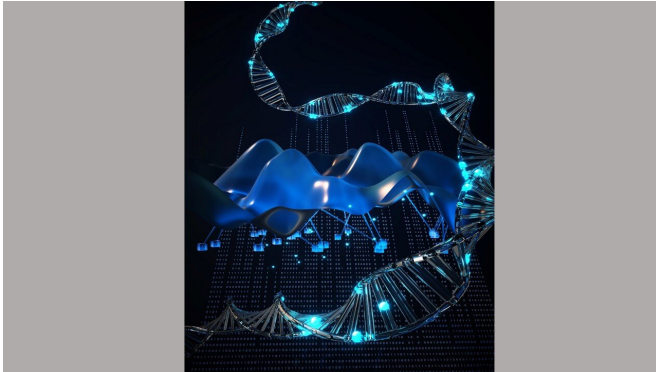


# 'Information theory' recruited to help scientists find cancer genes

20 April 2021



An illustration of the research concept that DNA methylation code can be analyzed using information theory, represented by strings of 0s and 1s. This analysis helps researchers understand the epigenetic landscape of cancer (pictured in blue) and identify genes that regulate this landscape, (noted as strings and posts underneath the blue landscape). Credit: Kate Zvorykina (Ella Maru Studio), Design: Michael Koldobskiy and Andrew Feinberg, Johns Hopkins Medicine

Using a widely known field of mathematics designed mainly to study how digital and other forms of information are measured, stored and shared, scientists at Johns Hopkins Medicine and Johns Hopkins Kimmel Cancer Center say they have uncovered a likely key genetic culprit in the development of acute lymphoblastic leukemia (ALL).

ALL is the most common form of childhood leukemia, striking an estimated 3,000 children and teens each year in the United States alone.

Specifically, the Johns Hopkins team used "[information theory](#)," applying an analysis that relies on strings of zeros and ones—the binary system of symbols common to computer languages and codes—to identify variables or outcomes of a particular process. In the case of

human cancer biology, the scientists focused on a chemical process in [cells](#) called DNA methylation, in which certain chemical groups attach to areas of genes that guide genes' on/off switches.

"This study demonstrates how a mathematical language of cancer can help us understand how cells are supposed to behave and how alterations in that behavior affect our health," says Andrew Feinberg, M.D., M.P.H., Bloomberg Distinguished Professor at the Johns Hopkins University School of Medicine, Whiting School of Engineering and Bloomberg School of Public Health. A founder of the field of cancer epigenetics, Feinberg discovered altered DNA methylation in cancer in the 1980s.

Feinberg and his team say that using information theory to find cancer driver genes may be applicable to a wide variety of cancers and other diseases.

Methylation is now recognized as one way DNA can be altered without changing a cell's genetic code. When methylation goes awry in such epigenetic phenomena, certain genes are abnormally turned on or off, triggering uncontrolled cell growth, or cancer.

"Most people are familiar with genetic changes to DNA, namely mutations that change the DNA sequence. Those mutations are like the words that make up a sentence, and methylation is like punctuation in a sentence, providing pauses and stops as we read," says Feinberg. In a search for a new and more efficient way to read and understand the epigenetic code altered by DNA methylation, he worked with John Goutsias, Ph.D., professor in the Department of Electrical and Computer Engineering at The Johns Hopkins University and Michael Koldobskiy, M.D., Ph.D., pediatric oncologist and assistant professor of oncology at the Johns Hopkins Kimmel Cancer Center.

"We wanted to use this information to identify

genes that drive the development of cancer even though their genetic code isn't mutated," says Koldobskiy.

Results of the study's findings, led by Feinberg, Koldobskiy and Goutsias, were published April 15 in *Nature Biomedical Engineering*.

Koldobskiy explains that methylation at a particular gene location is binary—methylation or no methylation—and a system of zeros and ones can represent these differences just as they are used to represent computer codes and instructions.

For the study, the Johns Hopkins team analyzed DNA extracted from bone marrow samples of 31 children newly diagnosed with ALL at The Johns Hopkins Hospital and Texas Children's Hospital. They sequenced the DNA to determine which [genes](#), across the [entire genome](#), were methylated and which were not.

Newly diagnosed leukemia patients have billions of leukemia cells in their body, says Koldobskiy.

By assigning zeros and ones to pieces of genetic code that were methylated or unmethylated and using concepts of information theory and computer programs to recognize patterns of methylation, the scientists were able to find regions of the genome that were consistently methylated in patients with leukemia and those without cancer.

They also saw genome regions in the leukemia cells that were more randomly methylated, compared with the normal genome, a signal to scientists that those spots may be specifically linked to leukemia cells compared with normal ones.

One gene, called UHRF1, stood out among other gene regions in leukemia cells that had differences in DNA methylation compared with the normal genome.

"It was a big surprise to find this gene, as its link to prostate and other cancer has been suggested but never identified as a driver of leukemia," says Feinberg.

In normal cells, the protein products of the UHRF1 gene create a biochemical bridge between DNA methylation and DNA packaging, but scientists have not deciphered precisely how alteration of the gene contributes to cancer.

Experiments by the Johns Hopkins team show that laboratory-grown leukemia cells lacking activity of the UHRF1 gene cannot self-renew and perpetuate additional leukemia cells.

"Leukemia cells aim to survive, and the best way to ensure survival is to vary the epigenetics in many genome regions so that no matter what tries to kill the cancer, at least some will survive," says Koldobskiy.

ALL is the most common pediatric cancer, and Koldobskiy says that decades of research on various treatments and the sequence of those treatments have helped clinicians cure most of these leukemias, but relapsed disease remains a leading cause of death from cancer in children.

"This new approach can lead to more rational ways of targeting the alterations that drive this and likely many other forms of [cancer](#)," says Koldobskiy.

The Johns Hopkins team plans to use information theory to analyze methylation patterns in other cancers. They also plan to determine whether epigenetic alterations in URFH1 are linked to treatment resistance and disease progression in patients with childhood [leukemia](#).

**More information:** Michael A. Koldobskiy et al, Converging genetic and epigenetic drivers of paediatric acute lymphoblastic leukaemia identified by an information-theoretic analysis, *Nature Biomedical Engineering* (2021). [DOI: 10.1038/s41551-021-00703-2](https://doi.org/10.1038/s41551-021-00703-2)

Provided by Johns Hopkins University School of Medicine

APA citation: 'Information theory' recruited to help scientists find cancer genes (2021, April 20) retrieved 24 April 2021 from <https://medicalxpress.com/news/2021-04-theory-scientists-cancer-genes.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.*