

New model helps identify mutations that drive cancer

June 20 2022



Credit: CC0 Public Domain

Cancer cells can have thousands of mutations in their DNA. However, only a handful of those actually drive the progression of cancer; the rest are just along for the ride.

Distinguishing these harmful driver [mutations](#) from the neutral passengers could help researchers identify better drug targets. To boost those efforts, an MIT-led team has built a new computer model that can rapidly scan the entire genome of [cancer cells](#) and identify [mutations](#) that occur more frequently than expected, suggesting that they are driving [tumor growth](#). This type of prediction has been challenging because some [genomic regions](#) have an extremely high frequency of passenger mutations, drowning out the signal of actual drivers

"We created a probabilistic, deep-learning method that allowed us to get a really accurate model of the number of passenger mutations that should exist anywhere in the genome," says Maxwell Sherman, an MIT graduate student. "Then we can look all across the genome for regions where you have an unexpected accumulation of mutations, which suggests that those are driver mutations."

In their new study, the researchers found additional mutations across the genome that appear to contribute to tumor growth in 5 to 10 percent of cancer patients. The findings could help doctors to identify drugs that would have greater chance of successfully treating those patients, the researchers say. Currently, at least 30 percent of cancer patients have no detectable driver mutation that can be used to guide treatment.

Sherman, MIT graduate student Adam Yaari, and former MIT research assistant Oliver Priebe are the lead authors of the study, which appears today in *Nature Biotechnology*. Bonnie Berger, the Simons Professor of Mathematics at MIT and head of the Computation and Biology group at the Computer Science and Artificial Intelligence Laboratory (CSAIL), is a senior author of the study, along with Po-Ru Loh, an assistant professor at Harvard Medical School and associate member of the Broad Institute of MIT and Harvard. Felix Dietlein, an associate professor at Harvard Medical School and Boston Children's Hospital, is also an author of the paper.

A new tool

Since the human genome was sequenced two decades ago, researchers have been scouring the genome to try to find mutations that contribute to cancer by causing cells to grow uncontrollably or evade the immune system. This has successfully yielded targets such as epidermal growth factor receptor (EGFR), which is commonly mutated in lung tumors, and BRAF, a common driver of melanoma. Both of these mutations can now be targeted by specific drugs.

While those targets have proven useful, protein-coding genes make up only about 2 percent of the genome. The other 98 percent also contains mutations that can occur in cancer cells, but it has been much more difficult to figure out if any of those mutations contribute to cancer development.

"There has really been a lack of computational tools that allow us to search for these driver mutations outside of protein-coding regions," Berger says. "That's what we were trying to do here: design a computational method to let us look at not only the 2 percent of the genome that codes for proteins, but 100 percent of it."

To do that, the researchers trained a type of computational model known as a deep neural network to search cancer genomes for mutations that occur more frequently than expected. As a first step, they trained the model on genomic data from 37 different types of cancer, which allowed the model to determine the background mutation rates for each of those types.

"The really nice thing about our model is that you train it once for a given cancer type, and it learns the mutation rate everywhere across the genome simultaneously for that particular type of cancer," Sherman says. "Then you can query the mutations that you see in a patient cohort

against the number of mutations you should expect to see."

The data used to train the models came from the Roadmap Epigenomics Project and an international collection of data called the Pan-Cancer Analysis of Whole Genomes (PCAWG). The model's analysis of this data gave the researchers a map of the expected passenger mutation rate across the genome, such that the expected rate in any set of regions (down to the single base pair) can be compared to the observed mutation count anywhere across the [genome](#).

Changing the landscape

Using this model, the MIT team was able to add to the known landscape of mutations that can drive cancer. Currently, when [cancer patients'](#) tumors are screened for cancer-causing mutations, a known driver will turn up about two-thirds of the time. The new results of the MIT study offer possible driver mutations for an additional 5 to 10 percent of the pool of patients.

One type of noncoding mutation the researchers focused on is called "cryptic splice mutations." Most genes consist of sequences of exons, which encode protein-building instructions, and introns, which are spacer elements that usually get trimmed out of messenger RNA before it is translated into protein. Cryptic splice mutations are found in introns, where they can confuse the cellular machinery that splices them out. This results in introns being included when they shouldn't be.

Using their model, the researchers found that many cryptic splice mutations appear to disrupt [tumor suppressor genes](#). When these mutations are present, the tumor suppressors are spliced incorrectly and stop working, and the cell loses one of its defenses against cancer. The number of cryptic splice sites that the researchers found in this study accounts for about 5 percent of the driver mutations found in tumor

suppressor genes.

Targeting these mutations could offer a new way to potentially treat those patients, the researchers say. One possible approach that is still in development uses short strands of RNA called antisense oligonucleotides (ASOs) to patch over a mutated piece of DNA with the correct sequence.

"If you could make the mutation disappear in a way, then you solve the problem. Those tumor suppressor genes could keep operating and perhaps combat the cancer," Yaari says. "The ASO technology is actively being developed, and this could be a very good application for it."

Another region where the researchers found a high concentration of noncoding driver mutations is in the untranslated regions of some tumor suppressor genes. The tumor suppressor gene TP53, which is defective in many types of cancer, was already known to accumulate many deletions in these sequences, known as 5' untranslated regions. The MIT team found the same pattern in a tumor suppressor called ELF3.

The researchers also used their model to investigate whether common mutations that were already known might also be driving different types of cancers. As one example, the researchers found that BRAF, previously linked to melanoma, also contributes to cancer progression in smaller percentages of other types of cancers, including pancreatic, liver, and gastroesophageal.

"That says that there's actually a lot of overlap between the landscape of common drivers and the landscape of rare drivers. That provides opportunity for therapeutic repurposing," Sherman says. "These results could help guide the clinical trials that we should be setting up to expand these drugs from just being approved in one [cancer](#), to being approved in

many cancers and being able to help more patients."

More information: Po-Ru Loh, Genome-wide mapping of somatic mutation rates uncovers drivers of cancer, *Nature Biotechnology* (2022).

[DOI: 10.1038/s41587-022-01353-8](https://doi.org/10.1038/s41587-022-01353-8).

www.nature.com/articles/s41587-022-01353-8

Provided by Massachusetts Institute of Technology

Citation: New model helps identify mutations that drive cancer (2022, June 20) retrieved 22 July 2023 from <https://medicalxpress.com/news/2022-06-mutations-cancer.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.