Meeting report
The integrated world of functional genomics
Itai Yanai

Address: Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, 76100, Israel. E-mail: itai.yanai@weizmann.ac.il

---

A report on the EMBO meeting 'Functional Genomics; The Future of Biology', Heidelberg, Germany, 13-16 October 2002.

Some writers believe that a dictionary is a graveyard for words, because as soon as a word is defined, it is limited to a particular meaning and consequently it is 'effectively dead'. It may thus be encouraging that the field of functional genomics is still unable to define itself. Is functional genomics limited to high-throughput experiments of entire genomic or proteomic complements, or can it focus on systems? How do genome comparisons and research on diagnostics and cures for diseases fit into this field? Does functional genomics really differ from molecular biology? The conference 'Functional Genomics: The Future of Biology', which took place at the European Molecular Biology Laboratory (EMBL) in rainy Heidelberg, gave representation to all sides of this nascent field, bringing together the top researchers from comparative genomics, transcriptome profiling, proteomics, structural biology, systems biology, and medical research. Speakers came mainly from academic institutes, but industry was also represented by a few talks. The talks spanned research in most of the model organisms: *Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Anopheles gambiae*, mouse and the humble human. Overall, the conference was an exciting and impressive view of the state of functional genomics today (Figure 1).

Although the range of subjects under the umbrella of functional genomics is wide, some of the properties of 'global research' are becoming clear, showing that the field is maturing. The unifying feature of functional genomics is its high throughput, or as Ewan Birney (EMBL European Bioinformatics Institute, Hinxton, UK) put it, functional genomics is "molecular biology in a 96-well format". More importantly, though, the familiar theme of integration ran through most of the talks. Many of the talks described research on the integration of datasets from high-throughput experiments. The

'future of biology', judging from this conference, seems to be the integration of data from various high-throughput experiments to obtain global views of biological processes.

Rick Young (Whitehead Institute for Biomedical Research, Cambridge, USA) presented his work on the transcriptional regulatory circuitry of *S. cerevisiae*. His group was able to derive a strain expressing a tagged version of each of 106 transcriptional regulators in yeast. Using a procedure involving chromatin immunoprecipitation and hybridization to microarrays, the strains were then used to retrieve all of the promoter sequences to which these regulators bind from the whole genome. This information was then used to examine the circuitry of the regulatory programs, leading to the detection of several 'network motifs', such as autoregulation, multi-component loops and feedforward loops. When the promoter-binding data were combined with expression data, a detailed picture of the yeast cell-cycle transcription regulatory network emerged. Young noted that it was the use of multiple data sources that was key to the insights into the cell cycle gained from the work. Esti Yeger-Lotem (Hebrew University, Jerusalem, Israel) described similar work on the discovery of complex regulatory circuits in *S. cerevisiae* by integration of data on the binding of proteins to DNA with protein-protein interaction data.

Marc Vidal (Dana-Farber Cancer Institute, Boston, USA) also spoke of developing hypotheses by integrating functional maps derived from high-throughput datasets, such as from expression profiling, protein interaction mapping, protein localization, biochemical genomics, structural genomics, and knock-out experiments. Vidal placed the notion of integration in the context of *C. elegans*, the model organism with which he mainly works, and gave examples of processes studied in this way, including vulval development.

Peer Bork (EMBL, Heidelberg, Germany) discussed the integration of three methods for the discovery of functional associations between genes by comparative genomics: two genes are likely to have related functions if they are found together
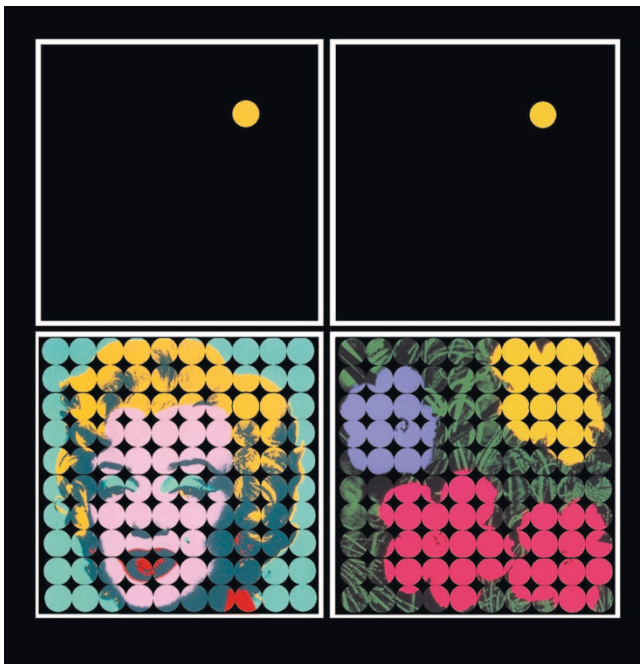
Figure 1
An illustration of the power of functional genomics, as used in the publicity for the meeting. The top two 'experiments' reveal the same results, but from the global views below it is clear that the pictures are different. By analogy, classical molecular biology gives information on only a single gene while functional genomics gives information on all of the genes in the organism. Reproduced with permission from Matthias W. Hentze and Petra Riedinger.

in the genome in several species (gene neighborhood), if they are both present in one set of species but absent from another (phylogenetic pattern), or if they are fused in one or more species (domain fusion). Taking these methods together has allowed Bork and colleagues to infer the function of an uncharacterized protein from its associations with other proteins. Bork also discussed another method, called anti-correlation, in which two entire protein families are linked if their presence and absence across genomes are complementary, in other words if one is always absent where the other is present and vice versa.

One of the most exciting aspects of functional genomics is the breathtaking view of the organism that it makes possible. Stuart Kim (Stanford University Medical School, USA) presented a new way to view clustering of genes according to their expression in many experiments. The display of the clustering is visually impressive: similarity of expression profile corresponds to proximity in a two dimensional plane and the gene density in the plane is shown in the third dimension. Application of Kim's algorithm to data on expression of 98% of *C. elegans* genes in over 500 experiments revealed a rugged mountain range, and an analysis of the genes that composed each mountain showed that they were generally genes known to be involved in the same

biological process, such as the response to heat shock, the development of the gut or of muscles, and so on.

The view from proteomics was equally impressive, with the arrival of protein chips. Michael Snyder (Yale University, New Haven, USA) presented what he called the 'yeast proteome version 1.0': a protein chip of 5,800 *S. cerevisiae* proteins (each fused to glutathione *S*-transferase-polyhistidine tags), corresponding to about 93% of the total genes. He described the various experiments his group has performed using these chips, including screening for protein-protein interactions, protein-lipid interactions, interactions between small molecules and proteins, and for post-translational modifications.

The general structure of a global view of the genome or proteome can itself reveal a major surprise, as shown by Giulio Superti-Furga (Cellzome AG, Heidelberg, Germany). Superti-Furga presented the results of his group's work on determining protein complexes using tandem-affinity purification and mass spectrometry. Most interestingly, his results led to the estimation that 85% of *S. cerevisiae* proteins form complexes that are connected in large networks by sharing of members between different complexes. As many as 40% of the proteins are a part of more than one complex. Frank Holstege (University Medical Center, Utrecht, The Netherlands) noted that, from his analysis, Superti-Furga's data contain the lowest number of false positives of the large-scale protein-protein interaction datasets currently available.

Holstege has used co-expression of proteins in microarray experiments to evaluate the quality of high-throughput protein-protein interaction results, which are known to include many false positives. As proteins that are known to interact do tend to be co-expressed, this seems to be a sensible screen. Holstege and his group identified several predicted interactions of products that show high co-expression and experimentally validated them.

He also showed that although all of the standard methods for normalization of microarray data assume that the RNA content of the cell remains constant between experiments, this is often not the case. Thus, a gene may be reported as expressed at a lower absolute level in tissue A than in tissue B, but if (on average) more RNA is present in B, it could appear after normalization as though the gene is expressed at higher levels in tissue A. Holstege presented a method for normalization using an external control that determines global mRNA changes between experiments.

Sven Bergmann (Weizmann Institute of Science, Rehovot, Israel) presented a robust 'signature' algorithm, developed by Naama Barkai's group, which can extract 'modules' of genes with similar expression profiles in all of the available expression data in *S. cerevisiae*. The algorithm has the unique feature of focusing only on the relevant conditions and is thus applicable to large datasets. Additionally, the

modules are context-dependent and thus genes can be present in more than one module; in this respect the algorithm differs from clustering algorithms that allow each gene to be in only one cluster.

Of the talks on structural biology, a particularly interesting one included a discussion of protein-function prediction from three-dimensional structure by Janet Thornton (European Bioinformatics Institute). Thornton presented methods for predicting whether a protein is multimeric, the location of its active site, and the type of ligand that is bound there, as well as its biochemical and biological functions. Joel Sussman (Weizmann Institute of Science) gave convincing evidence for the existence of unstructured proteins, with a case study of cholinesterase-like adhesion molecules (CLAMs). After being unable to crystallize the cytoplasmic domain of the CLAM gliotactin under many conditions, his group found that on a two-dimensional scale of mean charge and hydrophobicity, members of the family are actually predicted not to fold. Such unstructured proteins are presumed to assume a structure only when they interact with the appropriate 'partner in crime'. Also, Ana Rodrigues (University of York, UK) presented an informatics resource to the community for making an informed decision regarding which proteins to select and prioritize for structure determination, using raw genomic data; this is presently a crucial issue for structural genomics.

One of the most impressive parts of the field of functional genomics is systems biology, in which the complex behavior of a biological process is sought to be understood in terms of the simpler interactions between genes and their products. Naama Barkai (Weizmann Institute of Science) showed the power of mathematical modeling by presenting her group's work on searching genetic networks that can explain the patterning of the dorsal region of the *Drosophila* embryo. A model was developed that gave the greatest robustness to changes in gene dosage; this predicted that diffusion of the bone morphogenetic protein (BMP)-related ligand Decapentaplegic requires the BMP inhibitor Short gastrulation and the accessory protein Twisted gastrulation, which was verified experimentally.

Finally, Stan Leibler of the unusually named Laboratory for Living Matter (Rockefeller University, New York, USA) presented his work on natural systems in *Drosophila* and synthetic networks in *E. coli*, which have the capacity for some extremely complicated behavior. He showed that 'tinkering' with these circuits by combining five different types of promoters in front of the transcription factors LacI, TetR and Lambda CI in all 125 possible combinations can produce a wide range of behaviors.

If the topics described here are the work of a field without a definition, perhaps we should hope its definition remains elusive, so it can remain as productive in the years to come.

comment

reviews

reports

deposited research

refereed research

interactions

information