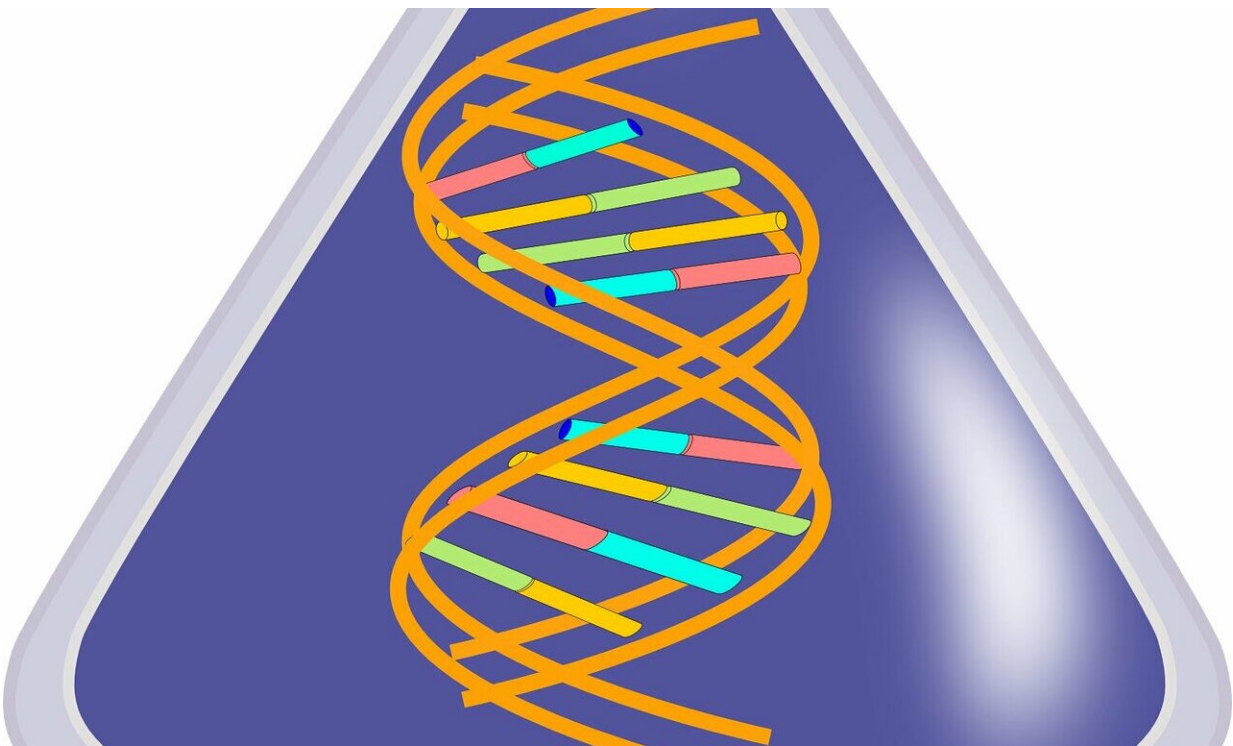# CICERO software enables detection of cancer-causing gene fusions

June 15 2020

Credit: CC0 Public Domain

After years of development, engineering and enhancement, researchers at St. Jude Children's Research Hospital have made publicly available a software system that enables better detection of gene fusions. The system, called CICERO, offers additional insights into cancers, as well as new targets for drug treatments. The latest version of CICERO was

published in *Genome Biology*.

"In both pediatric and adult cancers, gene fusions can be valuable targets for drug treatment," said Jinghui Zhang, Ph.D., St. Jude Department of Computational Biology chair. "In many pediatric cancers, they are the initiating genomic alterations that drives the tumorigenesis of a cancer, which means that drugs targeting that gene fusion can effectively treat the cancer."

The importance of gene fusions

Gene fusions are errors made during the replication of chromosomes as cells divide to make new cells. Hybrid genes are mistakenly formed when pieces of two previously independent genes are joined. Such fusions may consist of rearrangement, duplications, truncation or inversion of genes.

Genes are the blueprints of proteins in the cell, and cancer-causing fusions may activate genes to over-produce proteins. They may also create proteins that are continually switched on, driving the over proliferation of cancer cells. Additionally, disruption of tumor-suppressor gene through gene fusion could also trigger cancers.

In developing CICERO, Zhang and her colleagues took advantage of a technology called RNA sequencing. Using this technique, they analyzed cancer cells to detect fusion events. CICERO distinguishes fusion events by comparing the cells' RNA sequence, called the "transcriptome," with the published reference version of the DNA sequence of the human genome.

The RNA sequence of a cell is the template transcribed from the cell's DNA, which is the genetic blueprint of the cell. The cell uses the RNA template to manufacture the cell's workhorse proteins. Among these

proteins are cell switches that can function abnormally to drive cancers.

Zhang's lab programmed CICERO to look for RNA segments in the transcriptome with partial match normal DNA segments in the reference human genome but also contain fragments that do not match the reference human genome. Thus, CICERO could distinguish the telltale signatures of gene fusions.

The system at work

"The key to CICERO's ability to distinguish cancer-causing fusions from technical artifacts was a set of signal-to-noise-recognition procedures programmed into the system," said first author Liqing Tian, Ph.D., a bioinformatic research scientist in Zhang's laboratory. "These included assembly of the aberrant reads into a mini-genome that represent gene fusion; filtering technical artifacts; and performing extensive annotation to prioritize potential cancer-causing fusions."

In addition to automated analysis, a visualization tool, called FusionEditor, has been developed.

"By visually presenting the predicted gene fusions, FusionEditor enables investigators to incorporate their biological knowledge about this disease to sort out whether fusions discovered by CICERO are likely to be relevant from the disease point of view," Zhang said.

To test CICERO's ability to detect cancer-causing fusions, researchers conducted benchmark tests in which they applied the system to the transcriptomes of 170 pediatric leukemias, solid tumors and brain tumors whose driver fusions had been previously analyzed by alternative technology. CICERO outperformed other commonly used methods of detecting driver fusions.

To apply CICERO to an adult cancer, the investigators reanalyzed RNA sequence data from the glioblastoma, an adult brain cancer and compared the results with previously reported gene fusions in this disease.

"Finding any novel fusions in such a re-analysis would be difficult, but if we did it would prove that CICERO can also be a useful tool for adult cancer." Zhang said. "Indeed, we did discover several fusions that were missed by other methods that are definitely targetable by drugs."

The previously undiscovered glioblastoma fusions included those involving a gene called EGFR, one of the most frequently mutated genes in that brain cancer. The glioblastoma analysis revealed a surprising number of EGFR fusions that result in a protein missing one end, Zhang said. This could lead to loss of the protein's off-switch and drive the cancer.

St. Jude researchers have implemented CICERO on St Jude Cloud, named as RapidRNA-seq, to enable fast analysis of RNA-seq which may be important in delivery time-critical make it widely available.

  More information: Liqing Tian et al, CICERO: a versatile method for detecting complex and diverse driver fusions using cancer RNA sequencing data, *Genome Biology* (2020). DOI: 10.1186/s13059-020-02043-x

Provided by St. Jude Children's Research Hospital