

Machine Learning Model May Help Identify the Origin of Cancers of Unknown Primary

This approach may help guide personalized treatment decisions for patients diagnosed with these cancers.

April 20, 2026 By American Association for Cancer Research

A Machine Learning Model that Uses DNA Methylation Patterns May Help Identify the Origin of Cancers of Unknown Primary

San Diego – A machine learning model analyzing CpG-based DNA methylation accurately predicted the origin of many different cancer types in patients with cancers of unknown primary (CUP), according to research presented at the [American Association for Cancer Research \(AACR\) Annual Meeting 2026](#), held April 17-22.

CUP are metastatic malignancies in which the primary cancer site could not be identified. These cancers are often associated with poorer outcomes, in part because treatment decisions must be made without knowing the cancer's origin, and patients are typically treated with broad, nonspecific chemotherapy regimens rather than therapies targeted to a specific cancer type, according to presenter [Marco A. De Velasco, PhD](#), a faculty member in the Department of Genome Biology at Kindai University in Japan.

“Only between 15% and 20% of patients with CUP show features that allow physicians to treat them with site-specific therapies, which are associated with better outcomes,” explained De Velasco. “However, most patients, between 80% and 85%, receive more general chemotherapy, which is often less effective. Patients receiving site-directed therapy can survive up to 24 months, compared with six to nine months for those receiving standard treatment.”

Researchers have explored whether identifying the cancer's origin using molecular profiling could improve treatment decisions. These approaches analyze patterns in tumor biology, such as gene activity or chemical modifications to DNA, which differ between cancer types and may persist even after the cancer has spread. While some methods have shown promise, they have not demonstrated clear survival benefits in clinical trials, according to De Velasco.

In this study, De Velasco and colleagues, including colead investigator [Kazuko Sakai, PhD](#), and principal investigator [Kazuto Nishio, MD, PhD](#), developed a new approach focusing on CpG DNA methylation, a type of chemical modification that occurs at cytosine and guanine DNA bases. De Velasco noted that CpG methylation acts like a molecular “fingerprint” for different tissues in the body. By analyzing these patterns in tumor samples, the researchers developed a computational model capable of distinguishing among 21 different cancer types.

“Instead of relying on large and complex datasets, we aimed to identify a smaller, more practical set of markers that still retains strong predictive power,” said De Velasco. “The long-term goal is to create a tool that could support physicians in identifying the likely tissue of origin and helping inform more effective treatment decisions.”

The model was developed using methylation data from nearly 7,500 patients with 21 different cancer types obtained from The Cancer Genome Atlas program and other public datasets. Data were divided among training and test cohorts.

The researchers applied machine learning to identify sites of CpG methylation in the tumor DNA of the training cohort and build methylation profiles that were associated with different tumor types.

The study results showed that the model correctly identified the cancer type in about 95% of cases in the test cohort, and maintained strong performance—about 87% accuracy—when applied to an independent validation cohort from the researcher’s institution of 31 cases representing 17 different cancer types.

“One of the most important findings from our study is that we were able to accurately predict the origin of many different cancer types using a very small subset of DNA markers, about 1,000 CpG regions selected from hundreds of thousands across the genome,” said De Velasco. “This is important because it shows that we can simplify complex molecular data while still maintaining strong predictive performance.”

For patients with CUP, this model could help move physicians away from trial-and-error treatment approaches and instead select therapies tailored to a cancer’s likely site of origin, he added.

“Our findings suggest that DNA-based approaches can help identify where a cancer may have started, even when the original tumor is not visible. By using a much smaller and more focused set of markers, this approach could make these types of tests more practical and accessible in the future,” said De Velasco.

“Overall, we see this research as part of a broader effort to better understand cancer using molecular information, with the goal of supporting more informed and personalized care in the future. However, this work is still in the research stage. We next have to evaluate how well this approach performs in a prospective analysis of patients with true cancers of unknown primary,” added De Velasco.

One key limitation of this study is that the model was developed using cancers with known origins,

rather than true CUP, which means the model needs to be tested in actual patients with CUP to understand how well it performs in clinical settings. Another limitation is that not all tumors are easy to access for genetic testing, especially in the advanced stage setting. An important next step in this research, according to De Velasco, is adapting and evaluating this model using blood-based biopsy to analyze circulating tumor DNA instead of relying on DNA from tissue samples.

Funding for this study was provided by the Japan Society for the Promotion of Science. De Velasco reports no conflicts of interest.

This [news release](#) was published by the American Association for Cancer Research on April 20, 2026.

Click here for more news from [AACR 2026](#).

© 2026 Smart + Strong All Rights Reserved.

<http://beta.docker.cancerhealth.com/article/machine-learning-model-may-help-identify-origin-cancers-unknown-primary>