BUILDING TODAY THE FUTURE CADRE OF INFORMATICS-ORIENTED PHYSICIANS
Lessons from a New Biomedical Informatics Curriculum for Medical Doctors

Ronen Tal-Botzer[1,*], Rachel S. Levy-Drummer[1,*], Gidi Rechavi[2] and Ron Unger[1]
[1]*The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan, Israel*
[2]*Center for Cancer Research, Sheba Medical Center, Tel Hashomer, Ramat-Gan, Israel*
[*]*These authors contributed equally to this work*

Keywords: Biomedical informatics, Bioinformatics, Data mining, Medical education, Curriculum.

Abstract: Recent revolutions in bioinformatics, web and smartphone apps, and the initiative to significantly increase the implementation of electronic medical records, highlighted the need of physicians to be familiar with the newly emerging field of Biomedical Informatics.The huge potential of the field motivated us to educate a cadre of physicians with the required informatics skills, and consequently, improve the quality of medical care in Israel. Here we describe the curriculum of our certification study program in Biomedical Informatics for physicians, comprised of the following courses: Advanced Molecular Methods for Clinical Applications, Clinical Bioinformatics, Biomathematical Modelling, Biostatistics and the Design of Clinical Trials, Clinical Systems Biology and Medical Data Mining. We discuss the rationale of the program; explain our considerations in designing the curriculum; describe the content of each course; highlight the Medical Data Mining course which was specifically designed for this program; and discuss the feedback from the students. More information is available at the website of the program: www.bio-medical.info.

1 FORMATION OF A NEW DISCIPLINE

The interface between medicine and computer science has long been inspiring both researchers and clinicians from the two disciplines. Nevertheless, three recent revolutions, just from about the past five years, have led us to realize that the time has come to initiate a curriculum in Biomedical Informatics dedicated to students, who already have a Medical Doctor degree, and have worked as physicians for several years already. These revolutions, as we see it, are:

- **Bioinformatics:** The recent flourish in Bioinformatics databases and research, thanks to next generation sequencing technologies.
- **Web & Mobile:** The amazing popularity and intensive usage of social networks and smartphones in almost every aspect of our lives, including medical treatment and research.
- **EMRs:** The US health care stimulus package and funds of 2009 for physicians to use Electronic Medical Records (EMRs) and the construction of the National Health Information Network (NHIN), that will bring enormous medical datasets together for the first time.

The new study program, which began in 2010 and is now during its second year of operation, is a joint venture of Bar-Ilan University and the Sheba Medical Center in Israel. Its goal is to elucidate medical doctors' perspectives regarding the biomedical informatics revolution, as well as to provide them with both the theoretical basis and practical skills to enhance their medical practice and research with computational tools. Since the program is new and we consider it as a “pilot”, during its first two years of operation we decided to accept around one dozen students per year.

Bar-Ilan University was the first university in Israel, and probably one of the first ones in the world, to establish undergraduate and graduate study programs in Computational Biology already 15 years ago. Almost all universities in Israel have since opened similar programs, contributing to Israel’s success in Bioinformatics.

Tal-Botzer R., Levy-Drummer R., Rechavi G. and Unger R..
BUILDING TODAY THE FUTURE CADRE OF INFORMATICS-ORIENTED PHYSICIANS - Lessons from a New Biomedical Informatics Curriculum for Medical Doctors.
DOI: 10.5220/0003887603710377
In *Proceedings of the International Conference on Bioinformatics Models, Methods and Algorithms* (SSTB-2012), pages 371-377
ISBN: 978-989-8425-90-4

The Sheba Medical Center is the largest hospital and medical research facility in Israel. Both Sheba Medical Center and Bar-Ilan University are located in the city of Ramat-Gan, a fact that enables very fruitful cooperation between researchers and trainers from the two institutes, as well as convenience for Sheba's physicians to join the study program at Bar-Ilan's campus.

The program was approved by the Israeli Medical Association (IMA), which accredits it as "continuing education" for physicians. The complete information about the program, as well as details for prospective students, are available at its website: www.bio-medical.info.

2 THE NEED FOR INFORMATICS-ORIENTED PHYSICIANS

When building the new study program we first tried to understand what kind of new challenges future physicians and researchers will be required to face. The three above-mentioned revolutions (Bioinformatics, Web & Mobile and EMRs) are already changing the very basic foundations of medical treatment and of medical research (Stead et al., 2010). Following are just few examples.

2.1 The Change in Medical Treatment

A patient today is much more informed about his or her health condition and treatment options than he or she used to be just a decade ago. More and more patients consume medical information about their symptoms, diagnostic methods, diseases, medications, side effects, alternative treatment options, potential physicians, *etc.*, from either reliable or unreliable sources online.

Not only are health related search queries and websites gaining more popularity than ever before (Schmidt, 2008), but also unstructured and uncontrolled information is exchanged through social networks (Dreesman and Denecke, 2011; and Roblin, 2011). The "wisdom of the crowd" is now being formalized in web services like "Patients Like Me", where patients voluntarily report their health condition on a daily basis. This information is later being analyzed and sliced per disease per symptom per treatment per side effect (Wicks, 2011). Whether physicians like or dislike this phenomenon (indeed, many of them report their frustration from that situation) this revolution is here to stay, and physicians should be capable to react accordingly.

In addition to just better management of the information, smartphone apps and auxiliary gadgets help patients monitor health related signals on a 24/7 basis (Shetty, 2011), thus collecting more data and different kinds of data than before. For example, many people are using apps like the "Sleep Cycle" alarm clock, that analyzes their movements during sleep, in order to be woken up at optimal times, when they are not in a deep sleep mode (REM), hence waking up is easier.

But indisputably, the more interesting data is and will be coming very soon from the Bioinformatics and EMRs revolutions. Next generation sequencing technologies and patients' medical histories will drive medical treatment to the personalization era. Computer science plays a crucial role in combining the various information sources and kinds regarding an individual patient, analyzing them and outputting personalized diagnostics and recommended therapeutic pathways. Obviously, when combining the "official" genomic and medical history data with "softer" information, like patient's behavioural data, collected from one's mobile device or social network, a much deeper medical picture of the patient can be assessed (Shah and Robinson, 2011).

Interestingly enough, it is web algorithms, such as those that personalize feeds in social networks or deliver personalized ads to mobile devices, which are considered most promising in enabling better medical personalization and recommendations. A good example would be the Netflix movie rental web service, which is famous for its superior algorithmic ability to predict personal preference in movie selection.

2.2 The Change in Medical Research

Hand in hand with the above-mentioned advances in personalized medicine goes the need for adequate research and skilled researchers. The three revolutions described in the first chapter (Bioinformatics, Web & Mobile and EMRs) bring so many new kinds of information, at higher resolutions and much higher volumes, that they actually constitute a brand new platform for medical research by their own (Stead et al., 2010). The information is so different from what researchers used to know and work with, that it certainly holds a huge potential for new kind of medical discoveries and new innovative treatment solutions.

3 OTHER BIOMEDICAL INFORMATICS EDUCATION PROGRAMS

Clearly, due to this kind of revolution, it's being realized by many educators that there is a need to offer suitable training programs for physicians (Shortliffe, 2010). However, only very few medical schools offer significant computer science and informatics education within the course of regular medical studies. This is understandable due to the justified conservative nature of the medical discipline and the heavy load of medical programs, but also since these revolutions are quite recent. The first personal genome (of James Watson) was sequenced in 2007, The first iPhone being released in 2007, Facebook being opened to the public in 2006 and the US stimulus package for EMRs being legislated only in 2009.

Major medical schools and research institutes, like Stanford University School of Medicine and Columbia University Medical Center, have recently added some mandatory and elective courses in informatics related subjects. This is done in parallel to dedicated programs, offered by the Stanford Center for Biomedical Informatics Research (bmir.stanford.edu) and by Colombia University's Department of Biomedical Informatics (www.dbmi.columbia.edu/education). Both institutes offer large-scale dedicated programs for undergraduate and graduate students, which cover the full theoretical background required in computer science, biology and medicine. A thorough database and reviews on such programs can be obtained through the American Medical Informatics Association (Academic Informatics Programs, 2011).

In addition, some institutes also offer certification study programs for professionals, similar to our program, with differences in program focus, length and audience type. One of the most important programs is given by the American Medical Informatics Association (AMIA 10x10 Courses, 2011). Another notable curriculum of a certification study program, similar to the one described in this paper, is the University of Texas Certificate Program of Health Informatics. However, it is designed mainly for professionals working in the healthcare and information technology (IT) fields, rather than specifically to practicing physicians.

Since Biomedical Informatics is a new emerging field, it is not clear what should be included in the curriculum. For an interesting discussion, see a recent commentary in the Journal of American Medical Association (JAMA), (Shortliffe, 2010). Clearly, each institute has its own considerations when setting up such programs. In many cases, these are the result of specific fields of interest of current faculty members and/or the attributes of the medical services in each country. For example, Israel has quite a unique structure of only four HMOs, which combine the medical services together with the medical insurances for the vast majority of the population. The result is a relatively centralized management system of medical information, as well as standardized medical testing procedures and lab equipment. This situation should facilitate the emergence of the Biomedical Informatics field in Israel.

4 COMPOSING THE BIOMEDICAL INFORMATICS CURRICULUM

4.1 Guidelines

Since current curricula in medical schools are already too condensed and, as mentioned earlier, conservative by nature, we realized there is no realistic option to add significant computer science and/or informatics study units to current medical studies. A better option is to initiate a certification study program for young medical doctors during their resident years or as young attendants. This segment of prospective students enjoys the advantage of having real life experience as physicians, has access to clinical databases and the ability to apply their knowledge to the benefit of their patients. These guidelines led us to build a program that can fit into the schedule of practicing physicians.

4.2 Program Scope

Based on a priori discussions with candidate physicians, we understood that the way to achieve optimal learning atmosphere with minimal interference from other obligations physicians carry, is to assemble the course meetings in one full day per week. It also became clear that for practical reasons, given the career path of young physicians, the program duration cannot exceed the length of one year. Thus, we ended up designing a program of one day a week for two academic semesters. This framework allowed us to offer six courses, three in each semester. Each meeting of each course consists of 1.5 hours of frontal lecture, followed by an hour of hands-on and exercise session. Each semester lasts 14 weeks, therefore the whole curriculum spans over 210 hours in class. Both theoretical and programming assignments are given throughout the semester, and exams or final projects are given by the end of each course.

It is clear that within such limited time we cannot train physicians to become experts in biomedical informatics. Rather, our aim is to widen their horizon; make them aware of the new technologies and the new research and treatment opportunities; make them familiar with the main concepts and tools in the field; and enable them to collaborate effectively with experts. These goals led us to design a program that gives a broad perspective on the emerging field of biomedical informatics, rather than building a narrow program around one or two specific topics. In this way, each student will have enough background to continue further, if he or she chooses to, and deepen his or her knowledge in specific topics, if the need and opportunity arise.

It was clear that the program should start from the body of knowledge that is by now considered classic in the field, as a foundation to the more advanced topics. Thus, our current curriculum includes the following six courses: “Advanced Molecular Methods for Clinical Applications”, “Clinical Bioinformatics”, “Biomathematical Modelling”, ”Biostatistics and the Design of Clinical Trials”, “Clinical Systems Biology” and “Medical Data Mining".

We had a dilemma whether we should include computer programming as one of our subjects. Clearly, mastering biomedical informatics requires the ability to program and to grasp the fundamental concepts of computer science. However, computer programming is a “heavy” discipline, which requires far more teaching and exercise time than we can afford. Since, as we mentioned above, our goal was to give the students a broad introduction to the field, rather than make them practicing experts, we decided not to devote a course to programming. In order just to give a sense of programming, we teach the Perl script language, as part of the exercise sessions of the Clinical Systems Biology course.

4.3 Students’ Feedback and the Resulting Changes in the Program

By the end of each course we asked the students to fill a detailed questionnaire about their satisfaction levels and asked for constructive suggestions. In addition, by the end of the two semesters, we gathered all students and faculty together for a joint review of the past year.

It was not surprising to find out that some of the topics, which we initially considered as very relevant to physicians, easy to grasp, easy to implement and/or just interesting, turned out to be the opposite, and of course, vice versa. We used this feedback, although somewhat diverse, to fine tune the exact selection of topics in the following year.

We realized that there were opposing views among the students themselves regarding the balance between the theoretical/scientific elements and the hands-on/ exercise elements. While some students preferred the theoretical approach and said that they can later find web resources to practice by themselves, other students said that their main gain was the introduction of tools that they can implement in their own research.

As a result, we decided, first, to keep the hands-on topics that we considered as either essential or non-trivial. Such topics include the different bioinformatics and systems biology algorithms and tools, the SQL language, the Weka 3 suite; and more (see detailed syllabus in Chapter 5). Second, we decided to add to the existing curriculum a set of “satellite” sessions, in which students are not obligated to participate in. In these sessions we included topics from the original program that we considered as not being at the essential core of the biomedical informatics profession. Such topics include more challenging usage of the Perl script language, telemedicine, text mining methods and more. This differentiation of topics and lessons was also positively accepted by the students of the second class, and most of them showed interest in taking the satellite courses.

5 PROGRAM TOPICS BY COURSES

Below is a short description and notable insights regarding five of the courses in our curriculum, followed by a more detailed description of a sixth special course, “Medical Data Mining”. Courses similar to the first courses have been given in various versions for several years already within our Computational Biology programs for undergraduate and graduate students, and the challenge was to modify them towards clinical applications such that they will be suitable to an audience of medical doctors. However, the “Medical Data Mining” course was built from scratch, specifically designed to address the most recent advancements in this field.

5.1 Clinical Bioinformatics

This is a fundamental course in many biomedical informatics programs. The course introduces the students with the major bioinformatics databases for DNA, RNA and proteins, as well as more specialized databases of diseases, miRNA and microarray experiments. The computational methods described include sequence alignment, database searches, phylogenetic analysis, genome viewers, protein secondary and tertiary structure prediction, *etc.*. The clinical aspects are emphasized by using examples of genes and proteins involved in diseases, as well as explaining, for example, the flow of analysis that can lead us from a mutation in the genome to its clinical manifestation.

5.2 Biomathematical Modeling

This field of theoretical biology, which strives to model biological systems and their kinetic behaviour using differential equations, is usually not covered within classic bioinformatics programs. This field is very applicable to describe clinical situations, like the progression of a disease in the individual patient's body and the interplay between viruses and the immune system. Since we have the local expertise to teach such a course, we decided to include it in the program. A gentle reminder is given to the required mathematical knowledge; and we select examples where the level of the required mathematics is suitable.

5.3 Advanced Molecular Methods for Clinical Applications

This course covers new genomic technologies. It is a bit different from the other courses, in the sense that its main subjects are experimental methods rather than computational ones. However, we realized that physicians need a refreshment of their knowledge about the basic concepts of molecular biology; and especially need introduction to novel genomic technologies that emerged recently. Thus, the course presents the technologies of micro-arrays, deep sequencing and ncRNA analysis; and discusses their clinical applications, especially for cancer research. The computational aspects of the analysis of the results are highlighted.

5.4 Biostatistics and the Design of Clinical Trials

This course starts with a reminder on the basic tools in statistical analysis and moves to more advanced topics, like the proper statistical design of Clinical Trials. A special attention is given to the statistics of the analysis of large scale data. When data mining algorithms scan a huge number of hypotheses to explain the data, the issue of False Discovery Rate becomes much more critical than in regular statistical analyses.

5.5 Clinical Systems Biology

Systems biology deals with the organization and control of large and complex biological systems, like organs, diseases and genomes that consist of thousands of genes, proteins and metabolites. Systems biology aims to explain biological processes by the intricate networks of interactions between these elements. The course deals with both the theoretical aspects of the field and the practical tools that have been developed. A special emphasis is given to disease pathways and the possible implications of systems biology to clinical intervention.

5.6 Medical Data Mining

The use of computational methods to mine medical data, information and knowledge, as it is diverse and scattered among EMRs, medical journals, social networks and the web in general, has long been one of the most desired goals of Artificial Intelligence (A.I.). The potential applications in real-life situations and the implications to the medical industry, if and when algorithms function as good as we intend, are enormous and can hardly be grasped or imagined (Kim et al., 2011). Thanks to the growing spread of A.I. algorithms throughout the web, and its close and quick connection to large revenues (such as A.I. for online advertising, ranking search results, recommendation engines in e-commerce sites, *etc.*), data mining algorithms have recently evolved, and in a tremendous scale.

More and more academic institutes, as well as start-up companies, channel these advancements to the benefit of the medical field. We realized that although this subject is relatively new, and although most of the actual tools are commercial, this phenomenon should not be ignored by future physicians. On the contrary, they should understand the inherent limits and risks of such A.I. tools, as well as to be able to identify the circumstances, where such tools could be helpful, and might deliver intelligent and educated conclusions, even more than expert human physicians (Korhonen et al., 2009). Another important goal is to provide physicians the basic skills required to implement, tune and even participate in the development process of such tools.

The course is built as a dialog between horizontal and vertical subjects, being the medical applications of data mining algorithms, on one hand, and the computational methodologies that are intertwined in each of those applications, on the other hand. The topics regarding medical applications are: the automatic collection and management of medical data, information and knowledge; the semantic analysis of text, the individual-based vs. the population-based analysis of EMRs; the 4-dimension visualization of biological phenomena based on even sparse pieces of partial knowledge; the prediction of treatment outcome; the decentralization of medical assessment and treatment using telemedicine aided encounters; and the overall movement towards an era of a much more personalized medicine, which is based, among other types of clinical data, on the patient's own genome.

The other axis of topics - the computational methodologies, is comprised of the fundamental theories of computation and data mining, as well as basic algorithms in this field. The ones we found as most relevant are: data representation options, information retrieval (IR) engines, pattern recognition, motif discovery, clustering, classifiers, machine learning, multi-dimensional search, decision trees and basic graph theory. In addition to the theoretic lectures, we hold hands-on exercise sessions, in which we teach the SQL language and the 'Weka 3' data mining suite.

Two additional special chapters are dedicated to smartphone apps for physicians, such as: 'ePocrates', 'AOTrauma' and 'Calculate'; and to medicine-oriented social networks and web services, such as: 'Patients Like Me', '23andMe' and 'Knome'. Since these two topics can be considered as somewhat more "stand alone", not tightly connected to the other more classic data mining topics, they are taught in the framework of special seminar. Most of the seminar is built around real-life examples. They are demonstrated and discussed by a practicing physician, who uses such applications in his day-to-day work. In addition to our students, we invite to this special seminar other Israeli researchers, physicians and managers from the medical domain, who still do not practice biomedical informatics. We do that in order to harbour these easily explained yet impressive topics as an introduction to the field and our study program and encourage enrolment for the following year.

6 DISCUSSION

Computational Biology is a result of the marriage of Computer Science and Biology. Now that computational biology has come of age, the time is ripe for the next step of integrating computational biology with medicine. Clearly, there are many considerations when starting such a program and many compromises and adjustments have to be made in order to fit the program to the needs of the students. One of the problems we encountered is the diverse background and interest of the students, thus, it was not simple to come up with a program that would appeal to all students.

The feedback we received from the participants of the first year of the program was very encouraging. All students mentioned that the program opened their eyes to the new world of research and treatment opportunities. However, we could see again the diversity among our students. While some of the students wanted the program to be more hands-on and practical, others actually suggested concentrating more on theoretical issues. While some said they preferred to get a stronger mathematical and computational basis (including programming), other students complained that this material was already excessive and too difficult for them.

For us it was amazing to see how quickly the program paid off. One of the students, a children neurologist, told us that being inspired by the Systems Biology course, she changed the treatment of one of her patients. It is a patient who suffered from the rare Ruvalcabra syndrome, which is caused by mutations in the PTEN gene. The patient was treated for seizures and behavioural problems without much success. Led by insights she gained from the course, she tried off-label treatment with Rapamune, a drug that is often prescribed for Tuberous Sclerosis, which is another syndrome that shares the same pathway with Ruvalcabra. Surprisingly, the patient showed a significant cognitive improvement. While additional research is needed to follow up this and other patients, it is clear that the program already inspired clinical innovation.

A lot of research and collaborative thinking were put together in this initiative. Our hopes are that other institutes in Israel and around the world will follow with similar programs, and that together we will influence the building of the future cadre of physicians. According to our own scientific research experience, the exhibited emerging phenomenon in the world and the extremely positive feedback we received from physicians who graduated our program, we have no doubt that this future cadre will be much more informatics-oriented.

ACKNOWLEDGEMENTS

We thank Eran Eyal, Nineth Amariglio, Erez Levenon, Avidan Neumann, Yanay Ofran, Ramit Mehr, Sol Efroni, Hiba Waldman Ben-Asher and Niv Marom for their role in designing the courses and teaching in the program.

REFERENCES

Stead, WW., Searle, JR., Fessler, HE., Smith, JW., Shortliffe EH., 2010. Biomedical informatics: changing what physicians need to know and how they learn. In *Academic Medicine*. 86(4), 429-434.

Schmidt, E., 2008. Keynote speech. In *The Healthcare Information and Management Systems Society Conference*. www.youtube.com/watch?v=dTZKNcx9sBA.

Dreesman, J., Denecke, K., 2011. Challenges for signal generation from medical social media data. In *Studies in health technology and informatics*. 169, 639-643.

Roblin, DW., 2011. The potential of cellular technology to mediate social networks for support of chronic disease self-management. In *Journal of Health Communication*. Volume 16.

Wicks, P., Vaughan, TE., Massagli, MP., Heywood J., 2011. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. In *Nature Biotechnology*. 29, 411–414.

Shetty, V., 2011. Mobile phones: emerging tools for medical testing. In *Internet Evolution*. www.internetevolution.com.

Shah, SG., Robinson, I., 2011. Patients' perspectives on self-testing of oral anticoagulation therapy: content analysis of patients' internet blogs. In *BMC Health Services Research*. 3, 11:25.

Shortliffe, EH., 2010. Biomedical informatics in the education of physicians. In *Journal of the American Medical Association*. 304, 1227-1228.

Academic Informatics Programs, 2011. In *American Medical Informatics Association (AMIA) website*. www.amia.org/education/programs-and-courses.

AMIA 10x10 Courses, 2011. Training Health Care Professionals to Serve as Local Informatics Leaders and Champions. In *American Medical Informatics Association (AMIA) website*. www.amia.org/education/10x10-courses.

Kim, SN., Martinez, D., Cavedon, L., Yencken, L., 2011. Automatic classification of sentences to support evidence based medicine. In *BMC Bioinformatics*. 12, Suppl 2:S5.

Korhonen, A., Silins, I., Sun, L., Stenius, U., 2009. The first step in the development of text mining technology for cancer risk assessment: identifying and organizing scientific evidence in risk assessment literature. In *BMC Bioinformatics*. 22, 10:303.