

Supercomputer dramatically accelerates rapid genome analysis

February 19 2014



Beagle, a Cray XE6 supercomputer at Argonne National Laboratory, supports computation, simulation and data analysis for the biomedical research community. Credit: Argonne National Laboratory

Although the time and cost of sequencing an entire human genome has plummeted, analyzing the resulting three billion base pairs of genetic information from a single genome can take many months.

In the journal *Bioinformatics*, however, a University of Chicago-based

team—working with Beagle, one of the world's fastest supercomputers devoted to life sciences—reports that genome analysis can be radically accelerated. This computer, based at Argonne National Laboratory, is able to analyze 240 full genomes in about two days.

"This is a resource that can change patient management and, over time, add depth to our understanding of the genetic causes of risk and disease," said study author Elizabeth McNally, MD, PhD, the A. J. Carlson Professor of Medicine and Human Genetics and director of the Cardiovascular Genetics clinic at the University of Chicago Medicine.

"The [supercomputer](#) can process many genomes simultaneously rather than one at a time," said first author Megan Puckelwartz, a graduate student in McNally's laboratory. "It converts whole [genome sequencing](#), which has primarily been used as a research tool, into something that is immediately valuable for patient care."

Because the genome is so vast, those involved in clinical genetics have turned to exome sequencing, which focuses on the two percent or less of the genome that codes for proteins. This approach is often useful. An estimated 85 percent of disease-causing mutations are located in coding regions. But the rest, about 15 percent of clinically significant mutations, come from non-coding regions, once referred to as "junk DNA" but now known to serve important functions. If not for the tremendous data-processing challenges of analysis, [whole genome sequencing](#) would be the method of choice.

To test the system, McNally's team used raw sequencing data from 61 human genomes and analyzed that data on Beagle. They used publicly available software packages and one quarter of the computer's total capacity. They found that shifting to the supercomputer environment improved accuracy and dramatically accelerated speed.

"Improving analysis through both speed and accuracy reduces the price per genome," McNally said. "With this approach, the price for analyzing an entire genome is less than the cost of the looking at just a fraction of genome. New technology promises to bring the costs of sequencing down to around \$1,000 per genome. Our goal is get the cost of analysis down into that range."

"This work vividly demonstrates the benefits of dedicating a powerful supercomputer resource to biomedical research," said co-author Ian Foster, director of the Computation Institute and Arthur Holly Compton Distinguished Service Professor of Computer Science. "The methods developed here will be instrumental in relieving the data analysis bottleneck that researchers face as genetic sequencing grows cheaper and faster."

The finding has immediate medical applications. McNally's Cardiovascular Genetics clinic, for example, relies on rigorous interrogation of the genes from an initial patient as well as multiple family members to understand, treat and prevent disease. More than 50 genes can contribute to cardiomyopathy. Other genes can trigger heart failure, rhythm disorders or vascular problems.

"We start genetic testing with the patient," she said, "but when we find a significant mutation we have to think about testing the whole family to identify individuals at risk."

The range of testable mutations has radically expanded. "In the early days we would test one to three genes," she said. "In 2007, we did our first five-gene panel. Now we order 50 to 70 genes at a time, which usually gets us an answer. At that point, it can be more useful and less expensive to sequence the whole [genome](#)."

The information from these genomes combined with careful attention to

patient and family histories "adds to our knowledge about these inherited disorders," McNally said. "It can refine the classification of these disorders," she said. "By paying close attention to family members with genes that place them at increased risk, but who do not yet show signs of disease, we can investigate early phases of a disorder. In this setting, each patient is a big-data problem."

Beagle, a Cray XE6 supercomputer housed in the Theory and Computing Sciences (TCS) building at Argonne National Laboratory, supports computation, simulation and data analysis for the biomedical research community. It is available for use by University of Chicago researchers, their collaborators and "other meritorious investigators." It was named after the HMS Beagle, the ship that carried Charles Darwin on his famous scientific voyage in 1831.

Provided by University of Chicago Medical Center

Citation: Supercomputer dramatically accelerates rapid genome analysis (2014, February 19) retrieved 19 April 2024 from

<https://medicalxpress.com/news/2014-02-supercomputer-rapid-genome-analysis.html>

<p>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.</p>
--