T. YaoRIKEN Genomic Sciences Center
Yokohama, Japan

Synopsis

Bioinformatics and Systems Biology - towards Integrative Biology

In the Genomic/Post Genomic Era, Biology has been changing dramatically towards large scale and towards Integrative Biology.

With increasing full genome large scale measurements available, the field and the accumulation of biological data is expanding to embrace not only the genome, but also the transcriptome, the proteome, the metabolome and others.

New frontiers of research are emerging to include the:

- 1) Analysis of regulatory regions of genome sequences
- 2) Analysis of non-protein coding RNAs
- 3) Comparative genomics
- 4) SNP, Haplotype analysis
- 5) Gene expression analysis
- 6) Protein expression analysis
- 7) Protein interaction analysis
- 8) Structural Genomics/Proteomics
- 9) Metabolome analysis
- 10) Molecular Network analysis

In each area of research, bioinformatics plays an important or essential role in processing and analyzing the appropriate databases.

Additional important problems are emerging as follows.

- A) Annotation of gene and cDNA sequences.
- B) Integration, interconnection or reconstruction of various databases.
- C) Analysis of relationships between genotypes and phenotypes.
- D) Intercourse between Bioinformatics(BI) and Medical Informatics(MI).
- E) Standardization of concepts and words in BI and MI.
- F) Data-mining including literature information

Beyond this, we are now proceeding to a new frontier of research, elucidating various biological network systems which can be considered “design charts of life”, not only focusing on individual elements of life such as genes, proteins, metabolites etc. but looking at how they are integrated as we move from Molecular Biology to Systems Biology [1].

The following four papers were selected as the representative papers for the Yearbook of Medical Informatics. These are related to the above mentioned B), C), D) and E).

The first paper is a valuable integrated research report by 30 researchers on “The Synergy between Medical

Informatics and Bioinformatics”. Needless to say, one of the big outcomes of the genome sciences is the application to medicine. But bioinformatics researchers in the field of genome sciences usually face a lack of medical, clinical and patient data and information. On the other hand, medical informatics researchers recently have great expectations from the flood of genomics and post-genomics data and information and view collaboration between BI and MI to be beneficial for both sides. And the combination of both disciplines “Biomedical Informatics” should progress considerably by the fusion. This report will offer a strong background for the further development of Medical Informatics, Bioinformatics and Biomedical Informatics. It describes the expected impacts of a combination or fusion of MI and BI on Scientists/researchers, Clinical trials, Health-care professionals, Individual citizens, Health-care providers and systems, Policy-decision makers, Industry and Society.

The report also recommends 18 realistic research items, which should be done through collaboration between MI and BI researchers and practitioners. The 18 items are categorized into four groups. 1) MI in support of functional genomics, 2) BI in support

of individualized health care, and 3) Biomedical informatics in support of genomic medicine. Especially in 3), they emphasize the possibility to combine or fuse the molecular, physiological and pathological data and information by utilizing molecular / functional imaging technologies and modeling/simulation technologies. And in 4) Enabling technologies, they mention security, communication standards, integration of heterogeneous clinical and genetic databases, data and text-based knowledge discovery and the application of new type computing -Grid computing.

The second paper "From Phenotype to Genotype" presents a very unique approach to connect between genotype and phenotype. They started by listing questions by patients, doctors and researchers who are working on genetic diseases, and investigated how various and large scale genome-related databases could answer the questions.

They selected thirteen diseases such as *Huntington* disease, and listed up to nine questions on genetics, diagnostics and therapeutics. They selected 17 databases such as OMIM, GeneCards and KEGG from more than 300 biological databases to answer to the questions for the disease *Achondroplasia* as for one example.

As a result of this survey, many problems were pointed out. Especially, there are problems in terminology inconsistencies with gene names, in navigational inconsistencies due to the different updating cycles, and in inconsistent use of terms and methods. They say that the Gene Ontology will play important roles in these problems.

The third paper describes the evaluation result of "Ontology development tools for bioinformatics". This covers four tools for developing and maintaining Ontology -Protégé-2000, Chimaera, OilEd and DAG-Edit. They evaluated the performance and

user-friendliness of those tools using Gene Ontology as a test case. They conclude that no system is preferred in all situations, and all systems have their strengths and weaknesses. Further development will be needed. I dare say that another target besides GO should be added for a test case from the medical informatics side.

The fourth paper explains "caCORE" which was developed as the common information infrastructure for cancer research to collect and manipulate various and large-scale databases from many cancer projects in the USA which are supported by NIH. They show the EVS (Enterprise Vocabulary Services) in detail which controls and manages the data format and unified vocabulary "NCI thesaurus" (25,000 concepts and 70,000 words). They also explain the programming interface to use this caCORE which offers a very important tool for many users. An example of the application of this system to P53 signal pathway analysis in one project called CMAP (Cancer Molecular Analysis Program) is described.

This caCORE is one of the good examples of an object-oriented database. There are hundreds of biological databases now publicly presented through the Internet and some of them should be reorganized so as to be used as object-oriented databases (by disease, by organism etc.). But you need a lot of effort to coordinate scattered database activities for one direction and set up a thesaurus in each field. In this cancer case, the effort has been successful by considerable support from the cancer research fund. Many people from outside of cancer research would use the cancer DB, and the terminology of various fields should be consistent under the Ontology. So the authors should describe more clearly on the activity to translate Cancer thesaurus to Ontology.

Lastly, I think a comment on the importance of "Systems Biology" and "Integrative Biology" for the future of medical applications is in order.

As mentioned above, Bioinformatics is increasingly playing important and essential roles in handling a wide variety of large data sets in the genome/post genome era.

On the other hand, Systems Biology is now emerging to become an important approach which represents various life phenomena as systems and pursues an understanding of them and how this can help predict a system's behavior. The targets of systems biology range from the basic biological systems such as metabolic pathways, signal transduction pathways and circadian systems to whole microbial cells and particular cells such as liver cells, to organs and even to disease systems.

The ultimate target of systems biology might be to make life simulators or virtual organisms and "to understand the homeostasis of living systems".

In the process of systems biology research, we need to construct computer models to simulate the target systems using various kinds of data and knowledge of the systems and visualization of the simulation results would be helpful for understanding the phenomena and communications between experimentalists and theoreticians.

Systems biology will surely produce important tools for drug development, diagnostics and therapeutics [2,3]. There are especially strong research efforts in this field in the US, Europe and Japan.

In the United States, NIH, DOE and NSF have strongly promoted systems biology in these years and many projects or initiatives such as "Genome to Life" (www.doe-genome-to-life.org) are already ongoing. In 2004, NIH has announced to start an Integrative Cancer Biology project (http://dcb.nci.nih.gov->systems_biology). A

Biomedical Informatics and Systems Biology approach will be integrated in this project using the caCORE system.

In Germany, a large systems biology project “Systems of Life” started in 2003. They are focusing on the liver cells or hepatocytes as a challenging target (www.systembiologie.de).

In Japan, “The Genome Network Project” started in 2004 (Nature p.332, 27 May, 2004). They will pick up several human diseases as targets.

There are many horizontal Omics planes—genomics, transcriptomics, proteomics, metabolonomics, phenomics and vertical analyses which penetrate these planes to understand relationships “Genome vs. Phenome” are becoming more important [4]. Both Bioinformatics and Systems Biology will play important roles in this approach. We can view this whole approach as “Integrative Biology”.

Many centers or departments are being established throughout the world for the above purposes in recent years. Researchers or students from a wide variety of backgrounds are gathering in these centers or departments and they will explore new frontiers in the life sciences to solve exciting questions about biological and medical systems - and will open new applications of this knowledge to human healthcare and welfare.

Acknowledgements

I thank Casimir Kulikowski for giving me the opportunity to write this synopsis. I also thank Akiyoshi Wada and Tetsuro Toyoda at RIKEN for useful discussions.

References

1. Westerhoff H, Palsson B. The Evolution of Molecular Biology into Systems Biology. *Nat Biotechnol* 2004;22: 1249-52.
2. Butcher E, Berg E, Kunkel E (2004) Systems Biology in Drug Discovery. *Nat Biotechnol* 2004; 22:1253-59.
3. Hood L, Heath J, Phelps M, Lin B. Systems Biology and New Technologies Enable Predictive and Preventive Medicine. *Science* 2004;306:640-2.
4. Toyoda T, Wada A; Omic space: coordinate-based integration and analysis of genomic phenomic interactions. *Bioinformatics* 2004;20:1759-65.

Address of the author:

Toru Yao
RIKEN Genomic Sciences Center
1-7-22, SuehiroTsurumi
Yokohama, 230-0045, Japan
Tel: 81-45-503-9295
Fax: 81-45-503-9553
E-mail: yao@riken.jp
URL: www.riken.jp