Editorial

Guest Editor: Jugal Kalita

This issue of the *Journal of Bioinformatics Research and Applications* (IJBRA) showcases selected but considerably revised and extended papers from BIOT-04, the Bioinformatics and Biotechnology Symposium, held in Colorado Springs, Colorado, USA, in September of 2004. The Symposium brought together about 70 individuals from across the country for presentation of their research findings. The original papers are available at http://bioinfo.uccs.edu under the top-level menu item 'BIOT-04'. It also fostered fruitful discussions among the participating researchers, students and individuals from industries. Dr. Yi Pan, the Editor-in-Chief of IJBRA helped BIOT-04 as a member of the Program Committee and as an advisor. He was impressed enough with the quality of papers in the proceedings of BIOT-04 that he invited us to have an issue of IJBRA with selected papers for BIOT-04. BIOT-05 is being held in Colorado Springs in October of 2005 as a follow-up meeting. It is expected that starting 2006, the BIOT meetings will be held in locations outside Colorado as well.

This issue of IJBRA has six high-quality papers that cover a broad range of topics in bioinformatics and the companion area of computational biology. The first three papers are on topics that are traditionally considered bioinformatics. The next two papers are more in the area of computational biology, but they still deal with modelling and interpretation of biological data. The last paper deals with medical information represented in the form of images.

In the first paper, titled 'Predicting altered pathways using extendable scaffolds', the authors describe collaborative effort between the University of Texas M.D. Anderson Cancer Center and Rice University. Broom, McDonnell and Subramanian use Bayesian networks and expectation minimisation to learn optimised parameters for a network of scaffolds of metabolite fluxes, extended with genes that interact with those in the scaffold. They construct the models of genetic and metabolic processes incrementally using available gene expression data. This is an excellent first step in modelling and understanding the molecular basis of complex diseases such as prostrate cancer. The second paper is a collaborative work between the Computer Science and Integrative Biology Departments at Brigham Young University in Provo, Utah. In this paper, titled 'Jumpstarting phylogenetic analysis', Mecham, Clement, Freestone, Snell, Seppi and Crandall discuss how existing phylogeny of homologous sequences can be used in building bigger phylogenies, drastically decreasing the time required. This type of an approach is absolutely necessary for researchers to work with large Tree of Life type data sets as envisioned by the National Science Foundation over the past few years in its Call for Proposals. In the third paper, titled 'Discriminating TATA box from putative TATA boxes in plant genome', Loganantharaj of the University of Louisiana, reports on using linear (Naïve Bayes) and non-linear (three-layered artificial neural network) machine learning based classifiers to distinguish between putative TATA boxes and a 'real' TATA box. Many algorithms identify TATA boxes first in order to identify Transcription Start

Sites. A better identification of 'real' TATA boxes will lead to better detection of promoters. The results presented are very promising.

Flow cytometry is a technique used by research biologists and immunologists to collect useful data on large numbers of cells in an efficient manner. The attributes captured by flow cytometers are used to study cell behaviour and investigate treatments for diseases such as cancer, HIV and sickle-cell anaemia. Tools for analysis of flow cytometric data are proprietary and of limited functionality. They also process small amounts of data at a time. In the paper, titled 'A rich analytical environment for flow cytometry experimental results', Siebert and Cios of University of Colorado at Denver, Newell of the University of Colorado Institute of Bioenergetics, discuss how they handle voluminous data generated by flow cytometry experiments that attempt to find links between lipid availability and cell surface expression of Major Histocompatability (MHC) class II molecules. They report on parsing the flow cytometry data, uploading it into a database so that relational queries can be performed and also that the data can be analysed using various tools. This paper shows how biologists and computer scientists can come together so that large amounts of data generated by immunologists and other biologists can be interpreted and understood in an efficient and timely manner. Such data usually fill up reams of printed sheets and takes long and arduous hours for interpretation with considerable chances for error. In 'Computational modelling and simulation of the immune system', Kalita, Chandrashekar, Hans, Selvam and Newell, all of the University of Colorado discuss a software system called SIMISYS 0.3 that models, using the computational technique of cellular automata, many tens of thousands of cells and other entities that participate in the innate and adaptive immune systems of vertebrates. SIMISYS 0.3 also simulates the interactions these entities have with outside pathogens so that disease scenarios can be understood in silico. SIMISYS 0.3 can be an educational tool as well as a research tool. In the paper 'Clinically relevant medical compression', Zukoski of Wilkes University, Boult of University of Colorado and Iyriboz of the Penn State College of Medicine, write that application of technology in the health/medical area is governed by the needs of doctors and lawyers. What is possible technologically may not be acceptable for implementation in a production or real-life situation. Zukoski et al. discuss a novel model-based compression technique for images, which are produced in vast numbers in a clinical environment so that the images can be stored and moved around efficiently. They make use of clinically relevant regions of the images as defined by radiologists. They use lossless compression in clinically relevant regions and use lossy compression everywhere else to achieve a middle ground.

We hope the papers in this issue of IJBRA provide a cross-section of the research underway in bioinformatics and related areas. We thank IJBRA and Dr. Yi Pan of Georgia State University in particular, in giving this exposure to the BIOT series of symposia. I would take this opportunity to invite the readers of this issue of IJBRA to peruse the website of BIOT series, look at the proceedings that are online, and participate in the future if possible.