Synergy between medical informatics and bioinformatics: facilitating genomic medicine for future health care


Abstract

In this paper, we review the results of BIOINFOMED, a study funded by the European Commission (EC) with the purpose to analyse the different issues and challenges in the area where Medical Informatics and Bioinformatics meet. Traditionally, Medical Informatics has been focused on the intersection between computer science and clinical medicine, whereas Bioinformatics have been predominantly centered on the intersection between computer science and biological research. Although researchers from both areas have occasionally collaborated, their training, objectives and interests have been quite different. The results of the Human Genome and related projects have attracted the interest of many professionals, and introduced new challenges that will transform biomedical research and health care. A characteristic of the 'post genomic' era will be to correlate essential genotypic information with expressed...
phenotypic information. In this context, Biomedical Informatics (BMI) has emerged to describe the technology that brings both disciplines (BI and MI) together to support genomic medicine. In recognition of the dynamic nature of BMI, institutions such as the EC have launched several initiatives in support of a research agenda, including the BIOINFOMED study.

© 2003 Elsevier Inc. All rights reserved.

Keywords: Bioinformatics; Medical informatics; Genomics; Genomic medicine; Biomedical Informatics

1. Introduction

The Human Genome has created new opportunities for the study of monogenic and complex, more prevalent, multifactorial diseases [1]. Increased knowledge of the human genome supports the theory that diseases should be understood by considering the complex interactions between genes and environmental factors that initiate pathological processes and define the phenotype [2,3]. The large amounts of genetic and proteomic data [4] offer opportunities for new research targets and biomedical applications. New therapies for established diseases and novel interventions for preventing diseases are being developed and discussed [5]. These innovative approaches cannot be sustained without effectively dealing with the vast amounts of data generated in the laboratory in the areas of functional and structural genomics and proteomics (the world of BI) [6]. Within this framework, it is equally important to integrate clinical data generated by medical records (the world of MI) [7]. Such an integration will lead to innovative approaches to link patient care and public health. In such an environment, data-centered electronic health records (EHR), clinical decision support systems, and image and signal processing techniques and tools can be applied to a broader field of information. Biomedical Informatics (BMI) is the emerging discipline that aims to create this common conceptual information space to further the discovery of novel diagnostic and therapeutic methods in the rapidly evolving arena of genomic medicine.

The need for integrated post-genomic approaches in medicine has been aptly stated by the World Health Organization:

Some of the claims for the medical benefits of genomics have undoubtedly been exaggerated, particularly with respect to the time scales required for them to come to fruition. Because these uncertainties, it is vital that genomics research is not pursued to the detriment of well-established methods of clinical practice, and clinical and epidemiological research. Indeed, for its full exploitation it will need to be integrated into clinical research involving patients and into epidemiological studies in the community. It is crucially important that a balance is maintained in medical practice and research between genomics and these more conventional and well tried approaches. [8].

2. Motivation

The Conference ‘Synergy between Research in Medical Informatics, Bioinformatics, and Neuroinformatics’ (Brussels, December 14, 2001, http://www.ramit.be) was the kick-off point for a series of European-centered activities related to the analysis of the current status and interactions between MI and BI. After various meetings and discussions, a study named BIOINFOMED (http://bioinfomed.isciii.es/) (“Prospective Analysis of the Relationships and Synergy Between Medical Informatics and Bioinformatics”) [9] was approved and launched with support from the European Commission (http://www.cordis.lu/ist/) in March, 2002. It was aimed to analyse the relationships and potential synergies between MI and BI. Several goals were set for the project, including the elaboration of a White Paper, which should provide a number of clues for the European Commission itself and for researchers and professionals from these and other related areas.

Three groups, at the Institute of Health ‘Carlos III’ (ISCIII) (http://www.isciii.es/) (Madrid), at the Polytechnical University of Madrid (http://www.dia.fi.upm.es/) and at the University of Linköping (http://www.liu.se/en/), were the main contractors of the BIOINFOMED study. Thirty professionals from various European and US organizations were invited to participate and provide their expertise and knowledge. Experts from disciplines such as MI, BI, genomics, public health, clinical medicine, and bioengineering covered a wide range of interests and areas.

At the start of the study, a questionnaire was elaborated by the team at the ISCIII, including questions on the status and future of MI and BI, research directions, and the potential of already existing tools and methodologies that could be transferred between both disciplines and then be reused in BMI projects and topics. Responses were processed and a preliminary document was mailed to the experts. Two meetings took place in 2002, in Crete (Greece) and Valencia (http://bioinfomed.isciii.es/Bioinfosalud2002/Bioinfosalud2002-en.htm) (Spain), where experts discussed the contents of the document extensively. A draft for the White Paper was made and later refined and completed over the Internet by all participants.

This paper includes an abridged version of the final White Paper (http://bioinfomed.isciii.es/Bioinfomed/The White Paper/results/White Paper.pdf) delivered to the European Commission [10]. The paper is divided in various sections: (1) Background, (2) Expected impact, (3) Gaps and bridges, (4) Collaborative R&D agenda for BI and MI, and (5) Conclusions. Further details of the sections can be found below.
3. Background

MI and BI are two interdisciplinary areas located at the intersection of informatics with medicine and biology (genomics), respectively. Historically, MI and BI have always been separated and only occasionally have researchers of both disciplines collaborated in the past [11]. Almost from its very beginning MI focused mainly on the development of practical computer applications for health purposes. Recently, a debate has been initiated on its scientific content and future. BI, a younger discipline, has grown enormously due to its contribution to genomic research [12].

MI has traditionally been focused on the development of computer applications for representing and implementing health care [13]. As health care moves from patient/hospital-centered systems to citizen/community-centered systems, genetic data should be increasingly available for patient care, driving MI in a direction where synergy with BI could be naturally achieved [14,15]. BI has evolved to handle large amounts of sequence and structural data, generated in the laboratory [16]. While BI’s success in handling large data volumes is apparent, its weakness in data normalization will continue to cause problems for collaborative research.

When considering public-health issues, reasons for synergy become more compelling, and various reasons arise for interactions among epidemiology, clinical research and genomics.

Increasing awareness of BMI is evidenced by recent conferences [17–19]. For instance, the American Medical Informatics Association (AMIA) (http://www.amia.org/) has focused on the interaction between BI and MI at recent meetings. Panels, tutorials and sessions in periodic congresses (MEDINFO, MIE, PSB, RECOMB, ISMB, and ECCB) highlight this awareness, which has also been shown by various special issues published at different scientific journals. Other key activities include the inclusion of BI in university medical training programs [20] and collaboration with the pharmaceutical industry for physician training and technical support in genomics.

4. Expected impact

The biomedical community seeks to remove the walls between biological information and medical information. The interoperability of biological and medical information for all appropriately authorized users creates imperatives, opportunities, and challenges. Equally significant demands are made by the evolution from patient-centered systems to citizen-centered systems that actively engage citizen participation. The impact can be seen in various categories, such as:

4.1. Scientists/researchers

There is an implicit requirement for exchanging and sharing medical and biological information and knowledge in global (often virtual) work settings. Equally important is that professionals with background in either medical informatics or bioinformatics are better able to understand the basic principles and interrelations of all disciplines involved (the same requirement also applies to clinical health personnel who need to expand their background knowledge of pertinent topics).

4.2. Clinical trials

New clinical trials must be carefully designed to include genomic and proteomic data. High-quality biomedical databases are urgently needed to provide a sound scientific basis for diagnostic, treatment stratification and predictive tests. Ethical, legal, and privacy considerations should be given to tests that evoke anxiety in healthy people regarding their future [21].

4.3. Health-care professionals

4.3.1. New knowledge and technology

The very nature of BMI blurs some of the classical distinctions between clinical and molecular information. As we extend the concept of phenotype to encompass diseases, we also expand the ‘properties’ that are ‘visible’ to include sub-cellular structures and physiological processes. One of the major impacts of BMI will be a broader understanding of how combinatorial and quantitative variations in DNA sequences, protein synthesis, and subsequent protein function affect the evolution of diseases. Genomic and proteomic data analysis has already facilitated both the understanding of the underlying causes of the disease and the identification of new drug targets. As our knowledge about the molecular causes of disease increases, we can expect more elegant molecular interventions to diagnose, disrupt or ameliorate disease.

4.4. Individual citizens

4.4.1. The citizen ‘at risk’

New BMI approaches can give new dimensions to the concept of ‘citizen-at-risk.’ As the basis of the knowledge on genetic associations with illnesses increases, it is likely that this identifiable “at risk” group will grow, encompassing many asymptomatic citizens and therefore placing new demands on health-care systems.

4.4.2. Informed citizens

It will be important for health delivery systems to establish and publish standards for rational genetic testing. The average citizen must be able to understand
and evaluate the appropriate balance between the potential for health improvement and the potential drawbacks that could arise from such testing.

4.5. Health-care providers and systems

4.5.1. Technology diffusion and scientific evidence

With the emergence of novel BMI applications, health-care systems will face difficult challenges. Health-care providers should be prepared to carefully select technologies that have been proven safe and effective. If required, providers should limit the adoption of new technologies to appropriate scientific research settings.

4.5.2. Public health and disease prevention

Knowledge generated by very large biomedical databases will enable health-care organizations to identify citizens who are not only at ‘genetic’ risk for developing diseases but also whose risk of developing symptomatic disease could be reduced by one or more interventions.

4.6. Policy-decision makers

4.6.1. Investment for the future

Rational biomedical databases of the scope required for modern molecular research are expensive to establish and maintain. Programmed cooperation between government, academia and the industry is absolutely essential.

4.6.2. Prioritisation

Genetic testing and the associated concept of ‘citizen-at-risk’ will constitute yet another aspect to the current prioritization palette.

4.6.3. Legislative initiatives

Novel biomedical informatics applications will require clear and up-to-date legislation. Policy makers will have to foster a proactive and continuous legislative process that will keep up with the pace of current scientific developments and implementation plans. Such attempts have resulted in pertinent legislation in some countries (e.g., the Estonian Human Genes Research Act of 2000 (http://www.geenivaramu.ee/)).

4.7. Industry

In order for certain industrial efforts to succeed (such as pharmaceutical and biotechnology), more attention will have to be paid to how both clinical trials and exploratory analyses evolve and become successful. Industries taking advantage of the development and maintenance of large databases and knowledge bases by academic institutions should contribute to the financing of such public initiatives and collaborative efforts among different institutions.

4.8. Society

4.8.1. Consent to collect, view, and use information

Genomic and proteomic databases must be secured from unwanted intrusions. Correlation between clinical and genomic/proteomic profiles should only take place when informed consent has been obtained. Every citizens’ right to not know about his/her genetic risk should be respected.

4.8.2. Genetic discrimination

Scenarios of selection or exclusion on the basis of individuals’ ‘genetic profiles’ are not acceptable, and this fundamental principle should be guaranteed through pertinent legislation.

4.8.3. Racial profiling

Already, scientists are engaged in debates about ethnicity versus race, and one can see how ‘genetic assessments’ of this sort are invitations for misusing large biomedical databases. Great care must be taken to guarantee that biomedical databases are not subjected to unauthorized analyses of this sort.

4.8.4. Fetal testing and pregnancy termination

At the present time, pregnant women may elect to have their pregnancy terminated as a result of genetic testing of the fetus. As the knowledge about genotypes at risk for disease increases, more couples will be faced with the decision whether or not to have the fetus tested, and whether to act on the results of such testing.

Above we have stated some of the consequences that can result from the introduction of genomic information in various aspects of medicine. Numerous challenges can be envisioned for MI and BI, including some issues that must be analyzed and solved, as will be shown below.

5. Gaps and bridging solutions

5.1. Gaps

The tools and applications developed by MI reach a wide range of users including physicians, nurses, administrators, management, and researchers [22]. BI applications are characterized by a much more homogeneous user group dominated by researchers. Although the application domains differ, both MI and BI often use similar methodologies [23]. Both fields are active in machine learning, natural language processing, image analysis or database research. However, working on similar problems with related methods does not guarantee similar results because the application domains differ.

Another difference between the MI community and the BI community involves the degree of interaction
between research groups. In MI, collaborative efforts have been relatively scarce. In BI, collaborative research has been a key issue for success. This difference in sharing and exchanging research results has led to a significant number of open-source programs and information resources in BI, whereas efforts in MI have often been local and private.

The different application domains are also reflected in education. The typical MI trainee gets his/her education in a medical setting (often a medical school), whereas the focus in BI is in biology. Cognitive reasoning, teaching, terminologies, research environments, social circumstances, and other characteristics are different in medicine and biology, imposing some restrictions for integrated approaches.

5.2. Bridging solutions

Medicine benefits already from the achievements of biological research [24], whereas biology will benefit from the use of clinical data for research [25]. As the domains begin to overlap, both communities will increasingly explore additional interactions. Examples include, among others, the development of ontologies and taxonomies, the use of natural language processing and information retrieval.

Collaboration is driven by two principal factors. First, the results of research in molecular biology will increasingly move towards clinical research and clinical practice [26]. Second, the methodologies used by BI and MI will prove to have many similarities, allowing exchange of experience between the two fields. Finally, we should appreciate the changes that biomedical science in general is experiencing. There is a transition from a period of data starvation to a period of data overload—both in terms of research and patient data. We are standing on the threshold of a new era: we need computers not only to store the data we collect, but also to verify and expand the interpretations we are constructing. We suggest that future initiatives should fall into three categories: (1) stimulating integration at the informational exchange level, (2) initiating collaborations between the communities, and (3) training a new generation of scientists that speak both languages.

The proposed research agenda outlines the strategies and solutions coming from three different points of view or directions based on the flow of data and information. They are: (1) what can MI contribute to functional genomics, (2) what can BI contribute to individualized health care, and (3) how can the new area of BMI, including new combined approaches, contribute to genomic medicine. All these directions require the advancement of the enabling technologies necessary for the development of the solutions proposed in each of the above-mentioned areas. Fig. 1 shows a graphical representation of these different perspectives.

6. Collaborative agenda for BI and MI

Eighteen research lines have been identified and are grouped in various categories, as shown below.

6.1. MI in support of functional genomics

Genomic researchers are working to discover the molecular mechanisms of diseases. Access to and integration of data coming from the clinical setting is essential for functional genomics research. MI professionals need to accelerate the development of both the information models and the tools needed for these tasks.

(1) 'Phenotype' databases for clinical annotation of biological samples and for clinical validation of biological research results. To generate new knowledge from genomic and proteomic data, we need to combine the phenotypes, genotypes and proteotypes\(^1\) of very large numbers of patients, ideally from different parts of the world [27]. The medical community needs to adapt standardized annotations for biological samples, and to develop laboratory procedures that will facilitate comparison of genomic and proteomic test results [28]. When all data types of patient characteristics (phenotype, genotype and proteotype) are represented in a standardized, structured format, we will be able to realize the value of new genome-based technologies and to apply these techniques for the benefit of individual patients [29,30].

(2) Disease reclassification. The classification of diseases can now be reorganized, beginning at a molecular level, by using new insights in pathophysiology derived from functional genomics [31]. As mentioned above, the integration of complex databases from MI (clinical information) with those from BI (genome data) is required to validate functional genomic research [32]. For this validation process, issues such as data quality, appropriate sample sizes and common data models must be addressed [33].

(3) Informatics for supporting rational drug design and development. Post-genomic techniques and tools are already integrated into some of the key steps of the drug-development pipeline, including target identification and validation, lead compound finding and optimization, toxicity studies [34], patient typing and stratification for clinical phases [35]. The implementation of these new technologies is expected to increase efficiency, thereby reducing time to market and, ultimately, cost [36].

Even with more than 10,000 potential targets, the opportunity to dramatically transform the drug-discovery process through a combined in silico and lead compound development pipeline has so far been

---

\(^1\) Prototype: the total of all the proteins that are expressed within a cell, tissue or organism; by analogy with phenotype for physical, visible characteristics and genotype for the total of all the genes.
Fig. 1. A picture reflecting the viewpoints expressed in this paper. It reflects the interdisciplinarity of both MI and BI as well as of the new emerging disciplines of Genomic Medicine and the core of BMI. The arrows show the different perspectives related to potential synergies among the above-described areas.

Fig. 2. Representation of current and future technological developments in MI and BI.
overlooked. There are excellent opportunities to merge BI/cheminformatics and protein/DNA microarray technology with MI in studies of preclinical and clinical toxicity as well as of patient typing and stratification.

6.2. BI in support of individualized health care

(4) Including genetic data in the electronic health record. Electronic health records (EHR) are being increasingly used by practitioners for routine patient care; they contain an increasing amount of coded, structured data. Although genetic data are beginning to be included in EHRs, current records have not been designed to include specific requirements for representing genetic data [37]. As a result, genetic data are typically recorded as ‘laboratory data.’ Consequently, the use of the data is limited (e.g., family relationships are often recorded only minimally, limiting the possibilities for studying relationships among the phenotypes of relatives) [38]. If we expect to use genetic data in health care, models need to be developed that will support the optimal entry of and access to such data in electronic health records.

(5) Methods for personalized health care: guidelines and decision-making support systems. Clinical guidelines are standard means for disseminating clinical knowledge to support physicians in decision-making processes. Clinical guidelines are represented in different ways, e.g., as text documents (in paper or electronic form) or using specifications such as Guideline Interchange Format (GLIF) [39]. In all cases, guidelines incorporate various sorts of expert, professional recommendations for the diagnosis, treatment and prevention of particular diseases. An example of a guideline-based decision support is EON (http://smi-web.stanford.edu/projects/eon/) [40] or DIADOQ [41]. Using genetic information can further enrich the quality of clinical decision-making [42].

(6) Telegenetics. Telemedicine is already being employed to deliver a number of ‘genetic medicine’ services [43]. Telemedicine-enhanced services are being used by centers specializing in cancer genetics, clinical genetics, and reproductive genetics [44]. Many genetic centers that routinely use phone consultations with physicians and phone interactions with patients to help determine the need for genetic services or to prepare for an appointment are moving to Internet-based services that incorporate all of the requirements for security and confidentiality [45,46]. For genetic counsellors and medical geneticists, telemedicine has developed powerful tools that unite various kinds of distributed information: personal and family history, physical findings, and radiology and pathology results [47].

(7) Stratifying patients by their genetic profiles: molecular diagnosis, clinical trials, and pharmacogenomics. One of the benefits of the study of the human genome is the identification of the SNPs and haplotypes present in the human population. With this information, stratification of people based on their genetic profile will increase the knowledge regarding interactions between the environment and genetic traits and how such interactions affect the development of diseases [48]. Information on the different genotypes together with phenotypic and environmental information will allow us to improve the design of clinical trials and to optimize treatments [49]. Such new therapeutic approaches will blur the boundaries between diagnosis and pharmacology, leading to the potential for more ‘personalized medicine’ [50]. A robust BMI infrastructure is needed for the integration of genetic and environmental data into clinical studies (clinical trials), and for the design of personalized therapeutic interventions based on additional ‘stratification’ information.

(8) Point-of-care data collection and access. At present, genetic data are only being collected by major clinical research laboratories. New DNA/protein detection technologies are developing rapidly (e.g., biochips or lab-on-a-chip) and these technologies will not require complex laboratory environments to perform sophisticated tests. These new analytical devices will enable us to establish patients’ genetic profiles within a reasonable time (and at reasonable expense) at the point-of-care [51]. Such advances will generate significant challenges for data processing, handling, communication, security and storage [52]. Proper use of these data will require additional medical education and training.

(9) Complexity in characterizing genomic and phenotypic microbial diversity related to infectious diseases (Microbial genomics). Microbial genomics require whole-genome sequencing coupled with BI tools to facilitate assembly, gene prediction, and functional annotation. This approach has revolutionized our understanding of the biology of important human microbial pathogens [53]. Comparative genome analysis provides insights into adaptations of microbes to their ecological niches and allows the detection of factors that shape host–pathogen interactions [54]. There is considerable evidence that genetic polymorphism in both the microbial pathogen and host can affect both microbial virulence and host immune responses to infection. The elucidation of microbial pathogen genomes will contribute to the characterization of genomic and phenotypic microbial diversity as it relates to infectious diseases, will allow the rapid identification of microbial pathogens by means of genetic markers, and will shed light on the mechanisms of pathogenicity and antibiotic resistance [55].

6.3. Biomedical informatics in support of genomic medicine

Biomedical informatics will stimulate new approaches to diseases and health, in which all levels of information (from the molecule to the population, going through the
cell, the tissue, the organ, the patient, and the disease itself) will be integrated and processed. Techniques and methods may be variably applied, depending on the nature of the problem being addressed. Some methodology will come from BI and some from MI, including public health and epidemiology informatics.

(10) Molecular and functional imaging. Molecular imaging facilitates the characterization and measurement of biological processes in living animals—including humans—at tissue, cellular, and molecular levels [56]. Molecular imaging will continue to build on existing technologies in positron emission tomography (PET), computerized X-ray tomography (CT), high-field magnetic resonance (MR) and MR spectroscopy, optical imaging, and image analysis [57]. Significant informatics tools are needed to support molecular imaging. These fall into two types:

- Understanding correlations—biostatistics and machine learning to identify significant imaging, genomic, and clinical factors that can help answer clinical questions and make clinical predictions.
- Elucidating molecular disease pathology—integrated genomic and protein-interaction databases, pathway elucidation, analysis, modelling and simulation, and prediction [58].

While the bulk of molecular imaging research funding is currently focused on cancer, we foresee opportunities in cardiovascular disease as well as in neurological diseases such as Alzheimer’s.

(11) Modelling and simulation for an approach that integrates physiology and pathology. The discovery and evaluation of diagnostic and therapeutic agents will be accelerated and made less costly through the creation and use of integrated dynamic models of processes taking place in cells and tissues. These in silico models will combine, unify and reconcile genomic and proteomic data for better understanding of complex diseases that involve many molecular species and many cellular states. BI and MI professionals can make major contributions to these modelling and simulation activities. Such in silico approaches would not have been possible without results derived from applying theories of non-linear dynamical systems, recent advances in the measurement of dynamic processes in individual living cells, and characterizations of physical properties of biological objects. The characterizations range from the elasticity of DNA to mechanical properties of cells and tissues in different physio-pathological situations [59]. Models can be built by combining two complementary approaches: (1) top-down, from clinical manifestations to inner mechanisms, and (2) bottom-up, from molecules to clinical manifestations [60]. Only formal models can provide a unified abstraction for dealing with the inherent multi-scalar, complexity, non-linearity, and self-organization of living systems, the diversity of pathophysiological processes, and the design of optimal diagnosis and therapy [61]. Development of shared libraries of in silico models of molecules, interactions, pathways, and functions will be needed.

(12) Epidemiology: biobanks and population repositories. The development of new genetic information technologies will lead to cost-effective screening (genetic tests) at a population level. The intersection of genetic data from such testing with clinical data (such as is held in electronic health records), environmental data, and ‘lifestyle data’ should lead to an elucidation of polygenetic disease causality. When these data are included in population repositories or biobanks, it can be applied to public health projects such as disease-prevention programs that can be targeted to at-risk patients identified by genetic information. A more accurate assessment of the cost-efficacy of pharmacogenetics approaches in health systems will also be possible. Several initiatives in the US and in Europe have already started. Some examples are the CDC with the Human Genome Epidemiology Network (HuGENet) (http://www.edc.gov/genomics/hugenet/default.htm) and the National Cancer Institute in the USA. There is an ongoing project in Iceland (Decode) (http://www.decode.com/) that links health records with genealogical and genotypic information. Other on-going projects are also being carried out in the UK (http://www.wellcome.ac.uk/en/1/biovenpop.html) and Estonia [62].

(13) New methods for e-learning in genomic-based medicine. Due to the increasing amount of clinical knowledge derived from genomic-based medicine, physicians must update their knowledge of genetics and genomics [63]. Learning is enhanced when learners identify their own needs, select their own strategies, and evaluate their own learning outcomes [64]. Internet-based informatics tools can facilitate education regarding the changes in molecular medicine in a non-disruptive manner, minimizing physicians’ rejection [65]. The introduction of new learning technologies that provide open and flexible learning programs will be crucial for the improvement of doctor’s skills and knowledge [66].

6.4. Enabling technologies

(14) Security. Genomic medicine and the associated interplay between aggregated data and individual data have given rise to concern with regard to the proper collection, storage, communication, and processing of individually identifiable sensitive information. Besides the more traditional security issues dealing with confidentiality, integrity, availability, and accountability, additional attention should be given in particular to privacy and identity protection [67,68]. More advanced privacy enhancing and protecting measures using privacy enhancing techniques (PETs) need to be deployed. These techniques become of even more importance...
when storing, exchanging, and processing highly sensitive genetic information. Typical examples of privacy-related issues and techniques are: anonymization, pseudonymization, data linkage, gauging for direct and indirect reidentification risks in databases and GRID environments, proxy services, systems for controlled database dilution, and privacy-enhancing intelligent software agents [69].

(15) Communication standards—interoperability among clinical and genetic information systems. Communication between all information layers is necessary and has to be provided in a trustworthy way [70]. Services have to be developed, implemented and maintained for communication security and application security for heterogeneous distributed networks [71]. Interoperability is the prerequisite for communication and must be addressed in the following areas:

- Data and knowledge (structure, representation, terminology, etc.)
- Technology (architecture, hardware, and topology)
- Presentation of data and knowledge
- Security for systems, health-care professionals and patients.

(16) Knowledge representation techniques and novel approaches for the virtual integration of heterogeneous clinical and genetic databases. Biomedical information at many different sites is now accessible over the Internet. In order to leverage this increasing pool of data, researchers need novel methods to search and retrieve information that must be integrated, gathered, classified, and interpreted. To integrate distributed databases, two levels of heterogeneity must be considered: (1) databases may be located in various platforms, spread over the Internet, with different architectures, operating systems, and database management systems, and (2) databases can present different conceptual data models and different underlying database schemas [72]. Tools to solve these problems include standards for exchanging information such as XML (http://www.w3.org/XML/) [73,74] and HL7 (http://www.hl7.org/) as well as standards for connecting biomedical devices. Standard-based tools will facilitate the integration of databases in data warehouses, federated databases or virtual repositories [75]. Rather than focusing on the unlikely possibility of the development of a single terminology to cover all domains, the emphasis should be on semantic mapping between terminologies (including clinical and ‘non-clinical’). In this regard, the development of integrated ontologies will be essential for many research issues in BMI [76,77].

(17) Data and text-based knowledge discovery. Data mining is a step in the process of generating knowledge in databases [78]. It includes techniques for query databases, on-line analytical processing, and machine-learning algorithms. In the medical area, many applications have been created for decision support to address issues such as image and signal analysis and outlining clinical prognoses for patient conditions [79]. In biology, efforts have been centered on research issues such as the prediction of protein structures and drug studies [80]. Both types of predictive exercises present considerable challenges for future research. Text mining is a discipline that aims to extract data, information or knowledge from texts [81]. Finding information in biomedical databases using text mining and information-retrieval techniques is expected to leverage a substantial amount of biomedical information that has escaped analysis until now.

(18) Grid applications in biomedicine. Linking computers through Grid middleware will enable users to access additional computing power to retrieve information from heterogeneous and distributed sources without having to choose to which machine he/she wishes to connect [82]. In the last few years, the term ‘Grid’ has evolved to encompass a concept of ubiquitous and transparent computing. Today, the Grid implies a shared and coordinated knowledge structure as well as a vision for intensive computing [83]. Although Grid technology and standards are still being developed, the vision is to create a Grid-based environment for both knowledge discovery as well as for high-performance computing power. In such an environment, information at various levels (molecule, cell, tissue, individual, and population) can be associated to provide ‘personalized’ health care. Collaborative efforts leading to the creation of a Health-GRID (http://www.healthgrid.org/) community that will provide basic common services (web portals, computing resources) is a first step in this direction [84,85].

All of these technologies are required for MI and BI. While some of the technologies, such as probabilistic expert reasoning, standards development and vocabulary development, are mainly developed in MI, they are of use in BI. In turn, BI has developed other technologies, such as database integration and automatic annotation, that can be used by medical informaticists. The middle part of Fig. 2 shows technologies that have received a lot of attention because they are (or will be) required in both MI and BI. Genomic medicine is expected to use such technologies increasingly to improve and enhance health care, including personalized therapies, preventive medicine and molecular medicine. Table 1 shows a summary of the 18 priorities proposed for an R&D agenda in MI, BI, and BMI regarding applications for functional genomics, individualized health care and genomic medicine as well as their priorities for future programs and institutional actions.

Table 1 shows below a summary of the 18 priorities proposed by the BIOINFOMED team for a R&D agenda in MI, BI, and BMI, regarding applications in functional genomics, individualized health care, and genomic medicine, and their priorities for future programs and institutional actions.
7. Conclusions

In this paper, we have presented the results of the BIOINFOMED study, carried out by research groups from three different European centers and 30 experts in BMI and related areas. From the responses given to a preliminary questionnaire and two expert meetings carried out with support from the European Commission to discuss the contents of early drafts, a White Paper was elaborated. It has been presented here in a summarized version.

The success of the Human Genome Project (http://www.genome.gov/, http://www.ornl.gov/TechResources/Human_Genome/home.html) promises to introduce significant advances and challenges in biomedical research and practice. Applications of technologies such as microarrays are routinely used in many biomedical research settings, but some innovative approaches are still needed to change significantly current practices beyond laboratory tests or preventive plans. In this regard, BMI can contribute to accelerate this process by bringing computerized methods to collect, integrate and

| Table 1 |
| Summary of the research priorities proposed by the BIOINFOMED study |

<table>
<thead>
<tr>
<th>Barriers</th>
<th>Proposed solution</th>
<th>Priority</th>
<th>Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Enabling technologies</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High computational and data management requirements</td>
<td>Grid technologies</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Strong privacy issues associated to the nature of genetic data</td>
<td>Security technologies</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Need to expand current interoperability standards for new genetic data infrastructure</td>
<td>Data communication standards</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Heterogeneity of current clinical and genetic sources and databases. Different representation systems (i.e., ontologies) in medicine and biology</td>
<td>Ontologies and new approaches to integrate heterogeneous clinical and genetic databases</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Data and text growing exponentially. New tools demanded for analysis</td>
<td>Data and text mining</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td><strong>MI in support of functional genomics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient care data are not been systematically used in genomic research.</td>
<td>Phenotype databases suitable for genomic research</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Lack of accepted standards for clinical validation of results obtained from functional genomics research</td>
<td>Disease reclassification</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Lack of adequate matching between biomedical data and pharmaceutical targets</td>
<td>Pharmacogenomics</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td><strong>BI in support of individualized healthcare</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unavailability of models for including genetic data into electronic health records</td>
<td>Genetics data model for the EHR</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>Increased complexity in medical decision making due to new genetic knowledge</td>
<td>Clinical guidelines and decision making using genetic information</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>Scarcity and non-uniform geographic distribution of clinical genetics specialists and resources</td>
<td>Telegeneics</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Methods needed for stratifying patients by genetic profiles in the context of clinical research</td>
<td>New methods and information platforms to manage genetic data in clinical research</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Lack of interoperable devices to collect genetic data and include them in clinical information systems</td>
<td>Point-of-care data acquisition systems</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>Complexity in characterising genomic and phenotypic microbial diversity related to infectious diseases</td>
<td>Microbial genomics</td>
<td>Medium</td>
<td>Low</td>
</tr>
<tr>
<td><strong>BMI in support of genomic medicine</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of high resolution systems to correlate anatomical structures to physiological and genetic mechanisms</td>
<td>Molecular and functional imaging</td>
<td>Medium</td>
<td>Low</td>
</tr>
<tr>
<td>Lack of unified approaches to understanding and modelling the human body and human diseases.</td>
<td>Modelling and simulation</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>Linking environmental and lifestyle information to genetic and clinical data</td>
<td>Populational repositories</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Narrow view of genetics and genomics in health professionals and patients</td>
<td>e-Learning</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

*The priorities and risks arise from the results obtained from the questionnaires sent to the experts and the subsequent debates and discussions among them.*

*Risk refers to the risk of failure to deliver results.*
analyse combined clinical and genomic data, information, and knowledge.

The proposed agenda developed within the BIO-INFOMED study intends to provide different clues and directions to advance research in BMI, leading to new approaches and achievements in both functional genomics and genomic medicine. The development and implementation of the research priorities and the enabling technologies described in this paper would be facilitated and achieved by collaborative efforts between MI and BI. The members of the BIOINFOMED study believe that such interaction will lead to a synergy and significant developments in the new area of Biomedical Informatics.

Acknowledgments

While this report is produced under the sole responsibility of the Workgroup members and based on a consensus process and individual contributions to various sections of the report, we would like to acknowledge the support of Commission staff, in particular the DIRECTORATE GENERAL INFORMATION SOCIETY (Ilias Iakovidis2 and Sofie Nørager2), in advising and commenting on previous drafts. We thank FORTH for their hospitality during the workgroup meeting held in Crete, Greece (June, 2002). We thank the Spanish Health Informatics Society (SEIS) for sponsoring the EC-IST BIOINFOMED Project Meeting (November 18 and 19, 2002) under the umbrella of their Congress BIOINFORSALUD 2002, held in Valencia, Spain. We acknowledge Professor Ferrán Sanz Carreras of the University Pompeu Fabra of Barcelona, Spain and Professor Gregorio Martin-Quetglas of the Robotics Institute of the University of Valencia for their input during the feedback session by the experts in Valencia. We also acknowledge, and are particularly grateful, to the unfailing assistance of Francisco Javier Vicente and Isabel Hermosilla (ISCIII) for their assistance with the production of this report.

References


