caBIG: Seeking Cancer Cures by Bits and Bytes

Wendy Wolfson
DOI 10.1016/j.chembiol.2008.06.003

While our understanding of cancer is becoming increasingly nuanced, we still don't know why some people get certain cancers and others do not, nor why cancers behave differently in some patients. "Virtually every man will get prostate cancer if he lives long enough." says David Steffen, Ph.D., director, Bioinformatics Research, Department of Molecular and Human Genetics at Baylor University. "For most it isn't serious. For a few it is." According to Steffen, even with a careful statistical study, it is hard to verify biomarkers that can be used to distinguish the aggressive cancer from the watchful waiting variety. "Solving this problem would be fundamental to all cancer research," Steffen says. "It would be huge."

Is That an Elephant in the Room?

"We are literally at the precipice of a revolution," says Kenneth H. Buetow, Ph.D., director, Center for Biomedical Informatics and Information Technology, National Cancer Institute. "It is a revolution that is overdue in biomedicine."

Leaps in computing capacity as well as a flowering of genetic research have enabled laboratories to churn out a huge amount of data. For example, a plethora of genome-wide association studies are uncovering new genes statistically associated with disease, but researchers do not yet know how they function or how other genes and environmental influences act upon them. Buetow likens it to the metaphor of blind men trying to identify an elephant in a room by feeling its parts.

caBIG is allowing us to move from a set of isolated skirmishes to global strategy.

Redefining Cancer as an Information Problem

Steffen is hoping that the Cancer Biomedical Information Grid (caBIG) will provide answers by supplying a framework that allows Baylor to share data efficiently with the ten other National Cancer Institute (NCI) prostate cancer SPOREs (short for Specialized Programs of Research Excellence). caBIG was established by the NCI in 2004 to help cancer researchers use the internet to exchange information about genetics, treatments, molecular biology, tissue samples, and clinical trials, in some cases in real time. The consortium is also developing tools for imaging, classifying, and analyzing data. The caBIG consortium now includes over 1000 participants, including federal agencies, academic centers, and industry. Forty-six NCI-designated cancer centers are participating, as well as 16 community cancer centers. This $60 million information technology initiative completed its three year pilot phase about year and a half ago.

"There was a pent up demand for this," Buetow says. "We stepped into this because the community was asking for this."

According to Dr. H. Kim Lyerly, director of the Duke Comprehensive Cancer Center, researchers attempting to statistically correlate information gathered from sophisticated multigene biomarkers or gene expression arrays of 30,000 genes with relatively simple outcomes such as "would a patient live or die after a certain number of years" and "how did the pattern of gene expression relate to the treatment the patient received" were hitting a wall. By the time proteomic and metabolic data were layered in, "the complexity of trying to find correlations or associations between these typically disparate data sets made it pretty much impossible to do," says Lyerly. "It became increasingly apparent [that] a single strategy to interrogate across a system won't be informative as an unbiased approach where we use a variety of strategies. The tools [of caBIG] enabled new approaches. To fully utilize the power of the tools, we needed a new infrastructure." Duke University is now utilizing caBIG to analyze gene expression patterns in breast cancer patients and compare this to how they respond to therapy in real time.

Wrangling the Data

caBIG is a federated grid made up of layers of metadata so users can query data from different sources via a simple interface, no matter what software platform they happen to use. It also is a framework for, at current count, 60 different open source software tools for integrating clinical information with molecular data. Tools include image sharing and analysis, discovery, tissue banking and analysis, and clinical trial management. These adhere to common data standards and a shared, interoperable infrastructure so information can be exchanged while protecting privacy and data security. For example, cancer researchers can now access pooled clinical trial data from multiple places, including patient tissue samples, to investigate rare forms of cancer; a handy ability if they do not have specialized expertise in house. Tissue images can be annotated by different researchers, so overall patterns and relationships become apparent. caBIG's philosophy is open source development with "semantic interoperability" in mind, creating a layered set of standard vocabularies to ensure that everybody describes the elephant's trunk in precisely the same way. Early projects linked to caBIG include TCGA (The Cancer Genome Atlas) and REMBRANDT (Repository for Molecular Brain Neoplasia Data) for researchers to share information on rare brain tumors.

Rolling Out the Grid

In addition to the NCI translational research centers (SPORES), caBIG is now being rolled out to more community cancer centers, 50-odd cooperative groups, as the NCI-designated centers currently

treat only 15% of cancer patients in the US. The consortium is also partnering with the UK's National Cancer Research Institute (NCRI), which has undertaken a comparable project. Using caBIG, Duke University and the NCI also have partnered with the Peking University Health Science Institute to conduct cancer clinical trials in China.

If the framework for information sharing and collaboration can work for cancer, it could work for other diseases as well. Last year, the CardioVascular Research Grid (http://www.cvrgrid.org) was initiated with an $8.5 million, 4 year NIH grant. Based at Johns Hopkins University in Maryland, with partners Ohio State University and the University of California, San Diego, the grid will give researchers the ability to swap information about heart disease and analyze and model patient data. "The big picture with caBIG involves particular tools," says Dr. Joel Saltz, professor and chair of Biomedical Informatics at Ohio State University College of Medicine and the OSU Comprehensive Cancer Center. "The bigger picture is that caBIG will allow large teams of researchers to work together to develop increasingly effective treatments for cancer. Existing approaches to finding treatment have been piecemeal. caBIG is allowing us to move from a set of isolated skirmishes to global strategy."

Caveats

According to the "caBIG Pilot Phase Report 2003–2007," caveats include project definition and development issues such as buggy software and sample sizes too small to get real information. An underlying question is motivation for cooperation. Lyerly noted that while the caBIG community used wide vocabulary standards, there are "a number of people [who are] more methodical or slower adopters." Once software kinks are ironed out, caBIG's success will be measured by the rate of adoption and extent of collaboration within a community of researchers used to competing with each other, as well as by decreasing concerns about intellectual property and patient privacy. But the decentralized nature of caBIG is its strength, both for researchers and patients, because it can easily incorporate innovative ideas into the greater grid.

For Cancer Patients, It's about Time

Back in 1997, Joan Schreiner met Joanne Tyler when both women were undergoing breast cancer treatment. They became friends. Schreiner, the tech savvy former CFO of Shutterfly, had an idea that patients could use the internet to sign up for clinical trials where they could get novel therapies. They presented the concept to oncologists at the University of California, San Francisco, whom they knew in connection with their treatment—John Park and Debu Tripathy, as well as surgeon Laura Esserman—who agreed to sponsor the project at the university. "We knew we wanted a nonprofit and preferably an academic institution to at least initially sponsor this idea and see to its development. We didn't want it to go to a commercial for-profit entity," recalled Joanne Tyler. "We were trying to get a patient-centered service that had integrity and the interests of patients in mind." Schreiner and Tyler raised an initial budget of $30,000 including a grant from Amgen for $10,000. Tyler, now retired, volunteered time to developing Breastcancertrials.org (http://www.breastcancertrials.org) with Elly Cohen, the coordinator who came on board at UCSF. Then caBIG came into the picture and linked Breastcancertrials.org to caMATCH, its workspace to develop tools to match patients to clinical trials. The regional pilot, launched in 2005 and which will run until 2008, was cosponsored by UCSF and NCI, the California Breast Cancer Research Program, and the Department of Defense Breast Cancer Research Program. "It was a new thing for patients to be involved in a substantive way." Tyler says. The soft launch of the new nationwide service is imminent "It is due to the effort of a lot of people, Tyler says. "I'm really pleased. I'm only sorry Joan isn't still around to see how it has developed."

With almost no advertising, a thousand women signed on Breastcancertrials.org. "The internet lends itself to national things," says Dr. Laura Esserman, director, Carol Franc Buck Breast Care Center. "Most people are on the net. Let's make it so it is the norm rather than the exception." Safeway Corporation paid about $1 million to underwrite the national launch. The big expense for the project was the hand coding to put up the trials on the web. Esserman pointed out that drug companies spend billions on clinical trials. "In pediatric oncology, 70% of patients participate in clinical trials because of the way care is given; in adult cancer trials, only 2%–3%." Esserman says. "We want to get that up to 50%." In other words, they want to make it the norm for women to consider trial participation.

Breastcancertrials.org will work in tandem with another caBIG-linked program at UCSF called I-SPY, the next iteration of an adaptive-design clinical trial to determine biomarkers for the most aggressive cancers and how to tailor treatment to them. Under I-SPY, women with stage 2 or 3 breast cancer will be assigned to different arms of a study, receiving standard therapy or standard therapy plus novel agents. Based on ongoing results, regimens will be changed between groups of patients, so in subsequent iterations they will be reassigned to the study branches with the most favorable outcomes. The I-SPY-2 study will commence in collaboration with multiple drug companies, the FDA, and NCI at about 15 study sites in summer 2009. This study will treat patients when first diagnosed, instead of when they present with metastatic disease, so they have a chance to be helped by the drugs. "We are going to learn which agents are most effective, and which biomarkers predict that," says Esserman.

"What they can do is come to conclusions, one way or another, sooner than they could before." says Jane Perlmutter, who works with breast cancer patients in the I-SPY program. "As a patient advocate, I want to push researchers to do innovative and rigorous things."

Wendy Wolfson (wendywolfson@nasw.org) is a science writer based in the Bay Area.

