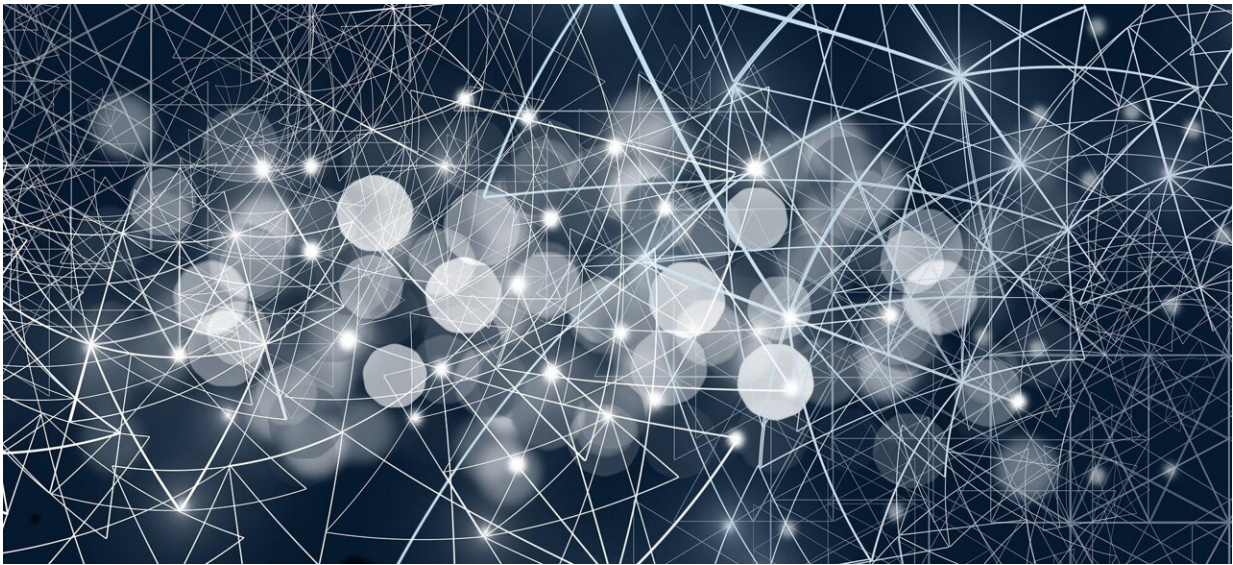


# AI model can help determine where a patient's cancer originated

August 7 2023

---



Credit: Pixabay/CC0 Public Domain

For a small percentage of cancer patients, doctors are unable to determine where their cancer originated. This makes it much more difficult to choose a treatment for those patients, because many cancer drugs are typically developed for specific cancer types.

A new approach developed by researchers at MIT and Dana-Farber Cancer Institute may make it easier to identify the sites of origin for those enigmatic cancers. Using machine learning, the researchers created

a [computational model](#) that can analyze the sequence of about 400 genes and use that information to predict where a given [tumor](#) originated in the body.

Using this [model](#), the researchers showed that they could accurately classify at least 40% of tumors of unknown origin with high confidence, in a dataset of about 900 patients. This approach enabled a 2.2-fold increase in the number of patients who could have been eligible for a genomically guided, targeted treatment, based on where their cancer originated.

"That was the most important finding in our paper, that this model could be potentially used to aid treatment decisions, guiding doctors toward personalized treatments for patients with cancers of unknown primary origin," says Intae Moon, an MIT graduate student in [electrical engineering](#) and computer science who is the lead author of the new study.

Alexander Gusev, an associate professor of medicine at Harvard Medical School and Dana-Farber Cancer Institute, is the senior author of the paper, which appears today in *Nature Medicine*.

## **Mysterious origins**

In 3% to 5% of [cancer patients](#), particularly in cases where tumors have metastasized throughout the body, oncologists don't have an easy way to determine where the cancer originated. These tumors are classified as cancers of unknown primary (CUP).

This lack of knowledge often prevents doctors from being able to give patients "precision" drugs, which are typically approved for specific cancer types where they are known to work. These targeted treatments tend to be more effective and have fewer side effects than treatments

that are used for a broad spectrum of cancers, which are commonly prescribed to CUP patients.

"A sizeable number of individuals develop these cancers of unknown primary every year, and because most therapies are approved in a site-specific way, where you have to know the primary site to deploy them, they have very limited treatment options," Gusev says.

Moon, an affiliate of the Computer Science and Artificial Intelligence Laboratory who is co-advised by Gusev, decided to analyze [genetic data](#) that is routinely collected at Dana-Farber to see if it could be used to predict cancer type. The data consist of genetic sequences for about 400 genes that are often mutated in cancer.

The researchers trained a [machine-learning model](#) on data from nearly 30,000 patients who had been diagnosed with one of 22 known cancer types. That set of data included patients from Memorial Sloan Kettering Cancer Center and Vanderbilt-Ingram Cancer Center, as well as Dana-Farber.

The researchers then tested the resulting model on about 7,000 tumors that it hadn't seen before, but whose site of origin was known. The model, which the researchers named OncoNPC, was able to predict their origins with about 80% accuracy. For tumors with high-confidence predictions, which constituted about 65% of the total, its accuracy rose to roughly 95%.

After those encouraging results, the researchers used the model to analyze a set of about 900 tumors from patients with CUP, which were all from Dana-Farber. They found that for 40% of these tumors, the model was able to make high-confidence predictions.

The researchers then compared the model's predictions with an analysis

of the germline, or inherited, mutations in a subset of tumors with available data, which can reveal whether the patients have a genetic predisposition to develop a particular type of cancer. The researchers found that the model's predictions were much more likely to match the type of cancer most strongly predicted by the germline mutations than any other type of cancer.

## Guiding drug decisions

To further validate the model's predictions, the researchers compared data on the CUP patients' survival time with the typical prognosis for the type of cancer that the model predicted. They found that CUP patients who were predicted to have cancer with a poor prognosis, such as pancreatic cancer, showed correspondingly shorter survival times. Meanwhile, CUP patients who were predicted to have cancers that typically have better prognoses, such as neuroendocrine tumors, had longer survival times.

Another indication that the model's predictions could be useful came from looking at the types of treatments that CUP patients analyzed in the study had received. About 10% of these patients had received a targeted treatment, based on their oncologists' best guess about where their cancer had originated. Among those patients, those who received a treatment consistent with the type of cancer that the model predicted for them fared better than patients who received a treatment typically given for a different type of cancer than what the model predicted for them.

Using this model, the researchers also identified an additional 15% of patients (2.2-fold increase) who could have received an existing targeted treatment, if their [cancer](#) type had been known. Instead, those patients ended up receiving more general chemotherapy drugs.

"That potentially makes these findings more clinically actionable

because we're not requiring a new drug to be approved. What we're saying is that this population can now be eligible for precision treatments that already exist," Gusev says.

The researchers now hope to expand their model to include other types of data, such as pathology images and radiology images, to provide a more comprehensive prediction using multiple data modalities. This would also provide the model with a comprehensive perspective of tumors, enabling it to predict not just the type of tumor and patient outcome, but potentially even the optimal [treatment](#).

**More information:** Machine learning for genetics-based classification and treatment response prediction in cancer of unknown primary, *Nature Medicine* (2023). [DOI: 10.1038/s41591-023-02482-6](https://doi.org/10.1038/s41591-023-02482-6)

Provided by Massachusetts Institute of Technology

Citation: AI model can help determine where a patient's cancer originated (2023, August 7) retrieved 28 April 2024 from <https://medicalxpress.com/news/2023-08-ai-patient-cancer.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>
--