

Bioinformatics – Lecture Notes

Class 1

1. Go over Syllabus
2. Get class list and backgrounds
3. Check if any other possible class time.
4. Go over NIH working definition of Bioinformatics and Computational Biology

NIH has a Bioinformatics Web Page

The Biomedical Information Science and Technology Initiative (BISTI)

<http://grants.nih.gov/grants/bistic/bistic.cfm>

5. Introduction to Bioinformatics

Types of Data –

Biological

Medical

Behavioral

Health

Bioinformatics encompasses a huge variety of areas.

Data Acquisition

Data Organization

Data Archives

Data Analysis

Data Visualization

6. Examples of Initiatives/Projects

Human Genome Project

<http://www.oml.gov/hgmis/>

Begun in 1990, the U.S. Human Genome Project is a 13-year effort coordinated by the Department of Energy and the National Institutes of Health. The project originally was planned to last 15 years, but effective resource and technological advances have accelerated the expected completion date to 2003. Project [goals](#) are to

- *identify* all the approximately 30,000 genes in human DNA,
- *determine* the sequences of the 3 billion chemical base pairs that make up human DNA,
- *store* this information in databases,
- *improve* tools for data analysis,
- *transfer* related technologies to the private sector, and
- *address* the ethical, legal, and social issues (ELSI) that may arise from the project.

Several types of genome maps have already been completed, and a working draft of the entire human genome sequence was announced in June 2000, with analyses published in February 2001. An important feature of this project is the federal government's long-standing dedication to the transfer of technology to the private sector. By licensing technologies to private companies and awarding grants for innovative research, the project is catalyzing the multibillion-dollar U.S. biotechnology industry and fostering the development of new medical applications.

The Visible Human Project[®]

http://www.nlm.nih.gov/research/visible/visible_human.html

The Visible Human Project[®] is an outgrowth of the NLM's 1986 Long-Range Plan. It is the creation of complete, anatomically detailed, three-dimensional representations of the normal male and female human bodies. Acquisition of transverse CT, MR and cryosection images of representative male and female cadavers has been completed. The male was sectioned at one millimeter intervals, the female at one-third of a millimeter intervals.

The long-term goal of the Visible Human Project[®] is to produce a system of knowledge structures that will transparently link visual knowledge forms to symbolic knowledge formats such as the names of body parts.

The Virtual Human Project – It is not just the NIH

<http://vhgallery.gsm.com/>

Images and animations made with their specified software are deposited at the site and are published for use by educators, students, healthcare professionals, publishers, and anyone else needing high-quality anatomical imagery. Over time, they expect the site to become a comprehensive collection that will rival the best of traditional anatomy publications.

Human Brain Project

<http://www.nimh.nih.gov/neuroinformatics/index.cfm>

The Human Brain Project is a broad-based initiative which supports research and development of advanced technologies, and infrastructure support, through cooperative efforts among neuroscientists and information scientists (computer scientists, engineers, physicists, and mathematicians). The goal is to produce new digital capabilities providing a World Wide Web (WWW) based information management system in the form of interoperable databases, and associated data management tools. Tools would include, and are not limited to, graphical interfaces, querying and mining approaches, information retrieval, data analysis, visualization and manipulation, integrating tools for data analysis, biological modeling and simulation, and tools for electronic collaboration. The Neuroscience database will be interoperable with other databases, such as genomic and protein databases, to create the capability to analyze functional interactions in greater depth. Tools will also need to be created to manage, integrate and share this resource via the WWW providing the capability for channels of communication and collaboration between geographically distinct sites. These databases and tools will be used by neuroscientists, behavioral scientists, clinicians and educators, in their respective fields, to understand brain structure, function, and development across the many levels and areas of data collection and analysis.

The Physiome Project

<http://www.physiome.org/frontpage.html>

The PHYSIOME PROJECT is an integrated multi-centric program to design, develop, implement, test and document, archive and

disseminate quantitative information and integrative models of the functional behavior of organelles, cells, tissues, organs, and organisms. The long-range goal is to understand and describe the human organism, its physiology and pathophysiology, and to use this understanding in improving human health. but much or most of what must be learned will come from other species. The project aims toward providing models that summarize information on physiological systems, integrating the observations from many laboratories into quantitative, self-consistent, comprehensive descriptions. The goal is to provide to the community of scientists, physicians, teachers, and to medical health professional and industrial communities, functional descriptions of human biological systems in health and disease. A fundamental and major feature of the program is the databasing of the basic observations for retrieval and evaluation.

The Rice Genome Project

<http://genome.sinica.edu.tw/>

An international rice genome sequencing project, IRGSP, is initiated three years ago, and the primary goal is the complete sequence of rice. The reasons people choose rice as the material as the first crop for genome sequencing project are: (1) rice is an important crop in the world; (2) the genome size of rice is 430 Mb, the smallest one among crops; (3) linkage maps and physical maps of rice have been established and many EST sequences have been registered; (4) the transgenic rice technology has been established; (5) rice shares a co-linear gene organization with other cereal grasses, thus rice is a key to knowledge of the genomic organization of the other grasses.

The Human Proteome Organization (HUPO) – “genes were easy”

http://www.healthtech.com/2001/hpr/hpr_pressrelease.htm

Proteomics is in essence the study of the function, regulation and expression of proteins in relation to the normal function of the cell and in the initiation or progression of a disease state. Proteomics is of particular importance as it is at the level of protein activity that most diseases are manifested. Consequently proteomics seeks to correlate directly the involvement of specific proteins and /or protein complexes in a given disease state.

The applications for proteomics are considerable:-

- Specific proteins can be identified as highly accurate and sensitive markers for disease at a very early stage of onset, thus ensuring their utility in a diagnostic capacity.
- Proteins are important in the prognosis and in the monitoring of therapeutic treatments, as the under or over expression of proteins identified as being disease markers reduces with the improvement in a disease condition. An important potential application here is in increasing the speed and efficacy of clinical trials.
- A knowledge of protein expression patterns can provide insight into potential toxic side-effects during drug screening and lead optimisation.
- Proteins identified as being relevant in specific disease conditions could be valid targets for therapeutic agents and thus could have an important role in the development of new therapeutic treatments.

7. Course Overview

This course will focus on a small subset of bioinformatics – computational molecular biology. This covers areas such as sequence alignment, evolutionary trees, protein structural prediction, and transcription data analysis (microarray analysis). For many, when using the term bioinformatics, this is what they mean. At present, in industry, there is a huge demand for people trained in these areas. Hence, the course will focus on these areas.

The National Center for Biotechnology Information (NCBI) has tools and databases for bioinformatics.

<http://www.ncbi.nlm.nih.gov/>

8. Biological References

1. *Molecular Biology of the Cell* by Bruce Alberts (1994)
2. *Molecular Cell Biology* by Harvey Lodish, Arnold Berk, S. Lawrence Zipursky, and Paul Matsudaira (1999)