V. Maojo, F. Martin-Sanchez[1]

Biomedical Informatics Group, Artificial Intelligence Laboratory. Polytechnical University of Madrid, Boadilla del Monte 28660. Madrid, Spain
[1]Bioinformatics Unit (BIOTIC), Institute of Health Carlos III, Majadahonda. Madrid, Spain

Review

Public Health Implications of Bioinformatics

Abstract: Epidemiologists are reformulating their classical approaches to diseases by considering various issues associated to "omics" areas and technologies. Traditional differences between epidemiology and genetics include background, training, terminologies, study designs and others. Public health and epidemiology are increasingly looking forward to using methodologies and informatics tools, facilitated by the Bioinformatics community, for managing genomic information. Future microarray developments will also facilitate the analysis of entire genomes on single arrays, enhancing genetic epidemiology research. The use of biomarkers, biobanks, and integrated genomic/clinical databases poses serious challenges for bioinformaticians in order to extract useful information and knowledge for biomedical research and healthcare. In this regard, there are various ethical, privacy, informed consent and social implications that should be carefully addressed by researchers, practitioners and policy makers.

1. Introduction

Diseases are not inherently uniform. Cancer, diabetes, allergies, rheumatologic or neurodegenerative diseases can no longer be understood without incorporating new molecular data. They can present different etiology, risk factors, markers, genetic susceptibilities, diagnostic conditions, therapy indications and patient management peculiarities. Until recently, epidemiology has not adequately considered these variations, and research is only beginning to show the whole scenario. Old approaches, based only on clinical definitions of diseases, are being redefined. New ideas arise that aim to characterize diseases at a systems biology level, using genetics, proteomics, physiology and population biology [1]. The new data collected from integrated studies based on these new perspectives can be analyzed and mined to generate scientific hypotheses that will be evaluated by researchers [2]. In this regard, there is a high potential of combined molecular and genetic epidemiology approaches in the prevention of various diseases (e.g., cancer, diabetes) through early identification of people "at risk", suggesting, for instance, dietary indications that can be quite useful in preventing the onset of the disease [3,4].

Epidemiology and genetics have traditionally been separated. Epidemiology can be defined as the study of the etiology, risk factors and distribution of a disease or physiological condition in human populations and of the factors that influence this distribution [5]. It aims to study disease patterns and their related circumstances in specific populations mostly from an observational perspective (since the subjects of study are not usually exposed to specific conditions designed by researchers). It involves the study of ill and "healthy" people, which clearly differentiates such studies from most clinical research [6]. Epidemiology has usually relied on evaluating a small number of genetic and phenotypic characteristics, which are often difficult to measure, such as mutations, metabolic or immunological markers, or ethnicity. New molecular technologies provide tools to directly track DNA sequences for analyzing pathogen distribution, or genetic variations related to drug resistance or specific morbidity [7]. Genetics has usually centered on rare diseases, caused by mutations in single genes, and affecting a small percentage of the overall population. Public health professionals have traditionally included genetics in their routine, but only in concrete cases such as birth defects prevention or inborn errors of metabolism [8].

Since genetics and epidemiology have followed separate paths, it is not surprising that terminologies are sometimes different for the same concept.

For instance, 'association studies' (genetics) and 'case–control studies' (epidemiology) refer to a similar type of study. Terms can also be misused, such as 'gene' and 'allele' in genetic literature and by epidemiologists [6]. Collaborations between epidemiology, molecular biology, bioinformatics (BI), medical informatics (MI), and genetics are currently being established, particularly since researchers have realized that the results of the Human Genome Project can provide new data sources useful for discovering the causes of diseases and to find new diagnostic methods and therapies. Since all of these disciplines are based on different backgrounds, foundations, study designs, training, assumptions, methodologies, tools, decision-making techniques and terminologies, interactions are not easy [2,9].

Besides the classical approaches of epidemiology devoted to the study of the factors that are responsible for the development of a disease, public health is confronting new challenges derived from a number of facts over the last few years: globalization, emerging infectious diseases, rise of antibiotic resistance and biodefense. The blurring of national borders for socio-economic reasons is facilitating the dissemination of infectious agents throughout the world. The global resurgence of infectious diseases presents a great threat to everyone's health. The incomplete or inadequate use of antibiotics leads to resistant microorganisms. Researching potential bioterrorism threats, like anthrax, tularemia or plague, requires a comparison of genome structures and functions to provide clues for the development of diagnostic tests and the design of antiviral and prophylactic vaccine strategies. In all these fields, advances in public health and epidemiology are increasingly dependent on sound methodologies and tools for acquiring, storing, integrating and analyzing genomic information from the human host and the different pathogenic microorganisms.

2. Genomic Research and Bioinformatics: Current Issues

The success of the HGP has attracted the interest of BI professionals, academicians and industry. Now that the first efforts have been completed, a transition from a *genomic* to a *post-genomic era* will appear [10]. The latter will imply a new agenda in health research and novel directions and applications in education and care [11,12]. Interactions with BI, bioengineering, cognitive science, epidemiology and biostatistics will be important interdisciplinary collaboration challenges that might be explored [13,14,15,16].

Over the last few years, the use of the term "omics" has spread across the fields of genomics, proteomics, metabolomics and others, aiming to provide a unified vision for understanding their underlying common mechanisms and the scientific methods and tools needed to advance research in these areas. This unified vision of "omics" areas is particularly adequate for providing multiple relationships with external disciplines such as BI, computational biology, genetic epidemiology and other related fields. In this regard, a workshop has recently been held to examine the links, challenges and limitations of the "omics" areas, when they are related to toxicology (including ecotoxicology) and epidemiology, as well as their ethical, social, and legal implications [17]. Four overarching issues emerged regarding the use of "omics" technologies:

(1) Risk assessment in the framework of toxicology and epidemiology.
(2) Effective application of "omics" to epidemiological studies will require suitable biological samples from large and diverse population groups at relevant time periods of exposure.
(3) From ethical, social, and legal perspectives, the "omics" technologies need the involvement of the research, academic, regulatory, public, and industry communities.
(4) A unified research agenda as applied to toxicology and epidemiology is urgently needed.

BI can be defined - adapted from [18] - as an area dedicated to the acquisition, processing, storage, distribution, analysis and interpretation of biological data by means of computing methods and tools. Some popular software programs have been available for searching and comparing databases and sequences [19]. GENBANK, the main DNA sequence database, was created in 1980 and SwissProt, its equivalent for proteins, became operational in 1987 [20]. Many BI programs have been developed in academic settings and have become commercial products, offering integrated sets of tools for molecular biology data management. The WWW has been generally accepted as the infrastructure for accessing biological databases and programs, designed for information exchange and retrieval in biological research. An outstanding inititative in this area is the Human Genome Epidemiology Network (HuGENet), http://www.cdc.gov/genomics/hugenet/default.htm, lead by the Office of Genomics and Disease Prevention from the Centers for Disease Control and Prevention (CDC).

Many tools have been developed to analyze and annotate DNA sequence information [21]. Some researchers [22,23] suggested that software tools for full-genome sequencing may successfully link genomics, population genetics, and epidemiology, leading to advances in the scientific discovery of

the underlying processes of diseases. Future developments related to microarrays may facilitate scanning and the analysis of entire genomes on single arrays, allowing new studies in genetic epidemiology [23].

BI and computational biology approaches and tools can be instrumental in the pharmaceutical industry for drug design and high-throughput screening of potential antimicrobial candidates. Rapid information exchange through specialized databases over the Internet can accelerate international health programs as it did for the Human Genome project [24]. New advances in pharmacogenomics may lead to a significant demand for clinical trial sites, transferring some studies to developing countries.

There has been a massive amount of data in epidemiology and public health, which has not attracted enough efforts at developing new computerized methods for analysis. With the advent of additional genomic data, epidemiologists must choose a high number of markers and be able to analyze the genotype for each participant in any study. Novel statistical and data mining methods will be needed. One solution is to increase the number of studies and participants involved, but this can be quite expensive. Methods to gather data from heterogeneous studies and databases can be used to solve this problem.

3. New Proposals / Initiatives in Genetic Epidemiology

The Environmental Genome Project was launched in 1997 at the US National Institute of Environmental Health Sciences to address research into the role of genetic variation in response to environmental exposure in large populations [25]. Susceptibility genes must be characterized in order to predict risk from environmental exposure and response to pharmaceuticals. Human diseases arise from many different factors, including diet, genetics, environmental conditions or exposures, aging, and others. Current models generally assume that mutations on a single gene or exposure to a single environmental condition may lead to a specific disease. Wider interactions are now being considered in research, given the recent changes in our understanding of the inherent mechanisms of disease (for instance, we now assume that many common human diseases seem to be polygenic rather than monogenic).

Molecular epidemiology will facilitate key information about microbial transmission and virulence patterns, whereas genetic epidemiology will identify the characteristics of individual and population susceptibility [26]. Genetic epidemiology may contribute to characterizing the molecular, metabolic, and disease profiles of different populations. It focuses on the distribution of diseases within families, and the distribution of genetic conditions in populations [27,28].

Correlations between genotypes and disease phenotypes in humans are constantly improving. Over one million SNPs are potentially highly useful markers for epidemiological studies [23]. Genetic epidemiology can bring new perspectives for improving experimental design in biology. At the same time, microarrays are new technologies for obtaining massive genetic variation or gene expression data.

At a recent meeting [17], held to create an unified research agenda for genomics technologies applied to toxicology, epidemiology, and risk assessment, experts stated various research recommendations for genetic epidemiology: *a)* characterize prevalence and background frequencies of genetic polymorphisms and their functions, *b)* focus on finding methods to assess gene expression in large populations, *c)* address the statistical and BI issues related to the field, and *d)* pursue a multidisciplinary approach to epidemiology research and development.

A topic under permanent debate in human genome epidemiology is the development of ambitious biobanks, for storing DNA specimens, that can be used in health research. Several issues, such as informed consent, privacy, economic implications, usefulness, quality control, data collection and analysis, health outcomes, and others arise. The best known project is the Iceland database, a computerized compilation of medical records at a national level [29]. Other countries, such as the UK and Estonia have developed or planned related projects. Developers aim to collect genetic information and link it with medical, lifestyle and genealogical information. In the end, researchers should be able to retrieve information on susceptibility genes for complex diseases, improving health care while obtaining a revenue from these activities. Researchers are discussing the value of using genetically homogeneous populations for genetic studies of complex disease susceptibility to reduce locus and allelic heterogeneity [28].

There are various ethical/privacy/informed consent/social implications for this kind of genetic epidemiology research. For instance, risks for many diseases can be estimated long before therapy is available. People may fail to take the adequate prevention measures, leading to social costs, or their lives may be negatively affected if they are acquainted with their risks and do not receive proper counseling or psychological advice. Data must remain private and, therefore, information stored in medical records and clinical and

epidemiological databases must be protected for confidentiality, leading to severe requirements and restrictions on computer security.

Novel genetic information technologies will facilitate the performance of cost-effective screening (genetic tests) in large populations. Gathering genetic, clinical, environmental and lifestyle data may provide clues for unraveling the causes and processes of polygenic diseases. Data included in populational repositories or biobanks [30,31] will be applied in public health, for issues such as disease prevention programs or the analysis of the cost-effectiveness of pharmacogenetics from a health care perspective.

The information infrastructure of biobanks raises various issues, given the different types of data collected over the decades. Linking genetic and clinical information implies variations in the methods of observation, patient management, and the scientific knowledge that medical practitioners considered at the time of patient evaluation. Genomic data have been obtained according to different scientific theories and models, also including a considerable degree of uncertainty [2]. In gene mapping studies, genetically homogeneous samples have been used to improve gene discovery. These projects are under strong discussion, and are being carefully monitored and studied to evaluate if they do indeed contribute new knowledge to biomedical research.

4. Potential Contributions of Bioinformatics to Public Health Practice, Research and Education

Genomics offers different perspectives on various limitations of epidemiology [8, 32], including informatics tools developed in the closely related area of BI. The Institute of Medicine (IOM), in a report named *The Future of Public Health* [33], defined three overall objectives for public health activities: (1) assessment, "the regular systematic collection, assembly, analysis, and dissemination of information, including human genome epidemiological information, on the health of the community", (2) policy development, "the formulation of standards and guidelines in collaboration with stakeholders, which promote the appropriate use of genetic information and the effectiveness, accessibility, and quality of genetic tests and services", and (3) "the assurance to constituents that genetic information is used appropriately and that genetic tests and services meet agreed-upon goals for effectiveness, accessibility, and quality".

According to this report, a priority for public health agencies is to develop initiatives to integrate genomics across their plans, including "maternal and child health, environmental health, chronic disease, laboratory services, and infectious disease". For instance, the American Public Health Association has supported the role of genomics in public health, expressing the need for research in the area [34]. It also encouraged public and private organizations to determine the genomics-related educational needs within public health activities and for professionals. Whereas public health entities should foster activities in this area, public health professionals must, at the same time, understand the relationships between genetic variations and disease risk and promote this information to reduce morbility and morbidity. A key objective is to evaluate the appropriate use of genetic tests [8].

It will be difficult to convince physicians that they should change their reasoning approaches to think in terms of genomic information [2, 35]. If genomics is going to contribute to achieving significant results in medicine, physicians and nurses will need to be educated on its scientific foundations, methods and tools. BI and MI can contribute to this objective by developing new tools that can be used in clinical routine [2]. In various European and US institutions, various projects are being carried out to facilitate the transfer of genetic tests to clinical and public health settings [36]. For this purpose, the education of health professionals is a central issue, since they usually have little knowledge of and background in genetics.

According to [23], the major challenge in the area of genetic epidemiology is "to devise computational methods for comparing massive amounts of genetic data across individuals and populations to complex phenotypes which may result from a variety of different genetic pathways". In this sense, epidemiologists and geneticists look forward to using results from the HGP as a reference catalog for variations in chromosomal DNA. Results may be applied to future applications in genomic medicine in areas such as patient diagnosis, monitoring, prognosis, therapy and management [10,37,38, 39]. Researchers are beginning to create and share standardized databases in genetic epidemiology. Data integration from multiple, reproducible studies can yield significant advances in epidemiological research, but some issues such as the study of methods to check integrity [40] and improve information retrieval and visualization must be considered. In this sense [22], technologies with binary outputs are more promising for data exchange than technologies that lead to banding patterns, such as PCR [23].

New achievements in BI can be fundamental for finding "predictive" gene sets. A significant objective in the joint efforts between toxicology and

epidemiology professionals is to correlate effect and exposure, and find the best biomarkers using "omics" technologies. Connections between gene expression data and toxicology will facilitate an integration of "omics" information with known toxicological measures for providing a better understanding of the mechanisms of chemical effects on biological systems, as shown in figure 1. There are no current guidelines to correlate quantitative or qualitative changes in gene/protein/metabolite expression with adverse effects [17, 23]. New statistical and data mining approaches will be needed to analyze and interpret data sets from epidemiological studies that include human genetic biomarkers. Since inconsistencies and errors in data can be significant, replication of studies will be important [17].

Joint BI/MI efforts should provide clues for planning the future of both disciplines, particularly related to genomic medicine. In this regard, MI has been dealing with the problems of structuring complex and large quantities of data and knowledge for decades. It should be expected that many standards, coding systems and ontologies and terminologies - such as, for instance, the UMLS [41] or gene-ontology [42] - will be shared between the two disciplines.

5. Conclusions

Collaborative efforts among bioinformaticians, molecular biotechnologists and epidemiologists in developing new methods to assess disease risk may rapidly lead to identifying genetic influences on complex human diseases. Relationships between genomics and diseases can transform health and society, identifying populations that are susceptible to environmental causes of diseases [43,44,45]. Since health is not the absence of a defect but a complete well-being, genomics can bring new perspectives to public health issues and practice, considering the wider aspects of health outcomes.

Given all the issues presented in this paper, associated with various aspects of biomedical research, a fundamental issue will be to redesign research studies, data analysis and interpretation of integrated genomic/clinical/epidemiological studies [46]. The classical methodologies used in epidemiology, public health, health services, clinical and laboratory research, data analysis (including data mining), and other areas must be redesigned to include integrated biological/clinical information [47]. The kind of cognitive skills, background, problem-solving approaches, data collection and analysis, and patient informed consent and advice needed in this genomic era are different from those used until recently [2]. Universities, industry and other private and public institutions are already transforming training, education, management and practice routine in this area. Whereas links between BI/genomics and epidemiology should give new clues for considering diseases in a broader sense and provide additional knowledge for health care, significant challenges lie ahead.

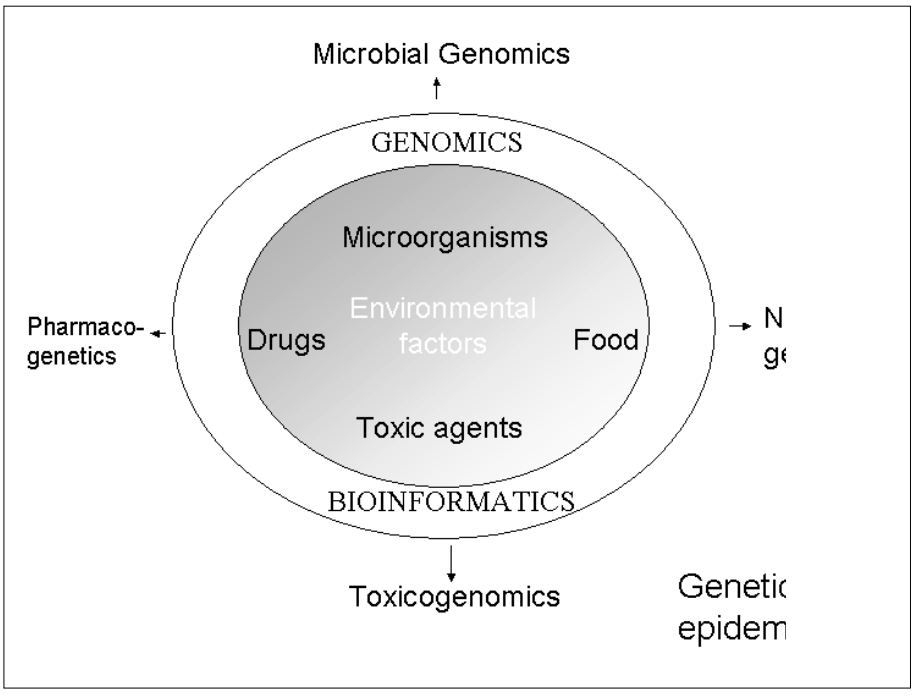

Fig 1. Various forces linking epidemiology, genomics and BI

Many techniques such as DNA microarrays for studying gene expression or advanced techniques for protein analysis are producing huge amounts of data that are being incorporated within clinical databases. Epidemiologists should be aware that errors are found in many databases, coming from the laboratory or from information processing, since BI tools used to predict the location of sequences can introduce important errors. In addition, many researchers trust their results, overlooking major inaccuracies. While some mistakes are tolerable in laboratory research, they are not acceptable if results are used for medical care. MI experiences in dealing with uncertain information when diagnosing and treating patients can be important in this regard [2].

The development of the World Wide Web and computer-based information systems has boosted information access and exchange in many areas, including biology and health. The results

and sanitary consequences of the HGP lead to educating practitioners and informing patients about the implications of genomics in new medical routines and practice. New computerized educational systems should contribute to enhancing this new scenario. Since BI professionals have centered on developing systems for biologists and research activities, there is no prior experience in developing systems for educational purposes. In this sense, the lessons learned in years of practice in MI should also be used for these new kinds of applications. Various reports have already pointed out opportunities for synergy between BI and MI [2,13,14,15,16,48], in a new discipline located at the center of both fields, called biomedical informatics, as shown in figure 2.

The introduction of genomic information for healthcare will oblige development of clinical applications in various areas from this novel, integrated, biomedical informatics perspective. Old systems will have to be updated. For instance, health records, standards and medical terminologies and vocabularies developed by MI professionals should be adapted to include "omics" data and information; on the other hand, innovative genomic-based medical records, where privacy, confidentiality and societal manipulation subject to economic and cultural forces raise significant challenges. Applications developed from the BI side, such as databases and information retrieval, should be also adapted and reused for various topics within biomedical research and clinical routine.

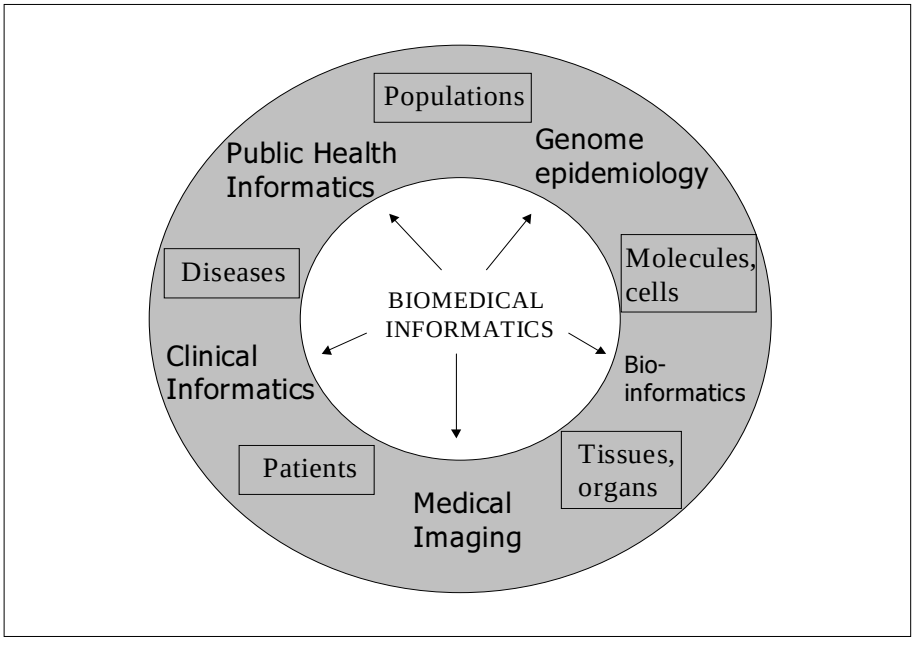

Fig. 2. Biomedical Informatics, an interdisciplinary new field.

Acknowledgments

We want to thank Prof. Kulikowski for his ongoing collaboration and thoughtful comments to the authors in the analysis of interactions between MI and BI. Dr. Maojo and Dr. Martin's research in related areas has been supported by funding from the EC BIOINFOMED Study, the INBIOMED project, Ministry of Health, Spain, and the MEDGENBASE project, Ministry of Science and Technology, Spain.

References

1. Butler D. Epidemiology set to get fast-track treatment. Nature 2001 Nov 8;414(6860):139.
2. Maojo V, Kulikowski CA. Bioinformatics and Medical Informatics: Collaboration on the Road to Genomic Medicine? J Am Med Inform Assoc. In press 2003.
3. Perera FP, Weinstein IB. Molecular epidemiology: recent advances and future directions. Carcinogenesis 2000 Mar;21(3):517-24.
4. Furberg AH, Ambrosone CB. Molecular epidemiology, biomarkers and cancer prevention. Trends Mol Med 2001 Nov;7(11):517-21.
5. Lilienfeld AM, Lilienfeld DE. The 1979 Health Clark lectures. 'The epidemiologic fabric'. I—Weaving the threads. Int J Epidemiol 1980 Sep;9(3):199-206.
6. Potter JD. At the interfaces of epidemiology, genetics and genomics. Nat Rev Genet 2001 Feb;2(2):142-7.
7. Foxman B, Riley L. Molecular epidemiology: focus on infection. Am J Epidemiol 2001 Jun 15;153(12):1135-41.
8. Beskow LM, Khoury MJ, Baker TG, Thrasher JF. The Integration of Genomics into Public Health Research, Policy and Practice in the United States. Community Genet 2001 Jul;4(1):2-11.
9. Maojo V, Iakovidis I, Martin-Sanchez F, Crespo J, Kulikowski C. Medical informatics and bioinformatics: European efforts to facilitate synergy. J Biomed Inform 2001 Dec;34(6):423-7.
10. Martin-Sanchez F, Maojo V, Lopez-Campos G. Integrating genomics into health information systems. Methods Inf Med 2002;41(1):25-30.
11. Collins FS, McKusick VA. Implications of the Human Genome Project for medical science. JAMA 2001 Feb 7;285(5):540-4.
12. Collins FS, Guttmacher AE. Genetics moves into the medical mainstream. JAMA 2001 Nov 14; 286(18): 2322-4.
13. Kohane, I. Bioinformatics and Clinical Informatics: The Imperative to Collaborate. J Am Med Inform Assoc 2000; 7, 5: 439-443.
14. Kulikowski, C. The Micro-Macro Spectrum of Medical Informatics. Challenges: From Molecular Medicine to Transforming Health Care in a Globalizing Society. Methods Inf Med 2001, 41: 20-4.
15. Miller, P. Opportunities at the Intersection of Bioinformatics and Health Informatics. J Am Med Inform Assoc 2000;7(5):431-8.
16. Altman, R. The Interactions Between Clinical Informatics and Bioinformatics. J Am Med Inform Assoc 2000; 7(5):439-43
17. Henry CJ, Phillips R, Carpanini F, Corton JC, Craig K, Igarashi K, et al. Use of genomics in toxicology and epidemiology: findings and recommendations of a workshop. Environ Health Perspect 2002 Oct;110(10):1047-50.
18. Benton D. Bioinformatics—principles and potential of a new multidisciplinary

tool. Trends Biotechnol 1996 Aug;14(8): 261-72.
19. Roberts RJ. The early days of bioinformatics publishing. Bioinformatics 2000 Jan;16(1):2-4.
20. Bairoch A. Serendipity in bioinformatics, the tribulations of a Swiss bioinformatician through exciting times! Bioinformatics 2000 Jan;16(1):48-64.
21. Mount, D. Bioinformatics: Sequence and Genome Analysis. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory; 2001
22. van Belkum A, Struelens M, de Visser A, Verbrugh H, Tibayrenc M. Role of genomic typing in taxonomy, evolutionary genetics, and microbial epidemiology. Clin Microbiol Rev 2001 Jul;14(3):547-60.
23. Dennise D. Dalma-Weiszhausz, Marina E. Chicurel, Thomas R. Gingeras Microarrays and Genetic Epidemiology: A Multi-purpose Tool for a Multifaceted Field. Genetic Epidemiology 2002;23:4 –20.
24. Lander, ES et al. Initial sequencing and analysis of the human genome. Nature 2001 Feb 15;409(6822):860-921.
25. Olden K, Wilson S. Environmental health and genomics: visions and implications. Nat Rev Genet 2000 Nov;1(2):149-53.
26. Fauci AS. Infectious diseases: considerations for the 21st century. Clin Infect Dis 2001 Mar 1;32(5):675-85.
27. Pang T. The impact of genomics on global health. Am J Public Health 2002 Jul;92(7):1077-9.
28. Austin MA. Ethical issues in human genome epidemiology: a case study based on the Japanese American Family Study in Seattle, Washington. Am J Epidemiol 2002 Apr 1;155(7):585-92.
29. Annas GJ. Rules for research on human genetic variation—lessons from Iceland. N Engl J Med 2000 Jun 15;342(24):1830-3.
30. Cambon-Thomsen A. Assessing the impact of biobanks. Nat Genet 2003 May;34(1):25-6.
31. Kaiser J. Biobanks. Population databases boom, from Iceland to the U.S. Science 2002 Nov 8;298(5596):1158-61.
32. Ellsworth DL, Manolio TA. The Emerging Importance of Genetics in Epidemiologic Research III. Bioinformatics and statistical genetic methods. Ann Epidemiol 1999 May;9(4):207-24.
33. Institute of Medicine. The Future of Public Health. Washington, DC: National Academy Press; 1988.
34. Steinberg KK, Gwinn M, Khoury MJ. The role of genomics in public health and disease prevention. JAMA 2001 Oct 3;286(13):1635.
35. Weatherall, D. Science and the Quiet Art. Medical Research & Patient Care. Oxford: Oxford University Press; 1995.
36. Noonan AS. Integrating genomics into US public health. Genet Med 2002 Nov-Dec;4(6 Suppl):68S-71S.
37. Mostowy S, Behr MA. Comparative genomics in the fight against tuberculosis: diagnostics, epidemiology, and BCG vaccination. Am J Pharmacogenomics 2002;2(3):189-96.
38. Schork NJ, Fallin D, Lanchbury JS. Single nucleotide polymorphisms and the future of genetic epidemiology. Clin Genet 2000 Oct;58(4):250-64.
39. Hoffman SL. Infectious disease. Research (genomics) is crucial to attacking malaria. Science 2000 Nov 24;290(5496):1509.
40. Bertone P, and Gerstein M. Integrative data mining: the new direction in bioinformatics. IEEE Eng Med Biol Mag 2001 Jul-Aug;20(4):33-40.
41. Lindberg, C. The Unified Medical Language System (UMLS) of the National Library of Medicine. J Am Med Rec Assoc 1990. May;61(5):40-2.
42. The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. Nature Genetics 2000, 25: 25-29.
43. Gilbert GL. Molecular diagnostics in infectious diseases and public health microbiology: cottage industry to postgenomics. Trends Mol Med 2002 Jun;8(6):280-7.
44. Broder S, Hoffman SL, Hotez PJ. Cures for the Third World's problems: the application of genomics to the diseases plaguing the developing world may have huge medical and economic benefits for those countries and might even prevent armed conflict. EMBO Rep 2002 Sep;3(9):806-12.
45. Goldie SJ, Levin AR. msJAMA. Genomics in medicine and public health: role of cost-effectiveness analysis. JAMA 2001 Oct 3;286(13):1637-8.
46. Visionary experimental designs. Nature 2001 Nov 8;414(6860):133. [No authors listed]
47. Millikan R. The changing face of epidemiology in the genomics era. Epidemiology 2002 Jul;13(4):472-80.
48. Maojo V, Martin-Sanchez F, Billhardt H, Iakovidis I, Kulikowski C. Establishing an agenda for biomedical informatics. Methods Inf Med 2003;42(2):121-5.

Address of the authors:
Victor Maojo
Biomedical Informatics Group
Artificial Intelligence Laboratory.
Polytechnical University of Madrid
Boadilla del Monte
28660 Madrid, Spain
E-mail: vmaojo@infomed.dia.fi.upm.es

Fernando Martin-Sanchez
Medical Bioinformatics Dept.
Institute of Health Carlos III
Majadahonda
Madrid, Spain