Bio-Intelligence: A Research Program Facilitating the Development of New Paradigms for Tomorrow's Patient Care

Sieu Phan, Fazel Famili, Ziying Liu, and Lourdes Peña-Castillo

Institute for Information Technology,
National Research Council Canada, Ottawa, Ontario, K1A 0R6, Canada
{sieu.phan,fazel.famili,ziying.liu,lourdes.pena-castillo}@nrc-cnrc.gc.ca

Abstract. The advancement of omics technologies in concert with the enabling information technology development has accelerated biological research to a new realm in a blazing speed and sophistication. The limited single gene assay to the high throughput microarray assay and the laborious manual count of base-pairs to the robotic assisted machinery in genome sequencing are two examples to name. Yet even more sophisticated, the recent development in literature mining and artificial intelligence has allowed researchers to construct complex gene networks unraveling many formidable biological puzzles. To harness these emerging technologies to their full potential to medical applications, the Bio-intelligence program at the Institute for Information Technology, National Research Council Canada, aims to develop and exploit artificial intelligence and bioinformatics technologies to facilitate the development of intelligent decision support tools and systems to improve patient care - for early detection, accurate diagnosis/prognosis of disease, and better personalized therapeutic management.

Keywords: Information technology, patient care.

1 Introduction

The emerging omics technologies have propelled the advancement of our understanding of disease development and progression mechanisms at the molecular basis. This new understanding could potentially lead to the development of new therapeutic targets, paradigms in disease treatment, and drug design. The Bio-intelligence program (Figure 1) at the Institute for Information Technology (IIT), National Research Council Canada (NRC) aims to develop novel IT methodologies to harness these emerging technologies to their full potential to medical applications. The Bio-intelligence program is a collaborative effort among NRC-IIT, NRC-BRI (Biotechnology Research Institute), NRC-IBS (Institute for Biological Sciences), the Ottawa Hospital, and Universidad Politécnica de Madrid. The research program consists of two principal components: knowledge discovery and knowledge integration.

M. Ulieru, P. Palensky, and R. Doursat (Eds.): IT Revolutions 2008, LNICST 11, pp. 250–253, 2009.

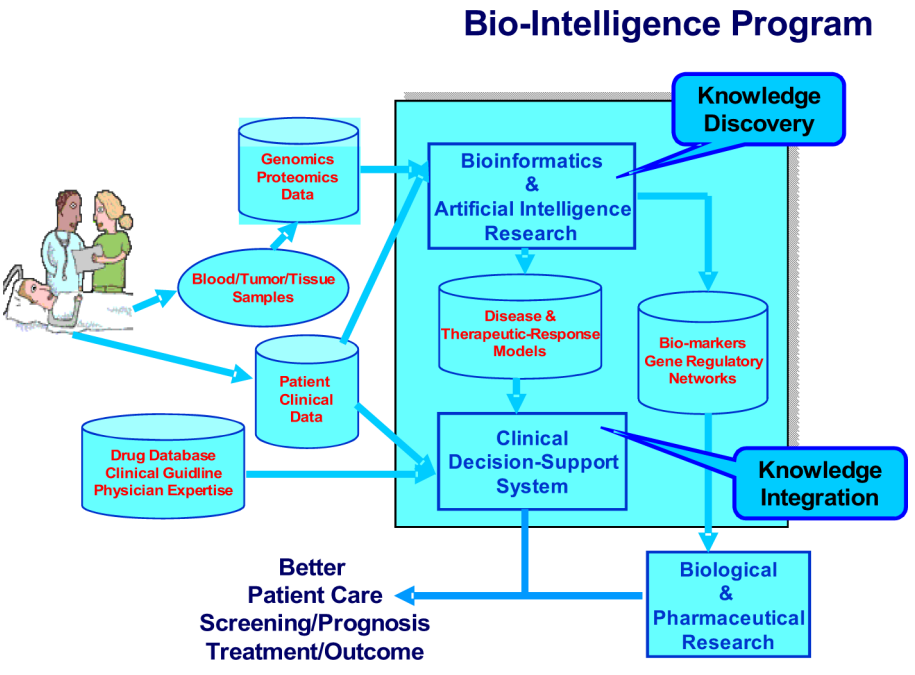

Fig. 1. Bio-intelligence program overview

2 Knowledge Discovery and Knowledge Integration

In the next two sections we discuss the two components of the Bio-intelligence program.

2.1 Knowledge Discovery

Cancer is one of the deadliest diseases. As reported by the Canadian Public Health Agency: "Two in five Canadians face a cancer diagnosis in their lifetime. In 2007 alone, an estimated 159,000 new cases of cancer and 72,700 deaths from cancer will occur in Canada. The burden of cancer in Canada is enormous, affecting the economic and social well-being of individual Canadians, their families and the country". To scale our research to a manageable level according to our resources, our initial piloting effort is focused on cancer treatments.

Cancer is a collection of many diseases in which cells of an organ or tissue in our body become abnormal, growing and multiplying out of control. Normal cells have a life cycle and they reproduce themselves throughout the body in an orderly and controlled manner. Normal growth continues throughout life to replace worn out tissue, to heal wounds, and to maintain healthy organs. When cells grow out of control, they usually form a mass, called a tumour, which results in one form or the other, in cancer. There is a chain of complex processes maintaining a cell's growth and life. In the process of making protein, the DNA

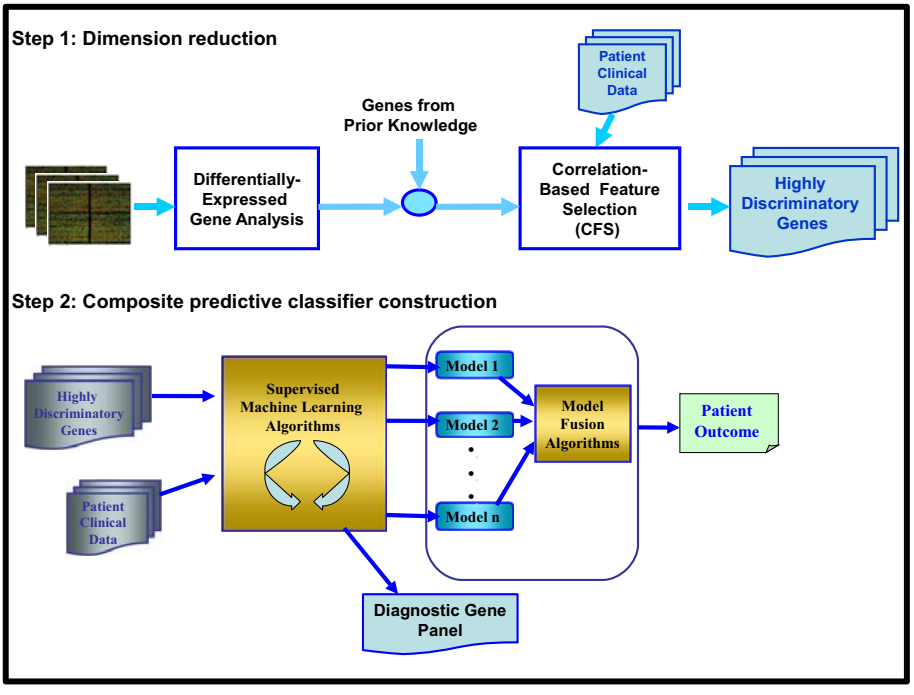

Fig. 2. Workflow for the identification of diagnostic gene panel from transcriptomics data and its companion decision support software module

in the cell nucleus is first transcribed to mRNA; mRNA is then translated into proteins. Some proteins will undergo further post-translational modifications. If the DNA gets mutated or something goes wrong in any of these processes, the cell could develop abnormal behaviour and trigger the development of diseases such as cancer.

Our goal is to develop IT methodologies coupling with the emerging omics technologies to identify discernible DNA/RNA/protein/metabolite variations that could be used to facilitate the fabrication of clinical test kits for early detection, accurate diagnosis/prognosis of cancer, and for the screening of patients' response to cancer therapy. A sample workflow for the identification of a dignostic gene panel from transcriptomics data and its companion decision support software is shown in Figure 2. Our expertise in artificial intelligence, knowledge discovery from data [1,2] and our previous successes in developing intelligent diagnostic and maintenance systems [3] position us very well to take a new approach (than the limited statistical approach) in tackling this life science challenge.

2.2 Knowledge Integration

The second objective of the project is knowledge integration. The plan is to exploit IT to develop a comprehensive clinical decision support system that

integrates several sources of knowledge and data. In addition to the usual capabilities such as disease screening, diagnosis, and prognosis, the system will be designed to help reducing medical errors and to facilitate in providing better and more personalized therapeutic management. For example, with the availability of drug databases and patient electronic health record, drug contradiction or adverse side effects could be avoided. With the availability of omics therapeutic response models and personal genetic profiles, more effective personalized treatment plan can be prescribed. Another example is, in the emergency department, where the clinical guideline module could help in speeding up the triage process in the less-frequently encountered situations. In addition, we would like to incorporate an intelligent look-up-and-reasoning module to support evidence-based medicine [4].

3 Conclusion

The emerging omics technologies have advanced biological research to a new frontier - from the limited view of the traditional approach of examining specific genes, proteins, or metabolites in isolation to the broader interaction view of the systems biology approach. This provides a more comprehensive means to study and understand disease development and progression and consequently promises new effective mechanisms to treat diseases and develop new drugs. Our research aims to develop the bridging IT methodologies to facilitate the translation of omics research to tomorrow's much more effective and efficient new paradigms in disease treatment and drug development.

References

1. Phan, S., Famili, F., Tang, Z., Pan, Y., Liu, Z., Ouyang, J., Lenferink, A., McCourt-O'Connor, M.: A Novel Pattern Based Clustering Methodology for Time-Series Microarray Data. Int. Journal of Computer Mathematics 84(5), 585–597 (2007)
2. Famili, F., Phan, S., Liu, Z., Pan, Y., Djebbari, A., Lenferink, A., O'Connor, M.: Discovering Informative Genes from Gene Expression Data: A Multi-Strategy Approach. In: The 18th European Conference on Machine Learning, Warsaw, Poland (2007)
3. Halasz, M., Dubé, F., Orchard, R., Ferland, R.: The Integrated Diagnostic System (IDS): Remote Monitoring and Decision Support for Commercial Aircraft - Putting Theory into Practice. In: Proceedings of AAAI 1999 Spring Symposium on AI in Equipment Maintenance, Palo Alto, California, USA (1999)
4. Centre for Health Evidence, `http://www.cche.net`