

Q&A: Paul Workman on Big Data and Large Biases

Integrating genomic and chemobiology data gives fresh views on cancer drug targets and gene networks

Located at The Institute of Cancer Research (ICR) in London, the 165-person Cancer Research UK Cancer Therapeutics Unit is the world's largest academic cancer drug discovery and development organization. "On average, we expect to produce one new preclinical drug candidate per year, and we've consistently more than met that target," says its director, Paul Workman, PhD, who is also deputy chief executive and head of cancer therapeutics at the ICR. Workman talked with *Cancer Discovery's* Eric Bender about the promise of integrating genomic, biologic, and chemical "Big Data."

What is distinctive about the way the Cancer Therapeutics Unit at ICR operates?

I would say technology, culture, and people. Modern drug discovery requires a large, highly integrated team of smart people. Our ability to collaborate creatively with industry and to establish public/private partnerships has been super valuable. We have a focus on what a drug will look like and in which patients it will be used from the outset. We're closely linked to ICR's basic science and have invested heavily in high-throughput screening robotics, crystallography, and high-performance computing for integrative genomics and chemobiology. Crucially, we're all under one roof, co-located with the Royal Marsden Hospital for clinical cancer insights and drug trials. We have a research culture that is really supportive of drug discovery in an academic environment, and we have long-term commitments from both our institution and Cancer Research UK as our main funder, with a research budget that runs for 5-year periods.

How are today's massive databases being integrated?

We're all totally inundated by huge quantities of data, which we must distill down to useable information. We're in the era of the exabyte, which is 10 to the 18th. Many people are now integrating genomics with biology and medicine, and a smaller but significant number are beginning to integrate chemistry as well, which is a really exciting new development. For example, Stuart Schreiber, PhD, from the Broad Institute, in Cambridge, MA, gave a keynote address at the International Chemical Biology Society meeting last fall about taking a computational integrative chemistry and biology approach to come up with intelligent drug combinations, using the 1,000-cell-line panel that the Broad and Novartis established.

What was the goal for your group's canSAR database?

canSAR [<https://cansar.icr.ac.uk/>] is an integrative biologic, chemical, and pharmacologic database. Its purpose is to integrate all these different types of data and make linkage easily accessible to cancer research scientists from multiple disciplines, for example to help with hypothesis generation and to support translational research. Its development was led by Bissan Al-Lazikani, PhD, our computational biology and

chemogenomics team leader. We built it to empower our own scientists to navigate easily and quickly across the links between genes, 3D [3-dimensional] protein structure, and chemical ligands.

How does canSAR's "objective assessment" tool work?

It allows you, in an unbiased way, to analyze large lists of genes and to objectively prioritize the best drug targets. Our experience—and it's how we used to do it ourselves—is that most investigators tend to make a very biased initial assessment of targets. They tend to look at their own data or published literature and say, "This looks like an interesting kinase," or "This looks like something I worked on before," or "This is the next target down the pathway," or "This is something our lab discovered, so we want to do drug discovery on it." We wanted to create the capability to look at very large numbers of genes simultaneously and try to make the initial target assessment much more systematic and objective, based on as much data as you can access on the biology, the actual or predicted 3D protein structures, and small-molecule chemical information.

What have you found so far?

We presented and exemplified our systematic objective large-scale assessment methodology by analyzing the gene list from the UK Sanger Institute's Cancer Gene Census [*Nat Rev Drug Discov* 2013;12:35-50]. Actually, the data surprised us very much. Of the total of 480 genes that we originally looked at, we found a particular group of 46 genes that we predicted would be druggable, are already known to be biologically linked to cancer, and yet have absolutely no chemistry associated with them as revealed by our database analysis. These are clearly overlooked potential drug targets.

Another surprise was that existing drugs and other chemical inhibitors of cancer proteins tend to cluster into distinct areas of the cancer gene network. We've been neglecting not only classes of proteins and individual ones but also clusters of proteins sitting in many gene subnetworks. So let's put some effort into exploring targets in these chemically unexplored areas. We might get better synergistic interactions by hitting different parts of the network.

This network aspect is really important when you think about the main challenge that we currently face in molecularly targeted cancer drugs: designing intelligent combinations that will overcome the problem of clonal heterogeneity and drug resistance mechanisms. The next wave of sophistication will be to understand how any cancer is operating as a dynamic network. My institution and many others have invested heavily in network biology systems approaches, and we will see that play out in the next 5 to 10 years in what I call network medicine. ■



"We wanted to create the capability to look at very large numbers of genes simultaneously" and make initial target assessment more objective, says Paul Workman, PhD.