

# Carnegie Mellon Algorithm Characterizes How Cancer Genomes Get Scrambled

Research Tool Identifies Interactions Involved in Genetic Changes

Byron Spice // Thursday, July 21, 2016 //

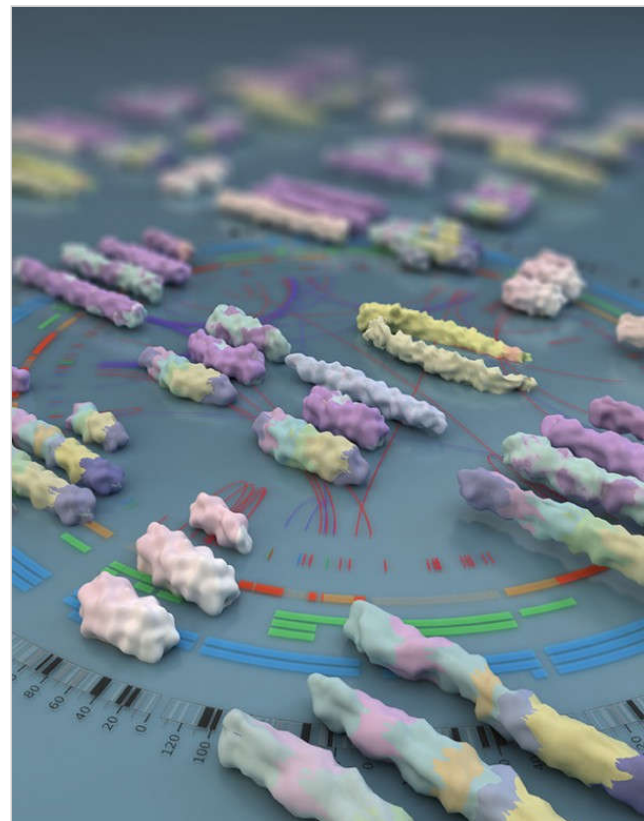
A new method for analyzing the scrambled genomes of cancer cells allows researchers, for the first time, to simultaneously identify two different types of genetic changes associated with cancers and to identify connections between the two.

[Jian Ma](#), associate professor in Carnegie Mellon University's [Computational Biology Department](#), said his new algorithm, called Weaver, could become an important tool for identifying interactions of the alterations within a cancerous cell's DNA that drive the disease.

"This work uses a rigorous and elegant approach to give a better picture of the genome changes that occur during the evolution of individual cancers," said [Robert F. Murphy](#), professor and head of the School of Computer Science's Computational Biology Department. "Having a clearer picture can help identify characteristics, such as responsiveness to drugs, that distinguish cancers and may contribute to developing more personalized treatments."

A report on this work by Ma and his former Ph.D. student at the University of Illinois at Urbana-Champaign, Yang Li, was published online today by the journal [Cell Systems](#).

The genetic code of a healthy cell is embodied by 23 pairs of chromosomes, "but in tumor cells, it's completely scrambled," Ma said. Many cancers are known to give rise to multiple copies of chromosomes, a type of mutation



An algorithm developed by CMU's Jian Ma simultaneously analyzes two types of changes that can scramble the genome in cancer cells. This artist's depiction shows extra copies of normally paired chromosomes. Variations in chromosome color show where DNA has become rearranged and duplicated. (Image courtesy of Ella Marushchenko.)

known as aneuploidy. Likewise, cancer cells are known to undergo another type of mutation, called structural variations, in which DNA sequences within the chromosomes are rearranged and duplicated.

Both of these genetic changes can occur within the same cell, but until now separate techniques had to be used to quantify them. Weaver, however, does these analyses simultaneously using the same input — whole genome sequencing data.

The advantage of doing these analyses simultaneously is that researchers can now see how aneuploidy is affected by structural variations and vice versa. When the analyses are done separately, these interactions are difficult to identify.

Weaver uses a probabilistic graphical model called Markov Random Field, which enables researchers to visualize not only the number of mutations but also how they connect with each other.

"This gives us a better view of the complexity of cancer genomes," Ma said. This may help researchers better characterize cancers or understand which combinations of genetic changes might affect cancer behavior for the same type of cancer or for different cancer types.

Ma said he anticipates Weaver could provide useful perspective for studies of [The Cancer Genome Atlas](#) (TCGA), a collaboration between the National Cancer Institute and the National Human Genome Research Institute that has described the genomes of tumor tissues and matched normal tissues from more than 11,000 patients. The publicly available TCGA dataset includes genomes for 33 different tumor types.

Ma and Li applied Weaver to whole genome sequencing data for two cancer cell lines and to ovarian cancers from the TCGA. In the case of the ovarian cancer samples, Weaver identified duplicated chromosomal regions caused by a specific type of structural variation — an association not previously reported.

David C. Schwartz and Shiguo Zhou of the University of Wisconsin-Madison joined Ma and Li in authoring the Cell Systems report. The National Institutes of Health and the National Science Foundation supported this research.

Byron Spice | 412-268-9068 | [bspice@cs.cmu.edu](mailto:bspice@cs.cmu.edu)

