Bioinformatics Meets Clinical Informatics

Jeremy Smith\textsuperscript{a}, Denis Protti\textsuperscript{b}

\textsuperscript{a}Calgary Health Region, Alberta, Canada
\textsuperscript{b}University of Victoria, British Columbia, Canada

Abstract

The field of bioinformatics has exploded over the past decade. Hopes have run high for the impact on preventive, diagnostic, and therapeutic capabilities of genomics and proteomics. As time has progressed, so has our understanding of this field. Although the mapping of the human genome will certainly have an impact on health care, it is a complex web to unravel. Addressing simpler “Single Nucleotide Polymorphisms” (SNPs) is not new, however, the complexity and importance of polygenic disorders and the greater role of the far more complex field of proteomics has become more clear. Proteomics operates much closer to the actual cellular level of human structure and proteins are very sensitive markers of health. Because the proteome, however, is so much more complex than the genome, and changes with time and environmental factors, mapping it and using the data in direct care delivery is even harder than for the genome. For these reasons of complexity, the expected utopia of a single gene chip or protein chip capable of analyzing an individual’s genetic make-up and producing a cornucopia of useful diagnostic information appears still a distant hope. When, and if, this happens, perhaps a genetic profile of each individual will be stored with their medical record; however, in the mean time, this type of information is unlikely to prove highly useful on a broad scale. To address the more complex “polygenic” diseases and those related to protein variations, other tools will be developed in the shorter term. “Top-down” analysis of populations and diseases is likely to produce earlier wins in this area. Detailed computer-generated models will map a wide array of human and environmental factors that indicate the presence of a disease or the relative impact of a particular treatment. These models may point to an underlying genomic or proteomic cause, for which genomic or proteomic testing or therapies could then be applied for confirmation and/or treatment. These types of diagnostic and therapeutic requirements are most likely to be introduced into clinical practice through traditional forms of clinical practice guidelines and clinical decision support tools. The opportunities created by bioinformatics are enormous, however, many challenges and a great deal of additional research lay ahead before this research bears fruit widely at the care delivery level.

Keywords:
Bioinformatics; Medical informatics; Clinical informatics; Genomics; Proteomics

1. Introduction

In their seminal paper, Maojo and Kulikowski compare and contrast medical informatics (MI) and bioinformatics (BI) and provide a viewpoint on their complementarities and potential for collaboration in various subfields [1]. They argue that the new field of biomedical informatics (BMI) holds great promise for developing informatics methods that will be crucial in the development of genomic medicine. In their view, the future of BMI will be influenced strongly by whether significant advances in clinical practice and...
biomedical research come about from separate efforts in MI and BI, or from emerging, hybrid informatics sub-disciplines at their interface.

What Maojo and Kulikowski fail to mention is that the field of bioinformatics has quickly become the focal point of the pharmaceutical industry. Where traditional biology and medical research has been practiced “in vivo” (in a living creature) or “in vitro” (in a glass container), increasingly research is being conducted “in silico” (in a computer) [2]. To accelerate drug development and create more effective treatments, pharmaceutical companies are finding it necessary to invest heavily in the computerization of their industry [3]. At the same time, the results of this research are helping to shift the approach to diagnosis and treatment from one of hypothesis and experimentation to one of direct analysis of underlying causes, prevention, optimal treatments, and analysis of both desired and undesired outcomes [4].

A great deal of work is already underway in the pharmaceutical and medical communities to bring together the clinical information of populations with the genetic profiles to enable scientists to conduct population-based research on the correlation between key genetic traits and the health status of the related individuals. Already, countries such as Iceland and Estonia are leading the way in merging electronic medical records for their population with their genetic profiles, and making this data available to researchers within tight codes of confidentiality and appropriate usage. With the advent of Canada Health Infoway and Genome Canada, both federally sponsored entities funded with a combined investment of roughly $1.5 billion, Canada could be poised to be on the forefront of this field in a few years’ time [5]. The recent creation of a new Bioinformatics Centre at the University of British Columbia is one manifestation of this movement [6].

Yet another educationally oriented initiative occurred in 2002–2003, when the American College of Medical Informatics undertook a study of the future of informatics training [7]. The study members viewed biomedical informatics as an interdisciplinary field, combining basic informational and computational sciences with application domains, including health care, biological research, and education.

Finally, the use of clinical and genetic information for research is clearly receiving a great deal of focus and investment - and narrow, focused successes have at least partially proved its potential. What is less clear, however, is when, and how, the promise of genomics, proteomics, and their related fields will turn into useful clinical tools at the point of care on a widespread basis. This paper is intended to briefly address the state of this field of research, and identify areas of potential application to clinical informatics and the Electronic Health Record (EHR). In doing so controversial issues related to ethical, legal, and social concerns will be referenced but excluded from detailed analysis.

2. Background

To date, bioinformatics has had the greatest impact within the pharmaceutical industry. Drug companies are looking to bioinformatics to improve their speed to market, increase the efficacy of their products, and reduce complications related to those medications. Boston Consulting Group estimates that the average drug now costs $880m to develop and takes almost 15 years to reach the market [8]. Proponents believe that genomics and proteomics will have a tremendous and direct impact on direct care delivery. Others believe that the impact will be slower to materialize and will be less direct. Early evidence suggests that, regardless of when and how, bioinformatics will have a very positive impact on diagnostic practices, analysis of prognosis, prevention of disease, screening, and gene therapy.

Much focus to date has been on genomics, and the sequencing of the human genome.
Although this research shows promise, scientists are increasingly skeptical of the direct and widespread relevance to care delivery. The results of the human genome project have, however, highlighted the importance “single nucleotide polymorphisms” (SNP). SNPs are single-letter variations in DNA sequences that correlate to key differences between individuals – appearance, predisposition to disease, reactions to medications, etc. SNPs correlate to a set of conditions known as Mendelian diseases, or single gene disorders. Although these diseases are relatively easy to identify with genetic testing, they are, unfortunately very rare. The most common and expensive conditions, particularly chronic conditions, are far more complicated that Mendelian diseases – due to “polygenic” disorders - and therefore require far more complex analysis [9].

The weakness of genomics is that genes, although at the root of proteins and cells, do not generally have a one-to-one relationship with disease or human traits. They interact, and are affected by many biological and environmental factors. Proteins, on the other hand, have a much closer relationship with a person’s health, as proteins themselves live within the cells of the body and regulate most of the functions of the human body. In fact, most drugs today act on proteins, not on genes, and proteins tend to show the first signs of ill-health, making them sensitive diagnostic tools.

To sequence the human proteome, however, is even harder than sequencing the human genome – if possible at all. Although the number of genes in the human genome is large (estimates vary between 35,000 and 150,000), the number of proteins expressed from those genes is far larger. A gene is made up of combinations of only four letters (AGCT). A protein, however, is made a string of 20 amino acids, making the permutations and combinations far more diverse. Furthermore, while only the linear sequence of the gene is relevant, proteins wrap themselves into 3 dimensional shapes which are just as relevant as the sequence of genes and amino acids within them. Most medications, for example, currently work by fitting into small gaps, grooves, or ridges in protein structures.

In addition, each type of cell has a different complement of proteins, and each of the 200 different types of cells in a human body has different protein patterns. For example, the proteins in the human brain are different from those of the pancreas. Furthermore, cells produce different protein structures over time, influenced by health, consumption of different foods, and other external factors.

The research does not, however, stop with genomics. Having succeeded in sequencing the human genome, researchers are turning to unraveling the multidimensional and dynamic collection of human proteins [10]. The goal of proteomics is to identify all the proteins, then understand their ranges of expression in different cell types, as well as characterize modifications, interactions, and structure. Though far from being fully mapped, the proteome is already yielding drug targets and information on disease states and drug response. Michael F. Moran, chief scientific officer of MDS Proteomics describes the proteomic challenge as "every state – plus or minus disease, plus or minus drug – is a different proteome [11].

The secret to unleashing both genomics and proteomics is high-throughput analytical devices. Gene chips contain thousands of probes, each of which fluoresces under different colors of laser light, showing which genes are present. The gene chips now have more than 500,000 interrogation points [12]. The biggest hurdle will be lowering the costs of these chips to a level that makes them viable for mass-market analysis. Due to the complexity and changing nature of proteins, the creation of “protein chips” is an even more challenging proposition. However a recent study revealed that a protein chip could be used to screen for a series of 16 prohibited drugs in urine samples [13].
3. Bringing Bioinformatics Research into Clinical Practice

There are two fundamental schools of thought regarding the translation of bioinformatics research into clinical practice. The proponents of genomics and proteomics believe that by combining large-scale databases of clinical data with the related individual genetic profiles, correlation analysis will permit the identification of key cause-and-effect relationships. This process is referred to as “bottom-up analysis” [14]. Others, however, believe that this approach will produce only minimal results due to the many interactions between genes and proteins, and the widely varying environmental factors. They, instead, believe that most significant findings will be identified by starting with known diseases and developing models which take into account the many related causal factors including genetics, environmental factors, etc. – known as “top-down analysis” [15].

In bottom-up analysis, pools of target genes are continuously evaluated against key diseases to establish relevance and test the ability to influence the disease with medications. This approach requires vast databases of medical records and genetic profiles, with the world’s most powerful computer tools for correlation analysis [16]. In the diagram below, bottom-up analysis starts at the bottom row, with the genomic or proteomic sequences, and moves upward, trying to identify the impact of the particular gene or protein on the biology of the individual.

```
Model

Gene/Protein  Gene/Protein  Gene/Protein
   \      /       \      /       \      /       \      /       \      /       \      /       \
Biol/Env Indicator  Biol/Env Indicator  Biol/Env Indicator
```

Top-down analysis uses the fact that a person with a particular condition will generally share a number of key traits with others who have the same condition, and develops models to describe the common patterns. In the diagram above, the process would start at the top, developing a model based on reliable biological and/or environmental indicators from the middle row. Genomic and proteomic profiles may, then be identified that are common to these individuals, and therefore become useful diagnostic tools [17].

According to a respected Canadian scientist:

“The main physiological pathways responsible for many diseases are probably identifiable with available data, top-down system analysis, and a tractable amount of experimentation (both virtually and clinically). The possible genes involved could be identified once the critical pathways are validated. However trying to do it from the bottom-up, starting at the genome level, without knowing which organ, which cell, and what time points, is like not even knowing which field and haystack to find a needle in. There will likely be relatively few diseases, and hence patients, that will be found to result from a single gene mutation/SNP. Most human diseases are likely to result from a combination of diet, multiple gene background, and their interaction” [15].
Implications & Future Direction

Only recently, with studies on entire national populations such as those being carried out on
the inhabitants of Iceland, are researchers beginning to overcome the problem of
insufficient data [18]. Even in that case, however, the data is only applicable to the unique
Icelandic population. Another project, in Estonia, is conducting similar research and since
the Estonian population is far more heterogeneous, it should be easier to generalize the
results [19].

The impact of genomics and proteomics on direct care delivery is starting to become
clearer, however, the exact tools and information that will be required to support the
individual care provider remain very unclear. If information about an individual’s genomic
or proteomic profile were known, clinical decision support tools could assist the physician
with contraindications for prescriptions, however, these tools are likely to use “rules” that
are not unlike those used in today’s software [20]. Likewise, following the “top-down”
approach, it is easily imaginable that the clinical practice guidelines used today will
continue their role in bringing clinical research to the hands of the front-line provider, but in
the future they may be significantly more accurate, tailored to the individual, and/or include
genetic or proteomic diagnostic tests and therapeutic substances. In the shorter term, more
complex decision support tools may be used to avoid medication errors by mapping
diagnostic, medication, and genomic/proteomic indicators that would point to efficacy or
adverse reactions.

5. Acknowledgements

The authors would like to thank the following individuals for sharing their time, insights,
and providing numerous sources for additional research: Dr. George Englebert, Chief
Information Officer, MDS Diagnostics, Toronto, Ontario; Richard Ho, Director of Clinical
Informatics, Johnson & Johnson Pharmaceuticals, San Diego, California; Roman Szumski,
VP Science and Technology, MDS International, Toronto, Ontario; Dr. J. Hans van de
Sande, Ph.D, University of Calgary Faculty of Medicine, Calgary, Alberta; and Dr. Janet
Wood, Professor of Biochemistry, University of Guelph, Guelph, Ontario.

6. References

[1] Maojo V and Kulikowski, CA. Bioinformatics and Medical Informatics: Collaborations on the Road to
[7] Friedman CP et al. Training the Next Generation of Informaticians: The Impact of "BISTI" and
Bioinformatics-A Report from the American College of Medical Informatics. J Am Med Inform Assoc.


**Addresses for correspondence**

Denis J. Protti  
Professor, School of Health Information Science  
Room A212, Human and Social Development Building  
University of Victoria  
Victoria, British Columbia, V8W 3P4 Canada  
Tel: 250.721.8814, Fax: 250.472.4751, http://hinf.uvic.ca

Jeremy Smith  
1436 Fairfield Road  
Victoria, BC  
V8S 1E5  
Tel: 250-415-0516  
eremy@jeremysmith.ca