© 2007 IMIA and Schattauer GmbH

Integrating Bioinformatics into Clinical Practice: Progress and Evaluation

Findings from the Section on Bioinformatics

E. Lang, Managing Editor for the IMIA Yearbook Section on Bioinformatics
University of Applied Sciences Darmstadt, Dept. of Information and Knowledge Management,
Darmstadt, Germany

Summary

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

Method: Synopsis of the articles selected for the IMIA Yearbook 2007.

Results: Current research in the field of bioinformatics is characterized by careful evaluation of methods and by improved integration of methods into clinical workflows. Ongoing research on genetic causes of diseases is performed on more and better sources of reference data (genome sets and respective annotations), but is still hampered by insufficient, lacking or biased patient data. The application area of bioinformatics has been broadened, leading to amendment or even replacement of traditional methods in fields like characterization of microorganisms. Researchers carry out thorough statistical analyses in order to ensure quality and methodological correctness of new methods based on bioinformatic approaches which are more and more competitive compared to well-established techniques.

Conclusions: The best paper selection of articles on bioinformatics shows examples of excellent research on methods concerning original development as well as quality assurance of previously reported studies. The crucial role of reliable and comprehensive data sources is affirmed, while technical development draws attention to the increasing problem of comparability of data derived some years ago with weaker equipment and those that are of up-to-date quality.

Keywords

Medical informatics, International Medical Informatics Association, yearbook, bioinformatics

Geissbuhler A, Haux R, Kulikowski C, editors. IMIA Yearbook of Medical Informatics 2007. Methods Inf Med 2007; 46 Suppl 1: 106-8

Introduction

Current trends in literature show ongoing integration of clinical bioinformatics into more and more fields of clinical practice. Analysis of genetic patient data helps in guiding diagnostic and therapeutic work, especially in determining between different causes of diseases with rather similar findings and signs on macroscopic level [1, 13, 2-4]. Substantial work based on differential coexpression analysis of microarray data [5, 6, 13] will increase the understanding of cancer and will hopefully lead to improvements in diagnosis and treatment. However, with the increasing use of vast amounts of existing data researchers have to pay attention to several traps caused by misfits of original and current purposes of referred data. As microarray techniques are easily available and can be performed quickly and on considerable quantities of cases, researchers are seduced to relate their new findings on the microscale level to existing clinical data on the macroscale level. [15] shows perils and shortcomings that can rise from this proceeding and gives hints based on recommendations given by Cox [8]. Thorough treatment of data using appropriate statistical methods, however, can lead to impressive results giving detailed insight into the mechanisms of cancer outbreak and growth [13] or on phylogenetic relationship [7, 12, 14]. Clinical bioinformatics is not only restricted to applications performed on the human genome, as there are considerable efforts on characterization of microorganisms playing a key role in infectouos diseases. Phylogenetic analyses help in identifying microorganisms and their relationship in more detail than traditional methods could [12].

Best Paper Selection

The best paper selection of articles for the section ‘bioinformatics’ in the IMIA Yearbook 2007 reflects these trends and follows the tradition of previous yearbooks ([9-11]) in presenting examples of excellent research on methods used for microarray analyses, handling and screening gene data annotations as well as comprehensive clinical studies based on statistical methods.

Four excellent articles representing the research in four different countries (two from Asia reflecting the growing influence of this region) were selected from four 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 2007 in the section 'Bioinformatics'. The articles are listed in alphabetical order of the first author's surname.

Section Bioinformatics
▪ Cai Z, Mao X, Li S, Wei L. Genome Comparison using Gene Ontology (GO) with statistical testing. BMC Bioinformatics 2006; 7:374. ▪ Choi JK, Yu U, Yoo OJ, Kim S. Differential coexpression analysis using microarray data and its application to human cancer. Bioinformatics 2005; 21:24; 4348-4355. ▪ Goh CS, Gianoulis TA, Liu Y, Li J, Paccanaro A, Lussier YA, Gerstein M. Integration of curated databases to identify genotype-phenotype associations. BMC Genomics 2006; 7:257. ▪ Mansmann U. Genomic profiling. Methods Inf Med 2005; 44:454-60.

Conclusions and Outlook

The best paper selection for the Yearbook section 'bioinformatics' can by no means reflect the broadness of the field that is still increasing impressively. The selected papers, however, shed light on some special aspects deserving particular attention as they concern methodological questions in the near future. A period of rapid development in the realm of laboratory equipment, sensibility of reagents and devices as well as improvement of algorithms has lead to extensive and fruitful experimental work. The current state shows a need for consolidation in terms of application purposes and, especially, in quality assurance mainly achieved by carrying out intensive statistical analyses on the primary results. This holds in particular for the typical situation that studies are combined from existing clinical data of patients whose DNA is investigated in current experiments in order to relate genomic findings to clinical outcomes. The years to come will probably show more and more studies that are fully designed in advance and performed by collecting clinical and genomic data simultaneously and with single-purpose restrictions. Future analyses will show if the current strategies of ex-post statistical correction are sufficient or if novel and coherent studies will lead to results that can not be anticipated at the moment.

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 greatly acknowledge the support of Martina Hutter and of the reviewers in the selection process of the IMIA Yearbook.

References

1. Calabrese P, Mecklin JP, Jarvinen HJ, Aaltonen LA, Tavare S, Shibata D. Numbers of mutations to different types of colorectal cancer. BMC Cancer 2005; 5:126-32.
2. Lai C, Reinders MJ, van't Veer LJ, Wessels LF. A comparison of univariate and multivariate gene selection techniques for classification of cancer datasets. BMC Bioinformatics 2006; 7: 236-45.
3. Roy M, Xu Q, Lee C. Evidence that public database records for many cancer-associated genes reflect a splice form found in tumors and lack normal splice forms. Nucleic Acids Res 2005: 33: 5026-33.
4. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, et al. The consensus coding sequences of human breast and colorectal cancers. Science 2006; 314: 268-74.
5. Brors B. Microarray annotation and biological information on function. Methods Inf Med 2005; 44: 468-72.
6. Tsou AP, Sun YM, Liu CL, Huang HD, Horng JT, Tsai MF, et al. Biological data warehousing system for identifying transcriptional regulatory sites from gene expressions of microarray data. IEEE Trans Inf Technol Biomed 2006; 10: 550-8.
7. Snitkin ES, Gustafson AM, Mellor J, Wu J, DeLisi C. Comparative assessment of performance and genome dependence among phylogenetic profiling methods. BMC Bioinformatics 2006; 7: 420-31.
8. Cox D. Cox DR. Planning of experiments. New York: Wiley & Sons; 1958.
9. 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.
10. 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.
11. Lang E. Bioinformatics and its Impact on Clinical Research Methods. Methods Inf Med 2006; 45: 104-6.
12. Cai Z, Mao X, Li S, Wei L. Genome Comparison using Gene Ontology (GO) with statistical testing. BMC Bioinformatics 2006; 7:374
13. Choi JK, Yu U, Yoo OJ, Kim S. Differential coexpression analysis using microarray data and its application to human cancer. Bioinformatics 2005; 21:24; 4348-4355
14. Goh CS, Gianoulis TA, Liu Y, Li J, Paccanaro A, Lussier YA, et al. Integration of curated databases to identify genotype-phenotype associations. BMC Genomics 2006; 7:257
15. Mansmann U. Genomic profiling. Methods Inf Med 2005; 44:454-60
16. van 't Veer L, Dai H, van de Vijver MJ, He YD, Hart AAM, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002; 415: 530-6.
17. Huang E, Cheng SH, Dressman H, Pittman J, Tsou MH, Horng CF, et al. Gene expression predictors of breast cancer outcomes. The Lancet 2003; 361: 1590-6.
18. Chang JC, Wooten EC, Tsimelzon A, Hilsenbeck SG, Gutierrez MC, Elledge R, et al. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. The Lancet 2003; 362: 362-9.

Correspondence to:
Prof. Dr. Elke Lang
University of Applied Sciences Darmstadt
Department of Information and Knowledge Management
Campus Dieburg
Max-Planck-Str. 2
D-64807 Dieburg
Germany
Tel: +49 6151 169412
Fax: +49 6151 169413
E-mail: lang@iuw.h-da.de

Appendix: Content Summaries of Selected Best Papers for the IMIA Yearbook 2007, Section Bioinformatics*

Cai Z, Mao X, Li S, Wei L
Genome Comparison using Gene Ontology (GO) with statistical testing
BMC Bioinformatics 2006;7:374

Gene Ontology (GO) is a sound and fast growing basis for exploring the genetic basis of differences in biological traits between species. [12] used GO annotations available for two complete genome sets of cyanobacteria, and developed a statistical approach to ensure the reliability of the differences detected. After assignment of GO terms to the genes in question using BLAST searches against genes with known GO assignments, the abundance of genes in the two genomes was compared using a chi-squared test and a subsequent false discovery rate correction. Different BLAST cutoffs were examined to distinguish variations in the sets of identified differences. Further the variations of results depending on subsets of genes or on complete genome sets were studied in the comparison of human vs. mouse and of Saccharomyces cerevisiae vs. Schizosaccharomyces pombe.

Choi JK, Yu U, Yoo OJ, Kim S
Differential coexpression analysis using microarray data and its application to human cancer
Bioinformatics 2005; 21:24;4348-55

Microarrays can help in identifying differential expression of individual genes as well as cluster genes that are coexpressed over various conditions. [13] studied alteration in coexpression relationships and proposed a model for finding differential coexpression from microarrays that can serve in testing biological validity with respect to cancer. They constructed a tumor coexpression network and a normal one, based on 10 published gene expression datasets from cancers of 13 different tissues. Cancer affected many coexpression relationships, mainly functional changes such as alteration in energy metabolism, promotion of cell growth and enhanced immune activity. The coexpression changes were not caused by differential expressions. Tumor-stage dependent differences were detected and discussed, but were not taken into account in the result presentation.

Goh CS, Gianoulis TA, Liu Y, Li J, Paccanaro A, Lussier YA, Gerstein M
Integration of curated databases to identify genotype-phenotype associations
BMC Genomics 2006;7:257

Fast detection of potentially harmful microorganisms is a crucial challenge in clinical practice and biodefense. Characterization of unknown microorganisms can be achieved by prediction of its phenotype based on the molecules encoded by its genome. [14] introduce a systematic approach that combines phenotypic information from a biomedical informatics database (GIDEON) with molecular information derived from NCBI's Clusters of Orthologous Groups database. Integrating the information in the two databases lead to correlation of the presence or absence of a given protein in a microbe. Its phenotype is determined by molecular characteristics or survival in particular growth media, as it is usually done. The authors could confirm 66 % of the associations by the literature with a 0.8 correlation score threshold; 86 % were positively verified at a 0.9 correlation threshold. They found possible phenotypic manifestations for proteins concerning sugar metabolism and electron transport and expect a possible extension of their approach to linking pathogenic phenotypes with functionally related proteins.

Mansmann U
Genomic profiling
Methods Inf Med 2005;44:454-60

The use of microarrays to generate prognostic profiles has become a widespread method, but literature shows severe shortcomings in methodology with respect to the six key issues of a good experimental design formulated by Cox [8] as early as 1958. [15] analyzed three original papers [16-18] on the prognosis of breast cancer using genomic profiling according to their compliance with Cox' key issues. He discusses the applied methods and data sets regarding the definition of relevant endpoints, avoidance of systematic bias, generalizability of results, appropriate sample size to achieve sufficient power, simple design as prerequisite for interpretability, and avoidance of artificial assumptions. Detailed discussion of accordance of the papers to the six principles ends up with finding severe violations in all of them. [15] proposes a strategy to assess whether a study has achieved a high level of quality and to establish a suitable protocol for future profiling projects.

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