From Functional Genomics to Systems Biology

Julio R. Fernández Masso

Centro de Ingeniería Genética y Biotecnología. Ave 31 e/ 158 y 190,
PO Box 6 162, Cubanacan, Cuba
E-mail: julio.fernandez@cigb.edu.cu

REPORT

After the Transcriptome conferences held in Paris (2000), Seattle (2002), and Tokyo (2003), Prof. Charles Auffray from the Centre National de la Recherche Scientifique (France), Pro. Zhu Chen, from the Chinese Academy of Science and Chinese National Human Genome Research Centre (China) and Prof. Leroy Hood, from the Institute of Systems Biology (USA) launched a new series of conferences specially designed to foster the transition of life science research from Functional Genomics to Systems Biology. The first of this new series of conferences was held in Shanghai, China from November 5th to 9th 2005, under the topic of Integrative Systems Biology for Health-Predictive and Preventive Medicine. The conference covers aspects such as Expression Profiling, RNAi, Proteomics, Metabolomics, Systems Biology on Model Organism, Systems Biology in Medicine, Computation and Modeling and Integrative Annotation of the Genome. Conferences were presented by the academic and industrial sectors.

*U*nderstanding the genome

The epigenetic interface between the genotype and the phenotype links the genetic information and gene expression of an organism. The state of DNA methylation is the most widely studied epigenetic mechanism. A number of presentations expressed the importance of measuring the epigenetic changes as a measure of disease condition. Prototype techniques for the study of DNA methylation is a rapidly evolving field that should be closely followed since it will mature rapidly in the scale of information and in performance.

The relevance in tracking mutations in the genome through the use of SNP or comparative genome hybridatization to complement gene expression studies in diseases that involve dynamic changes in the genome such as cancer or mental retardation, and to understand biological processes in stem cell research such as pluripotence or cell differentiation, was pointed out.

*T*he complexity of the transcriptome

Full length cDNA sequence has been used to explore the complexity of the transcriptome. Lukas Wagner from NCBI presented the results obtained from sequence 13000 human full length cDNA. While the coverage of alternative splicing was not a goal for the project, the large number of EST and fully sequenced cDNA provides an overview into the challenge of surveying splice variants. Sumio Sugano from the University of Tokio determined the 5´end of 1,145,855 cDNA isolated from 134 full length enriched cDNA libraries. They identified a large population of cDNA that code for small proteins, a large number of alternative splices and alternative promoters. FANTOM3: The comprehensive mouse full length cDNA collection and seque nce database was presented by Mutsumi Kanamori from RIKEN. The comprehensive analysis of this cDNA collection gave a great deal of information on the organization of mammalian transcriptome. The project found an unexpected number of alternative spliced transcripts and a significantly large amount of non coding RNAs (ncRNA). The expression analysis of a substantial subset of predicted ncRNAs indicates that these RNAs are regulated in a tissue-specific manner and that many of them may be involved in an antisense control mechanism. Their study concludes that ncRNAs are a major component of the transcriptome. The following step they propose is the analysis of expression regulatory regions, expression profiles and protein-RNA interactions, to unravel the Genome Network. Kanamori's group established a large scale Systems to identify promoter regions and transcriptional starting sites. More than 10 million sites were analyzed to discover new promoters and genes. From the combination of mapping full length cDNAs and promoter regions they identified low and high activity chromosomal regions.

One of the most surprising results of the human genome sequence is the scarcity of genes and the vast tentions of meaningless genomic deserts, with ultraconserved elements. Gustavo Glusman from the Institute for Systems Biology proposes a novel approach to gene prediction. The new algorithms he developed are based on the detection of genomic signatures of transcription accumulated over evolutionary time. He predicted thousands of additional human genes including many that do not code for proteins and genes with long introns and lacking sequence conservation. He expects these new tools to add valuable information when combined with algorithms based on gene structure and sequence similarity as a step forward in achieving the sensitivity and specificity required to fully automate whole genome annotation.

*U*p and down regulated genes: from expression data to Biological meaning

One of the problems in microarray experiments is that of developing tools to turn the data into mechanistic understanding. The consensus at the meeting was that most of the microarray papers published in foremost journals still present a list of up and down regulated genes, the validation of certain genes by Western blot or qPCR, and the formulation of new hypotheses on the biological significance of some of these findings, but lack a significant contribution to new biological understanding. The challenge in microarray studies has moved from the generation of high quality data sets to their analysis and interpretation.

Xuegong Zhang from Tsinghua University presents a subject-oriented literature mining tool developed in his lab for combining the information found in the literature with microarray gene expression data to construct reliable and manageable gene networks with specific biological processes. The comparison and possible advantages over other available tools was discussed.

Several talks refered to the importance of the combination of microarray studies with other technologies as CHIP to chip to increase the confidence of gene expression studies and to allow the definition of gene and protein regulatory networks. Technical advances in the field of microarray design for the genome, CHIP to chip and gene expression studies were presented during the meeting by Affymetrix and Agilent Technologies. Sequenom presented the data obtained from several papers using automated mass spectrometry for gene expression and promoter methylation studies.

The Proteome and the Metabolome

Few presentations at the meeting cover the use of proteomic or metabolomic tools. Nevertheless, it was clear that the understanding of the proteome and its integration with transcriptome, the phosphorylation profile and a network of protein-protein interactions clearly provide a strong insight on mechanisms of biological Systemss. Mark McDonald from Waters described a new label free method for quantitative protein profiling that is based on the LC-MS methodology that makes it possible to determine relative changes in the amount of peptides in complex digest mixtures. The combination of chromatographic separation and mass accuracy enables the identification of thousands of ions. The SCAPE method based on the selective capture of peptides using chromatographic separation was presented by Gabriel Padrón from CIGB. This technology simplifies complex mixtures of proteins for their identification in proteome analysis.

The expression of open reading frames is of special value to facilitate functional studies at the proteome level and to establish the biochemical and biophysical properties of individual proteins and their molecular interactions. Shuwei Yang from FulenGen showed their results on high throughput expression. He claims the expression of 16000 unique full length ORF of human genes into 15 types of expression vectors with eight different peptide tags in E.coli are available for protein purification and functional studies.

Functional Genomics: The impact of RNAi technology

Numerous studies demonstrated the power of RNAi as a tool for drug discovery and validation. The challenge now is whether this technology can become a new therapeutic agent. Patrick Lu from Intradign presents several developments to increase RNAi extracellular stability and targeting. Their group developed ligand directed nanoparticles to efficiently deliver siRNA oligos into the diseased tissues. Their developments in preclinical trials include products for oncology and infectious diseases. Conditional suppression could be important for therapeutic applications in which the stable inhibition of target genes could be deleterious for cells or could produce side effect on animals. Jacques Mallet from CNRS presented a novel polymerase promoter inducible by doxycycline minimal RNA III that permits the control of siRNA levels and RNA interference by induction activation.

The Taicor method as a tool in the use of moderate RNAi screens to discover genes that are differentially regulated between cell lines was presented by Anyndya Dutta from the Univ. of Virginia Health Sciences Center. Using this technology it was possible to identify genes that are essential for cell cycle progression in p53+ cells but dispensable in p53- cells.

Annotation, data integration, and mining

The process of genome annotation and the integration of different types of information generated by the use of high throughput technologies are challenges of today's biological sciences. Several approaches such as H-invitational and Gene Desk databases and data integration tools such as BioMart were presented at the meeting. Nevertheless the importance of developing new concepts, theories, computational and mathematical tools to explore and identify relevant interactions and networks was acknowledged by the participants.

Systems Biology

The concept of Systems Biology was used at the conference at different scales of biological complexity: from metabolic pathways to cells and model organisms. Charles Auffrey presented the Systemoscope Consortium and his vision of Systems Biology as an integrational and iterative process based on interdisciplinarity and networking. He proposed a working hypothesis that self organization of living systems result from the conjunction of a stable organization and chaotic fluctuations. Dr. Auffrey recognized the need to measure small variations of a large number of weak signals with high throughput technologies developed under quality assurance. To achieve the understanding of biological systems and stimulate applications on preventive and predictive medicine the members of the French Systemoscope Consortium have endeavored a trans disciplinary research and training program in mathematics, information technology, physics, biology and medicine. The Systemascope project designed a pilot project to measure the dysfunctions of energy metabolism and modulations of gene expression profiles in skeletal muscles of patients with lung diseases and heart failure before and after rehabilitation. While the presentation covered the general aspects of the Systemascope Project research, their results were not presented.

Conclusions

The meeting represents the first attempt to change the focus of previous meetings from Transcriptome to Systems Biology for Health-Predictive and Preventive Medicine. Hence, the conference attempts to cover all the current high throughput technologies and tools developed in recent years to study the functioning of living organisms. The conference presented the applications of these tools in the understanding of cancer,

infectious diseases, and stem cell biology. Applications focussed to discover the mechanism of actions of traditional Chinese medicine were also presented. The current limitations of the technology and the need to develop new tools to accurately measure small variations were extensively discussed during the meeting. The conference concluded with the announcement of continueing this series of meetings in 2007.