

Meeting report

Translating genome sequences into biological understanding

Vishwanath R Iyer

Address: Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX 78712, USA. E-mail: vishy@mail.utexas.edu

Published: 23 May 2003

Genome Biology 2003, 4:324

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2003/4/6/324>

© 2003 BioMed Central Ltd

A report on the Genomics, Proteomics and Bioinformatics Thematic Meeting during the 2003 American Society for Biochemistry and Molecular Biology (ASBMB) Annual Meeting, San Diego, USA, 11-15 April 2003.

April 2003 marked the fiftieth anniversary of the discovery of the double-helical structure of DNA and also the announcement of the finished human genome sequence. The period between the first outline of the DNA double-helix and the grand revelation of the human genome has yielded the complete genome sequences of many other organisms. It was apt that the theme of the 2003 ASBMB Annual Meeting was 'Translating the Genome', referring to the challenge of converting sequence information into biological insights; talks at the meeting highlighted various approaches being developed to meet this challenge.

Stephen Young (University of California, San Francisco, USA) described gene-trapping vectors for large-scale insertional mutagenesis and expression studies of mouse genes. The vectors contain the β -galactosidase gene (plus a resistance marker) with an upstream splice-acceptor site. A vector is electroporated into mouse embryonic stem (ES) cells, where it integrates randomly into the genome. Insertions within a coding sequence that allow the reporter to be functionally spliced are indicated by the growth of drug-resistant clones expressing β -galactosidase. Insertion sites can be identified by PCR sequencing of DNA flanking the insertion. The BayGenomics consortium [<http://baygenomics.ucsf.edu/overview/people.html>], headed by Young, has generated more than 7,000 cell lines representing insertions in 2,200 distinct genes, and they anticipate generating thousands more every year. The value of the insertional cell lines arises from the ability to generate from them mouse lines, which can be used in two main lines of investigation. First, the expression pattern of β -galactosidase in the developing embryo is a surrogate marker for the expression pattern of the gene targeted by the insertion.

Cell-type- or tissue-specific expression at specific developmental stages, which provides clues about gene function, can therefore be analyzed. Second, the ES cell lines can be used to generate mice with null alleles of the targeted genes. More than 150 knockout mice have been generated by the consortium and these too will be available for distribution.

Brian Seed (Harvard Medical School, Boston, USA) described an automated approach for identifying mammalian cDNAs that can activate specific signaling pathways or transcription factors. A reporter system was constructed, consisting of green fluorescent protein (GFP) controlled by a signaling-responsive promoter in an easily transfectable mammalian cell. In parallel, tens of thousands of clones from a library of cDNAs in expression vectors were individually grown in 384-well plates and pooled into several groups in two orthogonal groupings. DNA from each pool was used to transfect reporter cells. Reporter gene activity indicated the presence of a cDNA that could activate the particular signaling pathway to which the reporter gene was engineered to respond. The activating clone could be identified from the combination of pools in which it was found. One assay using an NF κ B-responsive GFP reporter revealed that G-protein-coupled receptors can activate NF κ B. Modifications of the reporter system will allow the identification of clones that activate other mammalian transcription factors or signaling pathways.

Genes with similar expression profiles may be co-regulated because they have similar functions, but it is also possible that co-expressed genes lack an underlying functional connection. This raises an important question: how can one distinguish sets of genes that are co-expressed because of functional relatedness rather than simply by chance? Stuart Kim and colleagues (Stanford University, USA) reasoned that if genes are co-regulated because they are functionally related, this relationship is likely to have been selected for during evolution and one is therefore likely to see co-regulation of those genes in different species. Kim and colleagues examined whether a set of genes whose expression was

highly correlated in, for example, *Drosophila* were also co-expressed in human. Regulation of genes encoding components of cellular complexes such as the ribosome and the protein translation machinery showed the highest degree of conservation, but others involved in signaling, transcription, ubiquitin-mediated protein degradation and the cell-cycle, as well as a number of uncharacterized genes, also showed conserved co-regulation. Kim's analysis suggests that two-thirds of the genes that appear to be co-regulated in a single species are in fact functionally related, given the conservation of their co-regulation.

One of the most exciting technologies for studying gene function to be developed in recent years is the use of RNA interference (RNAi) to abrogate function. Michael Boutros and colleagues from Norbert Perrimon's lab (Harvard Medical School) have extended this approach to the whole genome in *Drosophila*. Boutros described automated screens to test the effects on cell viability, cytokinesis, and signaling of inhibiting essentially every gene. RNAi for each *Drosophila* gene was achieved by *in vitro* transcription and then transfection of cultured *Drosophila* cells in 384-well plates. Phenotypes were determined by automated imaging of cell morphology or by quantitatively measuring the expression of a luciferase reporter gene. Using this approach, genes could be clustered on the basis of similar quantitative RNAi phenotypes. This strategy is also amenable to double-mutant analysis; for instance, Boutros reported that RNAi against the Ras GTPase-activating protein (GAP) inhibits signaling in response to lipopolysaccharide (LPS) in innate immunity, but inhibition of Ras1 is epistatic to the GAP effect.

Protein-interaction networks offer a means of deciphering how proteins function in concert to bring about biological functions. One concern with many high-throughput experimental approaches for identifying protein interactions is the high rate of false positives. Frederick Roth (Harvard Medical School) presented a probabilistic approach for validating protein interactions, which relies on the observation that true interactions tend to form small-world networks and are 'cliquish'. This means that the true interacting partners of a given protein are more likely also to interact with each other, compared to a random network. Roth's model was developed by first assuming that all proteins interact with each other with a certain weight (or probability), then assigning weights on the basis of experimental evidence. It allowed him to calculate the probability that a protein is part of a previously partially defined core complex or that it interacts with another protein.

An alternative method for identifying gene interactions is a synthetic lethal screen. If cells lacking the function of either of two genes are viable, but cells deficient in both genes are inviable, a functional relationship between the two genes is likely. Geneticists have used this approach for decades, but

Charles Boone and colleagues (University of Toronto, Canada) have brought it into the post-genome era by automating the testing of all pair-wise combinations of all the viable single-gene deletions in yeast (approximately 5,000 genes). This approach allows genes to be ordered into functional pathways: for instance, if genes X, Y and Z are all 'synthetic lethal' with gene A, it suggests that the former three genes function in the same pathway. Boone and colleagues have performed dozens of screens to identify thousands of interactions, and extended the use of the yeast deletion collection to perform chemical-genetic interaction screens. In these the idea is that deletion mutants sensitive to a drug should be 'synthetic lethal' with the targets of the drug, allowing the clustering of drug-response profiles with synthetic genetic profiles. Boone and colleagues could thus infer the action of the drug papuamide on Golgi transport through its effect on the genes *RIC1* and *YPT6*.

Marc Vidal (Dana Farber Cancer Institute, Boston, USA) provided an overview of integrated approaches to studying the proteome of *Caenorhabditis elegans*. His group began by cloning nearly 12,000 different open reading frames into Gateway vectors (from Invitrogen) by PCR amplification using gene-specific primers from representative cDNA libraries. This extensive resource, available through distributors, allows more representative genome-wide protein-interaction assays using techniques such as the yeast two-hybrid system or protein microarrays, and also aids in genome annotation. Vidal emphasized the importance of integrating different predictions of gene interactions to improve correlations; for instance, interaction networks derived from two-hybrid analysis could be compared to the linkages derived from gene-expression profiling studies and to 'phenome' clusters, which are generated based on the large-scale RNAi phenotypes that are being simultaneously collected.

The bold and creative approaches presented at the meeting engender optimism that the challenge of translating genome sequence information into comprehensive biological insights is being adequately tackled. Biologists of the future, celebrating the one-hundredth anniversary of DNA structure and the fiftieth anniversary of the Human Genome Project, will undoubtedly look back at the results of this and similar work as the source of much of their deep understanding of biology.