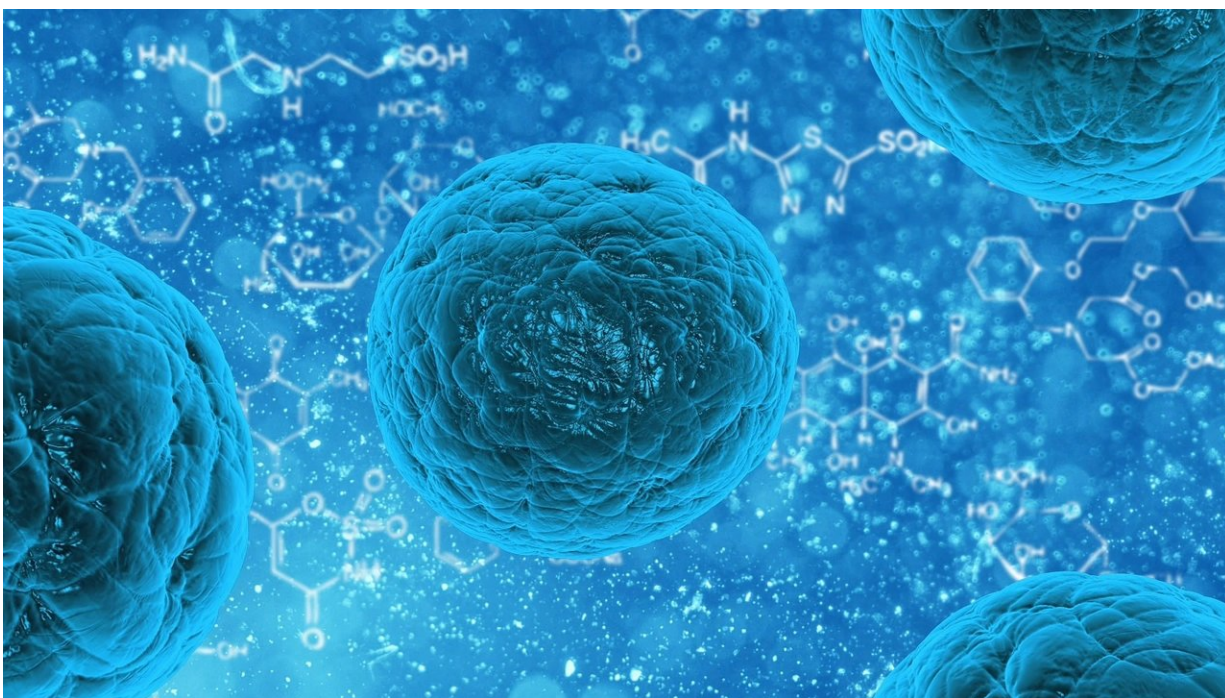# Machine learning offers high-definition glimpse of how genomes organize in single cells

October 11 2021

Credit: CC0 Public Domain

Within the microscopic boundaries of a single human cell, the intricate folds and arrangements of protein and DNA bundles dictate a person's fate: which genes are expressed, which are suppressed, and—importantly—whether they stay healthy or develop disease.

Despite the potential impact these bundles have on human health, science knows little about how genome folding happens in the cell nucleus and how that influences the way genes are expressed. But a new algorithm developed by a team in Carnegie Mellon University's Computational Biology Department offers a powerful tool for illustrating the process at an unprecedented resolution.

The algorithm, known as Higashi, is based on hypergraph representation learning—the form of machine learning that can recommend music in an app and perform 3D object recognition.

School of Computer Science doctoral student Ruochi Zhang led the project with Ph.D. candidate Tianming Zhou and Jian Ma, the Ray and Stephanie Lane Professor of Computational Biology. Zhang named Higashi after a traditional Japanese sweet, continuing a tradition he began with other algorithms he developed.

"He approaches the research with passion but also with a sense of humor sometimes," Ma said.

Their research was published in *Nature Biotechnology* and was conducted as part of a multi-institution research center seeking a better understanding both of the three-dimensional structure of cell nuclei and how changes in that structure affect cell functions in health and disease. The $10 million center was funded by the National Institutes of Health and is directed by CMU, with Ma as its lead principal investigator.

The algorithm is the first tool to use sophisticated neural networks on hypergraphs to provide a high-definition analysis of genome organization in single cells. Where an ordinary graph joins two vertices to a single intersection, known as an edge, a hypergraph joins multiple vertices to the edge.

Chromosomes are made up of a DNA-RNA-protein complex called chromatin that folds and arranges itself to fit inside the cell nucleus. The process influences the way genes are expressed by bringing the functional elements of each ingredient closer together, allowing them to activate or suppress a particular genetic trait.

The Higashi algorithm works with an emerging technology known as single-cell Hi-C, which creates snapshots of chromatin interactions occurring simultaneously in a single cell. Higashi provides a more detailed analysis of chromatin's organization in the single cells of complex tissues and biological processes, as well as how its interactions vary from cell to cell. This analysis allows scientists to see detailed variations in the folding and organization of chromatin from cell to cell—including those that may be subtle, yet important in identifying health implications.

"The variability of genome organization has strong implications in gene expression and cellular state," Ma said.

The Higashi algorithm also allows scientists to simultaneously analyze other genomic signals jointly profiled with single-cell Hi-C. Eventually, this feature will enable expansion of Higashi's capability, which is timely given the expected growth of single-cell data Ma expects to see in coming years through projects such as the NIH 4D Nucleome Program his center belongs to. This flow of data will create additional opportunities to design more algorithms that will advance scientific understanding of how the human genome is organized within the cell and its function in health and disease.

"This is a fast-moving area," Ma said. "The experimental technology is advancing rapidly, and so is the computational development."

More information: Jian Ma, Multiscale and integrative single-cell Hi-

Provided by Carnegie Mellon University