

Novel algorithm identifies DNA copy-number landscapes in African American colon cancers

July 23 2015

An algorithm dubbed ENVE could be the Google for genetic aberrations—and it comes from Case Western Reserve.

Remember the World Wide Web before the famed search engine? The web offered extraordinary amounts of information, but no consistently reliable way to secure relevant results.

Cancer researchers at Case Western Reserve encountered a comparable conundrum when considering reams of data about the body that new technological advances provide—how could they tell what parts of the information actually offer value. In this instance, the goal was to distinguish between distracting or even misleading material and evidence worthy of action.

Their answer—as well as remarkable findings involving genetic characteristics of African Americans with colon cancer—appears this week in the journal *Genome Medicine*.

ENVE—for Extreme Value Distribution based Somatic Copy-Number Variation Estimation—is designed to sort through what the researchers call "white noise" and actual signs of trouble. More specifically, the algorithm aims to eliminate once-common false positives or negatives and instead reveal areas where genetic changes have gone awry. This new degree of guidance could translate to more effective treatments for

multiple kinds of cancer, as well as other conditions.

"Our algorithm now resolves most of the noise issues," said senior author Kishore Guda, DVM, PhD, assistant professor of General Medical Sciences-Oncology, Case Comprehensive Cancer Center.

As part of their efforts, Guda and his colleagues tested the approach on two sets of colorectal cancer tissue samples—one consisting entirely of Caucasians and the other of African Americans. In the U.S., African Americans suffer higher rates of [colon cancer](#) incidence and deaths than any other racial or ethnic group. First, the researchers compared the ENVE model to an existing state-of-the-art algorithm, and found ENVE consistently outperformed its rival. Next they looked at differences between the two racial groups; while the major genetic changes were similar, there were also regions where they differed.

The researchers specifically applied ENVE to detect "copy number alterations" in DNA, the molecule that provides genetic instructions to each cell in the body. While DNA has the ability to replicate itself so that new cells get the appropriate orders, the copies are not always exact replicates. Some DNA copy-number alterations—structural deviations in the DNA duplication—can give rise to cancer or other illnesses.

DNA sequencing is a process that identifies the exact order of the molecule's instructions. Over the years, new technologies have allowed scientists to gain increasingly vast and precise information; the latest generation of DNA sequencing technology enables scientists to analyze thousands of genes at once, rather than one at a time. This technology zeroes in on exomes, part of the molecule that includes directions involving protein function—and also the source of a significant majority of deviations that lead to disease or other problematic conditions.

While this next-generation method—whole-exome sequencing—is

superior to previous technologies, it is also error-prone because of steps involved in the process of converting biological samples to data that computers can read. Such errors—inherent noise—can lead to an increase in both false positive and false negative rates during the detection of copy-number changes in genomic regions.

"Now we have a background noise modeling framework that can distinguish between deviations in tumor sequencing data that arise due to inherent noise versus those that arise due to real copy-number alterations in the tumor DNA," said lead author Vinay Varadan, PhD, assistant professor, Case Comprehensive Cancer Center.

Guda calls Varadan's mathematical framework "an ingenious application" of a probabilistic method previously used in modeling weather patterns and financial markets and applying that method to interpret copy-number alterations in tumor exome sequencing data. The idea is to distinguish between the variations arising due to the technical inconsistencies from those arising due to actual copy-number changes in tumor DNA.

"What makes our algorithm particularly robust is that it works without fixed, user-defined criteria that are often arbitrary. All that is required for ENVE is 15 to 20 normal tissue samples processed on the same sequencing platform to quantify the extent of inherent noise in the platform," Varadan said.

"Ours is the first study to characterize genome-wide copy-number landscapes in African American colon cancers," Guda said. Still, the focal copy-numbered regions identified by Guda, Varadan and colleagues need further validation.

"Our next objective is to compare even more cancerous colon tissue samples from African American and Caucasian patients, sequenced

using the same platform, to confirm these focal copy-number alterations selectively identified in African American colon cancers," Guda said. "Once we have that, we can then focus on figuring out if these copy-number alterations have a role in contributing to the aggressive colon tumor phenotypes in African Americans."

"We have already performed extensive pressure-testing of the algorithm. Still, as the software is being applied to new sequencing platforms and datasets by other researchers, we will be able to identify and actively resolve any software glitches that come up," Varadan said. "We therefore welcome interactions and feedback from researchers who are willing to apply the algorithm on their own datasets."

The two researchers are also currently optimizing ENVE to enable detection of copy-number alterations in formalin-fixed paraffin-embedded (FFPE) tissue samples instead of relying solely on fresh, frozen tissue sample sources. FFPE tissue samples are preserved with formalin, a solution that contains formaldehyde, and embedded in paraffin wax. It is the FFPE biospecimens that are archived most frequently.

"To date, there are no algorithms available for profiling such alterations in deep sequencing datasets derived from FFPE samples," Guda said. "We anticipate releasing a newer version of ENVE in the near future that incorporates this module such that researchers would be able to make use of the vast FFPE resources held in hospital pathology archives."

More information: Guda and Varadan invite the scientific community to test the ENVE algorithm by accessing a [special link](#). Readers can download and use the algorithm at no charge as long as it is used for academic, not-for-profit use.

Provided by Case Western Reserve University

Citation: Novel algorithm identifies DNA copy-number landscapes in African American colon cancers (2015, July 23) retrieved 26 April 2024 from <https://medicalxpress.com/news/2015-07-algorithm-dna-copy-number-landscapes-african.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.