

New research tools are too complex for easy answers, researchers say

December 27 2007

Scientists who study cancer may be prone to drawing simplistic conclusions from the powerful molecular tools now available because they don't appreciate how complex the data is that is being generated, said a team of Georgetown University Medical Center (GUMC) researchers in the January issue of *Nature Reviews Cancer*.

In a review article summing up the state of the field, they said cancer investigators should endeavor to better understand the issues these genomic and proteomic technologies create or conclusions from their research may be misleading.

"These tools have allowed us to see that nature is more complex than we thought, and while we don't yet know what the overarching biological rules are – such as the interrelationship between multiple signaling pathways that can lead to cancer development – we are trying to play the game like we do," said the review's lead author, Robert Clarke, Ph.D., D.Sc., professor of oncology and physiology & biophysics at the Lombardi Comprehensive Cancer Center at GUMC, where he co-directs the Breast Cancer Program. Clarke is the interim director of GUMC's Biomedical Graduate Research Organization, which is home to more than 60 percent of the University's biomedical research funding.

"The answers to our questions are probably there in the data," he said, "but the issue is whether we can get them using these complex tools and, also, how we will know they are right when we see them."

Clarke led the analysis with six other scientists from Georgetown and from Virginia Polytechnic Institute. GUMC is pioneering a field of systems medicine study designed to understand the theory and properties of the data generated by these new tools and how they may affect data analysis and interpretation.

“This review addresses the challenges in reducing high-dimensional molecular data and making the output relevant to cancer treatment,” said Dr. Howard Federoff, executive vice president for health sciences at GUMC. “There is no doubt that the integration of traditional clinical data alongside transcriptomic and proteomic data will result in a change in our understanding of disease mechanisms, likely drive a revision in nosology and have meaningful impact on patients with cancer. I place great value on this systems medicine approach because it heralds the future of medical practice and holds promise to transform healthcare.”

The genomic and proteomic technologies used in cancer research help provide a snapshot of the molecular workings of cancer cells.

Researchers hope to identify the genes that are active during cancer development and which transcribe the messenger RNA (mRNA) needed to produce the proteins that actually do the “work” of the cell. In theory, knowing the genes, mRNA, and proteins that are linked to specific cancers will help researchers build better predictive models of diagnosis, prognosis, and therapy.

But there are thousands of active molecules in a single slice of a tumor analyzed after surgical removal, Clarke said, and this produces “very high-dimensional data spaces.” That means that a molecular snapshot could “have 10,000 or so dimensions if you consider a molecule working along a pathway as a dimension. Think of a box which is described as having a height, width and length, but if you add color and the box’s fiber, you have two more dimensions. There are countless things going on in a cell that could describe it – this is the essence of multi-

dimensionality and these tools tell you all of that, ” said Clarke.

But there are perils in generating such large amounts of data, Clarke said, because the data being generated will not all be relevant to the question researchers are trying to ask since there are countless dynamic processes ongoing at one time within a tumor. “Some cells in a tumor are dying, some are not. Some are growing, others are not. Some are trying to spread and the rest aren’t,” Clarke said. “Everything is going on in a tumor at once, and all of these activities require coordination of different genes. So it may not be accurate to analyze these molecules as if they are all focused on performing a single function.

“We need to discover what specific genes perform which function. If we knew the rules – what genes are involved in which process – we should be able to understand some of the questions we have, but we are not there yet,” he said.

Despite the lack of understanding, many studies have been published that link specific “biomarkers” –genes, mRNA or proteins – with an aspect of cancer development or treatment, and the results often appear to be statistically valid, Clarke said. “But it is not clear that that solution is complete or is necessarily correct. It may be partly right and may be intuitively pleasing because you are getting what you expected to see from an experiment. That could be a trap, a self-fulfilling prophecy.”

And while the findings may “fit” in the tumor samples they are tested in, they may not if other tumor tissue is studied, and many times researchers don’t take that extra step, the researchers said in their article. “The lack of rigorous validation is a problem that currently plagues cancer research, Clarke added.

Another pitfall in using the new technology is the “curse of multi-dimensionality,” Clarke said. “You have a lot of measurements, and the

statistical model gets very complicated. So sometimes you don't have enough computing power to derive the right answer or you get an answer that is only true for part of the data.”

In other words, scientists don't always know what they don't know when looking at multi-dimensional data sets.

“We still don't always have enough knowledge to know whether we have the answers right or not.”

Source: Georgetown University Medical Center

Citation: New research tools are too complex for easy answers, researchers say (2007, December 27) retrieved 20 April 2024 from <https://medicalxpress.com/news/2007-12-tools-complex-easy.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.