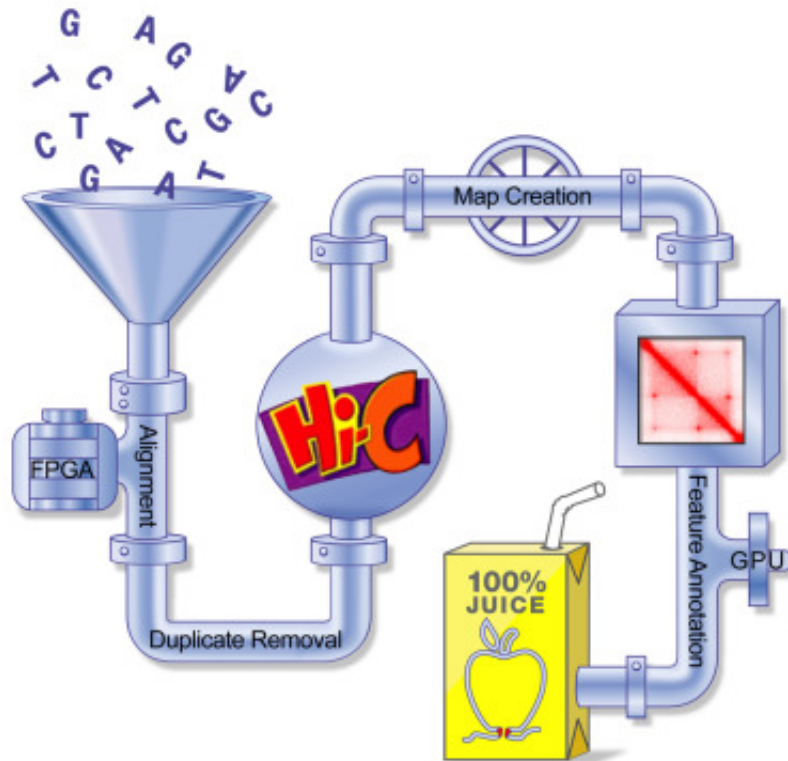


New studies explore 3-D structure of DNA

August 1 2016



Credit: *Cell Systems* (2016). DOI: 10.1016/j.cels.2016.07.002

In a set of papers published last week in *Cell Systems*, Dr. Erez Lieberman Aiden, assistant professor of molecular and human genetics and McNair Scholar at Baylor College of Medicine and director of the Center for Genome Architecture (TC4GA), and his colleagues introduce Juicer, an open-source tool used in three-dimensional (3-D) genome sequencing (Hi-C) processes.

Hi-C, invented by Aiden and collaborators in 2009, explores the three-dimensional structure of the [genome](#), creating terabases of sequencing data resulting in high-resolution contact maps that comprehensively chart the loops that form when the genome folds up inside the nucleus of a cell.

In previous Hi-C experiments, Aiden and his team identified the sheer bandwidth of the data as a central challenge. Existing hardware and software simply could not process and analyze the massive amounts of data produced in these experiments, with a single map spanning billions of reads and trillions of base pairs.

To alleviate this bottleneck in data analysis, Aiden and his team at Baylor, led by Dr. Neva Durand, Muhammad Shamim and Ido Machol, designed Juicer, a fully-automated pipeline that allows users with little to no computational background to transform raw sequencing data into genome-wide maps of looping with a single click. Juicer produces the Hi-C file with loops and contact domains automatically annotated, which facilitates the visualization and analysis of the map and its structural features.

"The studies published in *Cell Systems* describe our team's new, end-to-end system for analysis of 3-D genome sequencing data. It is the first system of its kind, making it possible to map the loops in a mammalian genome in a fully automated fashion," said Durand, a senior scientist at TC4GA and co-first author on both new studies.

As a demonstration of the power of the new tool, Aiden and his colleagues created the deepest 3-D maps of the genome to date, spanning over three terabytes of data drawn from a single experimental condition.

But improvements in software weren't enough: adequate hardware is also a central challenge. The researchers tracked the performance of Juicer

on four cluster systems, including a system based on Edico Genome's DRAGEN Bio-IT processing platform coupled with IBM's Power8 architecture.

Edico's DRAGEN platform accelerated the analysis of the massive data sets derived from this study of 3-D structures of DNA by nearly 20 fold, a dramatic speedup from all of other systems tested.

Machol, a co-author on both studies, noted that, "When we ran our pipeline on a hybrid DRAGEN/Power system, the data analysis was 20-fold faster than running the pipeline on an industry standard cluster. That kind of difference opens the door to many analyses that would have been very impractical before."

DRAGEN generates accelerated implementations of genome pipeline algorithms using a field-programmable gate array (FPGA). The platform is reconfigurable and flexible through remote downloads, allowing users to create custom algorithms and refine existing pipelines.

"Given the dramatic acceleration that we observed, we are excited about the extraordinary potential of FPGA technology in 3-D genomics." said Shamim, an M.D./Ph.D. student at Baylor and co-first author on the Juicer study.

Aiden, who is also a faculty member at Rice University, in the department of Computer Science and at the Center for Theoretical Biological Physics, commented on the experiment, saying, "The partnership between TC4GA and Edico Genome is a game-changer. The results that are possible using DRAGEN are more than a one-off exercise: they are a strong indicator of the future of the 3-D genomics field as a whole. We are confident that our collaboration will lead to a great deal of innovation both within the Texas Medical Center community, and beyond."

Added Pieter van Rooyen, Ph.D., chief executive officer of Edico Genome, "Dr. Aiden and his team's application of DRAGEN to accelerate Juicer is a great example of DRAGEN's effectiveness in processing massive amounts of raw sequencing data in minimal time, and without requiring any additional training or a post-graduate degree. We are continually working to optimize DRAGEN and expect the next version to be even faster than the speed we have already achieved."

Juicer is available as open source software and is compatible with multiple cluster operating systems, Edico's DRAGEN, and Amazon Web Services. It may be downloaded on the web at <http://aidenlab.org/juicer/>.

More information: Neva C. Durand et al, Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments, *Cell Systems* (2016). [DOI: 10.1016/j.cels.2016.07.002](https://doi.org/10.1016/j.cels.2016.07.002)

Provided by Baylor College of Medicine

Citation: New studies explore 3-D structure of DNA (2016, August 1) retrieved 10 July 2024 from <https://phys.org/news/2016-08-explore-d-dna.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.