The Promise of Systems Biology in Clinical Applications

Findings from the Yearbook 2008 Section on Bioinformatics

Y. L. Yip, Managing Editor for the IMIA Yearbook Section on Bioinformatics
Swiss-Prot group, Swiss Institute of Bioinformatics, Geneva, Switzerland
University of Geneva, Dept. of Structural Biology and Bioinformatics, Geneva, Switzerland

Summary

Objectives:: To summarize current excellent research in the field of bioinformatics.

Method: Synopsis of the articles selected for the IMIA Yearbook 2008.

Results: Current research in the field of Bioinformatics shows that the emergent field of systems biology is starting to offer innovative solutions to clinically-relevant problems. The approach used can be top-down, where models are created based on hypotheses to describe previously unexplained phenomena and then tested against experimental or clinical evidence. It can also be bottom-up, where mathematical models are built by harnessing existing information about the components (e.g. protein entities, interaction networks) of a system in order to discern critical system-level mechanisms that can be relevant for clinical applications. Progress in this area is aided by the ongoing development in data integration and management, whose current focus is on better semantics for facilitating translational research. Advances in other important areas, such as microarray technology, text mining and ontologies, are also noted.

Conclusions: The best paper selection of articles on bioinformatics gives examples of original research that exploits mathematical modeling to tackle medical problems and of improved semantic solutions for data integration. As new directions are explored and the technologies mature, these approaches are expected to be increasingly integrated into clinical practice.

Keywords

International Medical Informatics Association, IMIA Yearbook, bioinformatics, systems biology

Geissbuhler A, Kulikowski C, editors. IMIA Yearbook of Medical Informatics 2008. Methods Inf Med 2008; 47 Suppl 1:102-4

Introduction

Systems biology is an emerging field that aims at system-level understanding of biological processes. While systems biology is in its infancy, it is an area that promises to come into the main stream of biological sciences in this century. In a broad sense, systems biology involves different areas such as integrative systems biology, which seeks to integrate different types of information to advance our understanding, and dynamic systems biology, which aims to uncover how biological systems evolve over time. Many aspects of mathematical modeling, computer science and informatics are exploited in systems biology. These include the development of computational or mechanistic models, and new approaches to data integration and interoperability. Examples of direct application of systems biology in clinically relevant situations are still scarce. However, a survey of current literature suggests that the field has started to show its potential impact in the clinical domain. References [1-3, 15-16] demonstrate that mechanistic models can prove helpful in identifying drug targets, understanding the physiology of diseases, or establishing new hypotheses for pathogenesis or disease progression. There have been also developments of data standards and representation models for better managing the increasing amount of biomedical data [4, 17-18]. These activities have as their primary aims to facilitate data analysis and to accelerate translational research, i.e. the process of bringing the results of basic science research into clinical practice. Recently, more focus has been placed on data semantics, whether it is inherent to data models or related to the use of Semantic Web technology.

Beyond these new trends, the current literature shows ongoing progress in areas such as microarray experiments [5-8], text-mining and ontologies [9-10]. The quality of the data and the derived results from microarray experiments continue to improve as algorithms are compared and primary results are combined or re-analyzed [5-6]. New study designs are also being proposed to better exploit clinical data [7-8].

Best Paper Selection

The best paper selection of articles for the section on Bioinformatics in the IMIA Yearbook 2008 reflects these trends and follows the tradition of previous yearbooks [11-14] in presenting examples of excellent research. Four outstanding articles representing the research in the areas mentioned above were selected from international peer-reviewed journals in the fields of medicine, medical informatics, and bioinformatics. Table 1 presents the selected papers. A brief content summary of the selected best papers can be found in the appendix of this report.

Table 1 Best paper selection of articles for the IMIA Yearbook of Medical Informatics 2008 in the section 'Bioinformatics'. The articles are listed in alphabetical order of the first author's surname.

Section
Bioinformatics
▪ Kaplan S, Itzkovitz S, Shapiro E. A universal mechanism ties genotype to phenotype in trinucleotide diseases. PLoS Computational Biology 2007;3(11):e235.
▪ Luan D, Zai M, Varner JD. Computationally derived points of fragility of a human cascade are consistent with current therapeutic strategies. PLoS Computational Biology 2007;3(7):e142.
▪ Ruttenberg A, Clark T, Bug W, Samwald M, Bodenreider O, Chen H, Doherty D, Forsberg K, Gao Y, Kashyap V, Kinoshita J, Luciano J, Marshall MS, Ogbuji C, Rees J, Stephens S, Wong GT, Wu E, Zaccagnini D, Hongsermeier T, Neumann E, Herman I, Cheung KH. Advancing translational research with the Semantic Web. BMC Bioinformatics 2007;8(Suppl3):S2.
▪ Sioutos N, de Coronado S, Haber MW, Hartel FW, Shaiu WL, Wright LW. NCI Thesaurus: A semantic model integrating cancer-related clinical and molecular information. J Biomed Inform 2007;40:30.

Conclusions and Outlook

The best paper selection for the Yearbook section on Bioinformatics highlights the rapid developments in the field of systems biology and the growing importance of semantic data representation. One can see an increasingly important role for systems biology in healthcare from these early developments. It will probably not be long before the modeling of human diseases becomes a norm rather than an exception. This system-centered view of diseases will bring out new understanding of the physiological mechanisms as well as new therapeutic options that aim to target the disease in the context of the whole organism. Meanwhile, it is foreseeable that new avenues and application areas will be explored as mathematicians, engineers, computer scientists, biologists and clinicians increasingly interact. Efforts to accelerate translational research via more efficient data and knowledge integration strategies can also be foreseen. Development will not only be purely technical, but communities of scientists and practitioners will be called upon to actively participate in adopting standards and new technologies such as the Semantic Web. Up-to-date information about current and future issues of the IMIA Yearbook is available at http://www.schattauer.de/ index.php?id=1384

Acknowledgement

I would like to acknowledge the support of Martina Hutter and the reviewers in the selection process of the IMIA Yearbook.

References

1. Beerenwinkel N, Antal T, Dingli D, Traulsen A, Kinzler KW, Velculescu VE, et al. Genetic progression and the waiting time to cancer. PLoS Comput Biol 2007;3:e225.
2. Filipe JAN, Riley EM, Drakeley CJ, Sutherland CJ, Ghani AC. Determination of the processes driving the acquisition of immunity to malaria using a mathematical transmission model. PLoS Comput Biol 2007;3:e255.
3. Chan C, Kepler TB. Computational immunology-From bench to virtual reality. Ann Acad Med Singapore 2007;36:123-7.
4. Maojo V, Crespo J, de la Calle G, Barreiro J, Garcia-Remesal M. Using Web services for linking genomic data to medical information systems. Methods Inf Med 2007;46:484-92.
5. Baross A, Delaney AD, Li HI, Nayar T, Flibotte S, Qian H, et al. Assessment of algorithms for high throughput detection of genomic copy number variation in oligonucleotide microarray data. BMC Bioinformatics 2007;8:368.
6. English SB, Butte AJ. Evaluation and integration of 49 genome-wide experiments and the prediction of previously unknown obesity-related genes. Bioinformatics 2007;23:2910-17.
7. Nicolau M, Tibshirani, Borresen-Dale AL, Jeggrey SS. Disease-specific genomic analysis: identifying the signature of pathologic biology. Bioinformatics 2007:23:957-65.
8. Sun Y, Goodison S, Li J, Liu L, Farmerie W. Improved breast cancer prognosis through the combination of clinical and genetic markers. Bioinformatics 2007;23:30-7.
9. Van Vooren S, Thienpont B, Menten B, Speleman F, De Moor B, Vermeesch J, et al. Mapping biomedical concepts onto the human genome by mining literature on chromosomal aberrations. Nucleic Acid Research 2007;35:2533-43.
10. Hoehndorf R, Loebe F, Keiso J, Herre H. Representing default knowledge in biomedical ontologies: application to the integration of anatomy and phenotype ontologies. BMC Bioinformatics 2007; 8:377.
11. Knaup P, Ammenwerth E, Brandner R, Brigl B, Fischer G, Garde S, et al. Towards clinical bioinformatics: advancing genomic medicine with informatics methods and tools. Methods Inf Med 2004;43: 302-7.
12. Bott OJ, Ammenwerth E, Brigl B, Knaup P, Lang E, Pilgram R, et al. The challenge of ubiquitous computing in health care: technology, concepts and solutions. Findings from the IMIA Yearbook of Medical Informatics 2005. Methods Inf Med 2005; 44:473-9
13. Lang E. Bioinformatics and its Impact on Clinical Research Methods. Methods Inf Med 2006;45:104-6
14. Lang E. Integrating bioinformatics into clinical practice: progress and evaluation. Methods Inf Med 2007;46:106-8.
15. Kaplan S, Itzkovitz S, Shapiro E. A universal mechanism ties genotype to phenotype in trinucleotide diseases. PLoS Comput Biol 2007;3:e235.
16. Luan D, Zai M, Varner JD. Computationally derived points of fragility of a human cascade are consistent with current therapeutic strategies. PLoS Comput Biol 2007;3:e142.
17. Ruttenberg A, Clark T, Bug W, Samwald M, Bodenreider O, Chen H, et al. Advancing translational research with semantic web. BMC Bioinformatics 2007;8:S2.
18. Sioutos N, de Coronado S, Haber MW, Hartel FW, Shaiu WL, Wright LW. NCI Thesaurus: a semantic model integrating cancer-related clinical and molecular information. J Biomed Inform 2007;40:30-43.

Correspondence to:
Dr. Yum Lina Yip
Swiss-Prot group
Swiss Institute of Bioinformatics
1 Rue Michel-Servet
CH1211 Geneva, Switzerland
Tel: +41 22 379 5049, Fax: +41 22 379 5858
E-mail: lina.yip@isb-sib.ch

Appendix: Content Summaries of Selected Best Papers for the IMIA Yearbook 2008, Section Bioinformatics*

Kaplan S, Itzkovitz S. Shapiro E
A Universal Mechanism Ties Genotype to

* The complete papers can be accessed in the Yearbook's full electronic version, provided that permission has been granted by the copyright holder(s)

Phenotype in Trinucleotide Diseases

PLoS Computational Biology 2007;3(11):e235

The underlying mechanism of trinucleotide diseases, a family of hereditary disorders characterized by an abnormally large number of DNA trinucleotide repeats in a gene, remains unknown despite a decade of research. Indeed, while the diseases affect genes that are functionally and structurally unrelated, they share strikingly similar phenotypes including the correlation between the number of inherited repeats and the age of onset and disease progression. Using computational modeling and simulations, Kaplan et al. proposed a universal mechanism of length-dependent somatic mutations to explain the genotype-phenotype correlations common to all trinucleotide diseases. Their model concurs with both clinical and experimental data. Moreover, the model is capable to provide an explanation to previously unexplained onset-related phenomena. Their findings also highlight the potential of a common approach to treatment for these diseases.

Luan D, Zai M, Varner JD

Computationally Derived Points of Fragility of a Human Cascade Are Consistent with Current Therapeutic Strategies

PLoS Computational Biology 2007;3(7):e142

Mechanistic mathematical modeling has not yet proved their utility in molecular medicine and clinical development. This is mainly due to model uncertainty which is widely perceived as the limiting factor. Luan et al. explored the hypothesis that mechanistic models of human relevant cascades can be computationally screened for points of fragility despite model uncertainty. For this, they developed a mechanistic model composed of 92 proteins and 148 protein-protein interactions of the TF-FVIIa initiated coagulation cascade from literature sources. The model was validated using 21 published datasets generated from two different quiescent in vitro coagulation models. Luan et al. then used Monte Carlo sensitivity analysis to computationally screen for sensitivity mechanisms in the presence and absence of natural anticoagulants. They found that in the absence of anticoagulants, fluid and surface phase factor X/activated factor X (fX/FXa) activity and thrombin-mediated platelet activation were fragile, while those of fIX/FIXa and fVIII/FVIIIa were robust. Their results were shown to be consistent with experimental findings. Indeed, the predicted fragile mechanisms are molecular targets in current anticoagulation clinical therapies, preclinical development, and clinical trials. The authors concluded that despite model uncertainty, mechanistic models could be used to pinpoint critical elements in complex cascades and help uncover therapeutic targets.

Ruttenberg A, Clark T, Bug W, Samwald M, Bodenreider O, Chen H, Doherty D, Forsberg K, Gao Y, Kashyap V, Kinoshita J, Luciano J, Marshall MS, Ogbuji C, Rees J, Stephens S, Wong GT, Wu E, Zaccagnini D, Hongsermeier T, Neumann E, Herman I, Cheung KH

Advancing Translational Research with the Semantic Web

BMC Bioinformatics 2007;8(Suppl3):S2

The Semantic Web represents a promising information environment to facilitate the synthesis of interdisciplinary knowledge required for the advancement of translational research. Ruttenberg et al. described the structure and the activities of Semantic Web Health Care and Life Sciences Interest Group (HCLSIG), which had the present role to explore the use of Semantic Web technologies in working with biomedical knowledge, and to address the issues arisen from it. The authors presented a bench-to-bedside use case scenario of a clinical researcher attempting to develop immunotherapies for Alzeihmer's Disease (AD). Their results showed that Semantic Web technologies offer both promise and challenges. While existing standards, technologies and tools are already adequate to be applied to a range of bench-to-bedside use cases, the technology is still young and thus presents gaps in standards and implementations. Issues such as the scarcity of semantically annotated information sources, performance and scalability, representation of evidence and data provenance, and lack of a standard rule language have all affected the work of the HCLSIG. Ruttenberg et al. have further identified areas in which coordination of efforts is needed to advance translational research. Finally, the authors indicate that incentives to encourage appropriate representation of scientific and clinical data on the Web as well as adequate education and training of skilled workers are both important for the widespread adoption of the Semantic Web in health care and the life sciences.

Sioutos N, de Coronado S, Haber MW, Hartel FW, Shaiu WL, Wright LW

NCI Thesaurus: A Semantic Model Integrating Cancer-related Clinical and Molecular Information

J Biomed Inform 2007;40:30

The understanding and effective treatment of cancer can only be improved through the integration of information across the molecular and clinical levels. Sioutos et al. present the NCI Thesaurus, a semantic model which serves as a foundational layer for the integration of clinical and molecular knowledge within a unified biomedical informatics infrastructure caCORE. The NCI Thesaurus is designed to represent and integrate information from diverse areas such as drugs, therapies, genes, pathways and model organisms. It is a structured reference terminology implemented in a deductive logical framework. Sioutos et al. provided a comprehensive description of the disease and drug models to illustrate the utility of the NCI Thesaurus in supporting clinical and research applications. Portions of semantic model related to pathways, genes and gene products, as well as to anatomy in humans and mice, are also described. Sioutous et al. further discuss the limits of the current approach and directions for future work to meet the information needs of the cancer community. The NCI Thesaurus can be accessed on the web, through the caCORE APIs, or by file download.