

Guest Editorial

Bio*Medical informatics and genomic medicine: Research and training

The last half-century has seen tremendous progress in computing, genetics, and clinical care. In computing we have gone from the ENIAC in 1946 [1] to the explosion of the World Wide Web in 1996 to the Internet becoming part of the fabric of society today. In genetics we have gone from the elucidation of DNA as the mechanism of inheritance in 1953 [2] to the sequencing of the human genome completed in 2003 [3]. In clinical care we have gone from the first published randomized clinical trial in 1948 [4] to ever-increasing adoption of evidence based medicine [5]. The convergence of these domains brings with it the promise of genomic medicine [6] with personalized, preventive, and predictive healthcare [7].

Biomedical informatics has a long tradition of being involved with key aspects of this convergence. Bioinformatics (computational biology in particular) has been fundamental in applying developments in the computing and information sciences to helping manage and make sense of vast quantities of genomic information from sequencing projects. Clinical informatics has similarly long been applying (and advancing) the fields of computing and information sciences to help care providers manage more traditional medical knowledge and patient data. The informatics community has long been aware of the potential for information systems to help improve the quality of care, but there has been a lag between the recognition of this by our community and the broader recognition of this by the healthcare community. Reports of benefits related to electronic medical record adoption appeared as early as the 1960s [8–11] with an IOM report touting the electronic medical record in 1991 [12], but only relatively recently with the IOM report on Crossing the Quality Chasm, the Leapfrog Group's efforts, and federal programs to promote a National Health Information Infrastructure has this awareness become mainstream. The experience of our community in clinical informatics in terms of both research and system adoption will be vital as we move to bridge the gap between clinical and biological information systems to help enable the promise of genomic medicine.

As an example of the need to bridge the gap, a key piece of genomic medicine will be testing individuals for their genetic predisposition to disease and then using this predic-

tive information to provide personalized preventive and/or therapeutic care. Providers today already struggle with managing standard healthcare information, which is relatively simple compared to larger scale genetic testing information. Focusing on single-gene testing, we see healthcare providers today having to deal with tests for 1253 diseases (as of April 2006), with more being developed every month [13,14]. As testing covers more and more diseases that are managed by nongeneticists, it is becoming an issue for other specialists and for primary care providers as well. Larger scale testing such as genome-wide SNP, gene expression, and protein expression testing will be even more complex to interpret and apply. New ways of thinking about clinical decision-support tools that use molecular information and that look at both the clinical and biological aspects of a genomic approach to health care are needed.

Professional societies such as the American Medical Informatics Association are poised to play a key role in supporting research that will enable genomic medicine. It is also critical that such organizations facilitate discussions on how best to train the next generation of scientists, engineers, and healthcare professionals in translational bioinformatics. In recognition of the need to bridge bioinformatics and clinical informatics, the theme of the 2002 AMIA Annual Symposium was “Bio*Medical Informatics: One Discipline.” Building on the ideas presented at this meeting, the AMIA Genomics working group leadership led a 2003 effort to focus activities in this area. The mission of the AMIA Genomics working group is, “To focus on opportunities in biomedical informatics that arise from the storage, retrieval, analysis, and dissemination of molecular information in a clinical setting” (see <http://www.amia.org/mbrcenter/wg/gen/>). Within this mission specific opportunities for biomedical informatics research were identified by members of the working group, including “(a) unifying clinical and molecular databases, (b) connecting molecular information currently collected in research studies (e.g., microarray data) with information currently located in patient health records, (c) developing the electronic medical record of the future, in which molecular information will be fully integrated, (d) linking clinical

trial and drug discovery information with clinical/molecular databases, (e) supporting the development of benchmark clinical/molecular datasets, (f) developing clinical decision-support tools utilizing molecular information, (g) visualizing and modeling the molecular basis of disease.”

The theme of this issue, the mission of the AMIA’s Genomics Working Group, and the recently announced NIH Roadmap Initiative titled “Re-engineering the Clinical Research Enterprise” are focused on similar goals: incorporating modern information technology into clinical research, improving the integration of translational and clinical research, improving the training and coordination of a united clinical and informatics workforce, and supporting key components of the translational research infrastructure that is necessary to store, retrieve, analyze, and disseminate molecular information in a clinically useful manner.

The promise and challenges of genomic medicine thus provide us with the broader context of this special issue, which focuses on both informatics research in this area and the ways in which informatics training programs are responding to the challenge of educating the next generation of biomedical informatics researchers who can span the clinical and biological worlds. As we stated in the Call for Papers for this special issue, “The unification of clinical and molecular databases poses a complex and urgent challenge. The *Journal of Biomedical Informatics* will devote a special issue to papers on the state-of-the-art in research and education related to this timely topic. The goal of this special issue is to help develop a body of literature spanning the traditional clinical informatics and bioinformatics research arenas and to share experiences with different approaches to educating students to work in industry and academics in this important area.”

The following is an overview of the papers in this special issue and their organization. We start with a broad overview of some of the challenges in unifying clinical and molecular databases with a methodological review focused on data integration and genomic medicine [15]. In the second section we collect a range of original work, arranged roughly from broader to narrower, bridging the clinical and molecular information worlds starting with a general system for integrating diverse genomic and phenotypic (disease) databases [16] and concluding with an in-depth look at clinical gene sequencing data reporting [17]. In the third section we collect four papers that are case studies on approaches to training again starting broadly looking at training for biomedical informatics in the 21st century [18] and concluding with a look at an interdepartmental Ph.D. program in computational biology and bioinformatics [19].

The methodologic review by Louie et al. [15] introduces some of the opportunities and challenges presented by genomic medicine in the context of data integration. The paper provides a brief review of genomic medicine and of data integration systems in general. The authors then illustrate specific data integration challenges in the context of

genomic medicine and review the application of data integration concepts and approaches to genomic medicine, concluding with a summary of gaps remaining.

The first of the papers presenting original work is that by Alonso-Calvo et al. [16], who describe a specific system for integrating data in the context of genomic medicine, pulling together data from public genotypic and phenotypic databases. They describe the OntoFusion agent and ontology-based data integration system, using a case study to illustrate the system. This paper provides a concrete example of many of the issues related to data integration and genomic medicine introduced in the methodologic review and provides examples of why vocabularies and ontologies play such a critical role in data integration.

An example of a particular ontology of importance in the context of genomic medicine is provided by Sioutos et al. [20], who describe the current state of the National Cancer Institute (NCI) Thesaurus. As the name implies, the NCI Thesaurus was originally conceived as a controlled vocabulary. However, in response to user needs to integrate molecular and clinical cancer-related information, the NCI Thesaurus has incorporated a deductive logic framework to model the relationships between key concepts. Thus, today the NCI Thesaurus is a hybrid of a controlled terminology and an ontology. Sioutos et al. discuss examples of the shift toward ontological knowledge representation in the NCI Thesaurus, with an emphasis on the integration of molecular information in defining cancer concepts.

The article by Hoffman [21] discusses the challenges and opportunities in integrating molecular genomic information in the electronic medical record. Unique aspects of genomic results are summarized, such as their life-long value, emphasizing that they need to be stored to enable re-analysis as new knowledge and analysis methods become available. Genomic data are also unusual in that there is a clear potential to affect the decision making of family members if the EMR were to include structured family history information. Hoffman identifies three key developments needed for a genome-enabled EMR: (1) improved tools to support the capture of genomic results, (2) controlled vocabulary for the description of clinically significant genomic findings, and (3) decision-support applications to assist clinicians in using genomic results in patient care.

Mitchell and Mitchell’s article [17] points out the necessity for standardized electronic reporting of all the data for sequenced human genes associated with disease, as well as for comparator reference data. They address a key question of broad interest—could current sequence reporting methods cause data and information loss in the future? They conclude that laboratories should report all data generated (i.e., all bases sequenced) and all reference data (or identify their sources explicitly), not just data about positional nucleotide variations and interpretation of variations. They point out that genetic information is a dynamic body of knowledge that requires completeness to allow its linkage

to new biological information as it becomes known. Limited reporting of sequence variants, although meaningful today, may be irrelevant tomorrow. For this reason, clinical information systems must be able to store, retrieve, analyze, and disseminate complete DNA sequence data about the genes associated with human disease.

The next section consists of four invited papers that are case studies of the experiences of a range of biomedical informatics training programs. In the first of these training program papers, Altman and Klein [18] describe broadly the evolution of the Stanford biomedical informatics training program from its inception in the 1980s as the Medical Information Science program focused on clinical informatics (emphasizing decision support and artificial intelligence) to its current structure as the Biomedical Informatics program, which is a general program of biomedical informatics training with areas of emphasis including clinical informatics, bioinformatics, and imaging informatics. The authors present a range of training including both professional and research M.S. programs, a Ph.D. program, and certificate programs. They look forward to emerging trends in informatics research and training nationally, and at Stanford, with a focus on the challenges and opportunities they represent for a general biomedical informatics training program.

Johnson and Friedman [22] describe the biomedical informatics training offered at Columbia by the Department of Biomedical Informatics, including a comparison to the training offered by the Columbia Center for Computational Biology. The focus is on the challenges particular to the bridging of differences in culture between the disciplines of biological and clinical informatics and the impact this has had at Columbia on the evolution of training, with an emphasis on the opportunities and challenges posed by aspects of genomic medicine. The paper concludes with recommendations on other programs that are seeking to develop general biomedical informatics training programs spanning the biological and clinical cultures.

Kane and Brewer [23] summarize a component of a campus-wide graduate-level interdisciplinary program in biomedical informatics at Purdue University (the “Computational Life Sciences” program), which is a more focused program than the Stanford and Columbia programs. The Purdue program involves numerous departments and has the goal of encouraging students to develop skills outside their formal discipline. The paper describes a two-course sequence focused on the subdiscipline of bioinformatics that was designed for information technology students. A key feature of these courses is that they do not have any prerequisite life science courses. The first course introduces how information flows through a cell system and the second course applies skills in systems analysis and design to the biomedical informatics domain. The overarching goal of this training program is to support team-based bioinformatics system development in which information system specialists conduct requirements discovery.

Gerstein et al. [19] address more focused topics: a definition of computational biology and bioinformatics (CCB), why there is a national need for CCB, why Yale University chose to develop an interdepartmental program to address this need, what general concepts must be covered within any CBB program, and the structure of the CCB curriculum at Yale University (including examples of core courses and tailored programs, depending on the background of the student—i.e., biologist versus computer scientist). This thorough perspective and program overview provides a ready understanding of the key elements required to establish an outstanding training program in CCB.

Acknowledgment

We are grateful to Ms. Donna Knight for her excellent communication between the editors and the authors.

References

- [1] Goldstine H, Goldstine A. The electronic numerical integrator and computer (ENIAC) 1946. In: *The origins of digital computers: selected papers*. New York: Springer-Verlag; 1982. p. 359–73.
- [2] Watson J, Crick F. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 1953;171(4356):737–8.
- [3] Collins F, Morgan M, Patrinos A. The human genome project: lessons from large-scale biology. *Science* 2003;300(5617):286–90.
- [4] Medical Research Council Streptomycin in Tuberculosis Trials Committee. Streptomycin treatment for pulmonary tuberculosis. *Br Med J* 1948;ii:769–782.
- [5] Straus S, Richardson W, Glasziou P, Haynes R. Evidence-based medicine: how to practice and teach. EBM: Elsevier Churchill Livingstone; 2005.
- [6] Guttmacher AE, Collins FS. Genomic medicine—a primer. *N Engl J Med* 2002;347(19):1512–20.
- [7] Weston AD, Hood L. Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. *J Proteome Res* 2004;3(2):179–96.
- [8] Schenthal JE, Sweeney JW, Nettleton Jr W. Clinical application of large-scale electronic data processing apparatus. I. New concepts in clinical use of the electronic digital computer. *Jama* 1960;173:6–11.
- [9] Schenthal JE, Sweeney JW, Nettleton Jr W. Clinical application of electronic data processing apparatus. II. New methodology in clinical record storage. *Jama* 1961;178:267–70.
- [10] Schenthal JE. The electronic medical record. *Bull Sch Med Univ Md* 1962;47:53–5.
- [11] Greenes RA, Pappalardo AN, Marble CW, Barnett GO. Design and implementation of a clinical data management system. *Comput Biomed Res* 1969;2(5):469–85.
- [12] Institute of Medicine. *The computer-based patient record: an essential technology for health care*. Washington DC: National Academy Press; 1991 [Revised 1997].
- [13] GeneTests. Medical genetics information resource; 2006 [cited 2006 April]. Available from: www.genetests.org.
- [14] Pagon RA, Tarczy-Hornoch P, Baskin PK, Edwards JE, Covington ML, Espeseth M, et al. GeneTests-GeneClinics: genetic testing information for a growing audience. *Hum Mutat* 2002;19(5):501–9.
- [15] Louie B, Mork P, Martin-Sanchez F, Halevy A, Tarczy-Hornoch P. Data integration and genomic medicine. *J Biomed Inform* 2007;40(1):5–16.
- [16] Alonso-Calvo R, Maojo V, Billhardt H, Martin-Sanchez F, Garcia-Remesal M, David Perez-Rey D. An agent and ontology-based

- system for integrating public gene, protein and disease databases. *J Biomed Inform* 2007;40(1):17–29.
- [17] Mitchell D, Mitchell J. Status of clinical gene sequencing data reporting and associated risks for information loss. *J Biomed Inform* 2007;40(1):47–54.
- [18] Altman R, Klein T. Biomedical informatics training at stanford in the 21st century. *J Biomed Inform* 2007;40(1):55–8.
- [19] Gerstein M, Greenbaum D, Cheung K, Miller P. An interdepartmental Ph.D. program in computational biology and bioinformatics: the Yale perspective. *J Biomed Inform* 2007;40(1):73–9.
- [20] Sioutos N, de Coronado S, Haber M, Hartel F, Shaiu W, Wright L. NCI Thesaurus: a semantic model integrating cancer-related clinical and molecular information. *J Biomed Inform* 2007;40(1):30–43.
- [21] Hoffman M. The genome enabled electronic medical record. *J Biomed Inform* 2007;40(1):40–4.
- [22] Johnson S, Friedman R. Bridging the gap between biological and clinical informatics in a graduate training program. *J Biomed Inform* 2007;40(1):59–66.
- [23] Kane M, Brewer J. An information technology emphasis in biomedical informatics education. *J Biomed Inform* 2007;40(1):67–72.

Guest Editor

Peter Tarczy-Hornoch*

*Division of Biomedical and Health Informatics,
University of Washington, USA*

E-mail address: pth@u.washington.edu

Associate Guest Editor

Mia K. Markey

*The University of Texas,
Department of Biomedical Engineering, USA*

Associate Guest Editor

John A. Smith

*Department of Pathology,
University of Alabama at Birmingham, USA*

Associate Guest Editor

Tadaaki Hiruki

*Department of Pathology,
Queen's University at Kingston, Canada*

Available online 28 October 2006

* Corresponding author.