New tool predicts three-dimensional organization of human chromosomes

December 10 2019, by Eric Hamilton

The human genome consists of 23 pairs of chromosomes. A new study from UW–Madison's Sushmita Roy describes a computational tool for researchers to better predict the three-dimensional interactions between chromosomes. Credit: Wikimedia Commons

University of Wisconsin–Madison researchers have developed a computational tool that can accurately predict the three-dimensional interactions between regions of human chromosomes.

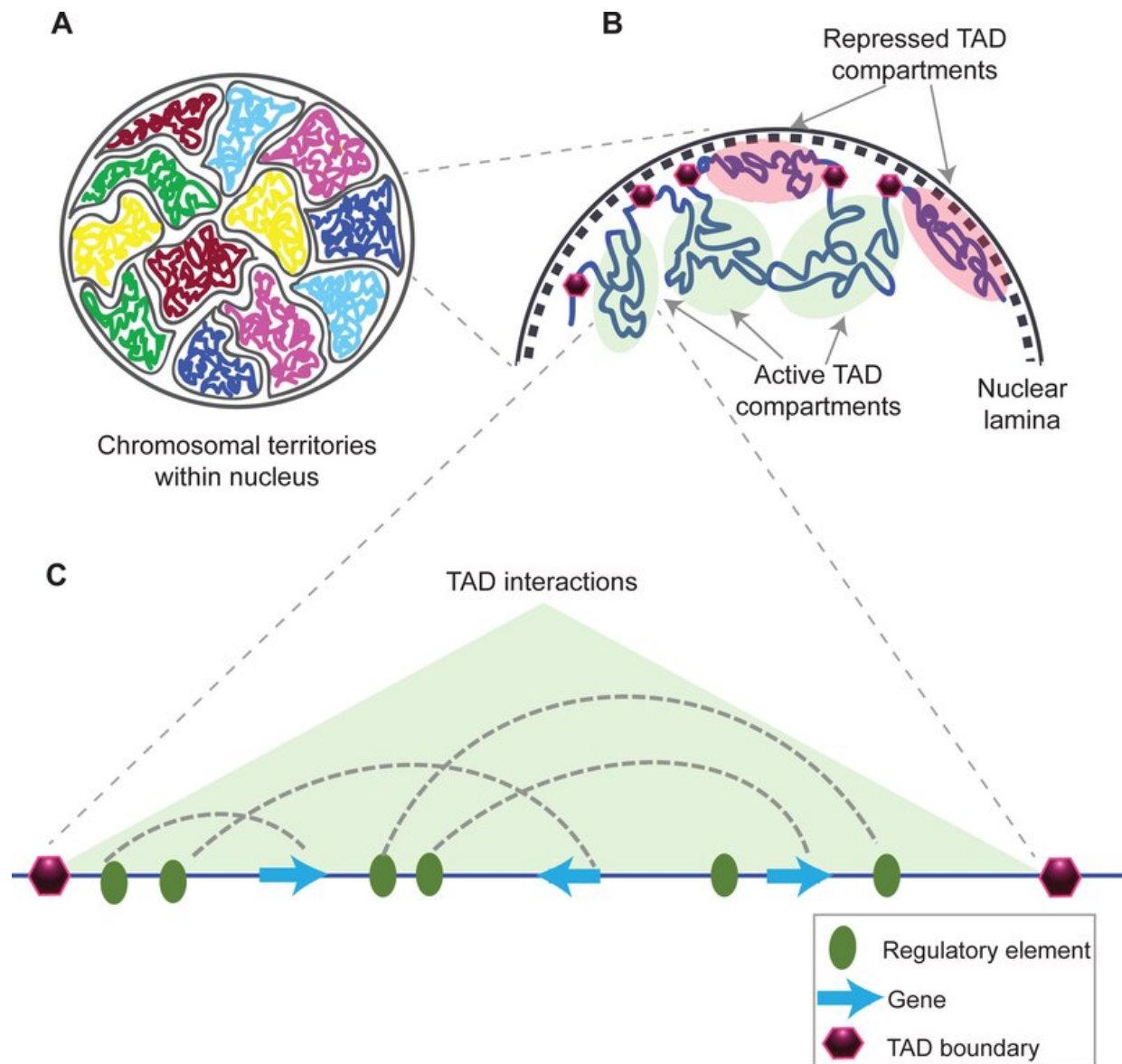The predictive tool is a boon for researchers studying how cells control

the activity of genes. The fine-tuned interaction between regulatory signals and the three-dimensional architecture of chromosomes helps explain how cells achieve their key functions, and how they go haywire, as happens in diseases such as cancer.

The experimental technique to measure these three-dimensional interactions, Hi-C, is expensive, which has limited high-quality data to just a few types of cells. In contrast, the new tool can predict these interactions using much more easily measurable and commonly available data. This could help biologists perform across many cell types more detailed research into tissue development, cancer and other diseases that are affected by this type of distant gene regulation.

UW–Madison researcher Sushmita Roy and her graduate student Shilu Zhang led the work, which was published Dec. 6 in *Nature Communications*. The researchers have made the tool freely available to other scientists and continue to improve the predictive power of the tool, which they named HiC-Reg after the resource-intensive experiments.

"We can very cheaply predict the output of Hi-C experiments, which can help us prioritize other regions of the genome to follow up with more fine-tuned experiments," says Roy, a professor in the Wisconsin Institute for Discovery and the UW–Madison Department of Biostatistics and Medical Informatics. "This can be used as a resource to interpret regulatory variation in the genome."

A far cry from the neat, straight lines of DNA pictured in textbooks, real chromosomes fold, twist and bend to fit several linear feet of DNA into a tiny cell nucleus. These loops also bring distant regions of a chromosome together. Some of these regions carry regulatory information that can promote or repress the expression of distant genes. This intricate gene expression magnifies the complexity of traits that organisms exhibit.

A Chromosomal territories within nucleus

B Repressed TAD compartments
Active TAD compartments
Nuclear lamina

C TAD interactions

Regulatory element
Gene
TAD boundary

The packing of chromosomes into a nucleus allows distant regions of a chromosome to interact and affect one another. Consistently interacting regions are known as topologically associated domains, or TADs. This kind of fine-tuned gene regulation produces more complexity in the traits that organisms express. Credit: Navneet Matharu and Nadav Ahituv

Roy and other researchers have previously developed models that could predict whether or not two distant regions of a chromosome would interact. HiC-Reg builds on that model and not only predicts whether two regions will interact but also how strong that interaction might be. It provides a more complex and realistic model of how chromosomal regions interact and potentially regulate gene expression.

To create HiC-Reg, Roy's team fed a series of commonly available genomic data, such as the presence of proteins and chemical modifications that activate or repress gene expression, into a machine learning algorithm. It also included Hi-C data from the few cell lines for which it is available. The tool then learned relationships that enabled it to predict the Hi-C measurements for a new pair of genomic regions.

"Let's try to use the data that's easy to measure to predict the information that's harder to gather," says Roy. The research was supported by the National Institutes of Health Big Data to Knowledge program, which allowed the team to mine this freely available but underutilized data. "We're trying to leverage publicly available datasets as much as possible."

HiC-Reg correctly predicted between 40 percent and 80 percent of regional associations. The [tool](link) is more accurate than estimating the strength of interactions based on chromosomal distance alone or just mapping the interactions from a pair of regions in one cell line to the same pair of regions in another cell line. But the interactions were harder to predict in some cell types than in others, a limitation the researchers are now working to overcome.

The computationally intensive work relied on UW–Madison's Center for High Throughput Computing, the UW Center for Predictive Computational Phenotyping and the Core Computational Technology research group at the Wisconsin Institute for Discovery.

Other researchers can now use HiC-Reg as-is to predict these three-dimensional interactions in their favorite cell line. Or, they can elect to re-train the program using their own datasets to improve its accuracy for their work.

Roy says that free access is consistent with the question that motivated this research: "How can we help biologists gather this data?"

More information: Shilu Zhang et al. In silico prediction of high-resolution Hi-C interaction matrices, *Nature Communications* (2019). DOI: 10.1038/s41467-019-13423-8

Provided by University of Wisconsin-Madison