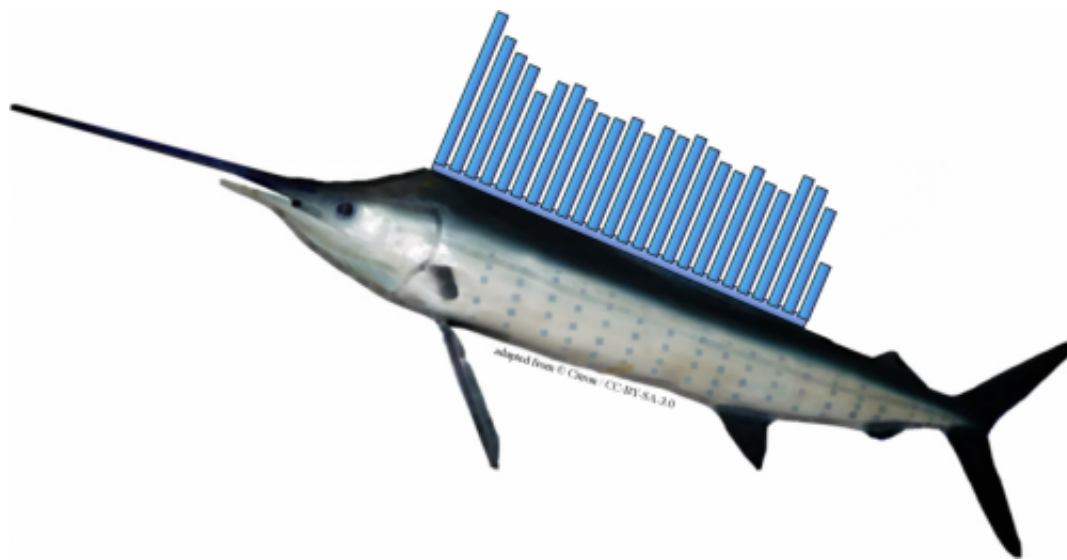


Computational method dramatically speeds up estimates of gene expression

April 20 2014



With gene expression analysis growing in importance for both basic researchers and medical practitioners, researchers at Carnegie Mellon University and the University of Maryland have developed a new computational method that dramatically speeds up estimates of gene activity from RNA sequencing (RNA-seq) data.

With the new method, dubbed Sailfish after the famously speedy fish, estimates of [gene expression](#) that previously took many hours can be completed in a few minutes, with accuracy that equals or exceeds

previous methods. The researchers' report on their new method is being published online April 20 by the journal *Nature Biotechnology*.

Gigantic repositories of RNA-seq data now exist, making it possible to re-analyze experiments in light of new discoveries. "But 15 hours a pop really starts to add up, particularly if you want to look at 100 experiments," said Carl Kingsford, an associate professor in CMU's Lane Center for Computational Biology. "With Sailfish, we can give researchers everything they got from previous methods, but faster."

Though an organism's genetic makeup is static, the activity of individual genes varies greatly over time, making gene expression an important factor in understanding how organisms work and what occurs during disease processes. Gene activity can't be measured directly, but can be inferred by monitoring RNA, the molecules that carry information from the genes for producing proteins and other cellular activities. RNA-seq is a leading method for producing these snapshots of gene expression; in genomic medicine, it has proven particularly useful in analyzing certain cancers.

The RNA-seq process results in short sequences of RNA, called "reads." In previous methods, the RNA molecules from which they originated could be identified and measured only by painstakingly mapping these reads to their original positions in the larger molecules.

But Kingsford, working with Rob Patro, a post-doctoral researcher in the Lane Center, and Stephen M. Mount, an associate professor in Maryland's Department of Cell Biology and Molecular Genetics and its Center for Bioinformatics and Computational Biology, found that the time-consuming mapping step could be eliminated. Instead, they found they could allocate parts of the reads to different types of RNA molecules, much as if each read acted as several votes for one molecule or another.

Without the mapping step, Sailfish can complete its RNA analysis 20-30 times faster than previous methods.

This numerical approach might not be as intuitive as a map to a biologist, but it makes perfect sense to a computer scientist, Kingsford said. Moreover, the Sailfish method is more robust—better able to tolerate errors in the reads or differences between individuals' genomes. These errors can prevent some reads from being mapped, he explained, but the Sailfish method can make use of all the RNA read "votes," which improves the method's accuracy.

More information: The Sailfish code has been released and is available for download at www.cs.cmu.edu/~ckingsf/software/sailfish/

Paper: Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms, *Nature Biotechnology*, DOI: [10.1038/nbt.2862](https://doi.org/10.1038/nbt.2862)

Provided by Carnegie Mellon University

Citation: Computational method dramatically speeds up estimates of gene expression (2014, April 20) retrieved 23 April 2024 from <https://phys.org/news/2014-04-method-gene.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.