

Meeting report  
**Digital drug discovery**  
Michael A Goldman

Address: Department of Biology, San Francisco State University, San Francisco, CA 94132-1722, USA. E-mail: goldman@sfsu.edu

Published: 30 September 2005

*Genome Biology* 2005, **6**:348 (doi:10.1186/gb-2005-6-10-348)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2005/6/10/348>

© 2005 BioMed Central Ltd

---

A report of the Cambridge Healthtech Institute conference 'Beyond Genome', San Francisco, USA, 13-16 June 2005.

---

Fallout from the completion of the Human Genome Project and the growing application of our knowledge to medicine includes a blurring of the borders between academia and industry. More than 1,000 individuals from universities, government agencies, financial, biotechnology and pharmaceutical companies - from established scientists and graduate students to financiers - came together in June at the Cambridge Healthtech Beyond Genome conference. This included the 14th annual Bioinformatics and Genome Research meeting, and this year's specialist topics included RNA interference, systems biology, proteomics and genomic variation.

Expectations are high. What we are looking at, according to Leroy Hood (Institute for Systems Biology, Seattle, USA) is the "digitalizing of biology and medicine - a revolution coming". Hood's long-term vision includes new platforms for the analysis of the billions of data points about human biology that we will acquire over the next ten years. These data will be of different kinds, and must be integrated. His institute's database can handle more than a dozen different types of data but, said Hood, "we will need a new math". Beginning with the analysis of model systems such as yeast and sea urchin, Hood ultimately envisions a study of neurobiology that starts from stem cells and moves up to nervous systems. Working with Eric Davidson (CalTech, Pasadena, USA), Hood has modeled a network of 35 genes that control early development in the sea urchin. On the medical front, he emphasized the need to analyze blood serum proteins efficiently as a means of diagnosing and staging disease. Within ten years, he predicted, we will be able to measure 1,000-2,000 serum proteins efficiently enough for use in routine diagnostics. Such analyses will need miniaturizing to achieve the necessary throughput, and the Alliance for Nanosystems Biology, which includes CalTech, the Institute for Systems Biology and the University of California at Los

Angeles, is working on microfluidics, mixing pumps and chambers comprising a nanolaboratory.

But how will we do this exciting new biology? Hood believes that the interdisciplinary work required is difficult within the traditional academic institution, and with the usual mechanisms of funding and publication. The National Institutes of Health (NIH) roadmap for research in the 21st century, published in 2003, clearly sees the importance of novel interdisciplinary projects, but, said Hood, the grant review panels have not caught up. The mainstream journals, too, have been resistant, and dedicated journals for systems biology are coming to the rescue.

Eugene Butcher (Stanford University, USA) believes that systems biology is "ready to be applied to drug discovery". The application of genomics has led to large numbers of potential targets for drug action, but Butcher thinks that target-based drug discovery is failing us: he estimates that no more than three, and sometimes fewer, innovative new drugs are produced each year and that most new drugs derive from previously existing drugs. He considers that a more sensible approach would be to match promising drug molecules to their cognate targets using computational biology. Butcher's systems approach to this uses the BioMAP disease model, which emphasizes the subset of regulatory networks involved in disease processes, and which is derived from a limited number of protein measurements taken from primary human cell cultures of various types. Starting with a database of drug molecules and data on the metabolic systems these drugs perturb, the analysis is automated and reproducible, and the model can be queried in much the same way as commonly used databases. Using BioMAP to look for drug molecules that perturb inflammatory pathways, Butcher's method detected most anti-inflammatory drugs presently on the market, as well as an anticancer drug that was later shown biochemically to affect inflammation. The potential for using the systems approach to identify novel applications for known, safe drugs is enormous and, not surprisingly, the US Food and Drug Administration

(FDA) is interested. Compound screening using a cell-based systems biology approach could shave 3 years and more than \$300 million off the cost of developing a new drug.

Building adequate computational models requires vast quantities of data, and these data must be reliable and reproducible. Microarray studies have been notoriously difficult to evaluate. The FDA lists "sensitivity, specificity, reproducibility, robustness, reliability, accuracy, precision" as some of the challenges in integrating microarray data into drug development and medicine. Investigators at Harvard Medical School and at the National Institute for Standards and Technology (NIST, Gaithersburg, USA) are among those trying to establish appropriate protocols. Zoltan Szallasi (Harvard's Children's Hospital, Boston, USA) pointed out the implications of using a single hybridization protocol for the thousands of distinct probes that comprise a microarray, resulting in widespread cross-hybridization. Szallasi cites several causes for the observed inconsistencies, including the use of incorrect probes, poor understanding of the sequence dependence of  $\Delta G$  (Gibbs free energy change) values of DNA-RNA hybridization, and the folding of labeled transcripts. "How can we trust the fate of patients to microarray measurements if we cannot reproduce the [...] classification with different microarray platforms?" he asked.

Marc Salit (NIST) sees NIST's role in this area as developing the tools needed to understand the performance of gene-expression microarrays. Such tools are likely to include standards, reference data, measurement methods, statistical methods, and thermodynamic models. The complete experiment, from sample preparation through to data analysis and interpretation, can be supported through a better understanding of the underlying measurement. Issues such as RNA sample integrity, microarray scanner performance, hybridization thermodynamics, and quantitative determination of measurement uncertainty will all contribute to that better understanding. One approach currently in use is the measurement of RNA degradation in samples, using fluorescence resonance energy transfer and PCR, to determine the integrity of the transcripts for 'housekeeping' genes. The traditional approaches of metrology - the science of measurement - will be applied to these problems to establish microarray measurements of known quality. Salit considers that the immediate goal is to enable users to understand the quality and meaning of array data.

Computational biology is still not sufficiently powerful to mimic every aspect of a biological system; for that, the cells themselves may still be the best machines. Two interesting approaches that were described at the meeting attempt to model liver cells and cardiac myocytes *in vitro*. About two-thirds of candidate drugs that fail do so because of toxicity or problems in absorption, distribution, metabolism and excretion, accounting for about one-fifth of the cost of drug development, according to Anand Sivaraman (Massachusetts

Institute of Technology, Cambridge, USA). In an attempt to make *in vitro* screening more efficient, his group is growing liver cells in the channels of a microchip, with the flow of culture medium mimicking blood flow in the liver. This three-dimensional bioreactor more faithfully replicates the *in vivo* complexity of the liver itself, resulting in an improved *in vitro* model. Bioreactors have been used to detect the induction of cytochrome P450, part of the liver's system for metabolizing drugs, by xenobiotic agents. While the ultimate goal of this engineering might be to build or repair livers, its immediate usefulness is in screening potential drug molecules for liver toxicity and other key aspects of drug metabolism.

Effects on heart rhythm are among the most prominent and deadly complications of drug treatment, accounting for about half of the pharmaceuticals withdrawn from the market. Simple cellular models for cardiac function have been limited because adult cardiomyocytes tend to dedifferentiate rapidly in culture. Timothy Kamp (University of Wisconsin, Madison, USA) described his team's development of human cardiac myocyte models. Non-human cell lines may not provide an adequate model because the ion-channel proteins are very variable from species to species. Kamp and colleagues have induced human embryonic stem cells (hESCs) to differentiate into cardiomyocytes, and have used these cells to screen for drug toxicity and related properties. Despite the restrictions on the use of hESCs in the US, the WiCells from which Kamp prepared the cardiomyocytes were approved for federally funded research under President George W. Bush's policy of August 2001. The hESC-derived cardiomyocytes beat in culture, and display an action potential characteristic of embryonic, rather than adult, heart cells. For example, elongation of the Q-T interval (which represents the total duration of electrical activity in the ventricles *in vivo*) can be observed; this occurs as a drug side-effect and resulted in the withdrawal of the allergy medication Hismanal in 1999. The Madison-based company Cellular Dynamics International, a spin-off from the University of Wisconsin stem-cell group, is developing ESC technology as a tool in pharmacological studies. But Kamp's ultimate goal is the use of stem cell-derived heart cells in direct therapeutic applications.

Gary Peltz (Roche, Palo Alto, USA) is exploiting quantitative genetics in mice to understand and treat human disease. He described his vision of extending today's healthcare paradigm of diagnosis and therapy to include predisposition screening, targeted monitoring, and an emphasis on preventive medicine. In an attempt to identify genes influencing osteoporosis, Pelz has defined 58 chromosomal regions influencing bone density and strength in mouse models. This work led to the identification of 15-lipoxygenase encoded by a gene on chromosome 11, which affects mesenchymal stem-cell differentiation, as a potential target for therapeutic drugs. To speed such work, Roche maintains an extensive public database of single-nucleotide polymorphisms (SNPs)

in mice [<http://mousesnp.roche.com>], and uses 19 commercially available mouse strains, all of which have been haplotyped, enabling the co-occurrence of quantitative phenotypic traits (traits determined by quantitative trait loci, QTLs) and markers to be determined in days. This is extraordinarily quick compared with a typical QTL analysis, which involves animal breeding for several generations to give thousands of F2 animals, and typically takes up more than ten scientist-years per trait. Pelz and colleagues are next aiming at narcotic addiction treatments, where they have already identified polymorphisms in the  $\beta_2$ -adrenergic receptor as showing a strong correlation with pain tolerance in animals undergoing narcotic withdrawal. The results suggest immediately the possible application of  $\beta_2$ -blocking agents to alleviate the symptoms.

There is general agreement that the pharmaceutical blockbusters of today will give way to medications tailored to common genetic profiles. Russ Altman (Stanford University, USA) called the genes that influence drug responses “pharmacogenes”, and the study of such genes is widely recognized under the banner of pharmacogenomics. Investigators seek to relate genetic variation to differences in drug effectiveness and safety. For example, the metabolism of 6-mercaptopurine, a purine analog used to treat lymphoblastic leukemia, is influenced by the genetically determined activity of the enzyme thiopurine methyltransferase (TPMT). Altman sees much promise in pharmacogenomics, but the science is still in its infancy. Only limited data on genetic variation in drug responses are available in the public domain, and genotype testing is still relatively expensive. Healthcare providers may not be ready to understand and use the information. The pharmaceutical industry, long accustomed to blockbuster drugs, is not fully receptive to the idea of drug markets fragmented by the genetic stratification of patients. Altman’s laboratory manages PharmGKB.org, a public database for pharmacogenomics [<http://pharmgkb.org>]. The site, used by an estimated 25,000 people a month, includes genomics, laboratory and clinical data, and links with Medline, the Protein Data Bank, the SNP database (dbSNP), and GenBank. Relevant pathways have been rendered by artists as Illustrator files and are freely available.

While most speakers referred to personalized medicine, Michael Liebman (Windber Institute and Walter Reed Army Medical Center, Washington DC, USA) considers that the “quality chasm in healthcare between bench and bedside” will be closed only when we recognize “personalized disease”. Invasive ductal carcinoma, for example, may actually represent 130 different diseases, and a disease is a process rather than a single state. Phenotypic analysis to define the type of ductal carcinoma can involve mammograms, ultrasound, positron emission tomography (PET)/computer tomography (CT) scans, and magnetic resonance imaging (MRI), in addition to tumor staging, DNA sequencing, SNP analysis, comparative genomic hybridization, loss-of-heterozygosity

analysis, gene expression and proteomic profiling. Liebman maps disease phenotypes as a function of genetics, lifestyle, and environment, and includes events like polio vaccination; he is working on Bayesian networks for the staging and diagnosis of breast disease.

Michael Heller (University of California, San Diego, USA) is looking forward to the \$1,000 genome sequence - the day when someone will be able to get his or her own personal genome data, paid for by health insurance as an ordinary preventive medical expense, on a DVD. Heller is a strong believer in personalized medicine. With development costs at a staggering \$800 million for a single new drug, we are “littered with failed drug corpses”. Reliable genotyping to divide patients into smaller groups that could benefit from a potential new drug (patient stratification) is essential. Heller is the founder of Nanogen, in San Diego, a company dedicated to the accurate and reliable use of microarrays for genotyping. Collaborations with workers at the University of Texas Medical School at Dallas have shown that some sequences that are difficult to resolve by traditional methods are accurately determined by Nanogen’s experimental microarray platform. Heller thinks that the \$1,000 genome may depend on the development of new nanotechnologies, such as nanophotonic switching devices using quantum dots conjugated to DNA probes. The \$1,000 genome will need minimal handling of the material, should avoid labeling, amplification and orientation procedures, and should ideally take only hours to days to run. The NIH is currently allocating funds for technical developments in this field.

The application of genomics, bioinformatics and systems biology in drug discovery and medicine holds tremendous promise. Vast stores of microarray data and whole-genome scans feed sophisticated digital models of human health and disease. It is clear that we are on the cusp of a revolution in healthcare, but we have yet to realize significant changes in the clinic. We can anticipate more exciting developments when Beyond Genome returns to San Francisco in 2006.