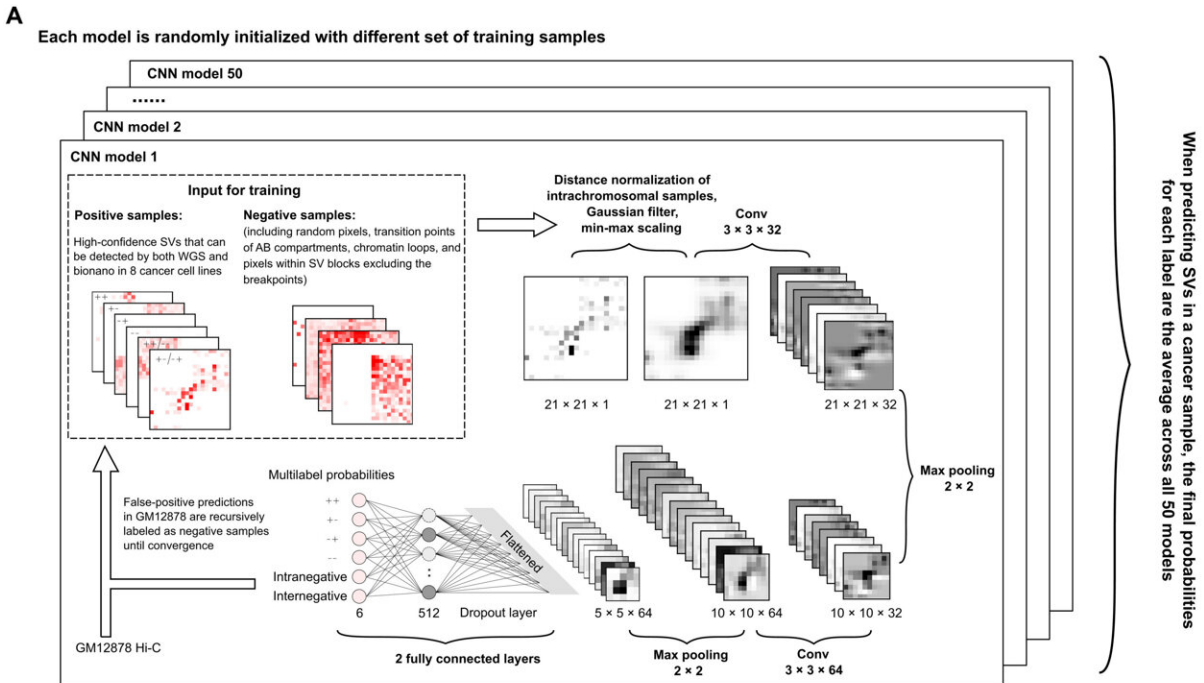
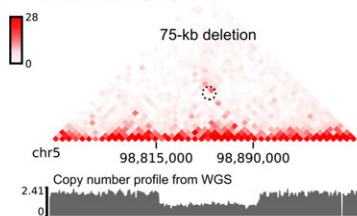


Deep learning empowers discovery of new genetic mutation in cancer

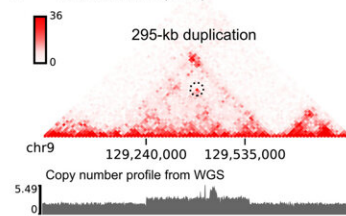
June 16 2022, by Will Doss



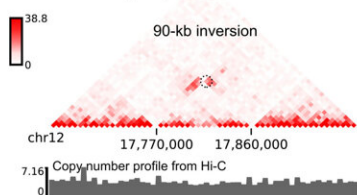
B SK-N-AS Hi-C (5kb)



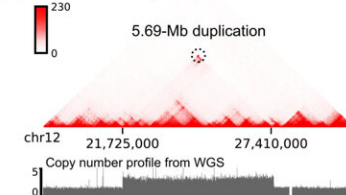
C SK-N-AS Hi-C (10kb)



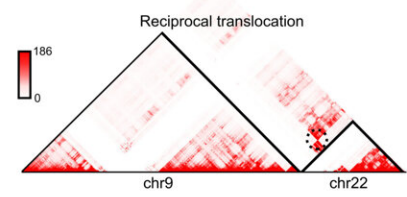
D C4-2B Hi-C (5 kb)



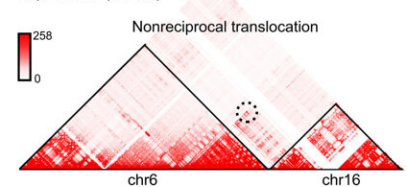
E SW480 Hi-C (50 kb)



F KBM7 Hi-C (250 kb)



G HepG2 Hi-C (250 kb)



EagleC predicts a full range of high-resolution SVs from chromatin interaction data. (A) Workflow of the EagleC framework. (B to G) Examples showing different types of SVs predicted by EagleC. The black dashed circle indicates the SV breakpoint position in each case. The resolution of each Hi-C map is labeled within the parentheses. (B) A short-range (75 kb) heterozygous deletion predicted in the SK-N-AS cells. (C) A short-range (295 kb) duplication predicted in the SK-N-AS cells. (D) A short-range (90 kb) inversion predicted in the C4-2B cells. (E) A long-range (5.69 Mb) duplication predicted in the SW480 cells. (F) A reciprocal translocation predicted in the KBM7 cells. (G) A nonreciprocal translocation predicted in the HepG2 cells. Credit: *Science Advances* (2022). DOI: 10.1126/sciadv.abn9215

A machine-learning model has helped scientists discover hundreds of genetic mutations in cancer that are undetectable by current genome sequencing, according to a study published in the journal *Science Advances*.

These findings provide new targets for [cancer](#) classification and potential therapy, according to Feng Yue, Ph.D., the Duane and Susan Burnham Professor of Molecular Medicine and senior author of the study.

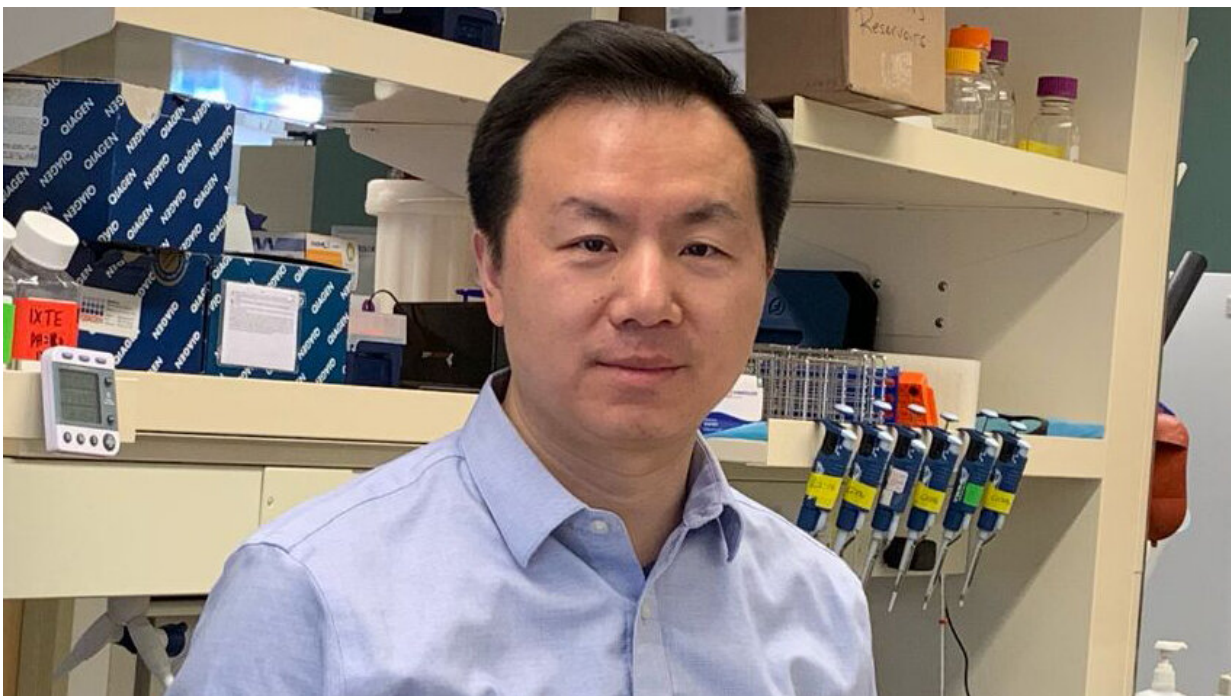
"Our work identified many previously unknown fusion events in [cancer genomes](#) and also captured novel regulatory mechanism for known oncogenes," said Yue, who is also an associate professor of Biochemistry and Molecular Genetics, of Pathology and director of the Center for Cancer Genomics at the Robert H. Lurie Comprehensive Cancer Center of Northwestern University.

Within each cell, long strands of DNA need to be precisely folded and organized so that they can fit inside the nucleus, which is usually only a few micrometers in diameter. Previously, Yue and his collaborators showed that [structural variants](#) in cancer genomes, such as inversions or

translocations, can be detected in genomic analytic tools such as Hi-C.

These patterns can be recognized by computer algorithms as indicators for structural variation. Further, these large structural variations are usually missed by whole genome sequencing (WGS) and even long read sequencing such as Nanopore, according to Yue.

"WGS is very good at detecting base pair mutations and short insertions or deletions, but has a hard time detecting larger variation," said Yue, who is also the director of the Center for Advanced Molecular Analysis at the Institute for Augmented Intelligence in Medicine.



Feng Yue, Ph.D., the Duane and Susan Burnham Professor of Molecular Medicine, was senior author of the study published in the journal *Science Advances*. Credit: Northwestern University

In this study, Yue and his collaborators collected a set of curated high-confidence structural variations of different types from eight cancer cell lines. These were used to train a [deep learning model](#)—named EagleC—to learn the hidden patterns buried in these signals. The results were generally concordant with traditional genome sequencing techniques, with 70 to 80% of genomic variation also found by either WGS or Nanopore sequencing.

However, EagleC found hundreds of additional fusion events that were missed by whole-genome sequencing or long-reads sequencing. These newly discovered events represent 10 to 20% of the total genetic variations detected by Hi-C, according to Yue.

Many of these fusion events cause linkage between an oncogene and a distal enhancer that is usually located on another chromosome. These events, called "enhancer-hijacking," can lead to upregulation of oncogenes.

In the study, investigators used EagleC to search for structural variation in more than 100 [cancer cell lines](#) and patient samples, finding additional fusion events that could be missed by whole genome sequencing. Using this model could expand knowledge of structural variation and their impact on cancer-related genes, according to Yue. This could be especially useful in prostate and [breast cancer](#), two of the most common cancers that also have a high frequency of fusion events.

"We could see if there are differences in therapy response in cancers with fusion events and cancers without," Yue said. "Our findings also present cancer researchers with many novel regulators that control essential oncogenes and pathways."

EagleC can also be used to detect structural variation in single-cell Hi-C analysis, where data is sparse. This allows scientists to examine

heterogeneity between individual cancer cells.

In the future, Yue said he hopes to apply this model to more cancer samples and look for potential drugs targeting the new [fusion](#) events discovered in the current study.

More information: Xiaotao Wang et al, EagleC: A deep-learning framework for detecting a full range of structural variations from bulk and single-cell contact maps, *Science Advances* (2022). [DOI: 10.1126/sciadv.abn9215](#)

Provided by Northwestern University

Citation: Deep learning empowers discovery of new genetic mutation in cancer (2022, June 16) retrieved 24 March 2023 from <https://medicalxpress.com/news/2022-06-deep-empowers-discovery-genetic-mutation.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>
--