"Synergy between Medical Informatics and Bioinformatics: Facilitating Genomic Medicine for Future Healthcare"

Martin-Sanchez F.[1] Iakovidis, I., Nørager, S.[2], Maojo, V.[3] de Groen, P.[4] Van der Lei, J.[5] Jones, T.[6] Apweiler, R.[7] Babic, A.[8] Baud, R.[9] Breton, V.[10] Cinquin, P.[11] Doupi, P.[12] Dugas, M.[13] Eils, R.[14] Engelbrecht, R.[15] Ghazal, P.[16] Jehenson, P.[17] Kulikowski, C.[18] Lampe, K.[19] De Moor, G.[20] Orphanoudakis, S.[21] Rossing, N.[22] Sarachan, B.[23] Sousa Pereira, A.[24] Spekowius, G.[25] Thireos, G.[26] Zahlmann, G. Abraham-Fuchs, K.[27] & Zvárová, J.[28]

[1] Institute of Health Carlos III. Madrid, Spain [2] European Commission. Directorate General Information Society. Brussels. Belgium [3] Polytechnical Univ. of Madrid. Madrid, Spain [4] Division of Gastroenterology and Hepatology, Mayo Clinic and Foundation Rochester MN. USA [5] Dept. Medical Informatics. Erasmus Medical Center. Rotterdam.The Netherlands [6] Oracle Corporation. Redwood Shores, CA, USA [7] EMBL/EBI. Cambridge, United Kingdom [8] Dept. of Biomedical Eng. Linköping University. Linköping, Sweden [9] University Hospitals of Geneva, Geneva, Switzerland [10] CNRS-IN2P3. Aubierre, France [11] Laboratoire TIMC. Université Joseph Fourier, Grenoble, France [12] Dept. Medical Informatics. Rotterdam. Erasmus Medical Center. The Netherlands [13] Department of Medical Informatics, Biometrics and Epidemiology; University of Munich, Germany [14] German Cancer Research Center. Heidelberg. Germany [15] GSF-Medis Institute. Munich. Neuherberg. Germany [16] Genomic Technology and Informatics Centre. University of Edinburgh. Summerhall, UK [17] Joint Research Center European Commission. Germany [18] Computer Sciences. Rutgers Univ. NJ, USA [19] Finnish Office for Health Care Technology Assessment, STAKES. Helsinki, Finland [20] Ramit VZW, Gent, Belgium [21] Institute for Computer Science - FORTH, Heraclion, Crete, Greece [22] Copenhagen Hospital Corporation, Copenhagen, Denmark [23] GE Research Center. Niskayuna, NY, USA [24] IEETA - University of Aveiro, Aveiro, Portugal [25] Philips Research Laboratories, Aachen, Germany [26] Institute of Molecular Biology and Biotechnology. FORTH. Heraclion, Crete, Greece [27] Siemens AG Medical Solutions. Erlangen, Germany [28] EuroMISE Center of Charles University and Academy of Sciences, Czech Republic

Fernando Martin-Sanchez
Area of Medical Bioinformatics
Institute of Health "Carlos III"
Ctra. Majadahonda a Pozuelo km2
28220 Madrid
e-mail: fmartin@isciii.es
fax: +34 91 509 79 17

Abstract

Medical Informatics (MI) and Bioinformatics (BI) are two interdisciplinary areas located at the intersection between computer science and medicine and biology, respectively. Historically, they have been separated and only occasionally have researchers of both disciplines collaborated. The completion of the Human Genome Project has brought about in this post genomic era the need for a synergy of these two disciplines to further advance in the study of diseases by correlating essential genotypic information with expressed phenotypic information. Biomedical Informatics (BMI) is the emerging technology that aims to put these two worlds together in the new rising genomic medicine. In this regard, institutions such as the European Commission have recently launched several initiatives to support a new combined research agenda, based on the potential for synergism of both disciplines. In this paper we review the results the BIOINFOMED study one of these projects funded by the EC.

Introduction

The complete sequencing of the Human Genome has opened, in this post-genomic era, new perspectives for the study of monogenic and complex multigenic diseases, the later being more prevalent. As our knowledge about the human genome increases so does our belief that to fully grasp the mechanisms of disease we need to understand the complex interplay among genes and environmental factors that initiate pathological processes. Therefore, allowing scientists to obtain those data needed to close the classical equation (genotype + environment = phenotype).

The new genetic and proteomic data has brought forth the possibility of developing new targets and therapies based on these findings, of implementing newly developed preventive measures and also of discovering new research approaches to old problems. To carry out the work it is important to be able to deal with the large amount of data generated in the laboratory by functional genomics and proteomics (Bioinformatics). Also, it is just as necessary to integrate the information derived from this data into the electronic health records together with the knowledge generated in the clinical setting (Medical Informatics). The coupling of BI with the tools and techniques that deal with clinical information (e.g. electronic health records, clinical decision systems, image- and signal-processing), will allow to correlate essentially genotypic information with expressed phenotypic information. Biomedical Informatics (BMI) is the emerging technology that aims to put these two worlds together in order to participate in the discovery and creation of novel diagnostic and therapeutic methods in the new rising genomic medicine.

The need for an integrated approach to medicine was mentioned by the WHO "*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 practices, 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*".

Background

Medical informatics (MI) and bioinformatics (BI) are two interdisciplinary areas located at the intersection between informatics and medicine and biology (genomics), respectively. Historically, they have been separated and only occasionally have researchers of both disciplines collaborated in the past. Almost from its very beginning MI mainly focused on the development of practical computer applications for health purposes. Recently, a debate has opened up about its scientific content and future. Meanwhile, BI is a younger discipline, which has grown enormously thanks to its contribution to genomic research.

BI and MI until recently had no reason for synergy. It is the elucidation of the human genome what has promoted the need for a synergy between the two. Classical epidemiology and clinical research on the one hand, and genomic research on the other, alone are no longer

enough for advancing in genomic medicine, and a new integrative approach is required. The integration of all the data and information generated at all levels requires synergy of Bioinformatics and Medical Informatics.

The history of the development of BI is different of that of MI. While the latter has been around since the introduction of computers in the hospitals and was developed mainly from an application-centred perspective, BI was developed to handle large amounts of data, mainly sequences, generated in the laboratories. Nowadays, there are a number of initiatives that combine elements of the two areas up to the point of an integrated approach on databases, standards, analysis, applications and education. This evolution towards joint actions is also observed in scientific meetings and publications.

At present the interaction between the two communities, medical informatics and bioinformatics, is increasing as shown by the number of congresses and conferences that deal with this subject. The need for this new interdisciplinary area has been realized by different institutions and entities. The American Medical Informatics Association, for example, has considered such interaction between MI and BI as the focus for its recent conferences. There are also cross programming of activities (panels, tutorials, sessions) in main congresses in both fields (MEDINFO, AMIA, MIE, ISMB, PSB, RECOMB), conferences and special issues of journals dedicated to the intersection and mutual interests, the inclusion of bioinformatics in training programs of health centers and medical schools, and the collaboration with the pharmaceutical industry for physician training and technical support in genomics.

The Conference "Synergy between Research in Medical Informatics, Bioinformatics and Neuroinformatics", (Brussels, Dec. 14[th,] 2001, www.ramit.be), organised under the Belgian Presidency of the European Union, by the Belgian Federal Ministries of Social Affairs and Public Health and the European Commission - Directorate General Information Society and Directorate General Research also followed this line of work. More than 400 participants gathered in Brussels, to discuss the scientific and developmental issues involved in MI and BI related to future healthcare.

This conference was the kick-off point for the activities related to the BIOINFOMED project ("*Prospective Analysis of the Relationships and Synergy Between Medical Informatics and Bioinformatics*"). The aim was to analyse the relationships and potential synergies between MI and BI. Several goals were set to be met during the duration of the project, among these there was the writing of a white paper were a research agenda was identified based on the potential for synergism of both disciplines. Several experts out of the almost 400 that attended the conference and that were particularly interested in the subject were invited to collaborate in the writing of a white paper with the group carrying out the project at the Institute of Health "Carlos III" (ISCIII), together with the Polytechnic University of Madrid and the University of Linköping. Two group work meetings were held, one in Crete (Greece) and another one in Valencia (Spain).

The group in charge of the project at the Institute of Health "Carlos III" (ISCIII) designed a questionnaire to identify research questions, areas, research directions, and potentials of already existing tools and disciplines within MI that could be applied in BI and vice versa. Once the team at the ISCIII processed the answers, these were used as a starting point for the paper, which included a brief background and an overview of the expected impacts that the

integration of clinical and genetic information facilitated by the synergy between BI and MI could have within the different sectors of society, as well as the proposed R&D agenda.

There are, however, barriers in the application and development of new activities required for the integrated approach. These barriers can be overcome by collaborative efforts between the two disciplines thus bridging the existing gaps.

Biomedical Informatics (BMI), rising from the synergy between BI and MI, provides the framework for developing and sharing new biomedical knowledge. New knowledge about the causes and treatments of disease will not be created as quickly without a dynamic, rational biomedical information environment. Since the creation of new knowledge is often accompanied by anxiety, BMI should provide clear ways of alleviating anxiety by properly informing all the stakeholders in the biomedical world of risks and realities. In order that the field develops at a proper pace, a dispassionate discussion of the impacts of the biomedical revolution is essential. In order that clinical care and basic biological investigations continue to address the health of the citizen, BMI must be effectively resourced.

Expected Impacts

In our view, the biomedical community seeks to remove the walls between biological information and medical information, to foster communication between clinician and scientist, and to enhance understanding between citizen and health care professionals. Our commitment to interoperability of biological and medical information for all appropriately authorised users creates imperatives, opportunities, and challenges. Equally significant demands are made by the evolution from patient-centred systems to citizen-centred systems that actively engage citizen participation.

a. Scientists / Researchers

They must become accustomed to exchanging and sharing medical and biological information and knowledge in global (often virtual) work settings. In addition, all workers will be challenged to more directly consider the ethical implications of research activities and to more deeply comprehend the repercussions of their work.

Clinical Trials

Biomedical databases are urgently needed to provide a sound scientific basis for what kind of genetic tests make sense and which tests just make healthy people anxious about their future. BMI bears a significant potential to clarify the sensitivity and sensibility of genetic testing.

b. Health Care Professionals

New knowledge and technology

The very nature of BMI highlights the blurring of hitherto comfortable distinctions between clinical and molecular information. As we extend the concept of phenotype ("the visible

properties of an organism that are produced by the interaction of the genotype and the environment"[1]) to encompass diseases as well as hair colour and body shape, 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 minute variations in DNA sequences, protein synthesis and subsequent protein function affect the evolution of diseases. Genomic and proteomic data analysis has already hastened both the elucidation of causes for disease and the development of drugs to combat disease. As our knowledge about molecular causes of disease increases, we can expect more elegant molecular interventions to diagnose, disrupt or ameliorate disease. BMI professionals may provide methods and tools for R&D in these issues. Greater magnification will be focused on how many different environmental changes affect phenotypic expression of genetic information.

Professionals in supportive role

Even as full-scale BMI exercises can create more knowledge, they can also create more anxiety. Genetic counselling will become an even more important part of clinical, hands-on care. We see a place for "culture brokers" – i.e. people who can translate between science and clinical care and between science and the "self caring" citizen.

c. Individual Citizens

The citizen "at risk"

New BMI approaches can result in the creation of a new role –"citizen-at-risk." As the knowledge base about genetic associations with illnesses becomes larger, it is likely that this identifiable "at risk" group will enlarge, encompassing many asyptomatic citizens and therefore placing new demands on health care systems.

Informing citizens

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

d. Health Care Providers and Systems

Technology diffusion and scientific evidence

European health care systems will face difficult challenges with the emergence of novel BMI applications: how and when to adopt them, particularly since new health technologies tend to be more expensive than old ones. In regard to several modern technologies, adoption has not followed a rational, evidence-based pattern. Biomedical informatics applications should not be

---

[1] Merriam Webster's Collegiate Dictionary – Merriam-Webster, Incorporated, Springfield, MA, 1995

brought into use as a result of market push. Rather, 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.

Feasibility of current care practices

Current treatment models may become obsolete as new biomedical knowledge is created. Health care providers should pay special attention to current developments in the field of BMI in order to be able to predict practical implications for everyday diagnostic/treatment routines.

Public health and disease prevention

Knowledge generated through very large biomedical databases will enable health care organisations to identify citizens who are not only at "genetic" risk for developing illnesses but whose risk of developing symptomatic illness could be reduced by one or more interventions. As we identify more and more genetically "at risk" citizens, more focused management programs must evolve.

e. Policy - Decision makers

Investment for the future

Rational biomedical databases of the scope required for modern molecular research are expensive to establish and maintain. Spending adequate time on system design and architecture may pay off in the long run. Programmed cooperation between government, academia, and industry is absolutely essential.

Prioritisation

No health care system is able to provide the best possible care for each and every condition; some level of prioritisation is done -either consciously or unconsciously. Genetic testing and the associated concept of "citizen-at-risk" will constitute yet another aspect to the prioritisation palette.

Legislative initiatives

Biomedical informatics will bring about several ethical and societal issues that warrant careful societal discussion. Policy makers should see to it that they initiate and foster such discussion at an appropriate time, thereby providing citizens with adequate information to participate. The use of 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.

f. Industry

In order for certain industrial efforts to succeed (such as pharmaceutical and bio technology), 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 data bases and knowledge bases by academic institutions should contribute to the financing of such public initiatives and collaborative efforts among different institutions.

g. Society

Consent to collect, view and use information

Genomic and proteomic databases must be secure from unwanted intrusions. Correlation between clinical profiles and genomic/proteomic profiles should only take place when informed consent has been obtained. Public health reasons for violating these principles must be explicit and must result from public debate. Further, every citizens' rights to not know about his/her genetic risk should be respected.

Genetic discrimination

 Any citizen would be uneasy about knowing that he/she was at risk for an illness that could lead to diminished job performance and very concerned about having such information available to third parties such as employers, insurers, financial institutions, etc. 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.

Racial profiling

Today, a "racial profile" based on genetic information derived from blood analysis is commercially available to anyone, although the meaning of such information is not at all clear. Scientists are already engaged in debates about ethnicity vs. race, and one can see how "genetic assessments" of this sort are invitations for misusing large biomedical databases. Great care must be taken to see that biomedical databases are not subjected to unauthorized analyses of this sort.

Fetal testing and pregnancy termination

At the present time, pregnant women may elect to have pregnancies terminated as a result of genetic testing of the fetus. As more knowledge is created about genotypes at risk for disease, more couples will be faced with decisions about whether to have fetal testing performed and whether to act on the results of such testing.

Gaps and bridging solutions

a. Gaps

Historically, MI and BI researchers have addressed different issues, used different methodologies, and got distinct sources of funding. MI matured in the broad and complex medical domain that is not only characterised by the delivery of patient care by many different specialists, but also involves, among others, research, administration, and management. Compared to MI, bioinformatics is more focussed, not only because of the human genome

project, but also because the main purpose of BI is to enable and support research. While the tools and applications developed by MI reach a wide range of users including physicians, nurses, administrators, management, and researchers. BI applications are characterised by a much more homogeneous user group dominated by researchers.

Although the application domains differ, both MI and BI will often use similar methodologies; both fields are active in machine learning, natural language processing, image analysis, or data mining in large databases. Working on similar problems with related methods, however, does not guarantee similar results because the application domains differ.

Another striking difference between the MI community and the BI community involves the degree of interaction between the research groups. In MI, collaborative efforts and research between different groups has 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 the medical school) whereas the focus in BI is in biology. The cognitive reasoning, classical teaching and terminologies are different in medicine and biology, limiting an immediate transfer or unification of courses.

b. Bridging solutions

Although the roots of BI and MI are located in different application domains, these domains will increasingly overlap. Results of research in molecular medicine will have an impact on clinical medicine. The shared application domain will provide a natural place to collaborate. Medicine will benefit from the achievements of biological research, and biology will benefit from the use of clinical data for research. As the domains begin to overlap, both communities increasingly will share a common goal, a common context, for exploring collaboration. Examples include the development of ontologies and taxonomies, the use of natural language processing, or information retrieval.

Two principal factors will drive the collaboration. First, the results of research in molecular biology will increasingly move toward clinical research and clinical practice providing a natural, shared issue: the application in daily practise. Second, the methodologies used by BI and MI will prove to have many similarities allowing exchange of experience between the fields. Finally, we should appreciate the changes biomedical science in general is going through. We are moving 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 desperately need computers not only to store the data we collect, but also to store, verify and expand the partial interpretations we are constructing of those data. Therefore, we suggest that future initiatives should fall into three categories: Stimulating information exchange, initiating collaborations between the communities and fostering a new generation of scientists that speak both languages.

Through our analysis we have seen the potential that both disciplines pose for an interaction. Not only do they share many interests, methods and tools but also each presents some complementary needs for the other.

The research agenda shows the strategies and solutions that come from three different points of view or directions based on the flow of data and information attending to the disciplines in which they are generated, processed and maintained. These are: what can MI contribute to functional genomics, what can BI contribute to individualised healthcare, and the new area of BMI, combined approaches included, and its contribution to genomic medicine. All these based on the enabling technologies necessary for the development of the solutions proposed in each of the above mentioned areas. This is further described below and graphically shown in figures 1 through 4.

- *MI in support of Functional Genomics*. Functional Genomics requires patient data coming from clinical information systems (laboratory tests, annotation of biological samples or familial history). MI can and should, therefore, play a role in facilitating this data for post-genomic research.

- *BI in support of individualized Healthcare*. The practice of medicine moves into the post-genomic era, there will be an increasing need for the practicing clinician, as well as for the medical informatician, to understand and use molecular level data. Knowledge of the concepts involved in acquiring, representing, analysing, and integrating such data falls in the scope of the bioinformatician. The collaboration between both disciplines will allow the real integration of genetic data of the patients in clinical information systems.

- *BMI in support of Genomic Medicine*. Which represents the development of a new discipline, BMI that deals with integrated approaches oriented towards analysing the knowledge of diseases or the personalization of clinical solutions using information coming from the different levels (molecular, clinical or environmental) that take part in disease development. BMI has to do with new perspectives that require the knowledge and the abilities to deal with multi-level information.

- *Enabling Technologies*. A sound and efficient computing, information and communication platform, a new generation of integrated analytical devices and virtual learning environments will play a key role in facilitating the implementation of all these scientific approaches upon which the research lines described in this work can be developed.
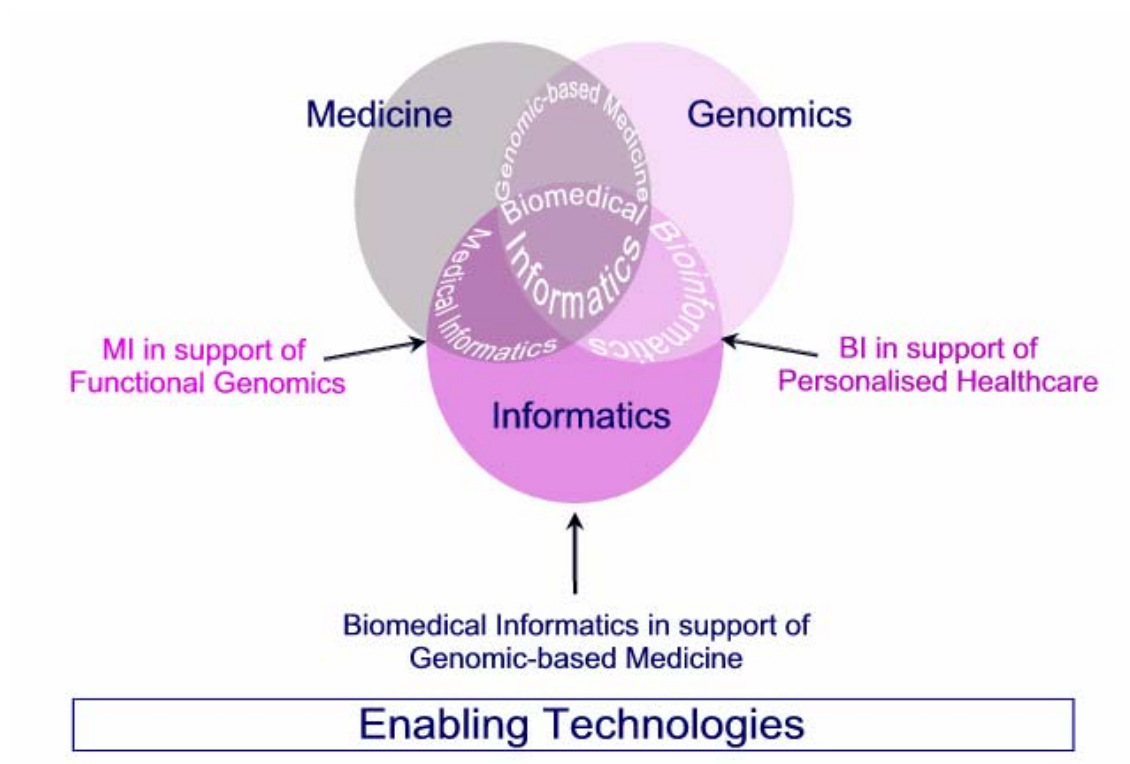
Figure 1

This diagram reflects the interdisciplinarity of both MI and BI as well as of the new emerging disciplines of Genomic Medicine and BMI. The arrows show the different perspectives related to potential synergies among the above described areas.
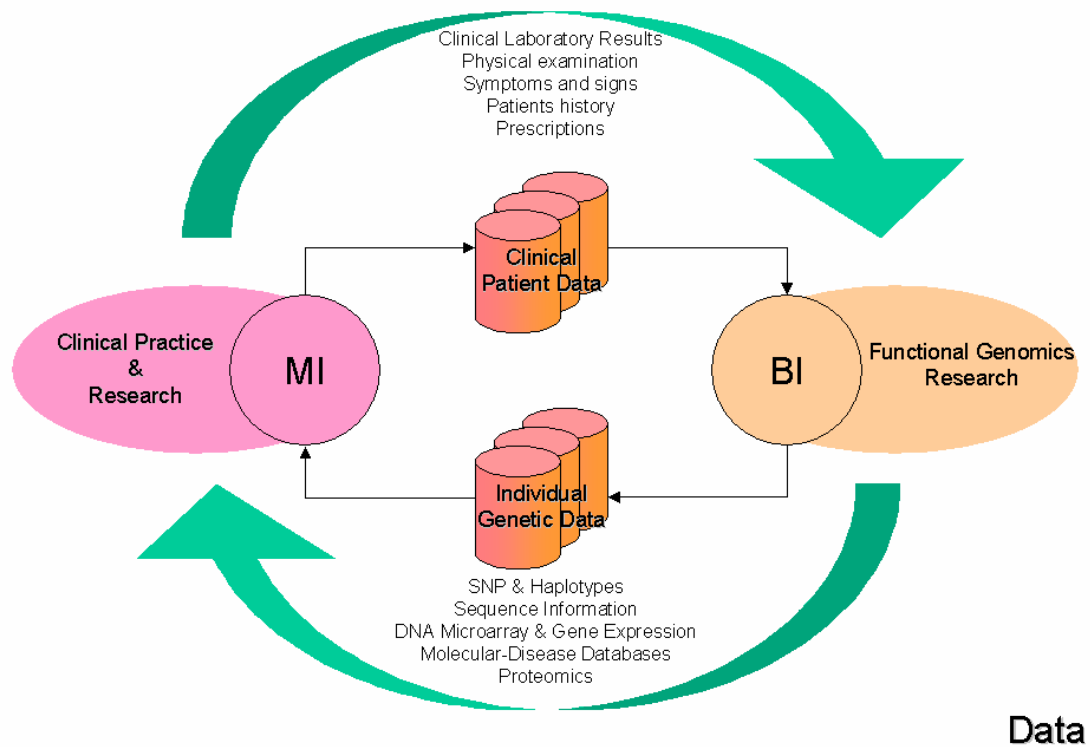
Figure 2

The diagram above shows the logical flow of data between MI and BI promoting synergy between the two disciplines. The top part of the diagram shows how data collected by medical informaticians during regular clinical practice and research, for example symptoms and signs or clinical laboratory data, can be made available in certain conditions (anonymized) to the bioinformaticians so they can utilize it in functional genomics. The incorporation of these data will allow to further advance in the research of the molecular bases of disease and to relate them to the genotypic characteristics of the patients. The bottom part of the diagram shows how data arising from research in functional and individual genomics processed and managed by bioinformaticians like, for instance, gene profiles, SNPs or haplotypes, can be incorporated in clinical information systems to complete patient records and to further care and personalized treatment of patients, as well as on the prevention of diseases and on clinical research.
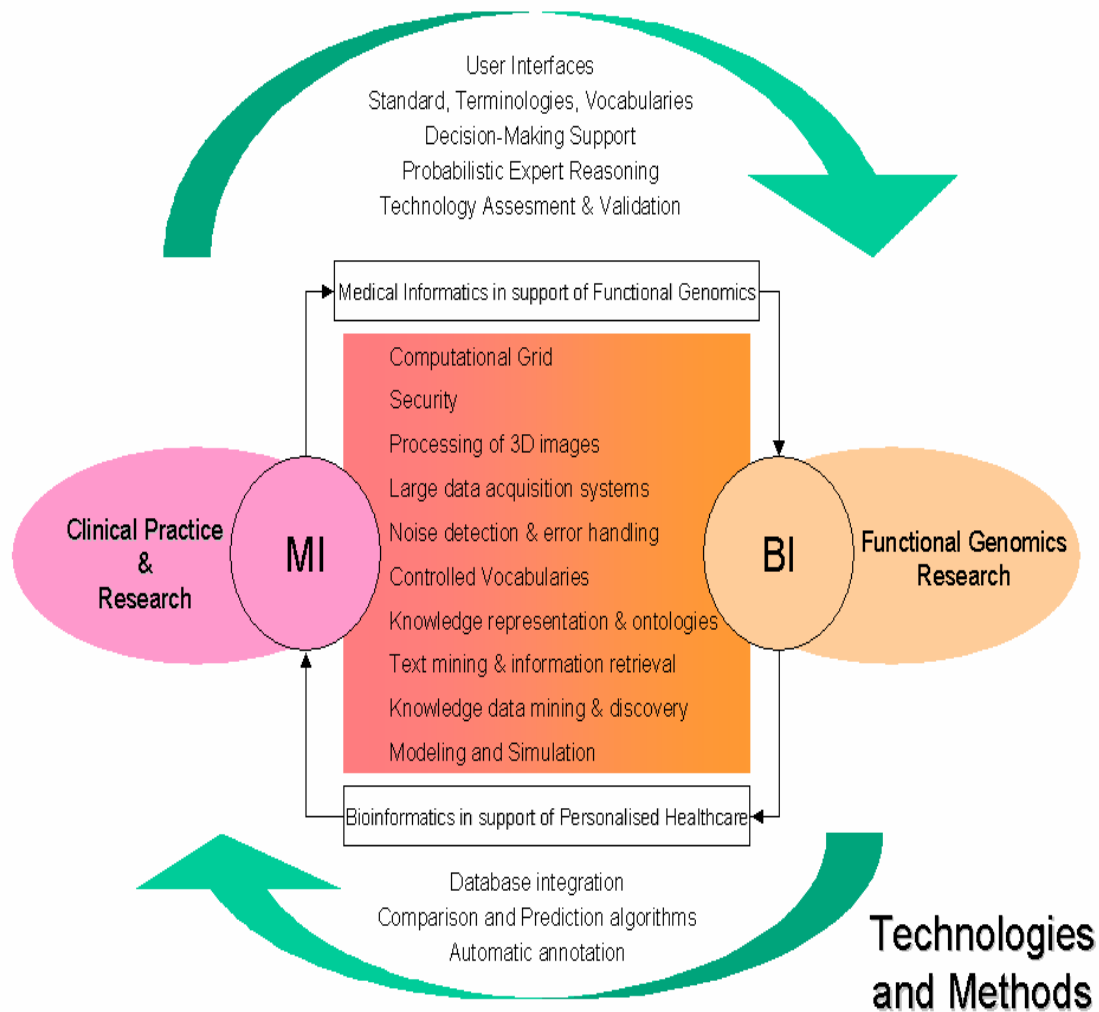
Figure 3

The diagram depicts the actual and future development of technologies used in these fields. All these technologies are necessary for MI and BI. However some of these technologies like probabilistic expert reasoning, standards or vocabularies among others, are mainly developed in MI but they are of use in BI. In turn BI have developed further certain technologies, for instance database integration and automatic annotation that can also be used by medical informaticians. The middle part of the figure shows technologies that have received recently big attention because they are or will be needed in both disciplines and will of course be utilized by genomic medicine to improve and enhance healthcare, including here personalized healthcare, preventive medicine and molecular medicine.
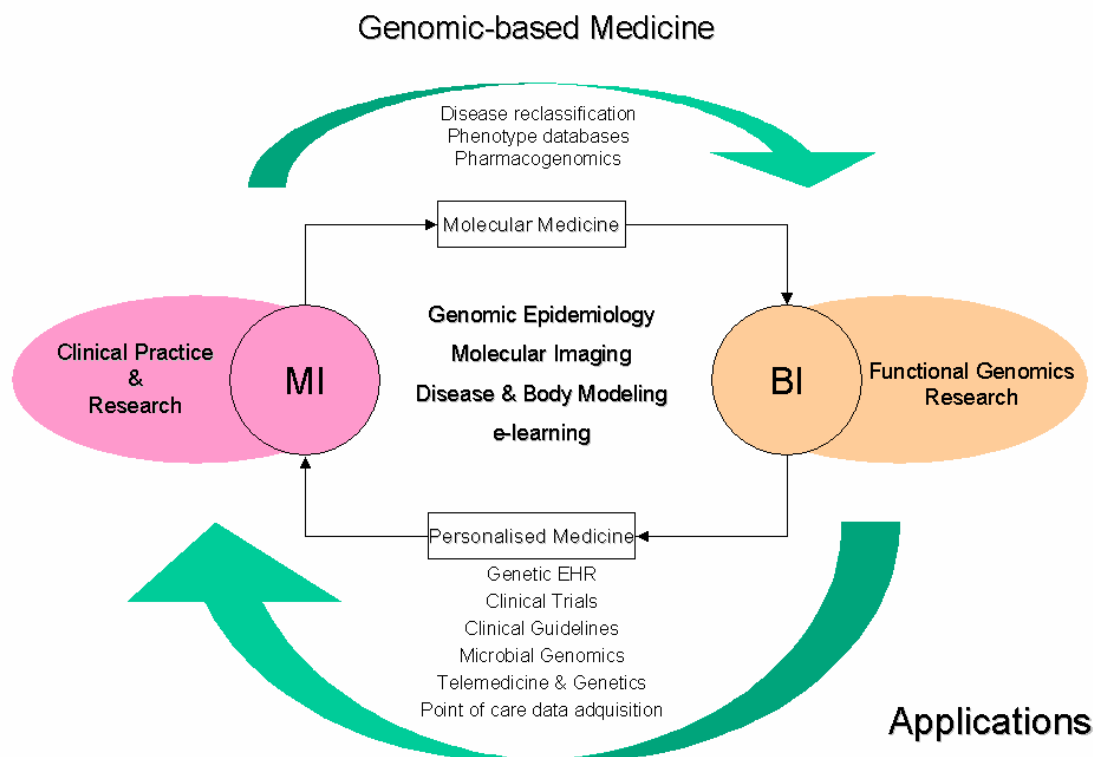
Figure 4

The top part of this diagram reflects that the use of data handled by MI and the incorporation of the technologies shown in Figure 3 enable the development of applications that could be included within molecular medicine, for example disease reclassification. The bottom part of the figure shows that the utilization of the data coming from functional genomics research processed by the technologies mentioned in Figure 3 give rise to new applications included in what is called personalized medicine based on genomics such as telemedicine or clinical trials. Applications that have emerged or will emerge in which the synergy between both disciplines is obvious are shown in the middle part of the diagram.

Collaborative agenda for BI and MI

A total of 18 research lines have been identified.

a. MI IN SUPPORT OF FUNCTIONAL GENOMICS

Genomic researchers will need to draw inferences about the molecular mechanisms of diseases. Therefore access and integration of data coming from the clinical setting is essential for functional genomics research. The challenge for medical informaticians is to adapt existing systems or to develop new ones to allow this exchange of data.

Phenotype databases for clinical annotation of biological samples and clinical validation of biological research results

The application of new technologies derived from the discoveries in genomics and proteomics requires (1) accurate definition of the clinical characteristics of each patient (the "phenotype") in a structured and computerized representation, (2) computational capabilities to interpret the large amounts of new, raw data and ability to store and retrieve the derived data in relational databases (the "genotype" and "proteotype"), and (3) processing tools and power to discover new relationships between the phenotype, genotype and proteotype, and create new knowledge.

To obtain new knowledge from the genomic and proteomic data we need to combine the phenotype, genotype and proteotype of very large numbers of patients, ideally from different parts of the world. This will only be possible if the medical community adapts standardized annotation of biological samples (description of the phenotype), and develops laboratory procedures that will allow comparison of genomic and proteomic test results. Standardized description of the phenotype can be achieved by structured, physician data entry (*a priori* definition of structured data elements), by computerized interpretation of the EMR content (*a posteriori* derivation of structured data elements from free text) or a combination of these two methods. The laboratory procedures, however, likely will require *a priori* guidelines for tissue handling as well as analytic protocols, and representation of the test results in a standardized format. Only if all data types describing patient characteristics (phenotype, genotype, and proteotype) are represented in a structured and standardized format, will we be able to assign value to the results of the new genome-based technologies, and apply these to the benefit of the individual patient.

Disease reclassification

Classification of diseases is enhanced to a molecular level by new insights in pathophysiology derived from functional genomics. It involves different patient characteristics like clinical findings and various diagnostic procedures. When classification of diseases is enhanced to a molecular level, knowledge from clinical research is being combined with functional genomics.

Because of the abundance of molecular markers, it is a challenge to distinguish between random and clinically relevant associations. The validation of results from functional genomics research involves the integration of complex databases from MI concerning clinical information and BI with respect to the genome data. High data quality, appropriate sample sizes and common data models are important success factors for this validation process.

Informatics for supporting rational drug design and development

Post-genomic tools are already well integrated into many of the key steps of the drug development pipeline, including target identification and validation, lead compound finding and optimisation, toxicity studies, patient typing and stratification for clinical phases. The implementation of these new technologies is aimed at increasing efficiency, reducing time to market and ultimately cost.

For historical pre-genomic and perhaps other pragmatic reasons the drug pipeline is geared for a "shot-gun" wet-lab approach to the target and lead compound discovery and development process. This approach fails to take full advantage of the post-genomic era. 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 overlooked. There exists excellent opportunity to merge fields such as BI/ cheminformatics, protein and DNA microarray technology with MI in preclinical and clinical toxicity, patient typing and stratification.

b. BI IN SUPPORT OF INDIVIDUALIZED HEALTHCARE

Bioinformaticians are playing a key role in the acquisition, processing and analysis of individual genetic information (SNPs, haplotypes). Therefore they can and should help to integrate genetic data obtained in functional and comparative (individual) genomics into the clinical information systems to aid in a true personalised healthcare. Knowledge of the concepts involved in acquiring, representing, analysing, and integrating such data will be essential in effectively applying molecular information in the diagnosis and treatment of complex medical disorders by the practicing clinician.

Including genetic data into the electronic health record

Current electronic health care records contain an increasing amount of coded, structured data. Although genetic data are beginning to be included in electronic health records, current records have not been designed to include the specific requirements of genetic data. As a result, the genetic data are typically recorded as "laboratory data" on the individual patient. Consequently, the use of the data is limited (e.g., family relationships are often recorded only minimally, limiting the possibilities to study relationships among the phenotypes of relatives). Based on the (expected) use of genetic data in health care, models need to be developed that will support the optimal use of the data in electronic health records. Optimal use will have to include the use of the data to provide decision support to the treating physician based on the available genetic data.

Methods for personalized health care: guidelines and decision making support systems

Clinical guidelines are standard means for dissemination of clinical knowledge and the support of physicians in the course of decision-making. Using genetic information can further improve decision-making quality.

Clinical guidelines are text documents (in paper or electronic form) containing various sorts of recommendations for the diagnosis, treatment and prevention of particular diseases. The task for which the computerized guidelines are most commonly used is the *support of the clinician in the course of decision-making*. Due to the safety-critical character of such online applications, they rely on complex knowledge representation and on combination of multiple inference strategies. Still, interaction with the medical staff is frequently needed. An example of a guideline-based decision support is the Stanford-based EON system. Combining clinical and genetic information and using nowadays decision- support and knowledge -based system decision-making quality can further improve quality of care in individual. Diagnosis and

therapy of diseases will be individualized with the support of genetic knowledge based systems and molecular expert systems.

Telegenetics

Telemedicine services are a reality nowadays, covering many scenarios (e.g. teleconsultation, remote monitoring, training and education, emergency, tele-surgery) and medical specialities (e.g. radiology, cardiology, obstetrics, pathology, psychiatry, genetics).

In the domain of genetic medicine there are currently a significant number of services being delivered using telemedicine. Services in the domains of cancer genetics, clinical genetics and reproductive genetics can be found in the literature. In fact, many genetic centres that routinely utilize phone consults 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, and incorporating all the needed security and confidentiality requirements. For genetic counsellors and medical geneticists telemedicine is a powerful tool bringing together multiple kinds of distributed information: personal and family history, physical findings, and radiology and pathology results.

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 at hand the stratification of people based on their genetic profile would allow to further the knowledge of the interactions between the environment and genetic traits and how they affect the development of diseases.

Information on the different genotypes together with phenotypic and environmental information would allow to better design clinical trials and to ultimately optimise treatments. This new therapeutic approach may facilitate the merging of diagnosis and pharmacology, hence the possibility of personalized medicine. There will be a need for a BMI infrastructure to make possible the integration and posterior management of this genetic and environmental data into clinical trials, and the design of personalised therapeutic interventions based on the available information.

Point-of-care data collection and access

At present, genetic data are typically collected by (larger) clinical or research laboratories. New DNA / protein detection technologies are developing rapidly (e.g. biochips or lab-on-a-chip) and will not require a complete laboratory environment to perform a test. The new analytical devices offer the possibility of accessing patients' e.g. genetic profiles within reasonable time and expenses at the point-of-care. These advances bring along a large number of challenges for the data processing, handling, distribution and storage.

- Interoperability and connectivity of point-of-care devices, data acquisition and analysis systems.
- Analytical devices as combined collectors of medical and genetic information, temporary repository of data, data query devices, and data entry point for MI-BI systems.
- Patient self testing and Web based software applications in the frame of individualized medicine.

❑ Support of individual diagnostic and therapy by genetic and proteomic data.

If general practitioners are going to be able to obtain these data, they will also need to access other complementary data, place them in context and assure their processing under quality criteria.

Complexity in characterising genomic and phenotypic microbial diversity related to infectious diseases (Microbial genomics)

Microbial genomics means whole-genome sequencing coupled with BI tools to facilitate the assembly, gene prediction, and functional annotation. This approach has revolutionised our understanding of the biology of important human microbial pathogens. Comparative genome analysis provides insights into adaptations of microbes to their ecological niches and allows the detection of factors that shape host-pathogen interactions.
There is considerable evidence that genetic polymorphisms in both the microbial pathogen and host can impact on microbial virulence or host immune responses to infection. The elucidation of microbial pathogen genomes will contribute to the characterisation of genomic and phenotypic microbial diversity related 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.

c. BIOMEDICAL INFORMATICS IN SUPPORT OF GENOMIC MEDICINE

A new approach to the processing of information about 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) would be integrated. The appropriate techniques and methods would be applied in each case; some would come from BI and others from MI and even from public health and epidemiology informatics. The objective is to process, as efficiently as possible, all the information coming from biological, clinical and environmental research and to advance in the development of Molecular and Personalised Medicine.

Molecular and functional imaging

Molecular Imaging is broadly defined as the characterization and measurement of biological processes in living animals -- including humans -- at the tissue, cellular and molecular level. In terms of healthcare, the dream is that pre-symptomatic diagnosis and treatment will be possible.

The challenge is to help medical doctors see a disease earlier than it is traditionally seen today, better diagnose, prescribe and monitor therapy. Molecular imaging will 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. Significant informatics tools are needed to support molecular imaging. There fall into two types:
❑ MI to Understand Correlations. This includes biostatistics and machine learning to identify significant imaging, genomic, and clinical factors to answer, predict and prognose important clinical questions.

- ❑ BI to Elucidate Molecular Disease Pathology. This includes integrated genomic and protein-interaction databases, pathway elucidation, analysis, modelling and simulation, and predication.

Much molecular imaging research funding is focused on cancer, but we see opportunities in cardio-vascular disease as well as neurological diseases such as Alzheimer's.

Modelling and simulation for an integrative approach of 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 understanding of complex diseases involving many molecular species and many cellular states. BI and MI professionals can largely contribute to it.

This approach will be strongly supported by results derived from the theory of non-linear dynamical systems, by recent advances in the measurement of dynamic processes in individual living cells, and by characterization of physical properties of biological objects, from elasticity of DNA to mechanical properties of cells and tissues in different physio-pathological situations.

These 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. Only formal models can provide a unified abstraction for dealing with the inherent multi-scale, complexity, non-linearity and self-organisation of living systems, diversity of patho-physiological processes, and design of optimal diagnosis and therapy.

- ❑ Development of multi-level dynamical models that would account for spatio-temporal organisation and adaptations from the molecular/cellular levels to the higher processing levels of tissue and organ physiology. While this is a very hard field where only preliminary models can be developed, it is at the centre of the scientific elucidation of relationships between information, regulation and organisation of organisms.
- ❑ Realistic, high resolution *in silico* models of the entire cell, its processes, and its environment. The capability of testing various hypotheses should be made available in *in silico* models that overcome the limitations of cell simulation models such as those of Electronic Cell (http://e-cell.org/) or Virtual cell (http:// www.nrcam.uchc.edu/ vcell_development/ vcell_dev.html).
- ❑ Development of shared libraries of *in silico* models of molecules, interactions, pathways and functions.
- ❑ Image processing and interpretation of bio, gene, protein or tissue-arrays data, in particular those coming from isotopic, fluorescent or ultrasound sources for understanding pathway mechanisms in relation to specific diseases. Molecular imaging creates new challenges and opportunities for combining imaging data with genomic and proteomic data, but only with an integrative model can this be realized.
- ❑ *In silico* modelling of genetic and metabolic networks should be designed to make specific and testable predictions about the key steps of either genetic regulations in operons or metabolic regulations in enzymatic pathways or in transportation chains.

Epidemiology: biobanks and populational repositories

Human genome epidemiology or genetic epidemiology is the new discipline that deals with collections of information on large number of tissues and samples stored in biobanks and populational repositories. Informatics, in this discipline, is applied to manage and analyse relevant data on gene-environment interactions that contribute to diseases of public health importance. Large amounts of molecular epidemiological data of different populations (both of patient and control individuals) are needed for this.

The development of new genetic information technologies will make possible to perform cost-effective screening (genetic tests) at the population level. The intersection of these genetic data with clinical data, electronic health records, environmental and lifestyle data will make possible, among other things, the unravelling of polygenetic disease causality, as well as the complex interactions existent in disease pathogenesis and causation. All these data obtained will be included in populational repositories or biobanks and this knowledge will be applied in public health, for instance disease prevention programs based on genetic data. 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 HuGENet (Human Genome Epidemiology Network) and the National Cancer Institute in the USA and in Europe there is an ongoing project in Iceland that will link health records with genealogical information and information about the genotype. Other on-going projects are also carried out in UK and Estonia.

New methods for e-learning in genomic-based medicine

Due to the increasing amount of medical knowledge in genomic-based medicine, physicians will have to update their knowledge on genetics and genomics. It seems unreasonable to think that this will change easily and rapidly. Research demonstrates that learning is enhanced when learners identify their own needs, select their own strategies and evaluate their own learning outcomes. Internet based informatics tools will be decisive to introduce these possible changes in molecular medicine in a soft manner, avoiding physicians' rejection. The introduction of new learning technologies, providing open and flexible learning programmes, will be crucial for the improvement of doctor's skills and knowledge.

d. ENABLING TECHNOLOGIES

Security

Regarding bio-medical information sciences, the next few decades look very promising and as always, with the promise of benefits also come the danger of abuse. Genomic medicine and the associated interplay between aggregated data and individual data have e.g. given rise to concerns about the proper collection, storage and processing of individually identifiable sensitive information. A focus is needed on privacy enhancing and protecting measures. Besides the more traditional security issues dealing with e.g. confidentiality, integrity, availability, accountability more advanced Privacy Enhancing Techniques (PETs) need to be

addressed. These techniques are of even more importance when storing, exchanging and processing not only medical but also genetic data.

With respect to threats against privacy, there are striking risk differences between genetic and medical data: genetic data concern not only individuals, but also their relatives, i.e. people who have not been tested directly; personal genetic profiles can be directly derived from tissue samples; medical data deal with the past or current health status of persons, whereas genetic tests also furnish indications about future health or disease conditions; an individual genotype is almost unique and stable.

Examples of privacy related issues and techniques are: anonymisation, pseudonymisation, data linkage, gauging for direct and indirect re-identification risks in databases and GRID environments, systems for controlled database dilution, privacy enhancing intelligent agents.

Communication standards – Interoperability among clinical and genetic information systems

Communication between all levels is necessary and has to be provided in a trustworthy way. This means services have to be developed, implemented and maintained for communication security and application security for heterogeneous distributed networks.

Interoperability is the prerequisite for communication and must be addressed in following areas:
❑ Data and knowledge (structure, representation, terminology,..)
❑ Technique (architecture, hardware, topology)
❑ Presentation of data and knowledge,
❑ Security for systems, health care professionals and patients

Standards used today include electronic health records (EHRs, CEN ENV13606), HL7, knowledge representation in GLIF (Guideline interchange format) and Arden syntax, health professional and patient cards, IP and other protocols. XML and XSL present the bigger potential to become the standard language for BMI. An integrated approach using a component based architecture will be an effective basis for further development in this new discipline.

Knowledge representation to facilitate the virtual integration of heterogeneous clinical and genetic databases.

Given the increasing availability of biomedical information located at different sites and accessible over Internet, researchers need new methods to integrate such information. Researchers also need novel methods to search, access, and retrieve this information, which must be gathered, classified and interpreted. To integrate distributed and heterogeneous databases two levels of heterogeneity must be considered: i) databases may be located at various platforms, spread over Internet, with different architectures, operating systems and database management systems, and ii) databases can present different conceptual data models and different underlying database schemas. Solutions for these problems include, for instance, standards such as XML, for exchanging information, or HL7, for connecting biomedical devices. Regarding the integration of databases, various approaches can be considered, such as the concept of data warehouses, federated databases or virtual repositories.

To this date, there is no integrated system of knowledge representation and management that can give answers to the new challenges that the new genomic medicine will bring about. Clinicians have their own systems (ULMS, SNOMED, MeSH, ICD…), in which the coverage of genetic terms (mutation, gene expression) is clearly insufficient. Bioinformaticians are developing several ontologies (MGED, GO, HUGO) but the clinical annotation of their samples (organ, pathology) is still a pending subject. Rather than focusing on the unlikely possibility of a single terminology to cover all domains, the emphasis should be on semantic mapping between terminologies (including clinical and "non-clinical".)

For useful biomedical development, multiple terminologies are required. Not only are multiple terminologies required to cover the words used to describe the clinical state (phenotype), but also additional terminologies are required to leverage genomic/post genomic information for many other uses.

Data and text–based knowledge discovery

Data mining is a step in the process of knowledge discovery in databases. It includes techniques for query databases, on-line analytical processing and machine learning algorithms, among others.
In the fields of medicine and biology, the enormous growth of information and databases, which are openly available for research, has led developers to focus on extracting knowledge from raw data. In the medical area, many applications have been created for decision support, in issues such as image and signal analysis or in clinical prognosis of patient conditions. In biology, efforts have been centred on research issues such as the prediction of protein structures and drug studies. Both offer considerable issues and challenges for future research.

Text mining is a discipline consisting of several methods oriented towards extraction of data, information or knowledge from texts. It is strongly emerging for two reasons: first, the multilingual natural language processing (NLP) tools have been improved and the computing power of any modern desktop computer make such an approach available to any end-user; second -especially with the development of the web and digital libraries - the increasing quantity of data available in electronic format challenges the human ability to handle the amount of knowledge.

Data and text mining are somewhat dependent on the natural language in use. When screening the scientific literature, the English language and the associated tools are adequate. This is basically the situation for BI. However, when screening patient medical records, all European languages are candidate and the necessary multilingual NLP tools are possibly not available. This is partly the situation of MI. In the future, the need to prepare and make available multilingual tools is recognized. To cope with structured as well as free-text repositories, bridges have to be built between national languages and the standardized vocabularies (like MeSH or SNOMED) or coding systems (HL7, TEI), in order to dispose for research purposes of a European corpus of EHR, for the benefit and crossfertilisation between BI and MI.

Health Grid, an infrastructure on which to build the synergy between BI and MI

The interconnection of computers using the Grid middleware enables the user to use computing power and retrieve information from heterogeneous and distributed sources without having to choose which machine he wants to connect to. Grids should be deployed to address the needs of the biomedical community using the state of the art of the middleware technology. Today, Grid technology is still under development and standards are just emerging. Based on the Grid technologies, the vision is to create an environment where information at the 5 levels (molecule, cell, tissue, individual, population) can be associated to provide individualised healthcare

In the last years, the term Grid evolved towards a concept of ubiquitous and transparent computing and encompassed the vision of intensive computing as well as of knowledge Grid, a sort of all-knowing magic mirror. The key question Grid might be able to answer is: How to make information on all levels from molecular to population accessible and understandable to the large variety of people, which could benefit from such knowledge.

Therefore, it is necessary that a pioneering work be done in the field of bio-informatics and MI on a Grid. The creation of a HealthGRID community and first collaboration through providing basic common services (web portals, computing resources) could be a first step in this direction.

A further step would be the development of generic grid metadata management tools, the services would be extended for instance to replication, mirroring and release management of biological data bases and remote medical data acquisition and storage. The design of specific data management and data analysis tools for biological and medical imaging data would open the door to data mining, distributed data management, modelling and processing of 3D and dynamic 3D structures, among others.

Conclusion

The research agenda proposed would allow to increase the knowledge and advance in the research of both functional genomics and genomic-based medicine through the development and implementation of the enabling technologies and of the other research lines described. We believe that this would be facilitated and best achieved by the synergy of Bioinformatics and Medical Informatics.

Priorities in R&D

Barriers	Proposed solution	Priority*	Risk*
ENABLING TECHNOLOGIES			
High computational and data management requirements	Grid	High	Low
Strong privacy issues associated to the nature of genetic data	Security	High	High
Need to expand current interoperability standards for new genetic data infrastructure	Data communication standards	High	Medium
Heterogeneity of current clinical and genetic sources and databases. Different representation systems (i.e. ontologies) in medicine and biology.	Knowledge representation to facilitate the virtual integration of heterogeneous clinical and genetic databases	High	Low
Data and text growing exponentially lacking tools to analyse them	Data and text mining	High	Low
MI IN SUPPORT OF FUNCTIONAL GENOMICS			
Patient care data are not been used systematically in genomic research.	Phenotype databases suitable for genomic research	High	Low
Lack of accepted standards for clinical validation of results obtained from functional genomics research	Disease reclassification	High	High
Lack of adequate matching between biomedical data and pharmaceutical targets	Pharmacogenomics	High	Medium
BI IN SUPPORT OF INDIVIDUALIZED HEALTHCARE			
Unavailability of models for including genetic data into Electronic Health Records	Genetics data model for the EHR	Medium	Medium
Increased complexity in medical decision making due to new genetic knowledge	Clinical guidelines and decision making using genetic information	Medium	Medium
Scarce and non-uniform geographic distribution of clinical genetics specialists and resources	Telegenetics	High	Low
Methods needed for stratifying patients by genetic profiles in the context of clinical research	New methods and information platforms to manage genetic data in clin. research	High	Medium
Lack of interoperable devices to collect genetic data and include them in clinical information systems	Point-of care data acquisition systems	Medium	Medium
Complexity in characterising genomic and phenotypic microbial diversity related to infectious diseases	Microbial genomics	Medium	Low
BMI IN SUPPORT OF GENOMIC MEDICINE			
Lack of high resolution systems to correlate anatomical structures to physiological and genetic mechanisms	Molecular and functional imaging	Medium	Low
Lack of unified approaches to understanding and modelling the human body and human diseases.	Modelling and simulation	Medium	Medium
Linking environmental and lifestyle information to genetic and clinical data	Populational repositories	High	Low
Narrow view of genetics and genomics in health professionals and patients	e-Learning	High	High

* The priorities and risks arise from the debates and discussions in the meetings of the project, the results of the questionnaires sent to the experts and from the opinion of the experts developing each of the lines.
 * Risk refers to the risk of failure to deliver results.

Acknowledgments

Research Line References

Phenotype databases for clinical annotation of biological samples and clinical validation of biological research results

- de Groen, P.C., J.A. Barry, and W.J. Schaller, (1998) Applying World Wide Web technology to the study of patients with rare diseases. Ann Intern Med,. 129 (2): p. 107-13

- Moorman, P.W., et al., (1995) Evaluation of reporting based on descriptional knowledge. J Am Med Inform Assoc, 2(6): p. 365-73

- Friedman, C., et al., (1999) Representing information in patient reports using natural language processing and the extensible markup language. J Am Med Inform Assoc,. 6(1): p. 76-87

- Alizadeh, A.A., et al., (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature, 403(6769): p. 503-11

Disease reclassification

- Schoch C, Kohlmann A, Schnittger S, et al., (2002) Acute myeloid leukemias with reciprocal rearrangements can be distinguished by specific gene expression profiles. Proc Natl Acad Sci 99(15):10008-13

- Pomeroy SL, Tamayo P, Gaasenbeek M, et al., (2002) Prediction of central nervous system embryonal tumour outcome based on gene expression. Nature 415(6870):436-442

- Alizadeh AA, Ross DT, Perou CM, van de Rijn M, (2001) Towards a novel classification of human malignancies based on gene expression patterns. J Pathol. 195(1):41-52

Informatics for supporting rational drug design and development

- Ulrich, R. Friend, SH, 2002. Toxicogenomics and drug discovery: will new technologies help us produce better drugs? Nat Rev Drug Discov. 1:84-8

- Roses, A.D., 2002. Pharmacogenetics place in modern medical science and practice. Life Sci. 15: 1471-80

- Argen, J., 2002. The evolving role of information technology in the drug discovery process. Drug Discov Today 7:315-23.

Including genetic data into the electronic health record

- Ford, JH 2nd, Turner, A, Yoshii, A. 2002. Information requirements of genomics researchers from the patient clinical record. J Healthc Inf Manag. Fall;16(4):56-61.

- Ibarrola, N.; Maojo, V.; Lopez-Campos, G.; Martin-Sanchez, F. 1999. Integration of Genetic Information in the computerized clinical record. EUROREC 99. III European Congress on computerized Clinical Records

Methods for personalized health care: guidelines and decision making support systems

- Musen M.A., Tu S.W., Das A.K., Shahar Y.: EON: (1996) A Component-Based Approach to Automation of Protocol-Directed Therapy. JAMIA 3, pp.367-388.

- Castillo E., Gutiérrez JM and Hadi AS: (1997) Expert Systems and Probabilistic Network Models. Springer, New York.

- Hofestadt R.: Bioinformatics. R.Haux, C. Kulikowski eds. (2002) Yearbook of Medical Informatics, Schattauer, Stuttgart, pp.581-583.

Telegenetics

- Information for Health: An information strategy for the modern NHS, NHS Executive, 1998.

- Gray J, Brain K, Iredale R, Alderman J, France E, Hughes H. (2000) A pilot study of telegenetics. J Telemed Telecare.;6(4):245-7.

- Gattas MR, MacMillan JC, Meinecke I, Loane M, Wootton R. (2001)Telemedicine and clinical genetics: establishing a successful service. J Telemed Telecare.;7 Suppl 2:68-70.

- Meck JM, Munshi G, Plempel J, Amato S, Macedonia C. (1999) Cytogenetic analysis using telemedicine consultation: an improved means of providing expert cross-coverage. Genet Med.;1(7):328-31.

- Schlag PM. (1997) On the Way to New Horizons: Telemedicine in Oncology. Oncologist.;2(2):III-IV.

Stratifying patients by their genetic profiles: molecular diagnosis, clinical trials and pharmacogenomics

- Roses, A.D. 2000. Pharmacogenetics and future drug development and delivery. Lancet. Apr 15;355(9212):1358-61.

- Mancinelli L, Cronin M, Sadee W. 2000. Pharmacogenomics: the promise of personalized medicine. AAPS PharmSci.;2(1):E4. Review.

- Stephens JC. 1999. Single-nucleotide polymorphisms, haplotypes, and their relevance to pharmacogenetics. Mol Diagn. Dec;4(4):309-17. Review.

Point-of-care data collection and access

- Ralph Wayne Mullins, (2002) Connectivity Industry Consortium Point-of-Care Communications in the Past, Present, and Future, POC Vol 1, No. 2, 114-116

- Chairholder Jeoffrey A. Dubois. POCT1A; Point-of-Care Connectivity, Improved Standard, , Published by the National Committee for Clinical Laboratory Standards, NCCLS

- McGowan JJ, Johnson-Lamarche H. The Senior Assessment Coupler: point-of-care decision support and data acquisition tool. Proc AMIA Symp. 1999;:868-72

Complexity in characterising genomic and phenotypic microbial diversity related to infectious diseases (Microbial genomics)

- Tang, CM., Moxon, ER. 2001. The impact of microbial genomics on antimicrobial drug development. Annu Rev Genomics Hum Genet 2 :259-69

- Subramanian, G., Mural, R., Hoffman, SL., Venter, JC., Broder, S. 2001. Microbial disease in humans: A genomic perspective. Mol Diagn 6 (4):243-52

- Haney SA, Alksne LE, Dunman PM, Murphy E, Projan SJ. 2002. Genomics in anti-infective drug discovery--getting to endgame. Curr Pharm Des 8 (13):1099-118

Molecular and functional imaging

- Sarachan, Simmons, Subramanian, Temkin, December 2001. "Combining Medical Informatics and Bioinformatics: toward Tools for Personalized Medicine," Conference Proceedings, Synergy between Research in Medical Informatics, Bio-Informatics and Neuro-Informatics, Brussels, 14

- Li, King (editor), 2002. Special issue of Journal of Magnetic Resonance on Molecular Imaging, GE Molecular Imaging Forum web site

http://apps.gemedicalsystems.com/geCommunity/nmpet/interest_groups/molecular_imaging/new_mi_home.jsp

- Weissleder R, Mahmood U. 2001. Molecular imaging. Radiology;219(2):316-33. Review.

Modelling and simulation for an integrative approach of physiology and pathology

- Kitano H., 2002. Computational systems biology, Nature 420, 206 - 210

- Moraru II, Schaff JC, Slepchenko BM, Loew LM. 2002. The virtual cell: an integrated modeling environment for experimental and computational cell biology.Ann N Y Acad Sci 971:595-6.

- Shafrir Y, Forgacs G, 2002. Mechanotransduction through the cytoskeleton. Am J Physiol Cell Physiol. 282(3):C479-486.

Epidemiology: biobanks and populational repositories

Kaiser, J. 2002. Biobanks. Population databases boom, from Iceland to the U.S. Science. 8;298(5596):1158-61.

New methods for e-learning in genomic-based medicine

- Goettner P. (2000) Effective e-learning for healthcare. Health Manag Technol.;21(12):64, 63.

- Soula G, Pagesy R, Giorgi R, Fieschi D, Gouvernet J, Daniel L, Fieschi M. (2001) An adaptive medical e-learning environment: the MEDIDACTE project. Medinfo.;10(Pt 2):1076-80.

- Broudo M, Walsh C. (2002) Sep MEDICOL: online learning in medicine and dentistry. Acad Med.;77(9):926-7. Review.

- Karim Qayumi A, Qayumi T. (1999) Computer-assisted learning: cyberPatient--a step in the future of surgical education. J Invest Surg.;12(6):307-17. Review

- Fung M. (2002) KOALA: An Internet-based learning and knowledge management system. Technology and Health Care 10 435.

Security

- Claudia Diaz, Stefaan Seys, Joris Claessens, Bart Preneel, 2002. Towards measuring anonymity. Proceedings of PET 2002, April 14-15, 2002, In Hannes Federath (Ed.), Designing Privacy Enhancing Technologies, Lecture Notes in Computer Science

- Regulation No 45/2001 of the European Parliament and the Council of the European Union. Official Journal of the European Communities, 12/1/2001

- Directive 95/46/EC of the European Parliament and the Council of the European Union. Official Journal of the European Communities, 24/10/1995

Communication standards – Interoperability among clinical and genetic information systems

- Blobel, B., Pharow, P., Spiegel, V., Engel, K., Engelbrecht, R.; (2000) Secure interoperability of Patient Data Cards in Health Networks, in: Medical Infobahn for Europe, Health Technology and Informatics, Vol. 77, IOS Press, Amsterdam, , p. 1059-1068

- SEISMED Consortium: Data Security for Health Care, (1996) Vol.1: Management Guidelines, Vol. 2 Technical Guidelines, in: Health Technology and Informatics, Vol. 31/32, IOS Press, Amsterdam.

Knowledge representation to facilitate the virtual integration of heterogeneous clinical and genetic databases.

- Rubin DL, Shafa F, Oliver DE, Hewett M, Altman RB. 2002. Representing genetic sequence data for pharmacogenomics: an evolutionary approach using ontological and relational models. Bioinformatics. Jul;18 Suppl 1:S207-15.

- Cheah YN, Abidi SS. 2000. Healthcare knowledge acquisition: an ontology-based approach using the extensible markup language (XML). Stud Health Technol Inform.;77:827-31.

- Billhardt, H., Crespo J., Mate, J.L., Maojo, V., Martin, F. 2001. A new method for unifying heterogeneous databases. In: Proceedings ISMDA, Madrid, Spain (2001) 2199:54-61

- Sujansky W. 2001 Heterogeneous database integration in biomedicine. J Biomed Inform. Aug;34(4):285-98.

- Klein TE, Chang JT, Cho MK, Easton KL, Fergerson R, Hewett M, Lin Z, Liu Y, Liu S, Oliver DE, Rubin DL, Shafa F, Stuart JM, Altman RB. 2001. Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenetics Research Network and Knowledge Base. Pharmacogenomics J.;1(3):167-70.

- Katehakis DG, Sfakianakis S, Tsiknakis M, Orphanoudakis SC. 2001. An infrastructure for Integrated Electronic Health Record services: the role of XML (Extensible Markup Language). J Med Internet Res. Jan-Mar;3(1):E7

Data and text–based knowledge discovery

- Cios, Krzysztof J., 2001 ed. Medical Data Mining and Knowledge Discovery. New York: Physica-Verlag,.

- Baldi, P. and Brunal, S. 2001. Bioinformatics: The Machine Learning Approach, Second Edition (Adaptive Computation and Machine Learning). The MIT Press.

- Hearst, M. 1999. Untangling Text Data Mining, in the Proceedings of ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics, University of Maryland, June 20-26.

- Fayyad U M, Piatetsky-Shapiro G, Smyth P and Uthurusam R. 2002. Advances in Knowledge Discovery and Data Mining. International Journal of Medical Informatics, Special Issue on Natural Language Processing in Biomedical Applications, Volume 67, Issue 1-3, 4

Health Grid, an infrastructure on which to build the synergy between BI and MI

- Foster and C. Kesselman. 1999. The Grid, blueprint for a new computing infrastructure, Morgan Kaufman, San Francisco.

- Foster, C. Kesselman, S. Tuecke. 2001. The anatomy of the Grid: enabling scalable virtual organizations. International Journal of Supercomputer Applications, 15(3).

- V. Breton, J. Montagnat & R. Medina, 2003. DataGrid, prototype of a biomedical Grid, Proceedings of the conference "Synergy between bioinformatics, medical informatics and neuroinformatics", December 2001,Brussel., Methods of Information in Medicine, 42 (2): 143-7.

General References

[1] Altman RB. 1998. Bioinformatics in support of molecular medicine. Proc AMIA Symp;53-61. Review.

[2] Altman RB. 2000. The interactions between clinical informatics and bioinformatics: a case study. J Am Med Inform Assoc.;7(5):439-43.

[3] Bayat A. 2002 Science, medicine, and the future: Bioinformatics. BMJ. 27;324(7344):1018-22. Review.

[4] Collins F. S. 1999. Medical and Societal Consequences of the Human Genome Project. New Eng. J. Med. 341: 28-37.

[5] Collins FS, McKusick VA. 2001. Implications of the Human Genome Project for medical science. JAMA;285(5):540-4.

[6] Dutton G. 1999. Computational genomics: the medicine of the future? Ann Intern Med.16;131(10):801-4.

[7] Freimer, N. & Sabatti, C. 2003. The Human Phenome Project. Nature Genetics; 34 (May): 15-21.

[8] Friedman CP, Ozbolt JG, Masys DR. 2001. Toward a new culture for biomedical informatics: report of the 2001 ACMI symposium. J Am Med Inform Assoc. Nov-Dec;8(6):519-26

[9] Friend, H. F. 1999. How DNA microarrays and expression profiling will affect clinical practice. British Medical Journal 319: 1-2.

[10] Ginsburg GS, McCarthy JJ. 2001. Personalized medicine: revolutionizing drug discovery and patient care. Trends Biotechnol;19(12):491-6.

[11] Guttmacher E, Collins SF. 2002. Genomic Medicine, a primer. N Engl J Med.; 347 (19): 1512-1520

[12] Kohane IS. 2000. Bioinformatics and clinical informatics: the imperative to collaborate. J Am Med Inform Assoc.;7(5):512-6.

[13] Kulikowski CA. 2002. The micro-macro spectrum of medical informatics challenges: from molecular medicine to transforming health care in a globalizing society. Methods Inf Med.;41(1):20-4.

[14] Liebman, M. 2001. From Bioinformatics to Biomedical Informatics. Genome Technology. N11, p64

[15] Luscombe NM, Greenbaum D, Gerstein M. 2001. What is bioinformatics? A proposed definition and overview of the field. Methods Inf Med.;40(4):346-58. Review.

[16] Rindfleisch, T.C. and D.L. Brutlag. 1998. Directions for clinical research and genomic research into the next decade: implications for informatics. JAMIA 5(5):404-11.

[17] Reifman J, Gilbert GR, Fagan L, Satava R. 2002. Military research needs in biomedical informatics. J Am Med Inform Assoc.;9(5):509-19.

[18] Maojo V, Iakovidis I, Martin-Sanchez F, Crespo J, Kulikowski C. 2001. Medical informatics and bioinformatics: European efforts to facilitate synergy. J Biomed Inform.;34(6):423-7.

[19] Martin-Sanchez F, Maojo V, Lopez-Campos G. 2002. Integrating genomics into health information systems. Methods Inf Med.;41(1):25-30.

[20] Martin-Sanchez, F. 2003. The role of informatics in Genomic Medicine. Celebrating 50 years of DNA. Business Weekly; special issue: 43-45.

[21] Miller PL. 2000. Opportunities at the intersection of bioinformatics and health informatics: a case study. J Am Med Inform Assoc.;7(5):431-8.

[22] Pollard TD. 2002. The future of biomedical research: from the inventory of genes to understanding physiology and the molecular basis of disease. JAMA. Apr 3; 287(13):1725-7.

[23] V. Maojo, F. Martin-Sanchez(1), H. Billhardt(2), I. Iakovidis(3) , C. Kulikowski(4). Establishing an Agenda for Biomedical Informatics. *Methods of Information in Medicine* 2003; 42: 121-12