

Meeting report

How will knowing genome sequences affect the life of the average cell biologist?

Janet Rossant

Address: Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, M5G 1X5, Canada.

Published: 27 April 2000

Genome **Biology** 2000, **1**(1):reports401.1-401.2

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2000/1/1/reports/401>

© Genome**Biology**.com (Print ISSN 1465-6906; Online ISSN 1465-6914)

A report on work presented at the 39th Annual Meeting of the American Society for Cell Biology, Washington DC, December 11–15, 1999

The impact of 'the genome project' on all aspects of biomedical research in the next century cannot be underestimated. We will know the complete sequence of the human genome and of the genomes of an ever-increasing number of model organisms, providing masses of new raw genetic data to sustain the study of any particular cellular process or pathway. But perhaps more importantly, the tools and mindset of the genomicist seem destined to alter the way we all think about and practice our science. Once we know the complete set of genes that exist in an organism, it will no longer be acceptable to focus on one gene at a time, to study its effects in terms of a read-out consisting of a few other related genes and cellular events, and then to conclude that we have understood its function. Conversely, if we want to understand a cellular process, we will need somehow to integrate all the molecular interactions that occur, rather than being satisfied with cataloguing the few major pathway components identified by genetic or biochemical screens. Genomics, and its newer siblings functional genomics, proteomics and bioinformatics, are teaching us all to think in a high-throughput, genome-wide manner and are providing the technologies to allow us to translate these thoughts into action.

The pace of technological progress in this area can be intimidating, to the established investigator and the new graduate student alike. No wonder, then, that there was a packed house at the opening symposium — entitled 'The Impact of Genome-wide Studies on Cell and Developmental Biology' — of the 39th American Society for Cell Biology meeting. David Botstein (Stanford University Medical School) began by emphasizing the importance of model organisms and comparative genomics in assigning meaning to the sequence of

As, Ts, Cs and Gs that will make up the final product of the Human Genome Project. The complete sequence of the genome of the yeast *Saccharomyces cerevisiae* has been available since 1996, and that of the worm *Caenorhabditis elegans* since 1998. As detailed by Gerry Rubin (University of California Berkeley), 'version one' of the 185 Mb *Drosophila* genome is also complete, as a result of a collaboration between academic centers in the United States and Europe and the company Celera Genomics. A draft sequence of the human genome is expected to follow shortly, with the complete sequence due to be finished by 2003 at the latest.

With the sequence information in the databases to date, it is already possible to query the genome in interesting ways. Botstein pointed out that comparison of yeast and worm genomes has identified a set of conserved genes encoding core biological functions, such as metabolism, DNA replication and protein trafficking. These genes make up 40% of the yeast and 20% of the worm genome. Alongside this core set of proteins — which are expected to be common to single-cell and multicellular eukaryotes — worms, flies and humans have developed a complex set of genes involved in signal transduction and gene regulation, in order to accomplish their elaborate developmental programs. These pathways show less absolute conservation of gene sequence, but general classes of proteins and protein domains can be recognized across evolution.

The extent of cross-species conservation across the whole range of gene function is, in fact, remarkable. Of the human genes known to date, 74% have related sequences in the worm. Cori Bargmann (University of California, San Francisco) illustrated how inter- and intra-species sequence comparisons of gene families can confirm conserved functions and identify new ones, by scanning the *C. elegans* genome for genes encoding ion channels and G-protein-coupled receptors. There is no voltage-activated sodium channel identifiable in the *C. elegans* genome, consistent with the results of electrophysiological studies, but voltage-regulated

comment

reviews

reports

deposited research

refereed research

interactions

information

calcium and potassium channels are found, some similar to vertebrate counterparts and some divergent. Clearly, some aspects of cell excitability are common and fundamental and some are subject to evolutionary pressure related to the adaptation of the organism to its particular environment. Despite the extensive prior mutational analysis of the worm, the genome sequence has thrown up many genes that were not identified by mutation. For example, Bargmann's lab cloned one odorant receptor gene, *Od-10*, through mutational analysis. It encodes a novel G-protein-coupled receptor. Analysis of the genome sequence revealed about 1,000 'orphan' G-protein-coupled receptors, many of which appear to be related to *Od-10* and which probably represent more odorant receptors. Overnight, then, a lab working on one gene finds itself working on hundreds.

The need for systematic, genome-wide methods for assigning gene function to this deluge of sequence information was a common theme of the symposium. Annotation by sequence similarity is the usual first step, but it is difficult and can be misleading. Genes of similar sequence may have acquired new functions during evolution. For example, mutational analysis of a *Drosophila* gene whose sequence suggested that it could be a chitinase enzyme revealed instead that it was a novel growth factor (Rubin). Developing a common language for gene annotation across different databases is one of the steps needed to improve the situation. This is the goal of the Gene Ontology (GO) project, a collaborative effort between workers on the yeast, fly and mouse genome databases. But even a common language and more sophisticated sequence analysis tools are not likely yet to replace the need for the seasoned eye of the expert biologist when interpreting genome sequence. The publication of the *Drosophila* genome sequence depends on the completion of a major effort to annotate all the genes identified in the sequence to date, by a group of biologists expert in different gene families, working alongside bioinformatics experts.

Much more information about gene function can, of course, be gleaned from knowing expression patterns and the effects of loss or gain of function. Genome-wide initiatives in assessing expression and function are underway for all model organisms. For example, an assay of the function of all *C. elegans* genes, using RNA interference, for their effects on early cell division processes is under way at the European Molecular Biology Laboratory (poster presented by Pierre Gönczy). The Berkeley *Drosophila* Genome Project is surveying the expression of all *Drosophila* genes by whole-mount *in situ* hybridization in embryos and attempting a catalog of gene mutations by accumulating insertions of P elements or Gal4 activation domains into as many different sites in the genome as possible (Rubin). Major efforts like

this in a few centers can provide enormous resources for the average cell or developmental biologist who wishes to pursue more detailed analysis of an individual gene or process of interest. Rubin noted that there are an estimated 15,000 *Drosophila* genes and 5,000 active *Drosophila* researchers, or three genes per biologist for future study. The general philosophy of the various academic genome initiatives - to share all information freely and rapidly via the internet - must surely apply to these next phases of large-scale functional genomics as well. A familiarity with database resources and the tools for database mining will therefore be essential in allowing biologists to gain the most from all the material made available by large-scale projects.

There are growing numbers of genome research applications, however, which the average biology lab will want to use directly for its own studies. None has generated more excitement than DNA microarrays or 'chips'. The cDNAs from all of the 6,000 or so yeast genes, a large proportion of the *Drosophila* and worm genes, and an increasing number of human and mouse genes, have been assembled onto glass microarrays in different labs world-wide. Hybridization analysis, combined with sophisticated clustering algorithms to analyse the results, allows analysis of gene expression changes in normal or abnormal states in unprecedented detail. Botstein has been a prime mover in validating this technology, along with his Stanford colleague, Pat Brown. Botstein described how yeast microarrays allow the identification of clusters of co-regulated genes whose expression changes together under a variety of different conditions. Because all the yeast genes are on a chip, complete sets of co-regulated genes can now be identified and studied. In larger organisms, microarrays are currently still sampling the genome, but a sample size of several thousand genes is sufficiently large to reveal unexpected patterns of co-regulated genes. Furthermore, microarrays are already finding utility in classifying tumor samples. 'Signature' patterns of gene expression can be discerned when many different tumor samples are compared, despite the potential heterogeneity imposed during sample collection (Botstein). These patterns can be used to define subclasses of tumors that are not apparent by more standard means. The possible applications of this technology to all aspects of biology seem bound to grow exponentially as the technology becomes more widely disseminated.

We are entering an exciting era, when genomes and genomics will provide the basis for a new form of truly integrative biological research. We cell and developmental biologists, who already think in integrative ways as a result of the questions our research addresses, need to be in the vanguard of this new biology.