This page is intentionally blank

Clinical bioinformatics: A merging of domains

E. Z. Silfen Senior Director, Biomedical Informatics Research, Philips Research North America

This article describes the emergence of the hybrid discipline of clinical bioinformatics. Specifically, it defines biomedical informatics as well as the four compositional domains of bioinformatics, imaging informatics, clinical informatics and public health informatics. Furthermore, it describes the relationship between bioinformatics, molecular medicine and clinical decision support, and offers a definition of clinical bioinformatics that arises from the integration of these domains. After establishing this background, the article discusses use cases for clinical bioinformatics, summarizes the field of clinical bioinformatics, and posits its potential for biomedicine research.

Biomedical informatics

Since the publication in 1959 of Ledley and Lusted's classic article, "Reasoning Foundations of Medical Diagnosis"[1] the scientific field of biomedical informatics has been evolving as a unique discipline that "studies the structure and properties of scientific information and the laws of all processes of scientific communication" [2]. After evolving over many years, Edward (Ted) Shortliffe, MD, PhD, the former chair of the Department of Biomedical Informatics at Columbia University and recent recipient of the ACMI Morris F. Collen Award for Excellence in the Field of Biomedical Informatics [3], defined biomedical informatics as the representation, storage, retrieval, presentation, sharing, and optimal use of biomedical data, information and knowledge for problem solving and clinical decision making that touches on all basic and applied fields in biomedical science and is closely tied to modern information technologies, notably in the areas of computing and communication [4].

From a use case perspective, biomedical informatics represents the intersection of information science, medicine and health care. The field addresses the resources, devices and methods required to optimize the acquisition, storage, retrieval and use of information in healthcare and biomedicine. Furthermore, biomedicine tools include not only computers but clinical guidelines, formal medical terminologies, and information and communication systems [5].

The four primary domains of biomedical informatics include: bioinformatics [6], imaging informatics [7], clinical informatics [8], and public health informatics [9]. In addition, sub-domains include nursing informatics, consumer health informatics, dental informatics, clinical research informatics, and pharmacy informatics [5]. Based upon these academic foundations, this article describes the discipline of clinical bioinformatics as used by Philips Research in the pursuit of best patient care.

Bioinformatics and computational biology

Bioinformatics is the study of biomedicine information at the cellular level. The field emphasizes the computational representation, analysis and presentation of biological information at the genomic, epigenomic, transcriptomic, proteomic and metabolomic levels [10]. Both bioinformatics and computational biology use methods associated with applied mathematics, informatics, statistics, computer science, artificial intelligence, chemistry and biochemistry to solve biological problems usually on the molecular level. Research efforts in computational biology often overlap with systems biology and include sequence alignment, gene finding, genome assembly, protein structure alignment, protein structure prediction, prediction of gene expression and protein-protein interactions, and the modeling of evolution [11].

The terms bioinformatics and computational biology are often used interchangeably. However, bioinformatics more properly refers to the creation and advancement of algorithms, computational and statistical techniques, and theory to solve formal and practical problems posed by or inspired from the management and analysis of biological data. Computational biology, on the other hand, refers to hypothesis-driven investigation of a specific biological problem using computers, carried out with experimental and simulated data, with the primary goal of discovery and the advancement of biological knowledge. A similar distinction is made by National Institutes of Health in their working definitions of Bioinformatics and

Computational Biology, where it is further emphasized that there is a tight coupling of developments and knowledge between the more hypothesis-driven research in computational biology and technique-driven research in bioinformatics. Computational biology also includes lesser known but equally important sub-disciplines such as computational biochemistry and computational biophysics [12].

Molecular medicine

Evolving from computational biology, molecular medicine is the application of the biological characteristics of disease in order to offer patients individualized medical care. By combining systems biology and informatics methods, molecular medicine creates an 'omics' medicine approach [13] that identifies and internally validates markers of cellular function for use in clinically enabling technologies.

Molecular medicine can be sub-divided into molecular diagnostics and molecular imaging. Based on the in vitro detection of biomarkers, molecular diagnostics focuses on testing for specific molecules associated with disease in order to determine an individual patient's predisposition to illness, screen patients for the presence of a medical condition, plan personalized therapeutic approaches, and monitor the patient's response to therapeutic interventions.

As a companion modality, molecular imaging emphasizes the identification of the in vivo location and extent of disease by gauging the presence of specific molecules associated with a disease. The core technology requires medical imaging equipment that can detect and quantify disease-specific, molecular contrast agents. When combined with molecular diagnostic methods, molecular imaging can offer important information for both the planning and monitoring phases of medical therapy.

The sine qua non for molecular medicine is the reliable and accurate identification and measurement of an individual's biomolecular pattern of disease as well as their treatment response profile [14]. This characterization enables the integration of the final component necessary for defining the discipline of clinical bioinformatics, clinical decision support.

Clinical decision support systems

Clinical Decision Support Systems (CDSS) is the branch of biomedical informatics that unifies knowledge discovery with engineering techniques to create expert systems. These systems require clinical evaluation, advanced design, usability testing, and heuristic evaluation prior to clinical deployment [15]. The hallmarks of CDSS include:

- The design, development and validation of the right databases and "intelligent" algorithms that can ask the right questions to mine the data [16];
- The integration of multiple streams of medical evidence in order to create a clinical environment that fosters personalized, predictive and pre-emptive medicine;
- Performing evaluation and comparative effectiveness studies that benchmark applications against ground truth and clinical domain expertise.

For patients, expert decision systems are essential for improving clinical outcomes while for clinicians, these solutions are necessary for workflow and care process improvement. In summary, CDSS relies upon computer-supported systems that enable physicians to optimally use biomedical data, information and knowledge for problem solving and clinical decision making [17].

Clinical Bioinformatics

Clinical Bioinformatics emerges from the intersection of the methods, techniques, applications, and solutions of bioinformatics, molecular medicine and clinical decision support. This "overlap" enables integration of evidence, thereby affording physicians more robust medical decision making capability [18] (Figure 1).

Specifically, as clinical decision support systems blend biomarker information into the expert systems necessary for evidence-based patient care, significant opportunities arise for clinical bioinformatics research [19]. Two of these are the use of previously identified genetic information for making medical care decisions, and the simulation of disease processes to guide the "type and timing" of therapeutic interventions. Examples of genetic applications include:

- Clinical correlation of genetic information. That is, genetic diseases might be caused by different mutations of a single gene or by mutations of different genes that are related, for example, because they code for different enzymes in a single metabolic pathway [20, 21, 22].
- Genetic variation may be correlated with different levels of severity of a disease or different presentations of signs or symptoms [23].
- Patients with different genetic makeup may

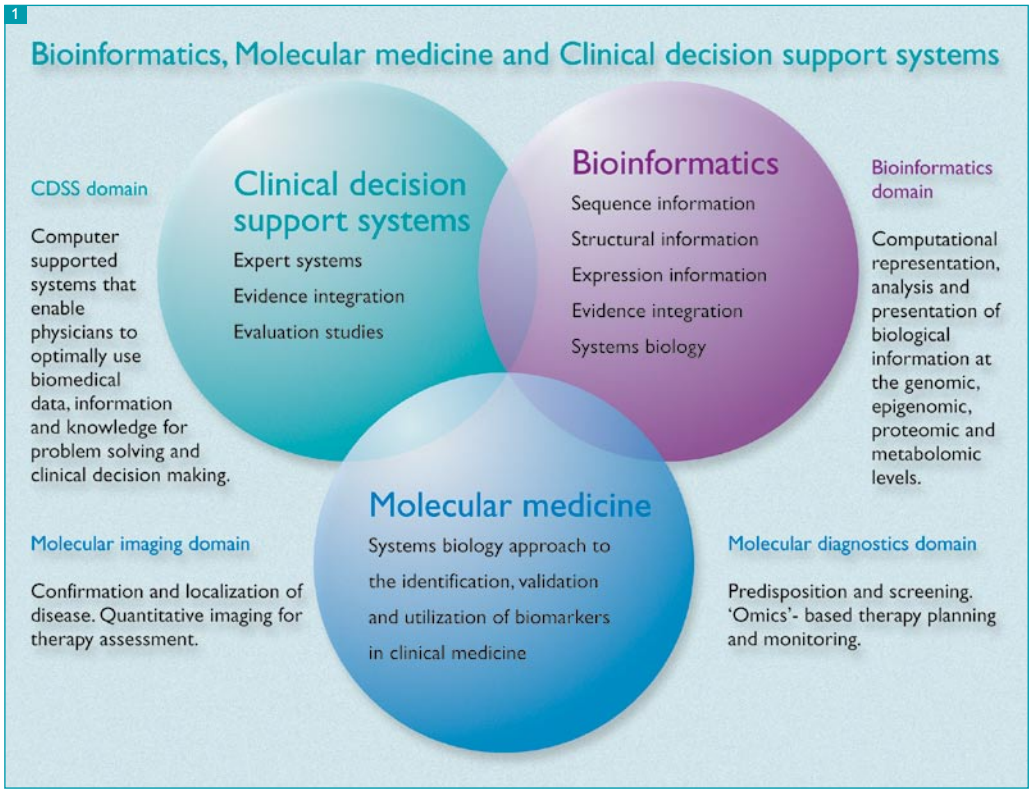

Figure 1: Clinical Bioinformatics as a merger of the classical domains of bioinformatics, molecular medicine and clinical decision support.

have different responses to treatment. The new field of pharmacogenomics is exploring the possibility of tailoring treatment of disease to a patient's underlying genetic makeup [24].

- A patient's genetic makeup may make the patient either more susceptible or relatively resistant to risk factors associated with disease [25].
- A patient's prognosis might differ depending on underlying genetic factors [26,27].

The second opportunity resides in the in-silico modeling of normal biological and disease processes in order to comprehend the underlying physiology. As high-throughput data become available, opportunities will arise for developing computational models of the inflammatory, coagulation, and adaptive immunological systems that can represent pathophysiology specific to various disease processes [28]. In addition, by using experimental data, these models can be refined and analyzed to suggest experiments that could be iteratively implemented in the clinical setting [29]. Essential to the function of in-silico modeling platforms are I/O design, simulation and optimization components. Specifically, I/O tools provide a user-friendly, graphical interface for developing and modifying the models, simulation tools give users a graphical way to fully describe the clinical study and to input datastreams into a computing cluster for distributed simulation, while optimization tools fit and tune the models to the selected data structures and clinical questions of interest.

Since it is often difficult or impossible to experimentally ascertain the values for many of the models' parameters in an in vivo environment, fixing parameters whose values are known and numerically optimizing the remaining parameters to obtain the best possible match to observed experimental results is essential for success of this methodology [30].

Clinical Bioinformatics and Translational Medicine

Colloquially referred to as "bench-to-bedside" medicine, the translation of basic scientific discoveries into usable therapeutic options has become a major challenge within the basic science, clinical research, medical practice, and public health communities for over the past 20 years [31]. At a high level of abstraction, these challenges can be classified as falling into two different "translational blocks", one from basic science discovery to the conduct of clinical research and the other from the derivation of evidence from clinical studies to the widespread adoption of validated therapies in common clinical practice [32]. Current difficulties in bridging these barriers have a significant and negative impact on the ability of clinicians to provide patient care and control healthcare costs [33].

Despite these challenges, one formal research instantiation of this translational approach is the use of biological information, molecular

measurement technology and expert decision solutions in advising patients of therapeutic options through the use of decision matrices. Specifically, in breast cancer management, the evaluation of patients with biopsy-proven, sentinel node negative, ductal carcinoma in-situ (DCIS) is quite difficult [34]. In terms of total tumor mass, DCIS is not usually extensive [35]. However, the tumor grades of individual patients have been associated with significantly different, long-term survival. Based upon the University of Southern California/Van Nuys scoring system (decision matrix), various combinations of tumor gross pathology and histological grade along with patient age result in prognostic scores (Van Nuys Prognostic Index) for which therapeutic recommendations [36] and five year survival are markedly different [37] (See Tables 1 & 2).

Currently, in addition to VNPI scoring for patients with DCIS, screening the individual patient as well as family members for BRCA1/BRCA2 HBC phenotype positivity [38], detecting immunohistochemical findings of cell differentiation with or without with evidence for comedonecrosis, evaluating DNA ploidy and S-phase fractions consistent with high grade lesions, determining hormone receptor positivity for estrogen (ER+) and progesterone (PgR+) [39] as well as epidermal growth factor receptor 2 positivity (HER-2/neu+) [40] play a significant role in influencing therapeutic recommendations and the necessary follow-up interventions.

Finally, it is hoped that the evaluation of the complexity of tumor genetic (DNA) mutation [41] could further refine decision matrices to better inform patients regarding the severity of their disease as well as guide the physicians and patients in terms of therapeutic options and choices. This translational medicine approach that integrates genotypic, phenotypic and pathologic information with best-evidence regarding prior clinical outcomes would create significant, "real-time", new knowledge that would enable oncologists to positively influence patient care [42].

Summary

To quote Elias Zerhouni, MD, Director, National Institutes of Health, we need to "spend money early in (the pre-clinical) life cycle of disease" [16]. In addition, we should "focus on translational research, whereby rapid cycle turnaround of patient-specific, medical research can be incorporated into the clinical environment, rather than traditional clinical research with its lengthy service and product cycle times. Finally, lessons from the world of systems biology must be generalized for all of medical science. (That is) science needs to accept a major shift in thinking, not looking at one molecule over and over again, but looking at systems of molecules." [43]

Merging bioinformatics methods into clinical practice through the use of innovative

Score	1	2	3
Size (mm)	≤ 15	16 – 40	≥ 41
Margin width (mm)	≥ 10	1 – 9	< 1
Pathologic classification	Non-high grade without necrosis (nuclear grades 1 or 2)	Non-high grade with necrosis (nuclear grades 1 or 2)	High grade with or without necrosis (nuclear grade 3)
Age (yr)	> 60	40 – 60	< 40

◀ Table 1. The USC/Van Nuys Prognostic Index scoring system. One to three points are awarded for each of four different predictors of local breast recurrence (size, margin width, pathologic classification, and age). Scores for each of the predictors are totaled to yield a VNPI score ranging from a low of 4 to a high of 12.

Patients	VNPI 4, 5, 6	VNPI 7, 8, 9	VNPI 10, 11, 12
Average age (yr)	57	53	48
Average size (mm)	8.3	18.0	38.2
Average nuclear grade	1.65	2.45	2.89
No. of recurrences	3 (1%)	78 (20%)	38 (50%)
No. invasive recurrences	0 (0%)	34 (44%)	15 (39%)
5 & 10-yr local recurrence-free survival	99%/97%	84%/73%	51%/34%
Breast cancer deaths	0	5	1
5 & 10-yr breast cancer specific survival	100%/100%	100%/98.1%	97.9%/97.9%

◀ Table 2. Tumor characteristics, recurrences, and breast cancer deaths by USC/Van Nuys Prognostic Index Groups.

molecular medicine technologies and 'expert' systems can be facilitated by this translational medicine approach [44]. By demonstrating clinical successes from this "fusion of domains", Philips Research can establish clinical bioinformatics as a discipline capable of answering Dr. Zerhouni's call [45] ■

References

[1] Ledley RS, Lusted LB. *Reasoning Foundations of Medical Diagnosis.* Science 1959; 130: 9-21.

[2] Mikhailov AI, Chernyi AI, Giliarevski RS. *Scientific Communications and Informatics.* Translated by Burger RH, Arlington, VA, 1984; from Nauchnye kommunikatsii i infomatika, Moscow: Nauka, 1976.

[3] Greenes RA, Buchanan BG, Ellison D. *Presentation of the 2006 Morris F. Collen Award to Edward H. (Ted) Shortliffe.* Journal of the American Medical Informatics Association. 2007; 14: 376-385.

[4] Shortliffe EH, Blois MS. *The Computer meets Medicine and Biology: Emergence of a Discipline. In EH Shortliffe & JJ Cimino (eds) Biomedical Informatics: Computer Applications in Healthcare and Biomedicine* (3rd ed.), p.24. New York: Springer-Verlag, 2006.

[5] http://en.wikipedia.org/wiki/Medical_informatics; (Last accessed June 1st 2007).

[6] *The Study of Biological Systems through the Development and Application of Data-Analytical and Theoretical Methods, Mathematical Modeling and Computational Simulation Techniques.* BISTIC Definition Committee of the National Institutes of Mental Health; Bioinformatics Definition Committee (Michael Huerta, MD, Chair). NIH Working Definition of Bioinformatics and Computational Biology. July 17th 2000.

[7] *The Study of Methods for Generating, Manipulating, Managing and Integrating Images into Biomedical Applications.* Kulikowski CA. *Medical Imaging Informatics: Challenges of Definition and Integration.* Journal of the American Medical Informatics Association 1997; 4,3: 252-253.

[8] *The Combined Application of Computer Science, Information Theory and Clinical Medicine to Assist in the Management and Processing of Clinical Data, Information and Knowledge to the Support of Medical Practice.* Doolan DF, Bates DW, James BC. *The Use of Computers for Clinical Care: A Case Series of Advanced U.S. Sites.* Journal of the American Medical Informatics Association 2003;10,1: 94-107.

[9] *The Systematic Application of Information and Computer Science and Technology to Public Health Practice, Research and Learning.* Yasnoff WA et al. *Public Health Informatics: Improving and Transforming Public Health in the Information Age.* Journal of Public Health Management & Practice 2000; 6,6: 67-75.

[10] Augen J. *Bioinformatics in the Post-Genomic Era: Genome, Transcriptome, Proteome and Information-based Medicine.* Boston: Addison-Wesley 2005.

[11] Ostell JM et al. The NCBI Data Model. In Baxevanis AD & Francis-Ouellette BF (eds) *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins* (2nd ed.), New York: Wiley-Interscience 2001.

[12] http://en.wikipedia.org/wiki/Bioinformatics (Last accessed June 2nd 2007).

[13] Maojo V, Kulikowski CA. *Bioinformatics and Medical Informatics: Collaborations on the road to Genomic Medicine?* Journal of the American Medical Informatics Association 2003; 10: 515-522.

[14] Finan JE, Zhao RY. *From Molecular Diagnostics to Personalized Testing.* Pharmacogenomics 2007; 8 (Jan 1): 85-99.

[15] Garg AX et al. *Effects of Computerized Clinical Decision Support Systems in Practitioner Performance and Patient Outcomes: A Systematic Review.* Journal of the American Medical Informatics Association 2005; 293: 1223-1238.

[16] *"Create the right databases and the algorithms that can ask the right questions to mine the data"* - Elias Zerhouni, MD, Director, National Institutes of Health in: Culliton B. *Extracting Knowledge from Science: A Conversation with Elias Zerhouni.* Health Affairs 2006; 25: w94-w103 (published online 9 March 2006; 10.1377/hlthaff.25.w94).

[17] Shortliffe EH, Blois MS. *The Computer meets Medicine and Biology: Emergence of a Discipline.* In EH Shortliffe & JJ Cimino (eds) *Biomedical Informatics: Computer Applications in Healthcare and Biomedicine* (3rd ed.), p.37. New York: Springer-Verlag 2006.

[18] Kohane I. *Bioinformatics and Clinical Informatics: The Imperative to Collaborate.* Journal of the American Medical Informatics Association 2000; 7: 439-443.

[19] Miller PL. *Opportunities at the Intersection of Bioinformatics and Health Informatics: A Case Study.* Journal of the American Medical Informatics Association 2000;7:431-437.

[20] Nadkarni PM, Reeder ST, Zhou J. *CECIL: A Database for Storing and Retrieving Clinical and Molecular Information on Patients with Alport Syndrome.* Proceedings of the 17th Symposium on Computer Applications in Medical Care 1993: 649-53.

[21] National Organization for Rare Diseases; http://www.rarediseases.org/ (Last accessed June 3rd 2007).

[22] Office of Rare Diseases; http://rarediseases.info.nih.gov/ (Last accessed June 3rd 2007).

[23] Quackenbush J. *Current Concepts: Microarray Analysis and Tumor Classification.* N Engl J Med 2006; 345: 2463-2472.

[24] Hicks et al. *Novel Patterns of Genomic Rearrangement and their Association with Survival in Breast Cancer.* Genome research 2006;16:1465-1479.

[25] Hedenfalk I, et al. *Molecular Classification of Familial Non-BRCA1/BRCA2 Breast Cancer.* Proceedings of the National Academy of Sciences 2003; 100,5: 2532-2537.

[26] Van't Veer LJ et al. *Gene Expression Profiling Predicts Clinical Outcome of Breast Cancer.* Nature 2002; 415: 530-536.

[27] Van de Vijver MJ et al. *A Gene-Expression Signature as a Predictor of Survival in Breast Cancer.* N Engl J Med 2002; 347,25: 1999-2009.

[28] Vodovotz Y, Clermont G, Chow CC, An G. *Mathematical Models of the Acute Inflammatory Response.* Current Opinions in Critical Care 2004; 10: 383-390.

[29] Kumar R, Clermont G, Vodovotz, Chow C. *The Dynamics of Inflammation.* Journal of Theoretical Biology 2004; 230:145–155.

[30] Clermont G et al. *In Silico Design Of Clinical Trials: A Method Coming Of Age.* Critical Care Medicine 2004; 32: 2061-2070.

[31] Bloom FE. *Science as a Way of Life: Perplexities of a Physician-Scientist.* Science 2003; 300(5626):1680-1685.

[32] Sung NS et al. *Central Challenges Facing the National Clinical Research Enterprise.* Journal of the American Medical Association 2003; 289,10:1278-1287.

[33] McGlynn EA et al. *The Quality of Health Care Delivered to Adults in the United States.* N Engl J Med 2003; 348,26: 2635-2645.

[34] Bland & Copeland. *The Breast: Comprehensive Management of Benign and Malignant Disorders.* 3rd edition. Vols. 1 & 2. W.B. Saunders 2004.

[35] Harris JR et al. *The Breast: Diagnosis, Imaging, Management and Pathology.* Lippincott, Williams & Wilkins 1999.

[36] *"The therapeutic options of lumpectomy, modified radical mastectomy, radiation therapy, adjuvant chemotherapy (hormonal therapy), and/or prophylactic contralateral breast irradiation or mastectomy must be weighed carefully against the best-evidence for balancing long-term survival, relative acceptable morbidity, and the risk of adverse events".*

[37] Silverstein MJ. *The University of Southern California/Van Nuys Prognostic Index for Ductal Carcinoma in situ of the Breast.* The American Journal of Surgery 2003;186:337–343.

[38] BRCA1; http://www.cancerindex.org/geneweb/BRCA1.htm (Last accessed June 12th 2007).

[39] Neven P, Huang HJ, Vanspauwen R et al. *The Prognostic and Predictive Value of the Progesterone Receptor in Women with an Oestrogen Receptor Positive Breast Cancer.* European Journal of Cancer 2004; 2: 46-48, (suppl; abstr 16).

[40] Olayioye MA. *Update on HER-2 as a Target for Cancer Therapy: Intracellular Signaling Pathways ErbB2/HER-2 and Family Members.* Breast Cancer Research 2001; 3,6: 385-389.

[41] Sebat J et al. *Large-scale copy number polymorphisms in the human genome.* Science 2004; 305, 5683: 525-528.

[42] Translational Research; http://nihroadmap.nih.gov/clinicalresearch/overview-translational.asp (Last accessed June12th 2007).

[43] Pollard TD. *The Future of Biomedical Research: From the Inventory of Genes to Understanding Physiology and the Molecular Basis of Disease.* Journal of the American Medical Association 2002; 287,13: 1725-1727.

[44] Payne P et al. *Breaking the Translational Barriers: The Value of Integrating Biomedical Informatics and Translational Research.* Journal of Investigative Medicine 2005; 53,: 192-200.

[45] Stead W. *The Challenge of Bridging between Disciplines* (Editorial). Journal of the American Medical Informatics Association 2001; 8: 105.