

New machine learning model predicts which base editor performs best to repair thousands of disease-causing mutations

June 12 2020



[SINGLE MODE](#) • [BATCH MODE](#) • [EDITOR EFFICIENCIES](#) • [USER GUIDE](#) • [ABOUT](#)

Target genomic DNA:

Base editor / cell type:

CRISPR protospacer: 

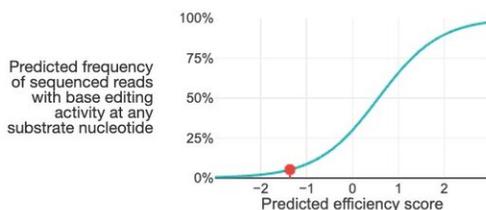
Amino acid frame:

Base editing outcomes among edited reads: DNA sequence

1	5	10	15	20	
TTTTATGTTAAATATCTTTT					
TGGATACGGGTGAAGATTACTTTT	A	TGTTAAATATCTTTT	TGGCCAACGGC		96.2%
TGGATACGGGTGAAGATTACTTTT	C	TGTTAAATATCTTTT	TGGCCAACGGC		1.3%
TGGATACGGGTGAAGATTACTTTT	G	TGTTAAATATCTTTT	TGGCCAACGGC		0.9%
TGGATACGGGTGAAGATTACTTTT	T	TGTTAAATATCTTTT	TGGCCAACGGC		0.3%
TGGATACGGGTGAAGATTACTTTT	G	TGTTAAATATCTTTT	TGGCCAACGGC		0.2%
TGGATACGGGTGAAGATTACTTTT	T	TGTTAAATATCTTTT	TGGCCAACGGC		0.1%
TGGATACGGGTGAAGATTACTTTT	C	TGTTAAATATCTTTT	TGGCCAACGGC		0.1%
TGGATACGGGTGAAGATTACTTTT	G	TGTTAAATATCTTTT	TGGCCAACGGC		0.1%
TGGATACGGGTGAAGATTACTTTT	A	TGTTAAATATCTTTT	TGGCCAACGGC		0.1%
TGGATACGGGTGAAGATTACTTTT	G	TGTTAAATATCTTTT	TGGCCAACGGC		0.1%

[Download table of predictions](#)

Base editing efficiency



This target has **below average base editing efficiency**.

Predicted Z-score: -1.36

Percentile: 8.6

If the average editing efficiency is 30.0%,  then this target's efficiency is 5.2%.

Adjust to your observed average editing efficiency in your experimental system. [Read more on how to do this accurately.](#)



[Shareable link to your results](#)

BE-Hive's machine learning model predicts which base editor performs best to repair thousands of disease-causing mutations. The library is free and available for public use. Credit: Liu lab

Gene editing technology is getting better and growing faster than ever before. New and improved base editors—an especially efficient and precise kind of genetic corrector—inch the tech closer to treating genetic diseases in humans. But, the base editor boom comes with a new challenge: Like a massive key ring with no guide, scientists can sink huge amounts of time into searching for the best tool to solve genetic malfunctions like those that cause sickle cell anemia or progeria (a rapid aging disease). For patients, time is too important to waste.

"New base [editors](#) come out seemingly every week," said David Liu, Thomas Dudley Cabot Professor of the Natural Sciences and a core institute member of the Broad Institute and the Howard Hughes Medical Institute (HHMI). "The progress is terrific, but it leaves researchers with a bewildering array of choices for what base editor to use."

Liu invented base editors. Fittingly, he and his research team have now invented a way to identify which are most likely to achieve desired edits, as reported today in *Cell*. Using [experimental data](#) from editing more than 38,000 target sites in human and mouse cells with 11 of the most popular base editors (BEs), they created a machine learning model that accurately predicts base editing outcomes, Liu said. The library, called BE-Hive, is available for [public use](#). But the effort produced more than a neat catalog of BEs; the machine learning model discovered new editor properties and capabilities that humans failed to notice.

"If you set out to use base editing to correct a single disease-causing

mutation," said Mandana Arbab, a postdoctoral fellow in the Liu lab and co-first author on the study, "you're left with a mountain of possible ways to do it and it is difficult to know which ones are most likely to work."

Base editors may be more precise than other forms of gene editing, but they can still cause unwanted, often unpredictable, edits outside the intended genetic target. Each editor has its own eccentricities. Different types operate within smaller or larger editing "windows," stretches of DNA about two to five letters wide. Some editors might overshoot or undershoot their targets; others might change just one of two As in a given window.

"If the sequence within the window is GACA," Liu said, "and you're using an adenine base editor to change one of those As, will one be preferentially edited over the other?"

The answer depends on the base editor, its paired guide RNA—the chaperone that ferries the editor to the appropriate DNA work site—and the surrounding DNA sequence. To corral all these complicating factors, the team first collected a massive amount of data. Over about a year, Arbab said, they equipped cells with over 38,000 DNA target sites and then treated them with the 11 most popular base editors, paired with guide RNAs. After the treatment, they sequenced the DNA of the cells to collect billions of data points on how each base editor impacted each cell.

To analyze this bounty, Max Shen, a Ph.D. student at the Massachusetts Institute of Technology's Computational and Systems Biology program, member of the Broad Institute, and co-first author designed and trained a machine learning model to predict each base editor's particular eccentricities. In a previous groundbreaking study, Shen and his lab mates trained a different machine learning model to analyze data from

another common gene editing tool, CRISPR, and dispelled a popular misconception that the tool yields unpredictable and generally useless insertions and deletions, Shen said. Instead, they showed that even if humans can't predict where those insertions and deletions occur, machine learning could.

Now, researchers can put a target DNA sequence into BE-Hive, Shen's beefed up [machine learning model](#), and see predicted outcomes of using each of the 11 base editors on that target. "BE-Hive predicts, down to the individual DNA sequence level, what will be the distribution of products that results from each of those base editors acting on that target site," said Liu.

Some of BE-Hive's predictions were surprising, even to the inventor of base editors. "Sometimes," Liu said, "for reasons that our primate brains aren't sufficiently sophisticated to predict, the model could accurately tell us that even though there are two Cs right in the editing window, this particular editor will only edit the second one, for example."

BE-Hive also learned when base editors can make so-called transversion edits: Instead of changing a C to a T, some base editors changed a C to a G or an A, rare and abnormal but potentially valuable quirks. The researchers then used BE-Hive to correct 174 disease-causing transversion mutations with minimal byproducts. And, they used BE-Hive to discover unknown [base editor](#) properties, which they used to design novel tools with new capabilities, adding a few more genetic keys to the ever-growing ring.

Provided by Harvard University

Citation: New machine learning model predicts which base editor performs best to repair thousands of disease-causing mutations (2020, June 12) retrieved 20 September 2024 from

<https://phys.org/news/2020-06-machine-base-editor-thousands-disease-causing.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.