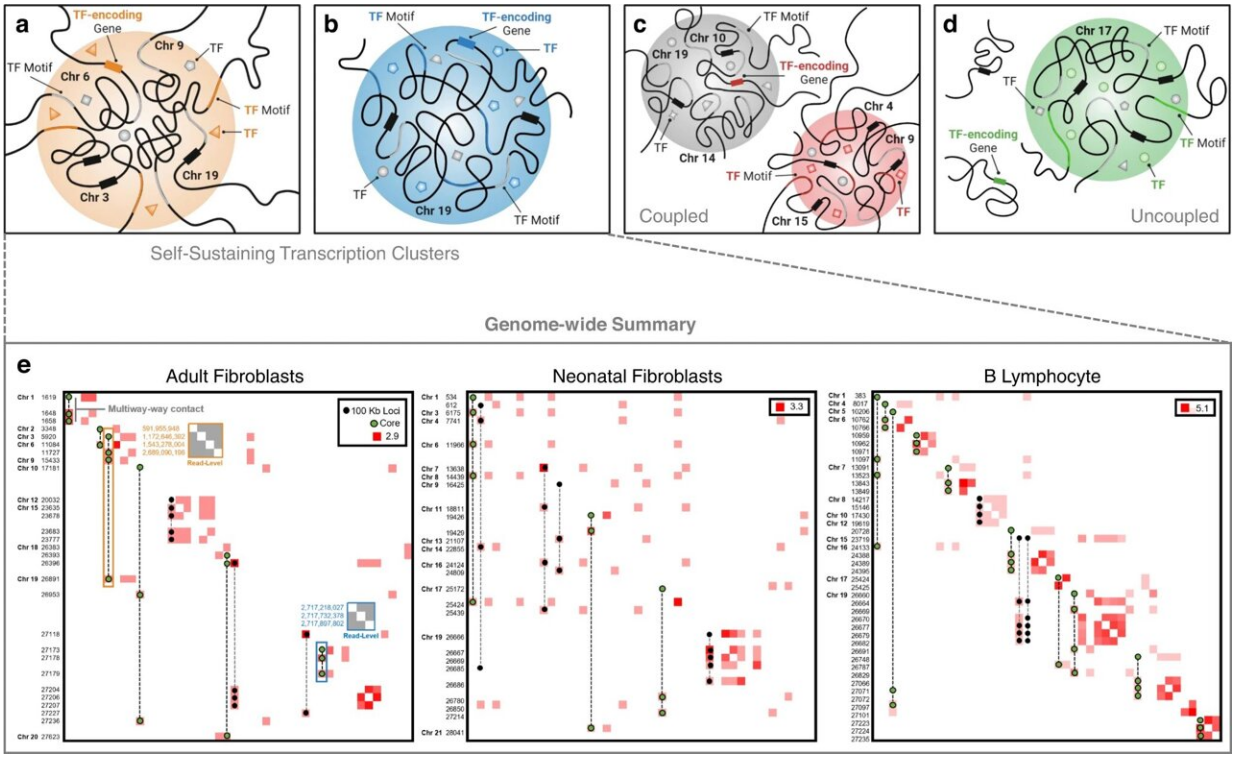
Mathematics enable scientists to understand organization within a cell's nucleus

September 20 2022



Classes of transcription clusters. In a self-sustaining transcription cluster, a TF and the gene encoding that TF are both present. The inter- and intra-chromosomal examples in (a) and (b), respectively, illustrate this phenomenon where in a we see the TF of interest (orange triangle) circulating at the cluster, its binding motif present on the chromatin (orange portion), and its corresponding gene expressed (orange rectangle on Chromosome 6). The gray shapes represent additional TFs with binding motifs (gray portion of chromatin) at the cluster. Black rectangles on Chromosomes 3, 9, and 19 represent additional genes present in the cluster. c An analog-independent class of transcription clusters

where we observe a TF (red square) bind at a transcription cluster (red cluster) and its corresponding gene expressed in a separate transcription cluster (gray cluster), yet not in the same cluster. d An analog-independent class of transcription clusters where we observe a TF (green circle) bind at a transcription cluster (green cluster) and its corresponding gene expressed but not within a transcription cluster. e Genome-wide cell type-specific self-sustaining transcription clusters extracted from multi-way contact data and decomposed into Hi-C contact matrices at 100 kb resolution. Contact frequencies are log-transformed for better visualization. Frequencies along the diagonal indicate interaction between two or more unique multi-way loci that fall within the same 100 kb bin. Axis labels are non-contiguous 100 kb bin coordinates in chromosomal order. Multi-way contacts that make up the self-sustaining transcription clusters are superimposed. Multi-way contacts with green-colored loci represent 'core' transcription clusters - transcription clusters containing a master regulator and its gene analog. An example read-level contact map for the inter-chromosomal FOXO3 self-sustaining transcription cluster is denoted by the orange highlighted box in the adult fibroblast contact matrix and a read-level contact map for the intra-chromosomal ZNF320 self-sustaining transcription cluster is denoted by the blue highlighted box. Values along the left axis of these read-level contact matrices are base-pair positions of the contacting loci in the genome. Credit: *Nature Communications* (2022). DOI: 10.1038/s41467-022-32980-z

Science fiction writer Arthur C. Clarke's third law says that "any sufficiently advanced technology is indistinguishable from magic."

Indika Rajapakse, Ph.D., is a believer. The engineer and mathematician now finds himself a biologist. And he believes the beauty of blending these three disciplines is crucial to unraveling how cells work.

His latest development is a new mathematical technique to begin to understand how a cell's nucleus is organized. The technique, which Rajapakse and collaborators tested on several types of cells, revealed

what the researchers termed self-sustaining transcription clusters, a subset of proteins that play a key role in maintaining cell identity.

They hope this understanding will expose vulnerabilities that can be targeted to reprogram a cell to stop cancer or other diseases.

"More and more cancer biologists think genome organization plays a huge role in understanding uncontrollable cell division and whether we can reprogram a cancer cell. That means we need to understand more detail about what's happening in the nucleus," said Rajapakse, associate professor of computational medicine and bioinformatics, mathematics, and biomedical engineering at the University of Michigan. He is also a member of the U-M Rogel Cancer Center.

Rajapakse is senior author on the paper, published in *Nature Communications*. The project was led by a trio of graduate students with an interdisciplinary team of researchers.

The team improved upon an older technology to examine chromatin, called Hi-C, that maps which pieces of the genome are close together. It can identify chromosome translocations, like those that occur in some cancers. Its limitation, however, is that it sees only these adjacent genomic regions.

The new technology, called Pore-C, uses much more data to visualize how all of the pieces within a cell's nucleus interact. The researchers used a mathematical technique called hypergraphs. Think: three-dimensional Venn diagram. It allows researchers to see not just pairs of genomic regions that interact but the totality of the complex and overlapping genome-wide relationships within the cells.

"This multi-dimensional relationship we can understand unambiguously. It gives us a more detailed way to understand organizational principles

inside the nucleus. If you understand that, you can also understand where these organizational principles deviate, like in cancer," Rajapakse said. "This is like putting three worlds together—technology, math and biology—to study more detail inside the nucleus."

The researchers tested their approach on neonatal fibroblasts, biopsied adult fibroblasts and B lymphocytes. They identified organizations of transcription clusters specific to each cell type. They also found what they called self-sustaining transcription clusters, which serve as key transcriptional signatures for a cell type.

Rajapakse describes this as the first step in a bigger picture.

"My goal is to construct this kind of picture over the cell cycle to understand how a cell goes through different stages. Cancer is uncontrollable cell division," Rajapakse said. "If we understand how a normal cell changes over time, we can start to examine controlled and uncontrolled systems and find ways to reprogram that system."

More information: Gabrielle A. Dotson et al, Deciphering multi-way interactions in the human genome, *Nature Communications* (2022). DOI: 10.1038/s41467-022-32980-z

Provided by University of Michigan