

# 'Junk DNA' no more: Investigators develop method of identifying cancers from repeat elements of genetic code

March 13 2024

---



New method called ARTEMIS uses machine learning to shed light on human genome "dark matter" involved in cancer and other diseases. Credit: Carolyn Hruban

Repeats of DNA sequences, often referred to as "junk DNA" or "dark matter," that are found in chromosomes and could contribute to cancer or other diseases have been challenging to identify and characterize.

Now, investigators at the Johns Hopkins Kimmel Cancer Center have developed a novel approach that uses machine learning to identify these elements in [cancerous tissue](#), as well as in cell-free DNA (cfDNA)—fragments that are shed from tumors and float in the bloodstream.

This new method could provide a noninvasive means of detecting cancers or monitoring response to therapy. Machine learning is a type of artificial intelligence that uses data and computer algorithms to perform complex tasks and accelerate research.

In [laboratory tests](#), the method, called ARTEMIS (Analysis of Repeat Elements in dISease) examined over 1,200 types of repeat elements comprising nearly half of the human genome, and identified that a large number of repeats not previously known to be associated with cancer were altered in tumor formation.

The investigators also were able to identify changes in these elements in cfDNA, providing a way to detect cancer and determine where in the body it originated. A description of the work is [published](#) in *Science Translational Medicine*.

"When you think about existing cancer genes and the DNA sequences around them, they're just chock full of these repeats," says Victor E. Velculescu, M.D., Ph.D., a professor of oncology and co-director of the Cancer Genetics and Epigenetics Program at the Johns Hopkins Kimmel Cancer Center, who led the study with Akshaya Annapragada, an M.D./Ph.D. student at the Johns Hopkins University School of Medicine, and Robert Scharpf, Ph.D., an associate professor of oncology

at Johns Hopkins.

"Until ARTEMIS, this dark matter of the genome was essentially ignored, but now we're seeing that these repeats are not occurring randomly," Velculescu says. "They end up being clustered around genes that are altered in cancer in a variety of different ways, providing the first glimpse that these sequences may be key to tumor development."

In a series of laboratory tests, the researchers first examined the distribution of 1.2 billion kmers (short sequences of DNA) defining unique repeats, finding them enriched in genes commonly altered in human cancers. For example, of 736 genes known to drive cancers, 487 contained an average fifteenfold higher than expected number of repeat sequences. These repeat sequences also were significantly increased in genes involved in cell signaling pathways that are commonly dysregulated in cancers.

Using [next-generation sequencing](#), technology that allows researchers to rapidly examine the sequences of entire genomes, the researchers also looked to see if repeat sequences were directly altered in cancers.

They used ARTEMIS to analyze over 1,200 distinct types of repeat elements in tumor and normal tissues from 525 patients with different cancers participating in the Pan-Cancer Analysis of Whole Genomes (PCAWG), and found a median of 807 altered elements in each tumor.

Nearly two-thirds of these elements (820 of 1,280) had not previously been observed as being altered in human cancers. Then, they used a [machine-learning](#) model to generate an ARTEMIS score for each sample to provide a summary of genome-wide repeat element changes that were predictive of cancer.

ARTEMIS scores distinguished the 525 PCAWG participants' tumors

from normal tissues with a high performance (AUC=0.96) across all cancer types analyzed, where 1 is a perfect score. Increased ARTEMIS scores were associated with shorter overall and progression-free survival regardless of tumor type.

The investigators next evaluated ARTEMIS' potential for noninvasive detection of cancer. They applied the tool to blood samples from 287 individuals with and without lung cancer participating in the Danish Lung Cancer Screening Study (LUCAS).

ARTEMIS classified patients with lung cancer with an area under the curve (AUC) of 0.82. But when used with another method called [DELFI](#) (DNA evaluation of fragments for early interception)—an assay previously developed by Velculescu, Scharpf and other members of their group that detects changes in the size and distribution of cfDNA fragments across the genome—the combination model classified patients with [lung cancer](#) with an AUC of 0.91.

Similar performance was observed in a group of 208 individuals at risk for liver cancer, in which ARTEMIS detected individuals with liver cancer among others with cirrhosis or viral hepatitis with an AUC of 0.87. When combined with DELFI, the AUC increased to 0.90.

Finally, they evaluated whether the ARTEMIS blood test could identify where in the body a tumor originated in patients with cancer. When trained with information from the PCAWG participants, the tool could classify the source of tumor tissues with an average 78% accuracy among 12 tumor types.

The investigators then combined ARTEMIS and DELFI to assess [blood samples](#) from a group of 226 individuals with breast, ovarian, lung, colorectal, bile duct, gastric or pancreatic tumors. Here, the model correctly classified patients among the different cancer types with an

average accuracy of 68%, which improved to 83% when the model was allowed to suggest two possible [tumor](#) types instead of a single cancer type.

"Our study shows that ARTEMIS can reveal genome-wide repeat landscapes that reflect dramatic underlying changes in human cancers," Annapragada says. "By illuminating the so-called 'dark genome,' the work offers unique insights into the cancer genome and provides a proof-of-concept for the utility of genome-wide repeat landscapes as tissue and blood-based biomarkers for cancer detection, characterization and monitoring."

Next steps are to evaluate the approach in larger clinical trials, says Velculescu. "You can imagine this could be used for early detection for a variety of cancer types, but also could have uses in other applications such as monitoring response to treatment or detecting recurrence. This is a totally new frontier."

**More information:** Akshaya Annapragada et al, Genome-wide repeat landscapes in cancer and cell-free DNA, *Science Translational Medicine* (2024). [DOI: 10.1126/scitranslmed.adj9283](https://doi.org/10.1126/scitranslmed.adj9283).  
[www.science.org/doi/10.1126/scitranslmed.adj9283](https://www.science.org/doi/10.1126/scitranslmed.adj9283)

Provided by Johns Hopkins University School of Medicine

Citation: 'Junk DNA' no more: Investigators develop method of identifying cancers from repeat elements of genetic code (2024, March 13) retrieved 27 April 2024 from <https://medicalxpress.com/news/2024-03-junk-dna-method-cancers-elements.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.