Clinical Bioinformatics

Phei Lang Chang, MD

Clinical bioinformatics provides biological and medical information to allow for individualized healthcare. In this review, we describe the uses of clinical bioinformatics. After the analysis of the complete human genome sequences, clinical bioinformatics enables researchers to search online biological databases and use the biological information in their medical practices. The data obtained from using microarray is extremely complicated. In clinical bioinformatics, selecting appropriate software to analyze the microarray data for medical decision making is crucial. Proteomics strategy tools usually focus on similarity searches, structure prediction, and protein modeling. In clinical bioinformatics, the proteomic data only have meaning if they are integrated with clinical data. In pharmacogenomics, clinical bioinformatics includes elaborate studies of bioinformatics tools and various facets of proteomics related to drug target identification and clinical validation. Using clinical bioinformatics, researchers apply computational and high-throughput experimental techniques to cancer research and systems biology. Meanwhile, researchers of bioinformatics and medical information have incorporated clinical bioinformatics to improve health care, using biological and medical information. Using the high volume of biological information from clinical bioinformatics will contribute to changes in practice standards in the healthcare system. We believe that clinical bioinformatics provides benefits of improving healthcare, disease prevention and health maintenance as we move toward the era of personalized medicine. *(Chang Gung Med J 2005;28:201-11)*

Key words: clinical bioinformatics, bioinformatics, genomics, proteomics, microarray, systems biology, personalized medicine, medical informatics.

During the early 1960s, computer sciences emerged as important tools in the study of molecular biology. Bioinformatics uses biological information and mathematical, statistical and computing methods to research living things. The exploration of sequence and protein structure information has led to explosive growth in bioinformatics during the last decade. The discipline is becoming increasingly important in the study of biomedical problems.[1-3] Along with the complete human genome sequences, analyses of genome sequences improves understanding of biological systems while at the same time demanding massive bioinformatics tools to aid in biological analyses.[4]

Recently, clinical bioinformatics has been studied for use in stimulating the development of post-genomic technologies in clinical research and the practice of medicine. Clinical bioinformatics provides the technical infrastructure and knowledge to allow individualized healthcare using relevant sources of medical information and bioinformatics.[5] Medical information focuses on the acquisition, stor-

From the Division of Urology, Department of Surgery and Bioinformatics Center, Chang Gung Memorial Hospital, Taipei; Chang Gung University, Taoyuan.
Received: Aug. 31, 2004; Accepted: Feb. 22, 2005
Address for reprints: Dr. Phei Lang Chang, Department of Urology, Chang Gung Memorial Hospital. 5, Fushing Street, Gueishan Shiang, Taoyuan, Taiwan 333, R.O.C. Tel.: 886-3-3281200 ext. 2137; Fax: 886-3-3274541; E-mail: henryc@cgmh.org.tw

age, and use of information in healthcare. Bioinformatics has a large impact on biological research in microarray technology, proteomics, pharmacogenomics, oncology and systems biology. In this review, we describe the field of clinical bioinformatics and the benefits in improving healthcare, disease prevention and health maintenance. Microarray technology has provided the means for global analyses of the expression of thousands of genes in a single assay. Consequently microarray experiments have produced enormous amounts of data. Microarray analysis, including data mining, involves data processing using various statistical methods to identify gene expression patterns. Various microarray analysis software packages have been developed by commercial and academic researchers.[6] The statistical analysis of microarray data is the most significant problem associated with the clinical use of this technique.[7-11]

Clinical bioinformatics in proteomics is expanding to handle large heterogeneous data sets and to strengthen the knowledge discovery process. High throughput acquisition of proteome data is possible. Recently proteomics bioinformatics platforms have emerged as data management systems and knowledge bases in proteomics.[12] Various tools have been used to explore the protein in clinical bioinformatics.[13] During the last two decades, two-dimensional electrophoresis (2-DE) gel has established itself as the de facto approach to separating proteins from cell and tissue samples.[14] The role of clinical bioinformatics in 2-DE gel and 3-D structure is an actively pursued topic in proteomics.

In pharmacology, clinical bioinformatics has been used in drug target identification, clinical trials, development of biomarkers and toxicogenomic and pharmacogenomic tools. A biomarker has a particular molecular feature that makes it useful in measuring the progress of disease and the effects of treatment. Toxicogenomics has been used to identify toxic substances in the environment based on gene and protein activity. Pharmacogenomics combines traditional pharmaceutical sciences with annotated knowledge of genes, proteins, and single nucleotide polymorphisms (SNPs). Integrated computational programs are being created with the aim of facilitating in silico pharmacology by linking the genome, transcriptome, and proteome to cellular pathophysiology.[15]

In cancer research, high-throughput genome technology has generated massive quantities of data, including genome sequences, SNPs and microarray gene expression. In clinical bioinformatics, the investigation of gene expression using genomic and computational approaches and identification of sequence motifs facilitates cancer researchers and medical practitioners.[16] Clinical bioinformatics has been used to assist in early cancer detection and management, risk identification, risk reduction, and cancer prevention.[17]

In systems biology, experimental studies have been generating the high-throughput quantitative data essential to support simulation-based clinical research. Combined with rapidly progressing genome and proteome projects, the data explosion has convinced increasing numbers of researchers of the importance of a system-level approach. At the same time, considerable advances in computational power in clinical bioinformatics have facilitated the creation and analysis of reasonably realistic yet intricate biological models.[18,19]

Clinical bioinformatics has been important in clinical practice, including correlations between genetic variation with clinical risk factors, disease presentation, and differential response to treatment. Clinical bioinformatics research can benefit from the technological expertise that medical clinicians have used widely in the clinical arena, including database organization and knowledge representation, data mining, and modeling and simulation.[20]

Clinical bioinformatics in genomics

Applying computational approaches to store, organize, archive, analyze and visualize sequences is the fastest developing technology in biology. Bioinformaticians develop accessible interfaces, which allow researchers to search databases and assess whole genomes. People working in the research and development department of drug companies who are using the genome for drug targets require bioinformatic analysis strategies to extract information from databases more efficiently.[21] With more advanced applications, clinical bioinformatics has enabled researchers to search online databases for a given gene's sequences, proteins, mutations, coverage in the scientific literature and annotation, and model gene regulatory networks and metabolic pathways in the medical practice.[22]

To handle most biological problems, biological data from many sources must be combined. However, providing access to multiple data sources is complicated. The most insidious difficulties include the lack of standard file formats and data access methods. Therefore, integrating biological data across different databases and visual combinations of data to facilitate medical decision making is crucial to clinical bioinformatics.

Genome browsers are powerful tools, onto which numerous disparate types of genome sequence data can be mapped. Tools that compare genomes from different species are also worthwhile. Eventually, clinical bioinformaticians hope to pull all available genomic data into electronic health records (EHRs) that also consider the effects of genetic mutations. The EHRs will allow researchers to assess the contribution of genomic variations to disease.[22] Achievements in this direction will be accelerated through the refinements and breakthroughs in biotechniques, spanning both biomedical and genomic methodologies. This will enable an easy transfer of information from bench to bed, to better focus on the ongoing process of disease.[23]

Sequence database

The imperative of powerful tools and databases became obvious after the human genome project, whose completion was established several years ahead of schedule. Following the sequencing of the first microbial genome in 1995, the genomes of more than 1000 organisms have been sequenced and large-scale genome sequencing projects have evolved into routine procedures.[24,25] As genome information becomes richer and more complex, more of the real, underlying data model is emerging in common representations such as GenBank files. The big genomic databases such as National Center for Biotechnology Information (NCBI), European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database, and DNA Data Bank of Japan (DDBJ) have been developed for further analysis of the collected data.[26-28] Entrez is NCBI's search and retrieval system that provides users with access to sequences, mapping, taxonomy and structural data. The ability to retrieve related sequences, structures and references is the unique feature of Entrez. The European Bioinformatics Institute's (EBI's) Sequence Retrieval System (SRS) integrates and links the main nucleotide and protein databases as well as numerous other specialized molecular biology databases. The SRS has become an integration system for both data retrieval and applied data analysis. It contains more than 130 biological databases, integrates more than 10 applications, and provides capabilities to search multiple databases by shared attributes and to query across databases fast and efficiently.[27]

Sequence analysis

In sequence analysis, a rapid sequence database search tool, i.e., the Basic Local Alignment Search Tool (BLAST), was introduced in 1990, and has become a valuable and essential tool in biological research.[29] The core of NCBI's BLAST services is BLAST 2.0 otherwise known as “Gapped BLAST”. BLAST is a set of similarity search programs and provides a powerful way to perform sequence homology analysis. Both functional and evolutionary information can be inferred from well designed queries and alignments. The nucleotide databases for the BLAST are Non-redundant GenBank (nr), expressed sequence tags (est), est_human, est_mouse, est_others, Genome Survey Sequence (gss), High Throughput Genomic Sequences (htgs), Patent division of GenBank (pat), Protein Data Bank (PDB), nr released in the last 30 days (month), Alu repeats from REPBASE (alu_repeats), EST from the STS division of GenBank (dbsts), complete chromosomes from the NCBI Reference Sequence project (chromosome), and Whole Genome Shotgun sequences (wgs). The common protein sequence databases of BLAST include nr, SWISS-PROT protein sequence database (swissprot), pat, month, and pdb. The BLAST encompasses five programs.[29-32] The user needs to select a suitable program for protein and nucleic acid sequence comparisons to locate areas of sequential resemblance with a view for comparing structure and function (Table 1).

Clustal W is a multiple sequence alignment program for DNA or proteins. It provides the best match for the selected sequences, and lines them up so that the similarities and differences can be identified. Other programs for sequence alignment include MAP, MAFFT for DNA or proteins, CAP Sequence Assembly for DNA only, and PIMA and SIM for protein only.[33-38] These programs are openly accessible for academic use.

The European Molecular Biology Open

Table 1. The BLAST Programs

Program	Function
blastn	Nucleotide sequence to a nucleotide sequence database
blastx	Nucleotide sequence to a protein sequence database
blastp	Amino acid sequence to a protein sequence database
tblastn	Protein sequence to a nucleotide sequence database
tblastx	Nucleotide sequence to a nucleotide sequence database in six-frame translations

Software Suite (EMBOSS) is a package of high-quality free open source software for multi-purpose sequences analysis. EMBOSS automatically copes with data in a variety of formats, even allowing transparent retrieval of sequence data from the web. EMBOSS is a platform, providing extensive libraries and allowing scientists to create and distribute software.[39]

Comparative genome analysis

Comparison of DNA sequences from different species is a method for identifying functional elements in genomes. With an increasing number of vertebrate genomes being sequenced, unique opportunities for decoding DNA sequences through comparative genome alignments have arisen. Some software was designed as comparative genomics tools to align long genomic sequences. Some novel tools and strategies encompass this large volume of genomic information, facilitating the transfer of predictions generated by comparative sequence alignment to researchers focusing on experimental genome function annotation.[40,41] As an example, the Evolutionary Conserved Regions (ECR) Browser is a tool that provides easy and dynamic access to complete genome alignments of human, mouse, rat and fish sequences.[42] Comparative genome sequences can address basic evolutionary issues and the power of this approach depends to a large extent on the amount and quality of data available.[43] Many websites have created species genome browsers for sequence similarity searches (Table 2).

The sequence of the mouse genome is a basic informational tool for understanding the human genome and a principal experimental tool for biomedical research. In 2002, many researchers contributed to a report presenting a comparative analysis of the mouse and human genomes and also discussed topics including analysis of the evolutionary forces shaping the size, structure and sequence of the genomes. In addition the number of protein-coding genes, the expansion of gene families related to reproduction and immunity, the evolution of proteins, and the identification of intraspecies polymor-

Table 2. Web Sites for Comparative Genome Analyses

	Web site	Comparative Genome
BMERC	http://bmerc-www.bu.edu/information/all_genomes.shtml	Completed Genomes collected resources
Broad Institute	http://www.broad.mit.edu/annotation/fungi/comp_yeasts/	Yeast Comparative Genomics for S. paradoxus, S. mikatae, and S. bayanus
Ensembl Genome Browser	http://www.ensembl.org/	Human, mouse, zebrafish, rat, chicken, mosquito, fugu, fruitfly, chimp, honeybee, dog, C. elegans, C. briggsae
HGMP Resource Centre	http://www.hgmp.mrc.ac.uk/GenomeWeb/comp-gen-db.html	Homology information for mouse, human and over 70 other mammalian species
IMB Jena	http://genome.imb-jena.de/	Structural genomics for human, mouse, chimpanzee, rat, Dictyostelium, Cyanidium, microorganism, fugu
MBGD	http://mbgd.genome.ad.jp/	Microbial Genome Database for Comparative Analysis
NCBI Mouse Genome Resources	http://www.ncbi.nlm.nih.gov/genome/guide/mouse/	Human, mouse, rat, zebrafish
VIDA	http://www.biochem.ucl.ac.uk/bsm/virus_database/VIDA.html	Homologous protein families from virus genomes

Abbreviations: BMERC: BioMolecular Engineering Research Center; HGMP: Human Genome Mapping Project; IMB: Institute of Molecular Biotechnology; MBGD: Microbial Genome Database; NCBI: National Center for Biotechnology Information; VIDA: Virus Database.

phism were also noted.[44] Some of the insights that can be gleaned from the two sequences were described.

The International Chimpanzee Chromosome 22 Consortium has sequenced chimpanzee chromosome 22 to a degree of completion and accuracy equivalent to that of the human genome assembly. The quality of this chimp chromosome sequence is sufficiently accurate to allow reliable comparisons with the human chromosome 21. A chimpanzee chromosome provides a unique angle from which to view the human genome and establish insights about its evolution because the sequences of these evolutionary near-neighbors began drifting apart some 6 million years ago. The longer-term hope for comparative genome analysis is to identify the sequence alterations that could account for the differences between chimps and the human.[45] The initial report showed that the genomic changes that occurred after speciation seems more complex than originally hypothesized.

Clinical bioinformatics in microarray

Microarray detects the presence and abundance of labeled nucleic acids in a biological sample, which will hybridize to the DNA on the array, and which can be discovered via the label. In the microarray investigations, the labeled nucleic acids are derived from the mRNA of a sample or tissue, and so the microarray measures gene expression. In microarray, the mathematics, statistics and computing are necessary to acquire, analyze and store microarray data. Many thousands of different DNA molecules bond to an array and researchers estimate the expression of many thousands of genes simultaneously in microarray. A microarray experiment can not be performed without bioinformatics involvement.[46]

In clinical bioinformatics, four conditions are commonly discovered from microarray data: (1) genes with similar expression patterns over all samples, (2) genes with unusual expression levels in a sample, (3) genes whose expression levels vary across samples and (4) samples with similar expression patterns.[47-51] The data obtained from microarray are extremely important and complicated, causing considerable challenges in medical practices. Therefore, selecting appropriate software to analyze the microarray data for medical decision making is crucial. All software generated by the medical institution may be downloaded and used for free by academic and other non-profit researchers. However, the commercial use of software and/or source codes normally need to be licensed, such as for ScanAlyze, Cluster and/or TreeView from Stanford University (Table 3).[52]

Clinical bioinformatics in proteomics

A common approach in proteomic analysis uses 2-DE gel to separate proteins and matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) to identify proteins.[53] In the develop-

Table 3. The Software for Analysis of Microarray Data

Program	Author	Provider	Platform	Description
Cluster	Michael Eisen	UC Berkeley	Windows	For clustering, SOM
TreeView	Michael Eisen	UC Berkeley	Windows	Graphically browse and analyses results of clustering
ScanAlyze	Michael Eisen	UC Berkeley	Windows	Processes fluorescent images of microarrays
ArrayMiner		Optimal Design	Windows Mac OS	Set of analysis tools
Expression Profiler		EBI	Web	Analysis & clustering of gene expression data
GeneX-Lite		NCGR	Windows Linux SunOS	Integrated toolset, provides an interface to RDBMS
BASE	Carl Troein	Lund University	Web	Microarray database and analysis platform
dChip	Cheng Li and Wing Hung Wong	Affymetrix analysis packages	Windows 2000	Analysis of oligonucleotide arrays
ArrayDB	Anthony Masiello	NHGRI	Windows	Mining and analysis of microarray data

Abbreviations: EBI: European Bioinformatics Institute; NCGR: National Center for Genome Resources; RDBMS: Relational Database Management Systems; BASE: BioArray Software Environment; NHGRI: National Human Genome Research Institute.

ing field of clinical bioinformatics, tools for protein analysis are of crucial to making the best use of the accumulated data in medical practices. Specialized databases concerning protein sequences, protein tertiary structure, and proteome analysis represent useful resources for analyzing protein sequences. Proteomics strategy tools usually focus on similarity searches, structure predictions, detection of specific regions, alignments, data mining, and protein modeling.[54] In clinical bioinformatics, the proteomic data only have meaning if they are integrated with clinical data.

The Expert Protein Analysis System (ExPASy) proteomics server of the Swiss Institute of Bioinformatics (SIB), created in August 1993, is devoted to the analysis of protein sequences and structures. ExPASy allows users to browse across many databases produced in Geneva, such as Swiss-Prot, PROSITE, SWISS-2DPAGE, SWISS-3DIMAGE, and ENZYME, as well as other cross-referenced databases (such as EMBL/GenBank/DDBJ, OMIM, Medline, FlyBase, ProDom, SGD, and SubtiList). ExPASy also allows access to many analytical tools for protein identifications, the analysis of their sequence and the prediction of their tertiary structure.[55]

Protein structure database (PDB)

The PDB is the only worldwide repository for the processing and distribution of three-dimensional (3-D) structure data of large molecules of proteins and nucleic acids. New structures are released each Wednesday by 1:00 am Pacific time.[56]

NCBI's structure database is called Molecular Modeling DataBase (MMDB). It is a subset of three-dimensional structures obtained from the PDB. Most 3-D structure data are obtained from X-ray crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy. They provide a wealth of knowledge about biological function, on mechanisms linked to the function, and on the evolutionary history of and relationships between macromolecules.[57]

The EBI Macromolecular Structure Database (MSD), the European project for the collection, management and distribution of data about macromolecular structures, is partly derived from the PDB.[58] In November 2003, the MSD, the Research Collaboratory for Structural Bioinformatics (RCSB) in the United States, and the Protein Data Bank Japan (PDBj) announced a collaboration to form the Worldwide Protein Data Bank (wwPDB).[59] The wwPDB is freely and publicly available globally. Each wwPDB website provides an independent view of the primary data, thus providing a variety of tools and resources for the global community.

The software to display the protein structure is extremely important. Protein Explorer, which was developed by the Department of Microbiology, University of Massachusetts, is a free software for visualizing the three-dimensional structures of protein, DNA, and RNA macromolecules, and their interactions and binding of ligands, inhibitors, and drugs.[60] Protein Explorer can work well in Windows, Macintosh, Linux or other Unix platforms. Version 2.0 of the Protein Explorer, released in increments in 2000-2002, is much easier to use, and much more powerful than RasMol and Chime. Chime is available only for Windows and Macintosh. Cn3D is a structure viewer and a helper application for web browser users allowing users to view 3-D structures from NCBI's Entrez retrieval service. Cn3D can run on Windows, Macintosh, and Unix, simultaneously displaying structures, sequences, and alignments, and now possesses powerful annotation and alignment editing features.[61]

Clinical bioinformatics in pharmacogenomics

Researchers of pharmacogenomics study how the construction of the genome of an organism influences the drug responses. The revolution of personalized medicine, where each patient would receive the most appropriate pharmacotherapy, based on his or her genetic make-up, will affect almost every medical discipline. Clinically, many types of adverse drug reactions are related to polymorphic gene alleles of drug metabolizing enzymes. Moreover, insights have been gained into reasons for inefficient drug efficacy, often related to SNPs or larger polymorphisms in genes encoding drug target proteins.[62-64] Clinical bioinformatics in pharmacogenomics, includes elaborate studies of bioinformatics tools, various facets of proteomics related to drug target identification and validation, and the evolving knowledge of the correlation of human genome diversity with drug efficacy disparity between members of different ethnic groups. The forthcoming generation of clinicians and researchers will be routinely acquainted with the most recent developments in

pharmacogenomics and clinical bioinformatics and will be capable of providing patients with the anticipated benefits of personalized medicine.[65,66]

SNPs are the alteration of a single nucleotide. It is considered that 93% of genes contain SNPs.[67] If they arose in gene coding for receptor proteins, drug metabolism enzymes, transporter proteins or any of the many other proteins involved in the drugs metabolism, they could potentially influence the patients responses to specific drugs. Therefore, genetic analysis of an individual looking for SNPs in specific areas known to code for important drug metabolism proteins could direct the clinician to specific drugs, or avoid other specific drugs as the individual's genetic make-up dictated. In collaboration with the National Human Genome Research Institute (NHGRI), the NCBI has established the dbSNP database to act as a central bank for both single base nucleotide substitutions and short deletion and insertion polymorphisms.[26]

The use of the DNA microarray technique in the field of pharmacogenomics has been developed in order to characterize and validate new therapeutic targets, their action mechanisms, metabolic pathways and undesirable secondary effects. mRNA expression profiling via microarray allows detection of genes which are expressed differentially resulting in different responses to the ingestion of drugs. These genes would be considered candidate genes that are involved in drug metabolism.[68]

Clinical bioinformatics in cancer research

In clinical bioinformatics, researchers apply computational and high-throughput experimental techniques to identify novel targets and agents in cancer diagnosis, treatment, prevention and control, and encourage the deployment of chemical, structural and biochemical methods to all basic and clinical cancer research.[69] A method to validate multiple microarray data sets, a Web-based cancer microarray database for biomarker discovery, and methods for integrating gene ontology annotations with microarray data to improve candidate biomarker selection have been developed.[70] Many clinical bioinformatics tools are being evaluated and applied in various medical fields, including early diagnosis, risk assessment, classification, and prognosis of cancer.[71]

The National Cancer Institute Center for Bioinformatics (NCICB) provides biomedical informatics support and integration capabilities to the cancer research community.[72] The NCICB provides direct support to four NCI research programs: The Cancer Genome Anatomy Project (CGAP), The Mouse Models of Human Cancer Consortium (MMHCC), The Director's Challenge: Towards a Molecular Classification of Cancer and Clinical Trials.[73,74]

In clinical bioinformatics, biomarkers are used to detect cancer during different stages, initiation, development, and progression. Biomarkers may be helpful in the diagnosis of lower-grade cancers, which have a low sensitivity of cytology. It also has the potential to provide a better predictive value and to monitor the effects of chemotherapy. Biomarkers should be easy to detect, measurable across populations, and useful for the detection of cancer at an early stage, identification of high-risk individuals, and detection of recurrence.[75] The NCI Biomarker Developmental Laboratories was created to identify techniques for finding molecular, genetic, and biologic early warning signals of cancer.[76]

Clinical bioinformatics in systems biology

Systems biology aims at system-level understanding of biological systems and is a new field in biology.[77,78] A system-level understanding of a biological system is derived from insight into four key properties: (1) system structures, (2) system dynamics, (3) control method, and (4) design method.[79] Systems biology represents the integration of computer modeling, large-scale data analysis, and biological experimentation. In clinical bioinformatics, computational modeling and analysis are now able to provide useful biological insights and predictions for clearly recognized targets, e.g. analysis of cell cycle and metabolic analysis.[80-82]

Systems Biology Markup Language (SBML), CellML language, and Systems Biology Workbench was aimed to establish a standard and open software platform for modeling and analysis.[83-85] Some databases involved in biological pathways allow them to develop machine executed models, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG), Alliance for Cellular Signaling (AfCS), and Signal Transduction Knowledge Environment (STKE).[86-88] The methods and concepts of systems biology will not only expand into all areas of biological science, its results are bound to have repercussions in and

inspire other sciences such as physics, engineering, mathematics, and social sciences.[89] The systems biology approach, with its combination of computational, experimental and observational enquiry, is relevant to drug discovery and response and the optimization of medical treatment regimens for individual patients.[90] In clinical bioinformatics, the application of systems biology in medical practices is the future of personalized medicine.

Clinical bioinformatics and medical informatics

Bioinformatics (BI) encompasses the development and application of informatics techniques in the biological sciences. Medical informatics (MI) has initiated the development and introduction of informatics methods in clinical medicine and biomedical research. BI and MI vary in their histories, scientific foundations, and methodological approaches to research, providing a basis for exchange of experiences in their different applications. Clinical bioinformatics cultivates informatic methods that will prove crucial in the development of genomic medicine. The future of clinical bioinformatics will be affected markedly by whether or not significant advances in clinical practice and biomedical research originate from separate efforts in BI and MI.[91-93]

Since the human genome project was completed in April 2003, the main interests in bioinformatics are now moving to post-genomic challenges, including functional genomics, comparative genomics, proteomics, metabolomics, pathway analysis, systems biology and clinical bioinformatics.[94] Computational analyses of human genes associated with diseases and their protein products are enhanced. Recently in clinical bioinformatics biological databases have been managed in EHR to improve health care and clinical practice moves toward personalized medicine.[95] A virtual patient system has been developed and used to model obesity, diabetes, and asthma. The day will soon come when we use these virtual patient models for routine clinical decision making with regard to clinical bioinformatics.[96]

Summary

Advances in molecular-based information have led to a deeper understanding of the complexity of life. Clinical bioinformatics using the high volume of biological information will contribute to changes in practice standards in healthcare systems. We believe that clinical bioinformatics will provide potential benefits to patients by means of improved healthcare, disease prevention and health maintenance as we move toward the era of personalized medicine.

REFERENCES

1. Hagen JB. The origins of bioinformatics. Nat Rev Genet 2000;1:231-6.
2. Gibas C, Jambeck P. Developing bioinformatics computer skills. 1st ed. California: O'reilly and associates, Inc., 2001:3-4.
3. Baxevanis AD, Ouellette BFF. Bioinformatics: a practical guide to the analysis of genes and proteins. 2nd ed. New York: John Wiley and Sons, Inc., 2001:1-2.
4. Yu U, Lee SH, Kim YJ, Kim S. Bioinformatics in the post-genome era. Biochem Mol Biol 2004;37:75-82.
5. de Groen PC. Synopsis: Towards clinical bioinformatics. Yearbook of Medical Informatics. Heidelberg, Germany: Schattauer publishing Co., 2004:223-5.
6. Krajewski P, Bocianowski J. Statistical methods for microarray assays. J Appl Genet 2002;43:269-78.
7. Eisen MB, Brown PO. DNA arrays for analysis of gene expression. Methods Enzymol 1999;303:179-205.
8. King HC, Sinha AA. Gene expression profile analysis by DNA microarrays. JAMA 2001;286:2280-8.
9. Li C, Wong WH. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. Genome Biology 2001;2:research0032.1-0032.11.
10. Schadt EE, Li C, Su C, Wong WH. Analyzing high-density oligonucleotide gene expression array data. J Cell Biochem 2001;80:192-202.
11. Murphy D. Gene expression studies using microarrays: principles, problems, and prospects. Adv Physiol Educ 2002;26:256-70.
12. Blueggel M, Chamrad D, Meyer HE. Bioinformatics in proteomics. Curr Pharm Biotechnol 2004;5:79-88.
13. Thongboonkerd V, Klein JB. Practical bioinformatics for proteomics. Contrib Nephrol 2004;141:79-92.
14. Dowsey AW, Dunn MJ, Yang GZ. The role of bioinformatics in two-dimensional gel electrophoresis. Proteomics 2003;3:1567-96.
15. Whittaker PA. What is the relevance of bioinformatics to pharmacology? Trends Pharmacol Sci 2003;24:434-9.
16. Yang HH, Lee MP. Application of bioinformatics in cancer epigenetics. Ann N Y Acad Sci 2004;1020:67-76.
17. Kapetanovic IM, Umar A, Khan J. Proceedings: the Applications of Bioinformatics in Cancer Detection Workshop. Ann N Y Acad Sci 2004;1020:1-9.
18. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science

2001;292:929-34.
19. Kitano H. Computational systems biology. Nature 2002;420:206-10.
20. Miller PL. Opportunities at the intersection of bioinformatics and health informatics: a case study. J Am Med Inform Assoc 2000;7:431-8.
21. Chicurel M. Bioinformatics: bringing it all together. Nature 2002;419:751, 753, 755 passim.
22. Schmidt CW. Data explosion: bringing order to chaos with bioinformatics. Environ Health Perspect 2003;111:A340-5.
23. Robert M, Alexander S, Michael M, Zlatko T. New trends in bioinformatics: from genome sequence to personalized medicine. Experimental Gerontology 2003;38:1031-6.
24. Janssen P, Audit B, Cases I, Darzentas N, Goldovsky L, Kunin V, Lopez-Bigas N, Peregrin-Alvarez JM, Pereira-Leal JB, Tsoka S, Ouzounis CA. Beyond 100 genomes. Genome Biol 2003;4:402.
25. Kanehisa M, Bork P. Bioinformatics in the post-sequence era. Nat Genet 2003;33 Suppl:305-310.
26. http://www.ncbi.nlm.nih.gov/
27. http://www.ebi.ac.uk/embl/
28. http://www.ddbj.nig.ac.jp/
29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403-10.
30. Zhang J, Madden TL. PowerBLAST: a new network BLAST application for interactive or automated sequence analysis and annotation. Genome Res 1997;7:649-56.
31. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389-402.
32. Wheeler DL, Church DM, Edgar R, Federhen S, Helmberg W, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Suzek TO, Tatusova TA, Wagner L. Database resources of the National Center for Biotechnology Information: update. Nucleic Acids Res 2004;32:D35-40.
33. http://dot.imgen.bcm.tmc.edu/multi-align/multi-align.html
34. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ. Multiple sequence alignment with Clustal X. Trends Biochem Sci 1998;23:403-5.
35. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 1997;24:4876-82.
36. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 2002;30:3059-66.
37. Huang X, Madan A. CAP3: A DNA Sequence Assembly Program. Genome Research 1999;9:868-77.
38. http://au.expasy.org/tools/sim-prot.html
39. http://www.hgmp.mrc.ac.uk/Software/EMBOSS/
40. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res 2004;32(Web Server Issue):W273-9.
41. http://bmerc-www.bu.edu/cgi-bin/blast_server.prl
42. Ovcharenko I, Nobrega MA, Loots GG, Stubbs L. ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. Nucleic Acids Res 2004;32(Web Server Issue):W280-6.
43. Weissenbach J. Genome sequencing: differences with the relatives. Nature 2004;429(6990):353-5.
44. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. Nature 2002;420:520-62.
45. The International Chimpanzee Chromosome 22 Consortium. DNA sequence and comparative analysis of chimpanzee chromosome 22. Nature 2004;429:382-8.
46. Murphy D. Gene expression studies using microarrays: principles, problems, and prospects. Adv Physiol Educ 2002;26:256-70.
47. Kostka D, Spang R. Finding disease specific alterations in the co-expression of genes. Bioinformatics 2004;20 Suppl 1:I194-I9.
48. Oh TJ, Kim CJ, Woo SK, Kim TS, Jeong DJ, Kim MS, Lee S, Cho HS, An S. Development and clinical evaluation of a highly sensitive DNA microarray for detection and genotyping of human papillomaviruses. J Clin Microbiol 2004;42:3272-80.
49. Affara NA. Resource and hardware options for microarray-based experimentation. Brief Funct Genomic Proteomic 2003;2:7-20.
50. Yeung KY, Medvedovic M, Bumgarner RE. From co-expression to co-regulation: how many microarray experiments do we need? Genome Biol 2004;5:R48.
51. Zhou XJ, Gibson G. Cross-species comparison of genome-wide expression patterns. Genome Biol 2004;5:232.
52. http://genome-www5.stanford.edu/resources/restech.shtml
53. Thongboonkerd V, Klein JB. Practical bioinformatics for proteomics. Contrib Nephrol 2004;141:79-92.
54. Pruess M, Apweiler R. Bioinformatics Resources for In Silico Proteome Analysis. J Biomed Biotechnol 2003;4:231-6.
55. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003;31:3784-8.
56. http://www.rcsb.org/pdb/
57. http://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml
58. http://www.ebi.ac.uk/msd/
59. http://www.wwpdb.org/
60. Martz E. Protein Explorer: easy yet powerful macromolecular visualization. Trends Biochem Sci 2002;27:107-9.
61. http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml
62. Nebert DW, Menon AG. Pharmacogenomics, ethnicity,

and susceptibility genes. Pharmacogenomics J 2001;1:19-22.
63. Licinio J. Pharmacogenomics and ethnic minorities. Pharmacogenomics J 2001;1:85.
64. Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A, Chen Z, Chu J, Carcassi C, Contu L, Du R, Excoffier L, Ferrara GB, Friedlaender JS, Groot H, Gurwitz D, Jenkins T, Herrera RJ, Huang X, Kidd J, Kidd KK, Langaney A, Lin AA, Mehdi SQ, Parham P, Piazza A, Pistillo MP, Qian Y, Shu Q, Xu J, Zhu S, Weber JL, Greely HT, Feldman MW, Thomas G, Dausset J, Cavalli-Sforza LL. A human genome diversity cell line panel. Science 2002;296:261-2.
65. Risch N, Burchard E, Ziv E, Tang H. Categorization of humans in biomedical research: genes, race and disease. Genome Biol 2002;3:1-12.
66. Gurwitz D, Weizman A, Rehavi M. Education: Teaching pharmacogenomics to prepare future physicians and researchers for personalized medicine. Trends Pharmacol Sci 2003;24:122-5.
67. McLeod HL, Yu J. Cancer pharmacogenomics: SNPs, chips, and the individual patient. Cancer Invest 2003;21:630-40.
68. Meloni R, Khalfallah O, Biguet NF. DNA microarrays and pharmacogenomics. Pharmacol Res 2004;49:303-8.
69. Greenwald P. Preface. The applications of bioinformatics in cancer detection. Ann N Y Acad Sci 2004;1020:ix-x.
70. Rhodes DR, Chinnaiyan AM. Bioinformatics strategies for translating genome-wide expression analyses into clinically useful cancer markers. Ann N Y Acad Sci 2004;1020:32-40.
71. Kapetanovic IM, Rosenfeld S, Izmirlian G. Overview of commonly used bioinformatics methods and their applications. Ann N Y Acad Sci 2004;1020:10-21.
72. http://ncicb.nci.nih.gov/
73. http://cgap.nci.nih.gov/
74. http://cmap.nci.nih.gov/
75. Wagner PD, Verma M, Srivastava S. Challenges for biomarkers in cancer detection. Ann N Y Acad Sci 2004;1022:9-16.
76. http://www.cancer.gov/newscenter/content_nav.aspx?viewid=0846bd05-b784-4fa8-bde3-4824249b055d
77. Kitano H, Imai S. The two-process model of cellular aging. Experimental Gerontology 1998;33:393-419.
78. Kitano H. Perspectives on systems biology. New generation computing 2000;18:199-216.
79. Kitano H. Systems biology: a brief overview. Science 2002;295:1662-4.
80. Borisuk MT, Tyson JJ. Bifurcation analysis of a model of mitotic control in frog eggs. J Theor Biol 1998;195:69-85.
81. Chen KC, Csikasz-Nagy A, Gyorffy B, Val J, Novak B, Tyson JJ. Kinetic analysis of a molecular model of the budding yeast cell cycle. Mol Biol Cell 2000;11:369-91.
82. Edwards JS, Ibarra RU, Palsson BO. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nat Biotechnol 2001;19:125-30.
83. http://sbml.org/index.psp
84. http://www.cellml.org/public/news/index.html
85. http://www.sbw-sbml.org/oldindex.html
86. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 2000;28:27-30.
87. http://www.signaling-gateway.org/
88. http://stke.sciencemag.org/
89. Ehrenberg M, Elf J, Aurell E, Sandberg R, Tegner J. Systems biology is taking off. Genome Res 2003;13:2377-80.
90. Kitano H. Computational systems biology. Nature 2002;420:206-10.
91. Maojo V, Kulikowski CA. Bioinformatics and medical informatics: collaborations on the road to genome medicine? J AM Med Inform Assoc 2003;10:515-22.
92. Kohane I. Bioinformatics and clinical informatics: the imperative to collaborate. J Am Med Inform Assoc 2000;7:439-43.
93. Maojo V, Martin-Sanchez F, Crespo J, Iakovodis I, Kulikowsli C. Medical informatics and bioinformatics: European efforts to facilitate synergy. Brief report. J Biomed Inform 2001;34:423-7.
94. Austin CP. The impact of the completed human genome sequence on the development of novel therapeutics for human disease. Annu Rev Med 2004;55:1-13.
95. Zhou M, Zhuang YL, Xu Q, Li YD, Shen Y. VSD: a database for schizophrenia candidate genes focusing on variations. Hum Mutat 2004 Jan;23:1-7.
96. http://www.entelos.com/

臨床生物資訊

張慧朗

「臨床生物資訊」結合生物與醫學資訊來促進個人化的健康照護，在此，我們回顧臨床生物資訊的重要性，同時重點討論生物資訊的發展。

人類基因體序列全部解開後，由於臨床生物資訊的快速成長，研究者可以由網路上找尋和使用生物資訊，應用於臨床醫學上。最近微陣列實驗產生了大量複雜的微陣列資料，臨床生物資訊必須選擇適當的工具做分析，讓微陣列資料可以做爲醫療決策的參考。

由於臨床生物資訊提供了臨床醫學資料，使得蛋白質體的研究發展更有臨床價值。藥理基因體學雖然剛起步，但基因體，蛋白質體，與單一核苷酸多形性在藥理學上的研究，也因臨床生物資訊分析的協助將日趨重要。癌症研究是生物學家長期抗癌的不變法則，由於臨床生物資訊的發展，使人類得以早期發現癌症同時減輕癌症的危險性。在系統生物學，臨床生物資訊研究已朝向基因、蛋白質、通路、及模組的系統性研究。

臨床生物資訊將結合生物資訊與醫學資訊，實際應用於臨床上來照顧病人。我們相信由於臨床生物資訊提供了病人詳細的臨床表徵、基因表現、蛋白質標記、與多形性的單一核苷酸，將使醫學達到針對個人發展的「個人化醫學（personalized medicine）」境界。(長庚醫誌 2005;28:201-11)

關鍵字：臨床生物資訊，生物資訊，基因體，蛋白質體，微陣列，系統生物學，醫學資訊，個人化醫學。

長庚紀念醫院　台北院區　外科部　泌尿科
受文日期：民國93年8月31日；接受刊載：民國94年2月22日。
索取抽印 本處：張慧朗醫師，長庚紀念醫院　泌尿科。桃園縣333龜山鄉復興街5號。Tel.: (03) 3281200轉2137 Fax: (03) 3274541；E-mail: henryc@cgmh.org.tw