

Two software tools that improve identification of cancer biomarkers earn certification

August 18 2009



Georgia Tech biomedical engineering associate professor May Dongmei Wang and graduate students John Phan and Todd Stokes (left-right) use the caBIG-certified omniBioMarker software to identify and validate cancer biomarkers from high-throughput gene expression data. Credit: Georgia Tech Photo: Gary Meek

The explosive growth of genomic and proteomic data has ushered in a new era of molecular medicine in which cancer detection, diagnosis and treatment are tailored to each individual's molecular profile. But this personalized medicine approach requires that researchers discover and link biomarkers -- such as genes or proteins -- to specific disease

behaviors, such as the rate of tumor progression and different responses to treatments.

Two new software programs that help address that challenge have recently earned silver-level compatibility certification from the National Cancer Institute's cancer Biomedical Informatics Grid®, also known as caBIG®. The programs improve the process of identifying cancer biomarkers from gene expression data.

Developed by May Dongmei Wang and her team in the Wallace H. Coulter Department of Biomedical Engineering at Georgia Tech and Emory University, the programs -- caCORRECT and omniBioMarker -- remove noise and artifacts, and identify and validate biomarkers from microarray data. Funding to develop the programs was provided by the National Institutes of Health, the Georgia Cancer Coalition, Microsoft Research and Hewlett-Packard.

"Certification by caBIG means the tools can be easily used by everyone in the cancer community to improve approaches to [cancer detection](#), diagnosis, treatment and prevention," said Wang, an associate professor in the Coulter Department and a Georgia Cancer Coalition Distinguished Cancer Scholar.

caBIG is a collaborative information network that enables researchers, physicians, and patients to share data, tools and knowledge to accelerate the discovery of new approaches that they hope will ultimately improve cancer patient outcomes. To become caBIG-certified, caCORRECT and omniBioMarker passed a rigorous set of requirements, ensuring the cancer research community that the software tools are high quality and interoperable with all other caBIG-certified systems for nationwide deployment.

caCORRECT -- chip artifact CORRECTION -- is a software program

that improves the quality of collected microarray data, ultimately leading to improved biomarker selection. Widely used Affymetrix microarrays contain thousands of probes, each including a 25-oligo sequence, which are used to detect mRNA expression levels.

"Once someone has collected microarray data, it is important to run quality control on it and remove any problematic points of data that could highlight incorrect biomarkers when analyzed," explained Wang, who is also director of the biocomputing and bioinformatics core in the Emory-Georgia Tech National Cancer Institute Center for Cancer Nanotechnology Excellence (CCNE).

Since each microarray chip contains thousands of spots, it is easy for a few spots to become marred by artifacts and noise. These unusable portions are typically the result of experimental variations by different laboratory technicians or errors that create scratches, edge effects and bubble effects on the data.

caCORRECT removes the noise and artifacts from the data, while retaining high-quality genes on the array. The software can also effectively recover lost information that has been obscured by artifacts.

In collaboration with Andrew N. Young, an associate professor in pathology and laboratory medicine at Emory University School of Medicine and clinical laboratory director at Grady Health System, Wang and graduate students Todd Stokes, Martin Ahrens and Richard Moffitt validated the caCORRECT software. A large-scale survey of public data and data from Young's laboratory demonstrated the ability of caCORRECT to assess and improve the quality of a wide array of datasets.

"caCORRECT is a quality assurance tool that allows researchers to utilize and trust imperfect experimental microarray data that they spent a

tremendous amount of time and money to generate," added Wang. "caCORRECT improves the downstream analysis of microarray data and should be used before conducting biomarker selection, therapeutic target studies, or pathway analysis studies in bioinformatics and systems biology."

Once the quality of the data is assured with caCORRECT, researchers can use the caBIG-certified omniBioMarker software to identify and validate biomarkers from the high-throughput [gene expression](#) data.

Candidate cancer biomarkers are typically genes expressed at different levels in cancer patients compared to healthy subjects. omniBioMarker searches these groups of patient data for genes with the highest potential for accurately determining whether a patient has cancer. However, because individual genes are not expressed independently, the software also identifies groups of genes that act in concert.

The advantage of the omniBioMarker software is that it fine-tunes biomarker selection to a particular dataset or clinical problem based on prior biological knowledge. It also applies unique analysis parameters for each specific clinical problem. The parameters are optimal when the software selects genes that are known to be relevant biomarkers based on clinical observations and laboratory experiments available in literature and public databases. Then the software finds new potential biomarkers for experimental validation.

Wang, graduate student John Phan and Young tested the ability of the software to identify biomarkers in clinical renal cancer microarray data. The researchers selected renal [cancer](#) for study because it has several distinct subtypes, which can appear in the same person in varying degrees and must be treated according to the diagnosed subtype to maximize treatment success. The results indicate that integrating prior laboratory and clinical knowledge with the microarray data improves

biomarker selection.

"Using omniBioMarker to create an optimal metric for ranking and identifying novel biomarkers reduces the number of false discoveries, increases the number of true discoveries, reduces the required time for validation and increases the overall efficiency of the process," noted Wang.

Since receiving caBIG silver-level compatibility certification for caCORRECT and omniBioMarker, Wang and her team have been working on getting two more software programs certified -- Q-IHC, a tool that analyzes and quantifies multi-spectral images such as quantum dot-stained histopathological images, and omniVisGrid, a grid-based tool that visualizes data and analysis processes of microarrays, biological pathways and clinical outcomes.

Source: Georgia Institute of Technology

Citation: Two software tools that improve identification of cancer biomarkers earn certification (2009, August 18) retrieved 2 May 2024 from <https://medicalxpress.com/news/2009-08-software-tools-identification-cancer-biomarkers.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>
--