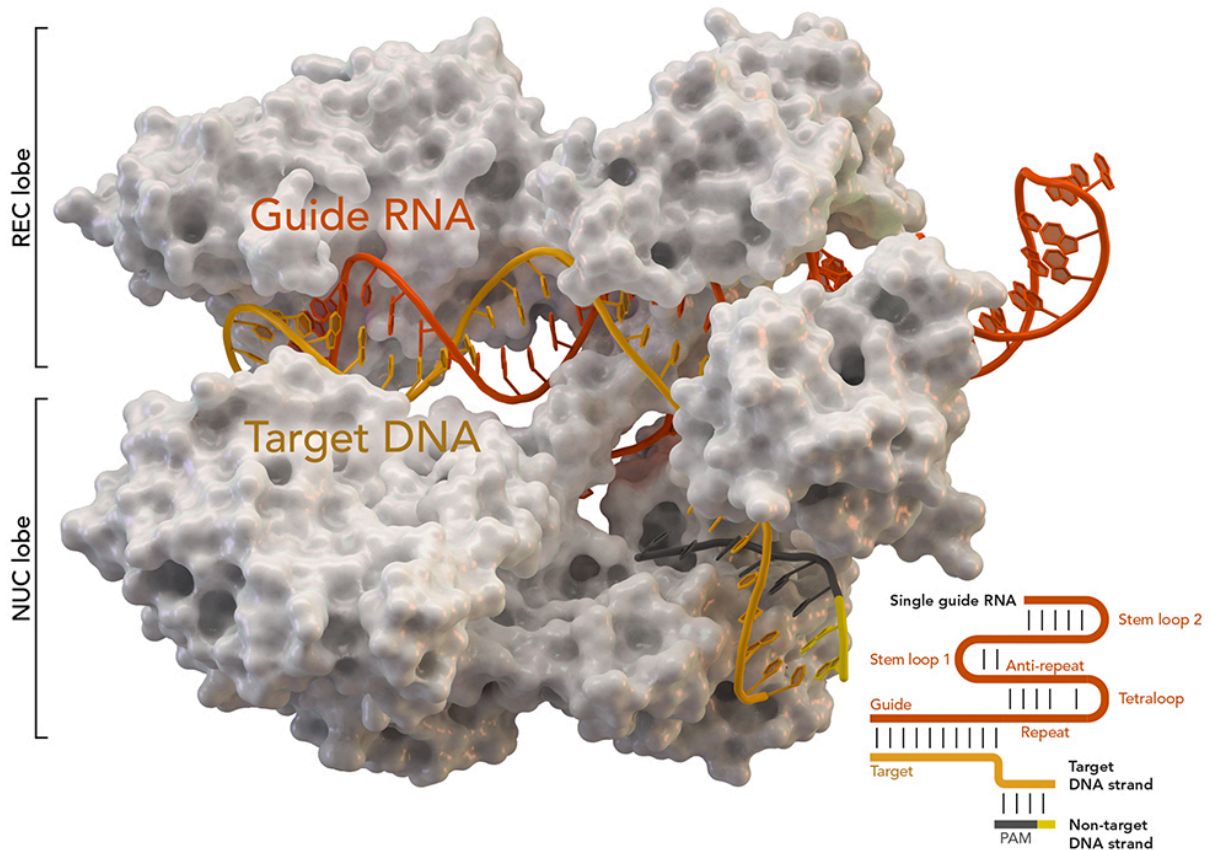# Machine-learning algorithm predicts how cells repair broken DNA

November 7 2018

CRISPR-associated protein Cas9 (white) from Staphylococcus aureus based on Protein Database ID 5AXW. Credit: Thomas Splettstoesser (Wikipedia, CC BY-SA 4.0)

The human genome has its own proofreaders and editors, and their handiwork is not as haphazard as once thought.

When DNA's double helix is broken after damage from, say, exposure to X-rays, molecular machines perform a kind of genetic "auto-correction" to put the genome back together—but those repairs are often imperfect. Just as your smartphone might amend a misspelled text message into an incoherent phrase, the cell's natural DNA repair process can add or remove bits of DNA at the break site in a seemingly random and unpredictable manner. Editing genes with CRISPR-Cas9 allows scientists to break DNA at specific locations, but this can create "spelling errors" that alter the function of genes.

This response to CRISPR-induced damage, called "end joining," is useful for disabling a gene, but researchers have deemed it too error-prone to exploit for therapeutic purposes.

A new study upends this view. By creating a machine-learning algorithm that predicts how human and mouse cells respond to CRISPR-induced breaks in DNA, a team of researchers discovered that cells often repair broken genes in ways that are precise and predictable, sometimes even returning mutated genes back to their healthy version. In addition, the researchers put this predictive power to the test and successfully corrected mutations in cells taken from patients with one of two rare genetic disorders.

The work suggests that the cell's genetic auto-correction could one day be combined with CRISPR-based therapies that correct gene mutations by simply cutting DNA precisely and allowing the cell to naturally heal the damage.

The study, published this week in *Nature*, was led by David Liu, the Richard Merkin Professor and director of the Merkin Institute of

Transformative Technologies in Healthcare, and vice chair of the faculty at the Broad Institute; David Gifford, professor of computer science and biological engineering at MIT; and Richard Sherwood, an assistant professor of medicine in the Division of Genetics at Brigham and Women's Hospital.

"Machine learning offers new horizons for the development of human therapeutics", said Gifford, "This study is an example of how combining computational experiment design and analysis with therapeutic goals can produce an unexpected therapeutic modality."

"We don't currently have an efficient way to precisely correct many human disease mutations," said Liu. "Using machine learning, we've shown we can often correct those mutations predictably, by simply letting the cell repair itself."

Many disease-associated mutations involve extra or missing DNA, known as insertions and deletions. Researchers have tried to correct those mutations with CRISPR-based gene editing. To do this, they cut the double helix with an enzyme and insert missing DNA, or remove extra DNA, using a template of genetic material that serves as a blueprint. The approach, however, only works in rapidly dividing cells like blood stem cells and even then it is only partly effective, making it a poor choice for therapeutics aimed at the majority of cell types in the body. To restore gene function without templated repair requires knowing how the cell will fix CRISPR-induced DNA breaks—knowledge that did not exist until now.

Evidence of a pattern to CRISPR repair outcomes had been noted previously, and Gifford's lab began to think that such outcomes might be predictable enough to model accurately; however, they needed much more data to turn those patterns into an accurate predictive understanding.

Led by MIT graduate student Max Shen and Broad Institute postdoctoral researcher Mandana Arbab, the researchers developed a strategy to observe how cells repaired a library of 2,000 sites targeted by CRISPR in the mouse and human genomes. After observing how the cell repaired those cuts, they poured the resulting data into a machine-learning model, inDelphi, prompting the algorithm to learn how the cell responded to cuts at each site—that is, which bits of DNA the cell added to or removed from each damaged gene.

They found that inDelphi could discern patterns at cut sites that predicted what insertions and deletions were made in the corrected gene. At many sites, the set of corrected genes did not contain a huge mixture of variations, but rather a single outcome, such as correction of a pathogenic gene.

Indeed, after querying inDelphi for disease-relevant genes that could be corrected by cutting in just the right place, the researchers found nearly two hundred pathogenic genetic variants that were mostly corrected to their normal, healthy versions after being cut with CRISPR-associated enzymes. They were also able to correct mutations in cells from patients with two rare genetic disorders, Hermansky-Pudlak syndrome and Menkes disease.

"We show that the same CRISPR enzyme that has been used primarily as a sledgehammer can also act as a chisel," said Sherwood. "The ability to know the most likely outcome of your experiment before you do it will be a real advance for the many researchers using CRISPR."

"We had hoped that we would be able to repair disease-associated genes to their native forms, and it was quite rewarding to see that our hypothesis was correct," said Gifford.

More information: Max W. Shen et al. Predictable and precise

template-free CRISPR editing of pathogenic variants, *Nature* (2018).
DOI: 10.1038/s41586-018-0686-x