HALF-DAY TUTORIAL
High Performance Computing in Biomedical Informatics

The last few years have witnessed significant developments in various aspects of Biomedical Informatics, including Bioinformatics, Medical Informatics, Public Health Informatics, and Biomedical Imaging. The explosion of medical and biological data requires an associated increase in the scale and sophistication of the automated systems and intelligent tools to enable the researchers to take full advantage of the available databases. The availability of vast amount of biological data continues to represent unlimited opportunities as well as great challenges in biomedical research. Developing innovative data mining techniques and clever parallel computational methods to implement them will surely play an important role in efficiently extracting useful knowledge from the raw data currently available. The proper integration of carefully selected/developed algorithms along with efficient utilization of high performance computing systems form the key ingredients in the process of reaching new discoveries from biological data.

This tutorial focuses on addressing several key issues related to the effective utilization of High Performance Computing (HPC) in biomedical informatics research, in particular, how to efficiently utilize high performance systems in the analysis of massive biological data. A major issue is how to design energy-aware models for executing computationally-intensive biomedical applications on HPC systems. Another key issue is how to develop innovative network filters for massive biological networks that would utilize parallel algorithms to construct networks samplers that would preserve original network structures while uncovering new ones. The integration between biomedical informatics and HPC will undoubtedly be a major driver in the next generation of biomedical research.

The tutorial is intended primarily for computational scientists who are interested in Biomedical Research and the impact of high performance computing in advancing Biomedical Informatics. Bio-scientists with some background in computational concepts represent another group of intended audience. Although some basic background in biomedical sciences would be useful, it is not necessary since the tutorial will provide a basic background of the needed concepts. Similarly some basic background in HPC would be useful but not necessary. Some understanding of basic algorithmic concepts would be helpful though.

The first part of the session covers the introduction, the background and an overview of key problems, algorithms and current tools in the area of Biomedical Informatics:

1. Introduction to Biomedical Informatics - Brief discussion on the various aspects of Biomedical Informatics that include Bioinformatics, Medical Informatics, Public Health Informatics, and Biomedical Imaging.

2. Background – The Bioscience aspect and the computational perspective, the need for efficient HPC models for addressing key problems in Biomedical Informatics.
3. Biomedical Informatics now – current state of the emerging discipline and overview of key Biomedical Research problems, plus an overview of selected current, first generation, data analysis tools

The second part focuses on introducing the audience to models for integrating HPC systems in Biomedical research with a focus on the concept of next generation data analysis and integration tools; that are Intelligent, Collaborative and Dynamic (ICD). The integration of HPC systems and Biomedical informatics will be presented, and then a focus on two specific case studies related to efficient utilization of HPC in biomedical research will be covered in details. These case studies are related to HPC energy-aware models and efficient parallel algorithms for sampling large biological networks:

4. The need for next generation data integration and analysis tools; Intelligent, Collaborative and Dynamic (ICD) Tools.

5. High Performance Computing (HPC) in Biomedical Informatics Research: current practices, pros and cons. A focus on HPC and new data integration and analysis tools.


Objectives of the Tutorial

The field of Biomedical Informatics has been attracting a lot of attention in recent years. The massive size of the current available biological and medical databases and its high rate of growth have a great influence on the types of research currently conducted and researchers are focusing more than ever to maximize the use of these databases. Hence, it would be of great advantage for researchers to utilize High Performance Computing (HPC) system to explore the data stored in the available databases and extract new information that would lead to better understanding of various biological and medical phenomena.

The Biomedical Informatics domain is rich in applications that require extracting useful information from very large and continuously growing sequence of databases. The marriage between the bioinformatics domain and high performance computing is a natural one; the problems in this domain tend to be highly parallelizable and deal with large datasets, hence using HPC is a natural fit.

In addition, from the IT point-of-view, the problem of efficiently collecting, sharing, mining and analyzing the wealth of information available in a growing set of the biological and clinical data has common roots in many IT applications. This is particularly critical in managing biological and clinical data since relevant data is available in different shapes and forms, and hence, employing all available data to extract meaningful properties is an enormous task. Heterogeneous data, obtained from microarrays, high throughput sequencers, mass spectrometry experiments and clinical records, can all be used to find potential correlations between genes/proteins and the susceptibility to have a particular disease. Addressing these issues require significant computational facilities; hence the need to integrate HPC research. How to efficiently manage the utilization of HPC systems in Biomedical Informatics is quickly emerging as one of the most urgent and critical problems in advancing biomedical research.
The tutorial will address these issues with a particular focus on the following objectives:

1- Provide an overview of the exciting disciplines of Biomedical Informatics, including medical, public health and bio informatics with a focus on the computationally intensive data mining/data analysis problems and their growing need for HPC systems.

2- Introduce the main computational problems in biomedical research with a focus on the current available algorithmic tools and address the advantages and the shortcoming of each tool.

3- Introduce the audience to the concept of intelligent data integrating and analysis tools with a focus on the need to incorporate HPC systems. Such tools are critical to leverage data collected from different resources to produce useful information in a timely manner. The success of these tools can further advance biomedical research and has the potential lead to new discoveries directly related to efficiencies and innovations in Healthcare.

Leader
Hesham Ali
UNO Bioinformatics Core Facility
College of Information Science and Technology
University of Nebraska at Omaha
Email: hesham@unomaha.edu

Hesham H. Ali is Professor of Computer Science and the Lee D. and Wilma Seaman Distinguished Dean of the college of Information Science and Technology (IS&T), at the University of Nebraska at Omaha (UNO). He is also the director of the UNO Bioinformatics Core Facility and the deputy director for computational sciences of the Nebraska Informatics Center for Life Center (NICLS). He received his Ph.D. from the University of Nebraska-Lincoln in 1988, and his BS and MS from the University of Alexandria, in 1982 and 1985, respectively, all in Computer Science. He has published numerous articles in various areas of Computer Science including scheduling, distributed systems, wireless networks, and Bioinformatics. He has also published two books in scheduling and graph algorithms and several chapters in Bioinformatics texts. He is currently serving as the PI or Co-PI of several projects funded by NSF, NIH and Nebraska Research Initiative (NRI) in the areas of wireless networks and Bioinformatics. He has been leading a Bioinformatics Research Group at UNO that focuses on aging research through mobility studies and mining data from temporal biological networks. The group is also developing computational approaches used to identify and classify biological organisms, with the goal of complementing the current experimental approaches in addressing a wide range of Biomedical Informatics problems. He has also been leading an NSF funded project with the goal of developing secure wireless infrastructure for medical environments.