Bioinformatics: Application of Information Technology in Biological Science

Dr. Mohammad Alamgeer

Bioinformatician & Software Engineer
Assistant Professor, Department of Information Systems, King Khalid University, Kingdom of Saudi Arabia (KSA).
Corresponding author: Dr. Mohammad Alamgeer

Abstract: *Bioinformatics was applied in the creation and maintenance of a database to store biological information at the beginning of the "genomic revolution", such as nucleotide and amino acid sequences. Development of this type of database involved not only design issues but the development of complex interfaces whereby researchers could both access existing data as well as submit new or revised data.*
Bioinformatics and computational biology are frequently integrated with experimental studies as well, with bioinformatics emphasizing informatics and statistics, while computational biology emphasizes development of theoretical methods, mathematical modeling, and computational simulation techniques to find solution for genetic problems. Bioinformatics studies include analysis and integration of genomics data, prediction of protein function from sequence, and comparisons of protein ligands to identify target effects of drugs. Computational biology includes simulation of protein motion and folding and how proteins interact with each other.
Keywords: *Bioinformatics, Computational Biology, Information Technology,Human Genome Project*

I. Introduction:

It is difficult to recall that only in 1953 was the famous double helical structure of DNA determined. Since then a stupendous series of discoveries has been made. The unraveling of the genetic code was only the beginning. Learning the details of genes and their discontinuous nature in eukaryotic genomes like ours has led to the ability to study and manipulate the material of that abstract concept of Mendel's, the gene itself. Learning to read the genetic material more and more rapidly has enabled us to attempt to decode entire genomes [1, 2].

The research in genomics and proteomics is challenging task to work on genomics big data. The collection of data has essential for nucleic acids, for proteins, and for individual organisms and for chromosomes. The size of nucleic acid databases, has an exponential growth rate. Consequently, a new subject and a new area of expertise is being created, combining the biological and information sciences called <u>Bioinformatics</u> - Marriage of Biology and Computer Science [3, 4, 5].

II. Bioinformatics:

Bioinformatics is a new growing area of science that is computational approaches to solve the biological problems. Solving these problems requires that researchers take advantage of large, complex data sets in a rigorous function to reach biological conclusions. The potential of such an approach is beginning to change the fundamental way in which basic science is done. The term <u>Bioinformatics</u> is relatively new, and as defined here, it encroaches on such term as "<u>Computational Biology</u>" and others [6, 7].Bioinformatics, a rapidly evolving discipline, is the application of computational tools and techniques to the management and analysis of biological data [8].

Bioinformaticians developing analytical tools and database with computational methods, that help scientists interpret experimental data especially in the content of biological systems. Such analytical tools have broad advantages throughout the research and development. The main contribution of bioinformatics is not only to developing the most accurate algorithms or the most sequential analyses; the goal is to find out how living things work.

It helps to identify uncovering disease related genes and pathways to predicting pathways perturbed by therapeutic compounds. Now scientists can concurrently analyze multiple database different platforms and identify key functions to compare resultant data that distinguish biological states. This innovative technology is especially useful for dose response assessment, patient stratification for clinical tasks, time course experiments, and simultaneous analysis of genes list [9].

III. Information TechnologyEnhanced Biology:

Similarly, to other aspects of our society [10], biology is undergoing a major paradigm changeunder the influence of information technology [11]. The database concept of information technology has become an increasing share of biological research.

We had searched for many years and found 1 human pheromone receptor. When thehuman genome became available, in one week we found 110 new genes! Then we wentback to the bench to analyze their activity with classical techniques. [12]. Rapidly increasing capacity of web programming, datamining, cloud computing, and hardware devices enables detailed simulations of biological systems and data management. The computational tools giving accurate and very fast result of gene comparison, motif finding, and target identification on lengthy genomes data. DNA microarrays analysis tools helping in global and parallel analysis of different cellular processes.

The area of bioinformatics work by experts is in statistical methods and pattern recognition in genomic sequences. Sequences alignment tools are helping to comparing two (pair-wise alignment) or more (multiple sequences alignment) sequences by searching for a series of individual characters or character patterns that are in the same order in the sequences. It helps to find the similarities and dissimilarities between two or more than two species.

IV. Biology in the Computer Age:

From the interaction of species, to the function of cells within an individual organism, biology is defined as the study of living things. In the course of that study, biologists collect and interpret data. Now, at the beginning of the 21stcentury, we use sophisticated laboratory technology that allows us to collect data faster than we can interpret it. We have huge volumes of DNA sequences. But to identify which parts of that DNA control the various biochemical processes is not an easy task. We know the function and structure of some proteins, but how do we determine the function of new proteins? And how do we predict what a protein will look like, based on knowledge of its sequence? We understand the relatively simple code that translates DNA into protein. But how do we find meaningful new words in the code and add them to the DNA-protein dictionary? [14]

Researchers come to bioinformatics from many fields, including mathematics, physics, chemistry, computer science, and linguistics. Unfortunately, biology is a science of the specific as well as the general. Bioinformatics is the solution for those who look for motif finding, patterns match, and make predictions without a complete understanding of where biological data. By providing algorithms, databases, interfaceand statistical tools, bioinformatics makes it possible to solve complex biological potentially significant innovations. Once we gain the understanding and become an intelligent user of bioinformatics methods, the speed at which our research progresses can be truly amazing.The growing technology and development has greatly enhanced our comprehension of cell signaling networks and pathways.

V. How Is Computing Changing Biology?

Hereditary and functional information of an organism is stored as DNA, RNA, and proteins. All of these are in form of linear chains which composed of smaller molecules. These macromolecules are assembled from a fixed alphabet of well-defined chemicals:i) DNA is made up of four deoxyribonucleotides (adenine, guanine, thymine, and cytosine), ii) RNA is made up from the four ribonucleotides (adenine, guanine, uracil, and cytosine), and iii) Proteins are made from the 20 amino acids. Due to the linear chain arrangement of defined components, the macromolecules can be represented as sequences of symbols. These sequences can be compared to find similarities of forms and functions of the molecules.

Sequence comparison is the most useful computational tool developed for molecular biologists. The World Wide Web, Web 2.0, and Cloud Computing has made it easy for a single public database of genome sequence data to serve data extraction through a uniform interface to a worldwide community of users. Molecular biologists compare an uncharacterized DNA sequence to the entire publicly collection of DNA sequences by using common computer program called BLAST. It helps to find suggested molecules of proteins for a particular disease based on probability-basedmotif pattern, sequence analysis, sequence compression, and protein modeling with the help of biological tools and databases.

VI. Computational Goals of Bioinformatics:

- **Motif and Pattern Search:**Most common thigs that we are doing in bioinformatics is to find the motifs and short sub-sequences of nucleotide or protein that are of particular interest. These are may be regulatory elements of nucleotide or short stretches of protein that are conserved across multiple sequences of many species. Helps to discover conserved patterns of sequences, structures, metabolism, interactions, and chemistries from various examples.

- **Prediction**: Structure and functions of newly sequenced genes, genomes, proteins or proteomes can be predict based on properties of existing matched sequence.

Proteins are essential components for our life which virtually participate in every metabolic activity within a cell.The prediction of protein structures is essential to determine the function of a protein. Protein structure prediction from amino acid sequence is one of the challenging problems in bioinformatics. Because the biological function of the protein is determined by its threedimensional structure. Thus, protein structure prediction is a fundamental area of computational biology.

- **Organize and Integration**: Information technology helps to develop a systematic and genomic approach to molecular interactions, metabolic pathways, gene expression, and cell signaling.

- **Simulation**: Model gene expression, gene regulation, protein folding, protein-protein interaction, protein-ligand binding, catalytic function, metabolism.

- **Gene Therapy:** Target specific genes, or gene mutations, RNAi to change a disease phenotype.

VII. The Human Genome Project:

At the beginning of the 1990's there was an ambitious undertaking to document the entire human genome. A genome contains the hereditary information for a species. It is composed of DNA and is held in the chromosomes of cells[15]. The Human Genome Project (HGP) was an international scientific research project with the goal of determining the sequence of nucleotide base pairs that make up human DNA, and of identifying and mapping all of the genes of the human genome from both a physical and a functional standpoint [16].

It remains the world's largest collaborative biological project [17]. After the idea was picked up in 1984 by the US government when the planning started, the project formally launched in 1990 and was declared complete in 2003 [18]. Funding came from the US government through the National Institutes of Health (NIH) as well as numerous other groups from around the world. A parallel project was conducted outside government by the Celera Corporation, or Celera Genomics, which was formally launched in 1998. Most of the government-sponsored sequencing was performed in twenty universities and research centers in the United States, the United Kingdom, Japan, France, Germany, Spain and China [19].The main goals of the Human Genome Project were to provide a complete and accurate sequence of the 3 billion DNA base pairs that make up the human genome and to find all of the estimated 20,000 to 25,000 human genes [20].

VIII. Challenges Understanding Genetic Information:

- Genetic information is redundant.
- Structural information is redundant.
- Genes and proteins are meta-stable.
- Single genes have multiple functions.
- Genes are one dimensional but function depends on three-dimensional structure.

Acknowledgement:

References:

[1]. James D. Watson and Francis Crick (1953). "A structure for deoxyribose nucleic acid". Nature 171: 737–8.
[2]. Barnes, M.R. and Gray, I.C., eds., Bioinformatics for Geneticists, first edition. Wiley, 2003. ISBN 0-470-84394-2.
[3]. Lawrence D'Antonio, Arthur Felix, Amruth Kumar, William Mitchell Edward, Saiff, Robert Shine. School of Theoretical and Applied Science. Catalog 2003-2004, Bioinformatics (BS).
[4]. Aluru, Srinivas, ed. Handbook of Computational Molecular Biology. Chapman & Hall/Crc, 2006. ISBN 1584884061 (Chapman & Hall/Crc Computer and Information Science Series).
[5]. Baldi, P and Brunak, S, Bioinformatics: The Machine Learning Approach, 2nd edition. MIT Press, 2001. ISBN 0-262-02506-X.
[6]. Baxevanis, A.D. and Ouellette, B.F.F., eds., Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, third edition. Wiley, 2005. ISBN 0-471-47878-4.
[7]. Baxevanis, A.D., Petsko, G.A., Stein, L.D., and Stormo, G.D., eds., Current Protocols in Bioinformatics. Wiley, 2007. ISBN 0-471-25093-7.
[8]. Tisdall, James. "Beginning Perl for Bioinformatics" O'Reilly, 2001. ISBN 0-596-00080-4.
[9]. Claverie, J.M. and C. Notredame, Bioinformatics for Dummies. Wiley, 2003. ISBN 0-7645-1696-5.
[10]. Baron, G.-L. and E. Bruillard, Technologies de communication et formation des enseignants, ed. G.-L. Baron and E. Bruillard. 2006, Paris: INRP.
[11]. Lombard, F. L'actualité de la biologie :vulgariserouautonomiser ? in JIES : XVIIIesJournéesinternationales sur la communication, l'éducation , la culture scientifiques techniques, et industrielles. 2007. Chamonix.
[12]. Rodriguez, I. (2008) La recherche sur les phéromones à l'UniGE. Bio-Reviews Volume.
[13]. Young, R. and S. Harmony, Working with faculty to design undergraduate information literacy programs. 1999: Neal-Schuman
[14]. Developing Bioinformatics Computer Skills by Cynthia Gibas, Per Jambeck

[15]. Wikipedia. (2011, November 1). Genome. Retrieved November 15, 2011, from Wikipedia.org: http://en.wikipedia.org/wiki/Genome
[16]. Robert Krulwich (2001-04-17). Cracking the Code of Life (Television Show). PBS
[17]. "Economic Impact of the Human Genome Project – Battelle" (PDF). Retrieved 1 August 2013.
[18]. https://www.genome.gov/11006943/human-genome-project-completion-frequently-asked-questions/. Missing or empty |title= (help)
[19]. "Human Genome Project Completion: Frequently Asked Questions". genome.gov.
[20]. Genetics Home Reference: https://ghr.nlm.nih.gov/primer/hgp/goals