DDT Vol. 7, No. 3 February 2002

Continuing the 'ome' revolution

Krishnarao Appasani, PerkinElmer™ Life Sciences, Research and Development, 549 Albany Street, Boston, MA, USA, tel: +1 671 350 9034, fax: +1 617 350 9658, e-mail: krishnarao.aapasani@perkinelmer.com

The XIIIth *Genome Sequencing and Analysis* Conference hosted by The Institute for Genomic Research (25–28 October 2001, San Diego, CA, USA) covered the sequence analysis, comparative genomics and molecular evolution of several organisms including microbes, *Drosophila*, *Arabidopsis*, the rat, the mouse and humans. Chair of the meeting, Claire Fraser (TIGR, Rockville, MD, USA), kicked off the conference by welcoming the crowd of ~2200 delegates and said she was optimistic that the sequence information obtained from nearly 100 species would revolutionize our understanding of basic biology, medicine and agriculture.

More than 150 posters covering sequencing, nucleic acid preparation, DNA and protein array manufacturing, SNP genotyping, lab-on-a-chip microfluidics, bioinformatics and probe-labeling technologies were presented, in addition to a lively exhibition comprising 15 bioinformatics companies and ~100 reagent, instrument and service-oriented companies.

Microarrays

In the first plenary session, chaired by Mark Adams (Celera Genomics, Rockville, MD, USA), Stephen Friend (Rosetta Inpharmatics, Kirkland, WA, USA) described Rosetta's molecular profiling platforms, which include Sensors™, References™, and Rosetta Resolver Decoder™. Rosetta used FlexJet™ microarray technology to analyze the expression profiles of several human disease tissues using arrays of 30,000–40,000 genes. Using a post-genomic expression profiling approach Rosetta generated and catalogued many genes or exons (70 prognostic markers) identified from breast cancer. In addition, Rosetta also developed models to study alternative mRNA splicing. Friend also described The Pathway Discovery Program (TPDP), which has classified several expressed sequence tags (ESTs) within their functional cellular pathways.

Daniel Shoemaker (Rosetta) presented the company's gene discovery program using inkjet 60mer oligonucleotide tiling arrays. He demonstrated kinetic and hybridization analysis of 20–60mer oligonucleotide arrays by detailing Rosetta's 10 bp length edge-detection method (monitoring of junction sequences between introns and exons). Rosetta has also verified transcripts by using junction arrays, where various lengths of oligonucleotides with a 10bp difference were studied for kinetic and hybridization properties, and classified them into various functional groups.

The use of gene chips from Affymetrix (Santa Clara, CA, USA) to study bacterial adhesion, *Escherichia coli* chemotaxis and yeast haploinsufficiency was described by Adam Arkin (University of California, Berkeley, CA, USA). Arkin also commented that comparative analysis, knowledge-based modeling and cell circuitry are important for the study of gene regulatory motifs and their control in cells.

Continuing the array theme, Krishnarao Appasani (PerkinElmer Life Sciences) presented a poster that demonstrated the application of MICROMAX™ cDNA arrays for the study of gene expression in HL60 and Jurkat cell lines and the subsequent characterization of a T-cell acute lymphoblastic leukemia specific gene.

Comparative genomics

In the second plenary session, Ian Bancroft (John Innes Institute, Norwich, UK) presented data on the highly conserved (85%) evolution of polyploid genomes, such as *Brassica* species. His group sequenced the 115 Mb genome of *Arabidopsis* and demonstrated that 60% of the genome is duplicate sequence. Bancroft suggests that this type of segmental duplication evolved from a common ancestor. His group also compared the genomes of the distantly related tomato, rice and *Arabidopsis* plants and identified five segments of homologous genes that are highly conserved.

Claire Fraser described the 57 microbial genome sequencing efforts at TIGR and identified ~134,000 annotated genes, which are deposited in the TIGR database. Fraser's group has used microarrays to reveal the differences between genome content among *Streptococcus* strains and has also identified gene duplication phenomena in *Chlorobium*, *Streptomyces* and *Vibrio* species. Most microbial genomes are evolved either by vertical or horizontal gene transfer, and horizontal transfer mechanisms were identified in *Nisseria*, *Vibrio* and several *Archaea* species. Fraser also described how genome reduction and degradation plays a role in shaping species diversity, especially in *Mycoplasma* and *Borrelia* species. Additionally, her group has focused on environmental genomics (identifying microbes that grow under abnormal conditions) and has identified a proton-pump protein called protorhodopsin. Their comparative genomics approach has revealed that the phylogeny of a single gene does not predict the overall species relationship to biology and that multiple mechanisms are involved.

Susan Celniker (University of California, Berkeley) described how fruitfly embryology has been integrated with bioinformatics as part of the *Drosophila* genome project and has identified and clustered several ESTs and deposited the data in the *Drosophila* Gene Collection. The project has three such data releases available for several fruitfly clones and is now using this sequence information

 1359-6446/02/$ – see front matter PII: S1359-6446(02)02156-0

to identify transcription factors and transcriptional enhancers.

A comparison of human and mouse microsatellite distribution data was presented by Inyoul Lee [Institute for Systems Biology (ISB), Seattle, WA, USA], who concluded that the mouse has simpler, longer dinucleotide and tetranucleotide repeats. In addition, Lee spoke of the orthology relationships between T-cell receptor variable(s), diversity and joining gene-segments.

Genome sequencing

The Chair of this session, Richard Gibbs (Baylor College of Medicine, Houston, TX, USA) and his colleague, Michael Metzker, presented an overview of the rat genome sequencing consortium efforts. The consortium has integrated bacterial artificial chromosome (BAC), shotgun whole genome and clone restriction-enzyme fingerprinting approaches and has sequenced the rat genome in only eight months. Using the results obtained from these strategies they are now attempting to understand diseases such as renal failure, hypertension, diabetes and arthritis.

Eugene Myers and Richard Mural (Celera Genomics) discussed the mouse genome assembly, including the Genescan program and Unitags – DNA sequences that are common in man and mouse. Mural's talk compared the mouse and human sequences and discussed the use of syntenic anchors for the identification of orthologous pairs of proteins. Picking up the theme of the human genome, Evan Eichler (Case Western Reserve University, Cleveland, OH, USA) discussed gene duplication and the dynamic changes in the mutation rate of the human genome.

An 'inverse genomics' technology platform was presented by Flossie Wong-Staal (Immusol, San Diego, CA, USA) that uses a combinatorial library of >10 million unique gene-inactivating hairpins and ribozymes to isolate drug discovery targets.

Epidemiology and polymorphisms

The use of biostatistics tools (such as synonymous and non-synonymous divergence patterns and poisson distribution) for the study of human polymorphisms was described by Andrew Clark (Celera Genomics), who showed that human and mouse genomes are 85–87% homologous. Clark illustrated various models that can be used to study human and mouse divergence.

Ivo Glynnegut (Centre National de Genotypage, Evry, France) presented on genomic epidemiology and the integration of blood-based epidemiology data with integrative physiology. His group developed high-resolution proteomic and metabolomic phenotypes based on genotypes from 70,000 patients. The group also used mass spectrometry (MS) to identify single nucleotide polymorphisms (SNPs), and developed Genalys™ – an all-in-one software module for SNP and haplotype discovery.

Proteomics

In this session, which focused on proteomics and its emerging applications in biomedical science, Denis Hochstrasser (Geneva University and GeneProt™, Geneva, Switzerland) presented a large-scale industrial approach to protein sequencing that has been used to identify protein markers for several diseases. Using HPLC and matrix-assisted laser desorption ionization (MALDI)-MS analysis of blood samples, the group identified a spectrum of proteins that could be used as diagnostic markers. Hochstrasser also discussed peptide fingerprinting and protein molecular scanning methodologies and concluded that the amount of clinical samples and the availability of bioinformatics expertise are the two bottlenecks in proteomics research.

Scott Patterson (Celera Genomics) discussed Celera's proteomics research using two-dimensional (2D) gel electrophoresis and time-of-flight (Tof)-based MS. Celera's platform technology includes a combination of both protein chemistry and cell biology integrated with MS. Celera is focusing on the identification of surrogate and bio-markers, small-molecule drugs and therapeutic antibody targets.

In the field of malaria, John Yates (The Scripps Research Institute, La Jolla, CA, USA) has used peptide mapping, 2D gels, MS–MS, and SciQuest™ software analysis to identify several hundred proteins from different stages of the *Plasmodium falciparum* life cycle. His group compared the malaria proteome across different infected cell types and identified a class of interesting proteins that included erythrocyte surface antigen and the cytoadherence molecules (spectrin-α, β-ankyrin and band 3). The goal of Yates' team is to use this technology to identify vaccine candidates for malaria infection.

Reudi Aebersold (ISB) presented a 'reduction genetics' approach to quantitative proteome analysis by highlighting ISB's proprietary technology platform that uses a new class of chemical reagents called Isotope-Coded Affinity Tags™. Using these reagents in combination with a tandem MS-based analytical process, ISB has identified and quantitatively measured proteins expressed in yeast cells.

Rue Mei (Affymetrix) presented data on high-density oligonucleotide Chip I and II arrays, which contain ~100 yeast and human genes, respectively. Mei presented correlation data between the predicted and observed melting temperatures for yeast and human sequences. The best probes were selected based on GC content, selectivity and specificity, and the probe quality metrics of linear and sigmoid models were compared; these give different oligo hybridization patterns when plotted as intensity versus size.

Mathias Uhlen (Royal Institute of Technology, Stockholm, Sweden) detailed a high-throughput proteomics approach for the analysis of human chromosome 21 and compared it with the gene expression data.

Bioinformatics

Chang Liu (University of Minnesota, Minneapolis, MN, USA) demonstrated the clustering of 2000 genes differentially expressed in yeast by the K-mean algorithm and cosine values (similarity metrics) using data mining tools. Over-represented hexamers were identified

using hypergeometric probability and 168 ribosomal genes were grouped into different clusters.

Homology based algorithms for the alcohol dehydrogenase gene structure of *Drosophila melanogaster* and prediction of gene variants were discussed by Michael Sievers (Paracel, Pasadena, CA, USA) and Tony Frudakis (DNAPrint Genomics, Sarasota, FL, USA). Using re-sequencing procedures and proprietary software tools Frudakis provided the data for several SNPs and haplotype maps for various genes involved in xenobiotic disposition.

Final thoughts

To conclude, the meeting provided an overview of the current status of world-wide genomic sequencing efforts and the wide range of multidisciplinary technologies that is available to exploit this data. Large-scale data analysis and target validation continue to be major bottlenecks in drug discovery, and so the take-home message from this conference has to be that the integration of genomics, proteomics and informatics is the key to fulfilling the promise of the post-genome era.

These opinions are exclusively of the author and do not reflect those of PerkinElmer.

Drug Targeting Technology: Physical – Chemical – Biological Methods

Edited by Hans Schreier, Marcel Dekker, 2001, Price US$150.00, 294 pages in hardback, ISBN 0-8247-0580-7

Rapid advances in genomics, combinatorial chemistry and HTS should result in vast numbers of new therapeutics in coming years. These new drugs hold the promise of being more effective for the treatment of disease. With increased biological activity, however, the potential for side effects can also increase. To avoid undesirable side effects from these highly active drugs, and to keep the dose and expense low, it would be advantageous to target them only to the diseased organ or tissue and to avoid the remaining periphery. However, at this point there are few, if any, drugs that are 'magic bullets', which unfailingly hit only the target site of action. Drug targeting itself can be looked at in several different ways, ranging from simply protecting the drug as it passes through the gastrointestinal tract, to direct targeting to a specific organ or tissue. Equally as challenging are strategies that seek to avoid adverse effects of the drug on tissues or organs that are not the intended site of action.

This book is a good introduction to the various methods of drug targeting. According to the preface, the intention of the Editor was not to cover comprehensively the topic of drug targeting, which would be difficult to accomplish in such a short volume. Rather, this book is intended to be a 'reader for interested scientists, experts and students who are open for lateral views beyond the boundaries of their own field of interest'. The book gathers together monographs that cover many of the most widely used drug targeting techniques. The old-fashioned concept of time-release capsules through the use of enteric coatings is included in the same volume as the more modern concept of modified viruses to deliver gene therapy. It is also the Editor's hope that the book will cause pharmaceutical practitioners to consider crossing the boundaries between these various approaches and to come up with better therapies. It does achieve its purpose in this regard; this rather short volume could easily serve as an introduction to the topic for, say, a relatively new medicinal chemist. For those more familiar with the concepts, multiple approaches are presented here in a fairly homogeneous format so boundary crossings do come to mind.

There are over 1000 references included, so if the reader has an additional interest in a particular topic this is an excellent jumpstart for research. Moreover, the majority of the references are less than 10 years old. The field is moving quickly, so timely references are particularly useful. Despite all the information, the writing is such that it is relatively easy to read; I managed to read the entire book in two weeks of short evenings. The chapters are homogeneous enough to be taken together as a whole, yet complete in themselves so that each can serve as an introduction to the topic at hand. The index is not particularly extensive, but the major topics are easy to find.

In summary, this book would make a good addition to the library of any practicing scientist in the pharmaceutical or biotechnology industry, in particular those that are directly involved in drug delivery or drug targeting. For those of us that are already involved in the field, it serves to make us think about the traditional boundaries between various aspects and the possible ways of crossing them. For those with a peripheral interest, I believe this book could serve as an excellent introduction.

John Weidner
Scientist, Parallel Synthesis
Medicinal Chemistry
Emisphere Technologies
765 Old Saw Mill River Rd
Tarrytown, NY 10591, USA
tel: +1 914 785 4792
fax: +1 914 593 8250
e-mail: Jweidner@emisphere.com

  PII: S1359-6446(02)02127-4