



## Applications of Bioinformatics in Medical Renovation and Research

<sup>1</sup>Aditi S. Pandey, <sup>2</sup>Viplove Divyashesh

<sup>1</sup>Department of Applied Chemistry, Priyadarshini Institute of Engineering and Technology, Nagpur, Maharashtra, India

<sup>2</sup>Invertis University, Bareilly, Uttar Pradesh, India

**Abstract:** *Using information technology to modernize our health care system will lead to improvements in medical research. In recent years new sciences have risen up due to the demand in understanding more the world around us like Bioinformatics, Biotechnology, Computational Biology, Biochemistry and others. Health informatics will allow medical researchers to determine the effectiveness of a particular treatment for a given population or to discover the harmful side-effects of a drug. Bioinformatics is a new multidisciplinary field that comes out from the combination of other sciences and fields like biology, computer science, statistics, chemistry, mathematics and even more. Bioinformatics have a huge influence in the medical field and also can play a key role in other fields like agriculture, livestock and even space explorations. Our contribution in this paper is to gather all the distributed fundamental information about Bioinformatics and summarize them in a systematic fundamental way, specifically for the cancer research.*

**Keywords:** *Bioinformatics, cancer research, DNA, databases, technologies*

### I. INTRODUCTION

(Molecular) bio – informatics: bioinformatics is conceptualising biology in terms of molecules (in the sense of physical chemistry) and applying "informatics techniques" (derived from disciplines such as applied maths, computer science and statistics) to understand and organise the information associated with these molecules, on a large scale. In short, bioinformatics is a management information system for molecular biology and has many practical applications. Bioinformatics, the subject of the current review, is often defined as the application of computational techniques to understand and organise the information associated with biological macromolecules. This unexpected union between the two subjects is largely attributed to the fact that life itself is an information technology; an organism's physiology is largely determined by its genes, which at its most basic can be viewed as digital information [1]. Bioinformatics deals with the exponential growth in biological data has led to the development of primary and secondary databases of nucleic acid sequences, protein sequences and structures.

It was a big challenge for researchers to describe this new field in a systematic scientific way and bring out the attention of its applications and services; one of these important services that Bioinformatics can be applied in, is the cancer studies, research and therapies for many beneficial reasons. This paper will give a clear glance overview of bioinformatics, aims, applications, technologies, the large amount of data produced in the biological field and how bioinformatics can organize, analyze and store them, discuss some algorithms that can be implemented over bioinformatics data, and how to apply bioinformatics to discover and diagnose diseases like cancer.

### II. HISTORY

Health care is becoming an increasingly data-intensive field as doctors and researchers generate gigabytes of medical data on patients and their illnesses. While a patient visiting the doctor before 15 years may have only generated a few data points basic information such as weight, blood pressure, and symptoms a medical encounter today may leave a long trail of digital data from the use of high-definition medical imaging to implantable or wearable medical devices such as heart monitors. More importantly, as doctors and hospitals transition away from paper medical records, this data is increasingly being collected and made available in an electronic format. The availability of large data sets of digital medical information has made possible the use of informatics to improve health care and medical research. Often referred to as "in silico" research, informatics offers a new pathway for medical discovery and investigation. The field of bioinformatics has exploded within the past decade to keep pace with advancements in molecular biology and genomics research. Researchers use bioinformatics to gain a better understanding of complex biological processes by, for example, analyzing DNA sequences or modeling protein structures.

The origin of bioinformatics goes back to Mendel's discovery of genetic inheritance in 1865. Since the 1953, big revolution achievements took place by James Watson and Francis Crick as they determined the structure of DNA [2]. Bioinformatics is by nature a cross-disciplinary field that began in the 1960s with the efforts of Margaret O. Dayhoff, Walter M. Fitch, Russell F. Doolittle and others and has matured into a fully developed discipline. In 1960s, the hard work of bioinformatics research started, symbolized by Dayhoff's atlas of protein sequences and the early modeling analysis of protein and RNA structures [3]. After a while, the term Bioinformatics came to sense and use in around 1990s and was described by the management and analysis of DNA, RNA, and protein sequence data. Later in 2000 a big achievement took place which is the announcement of the initial draft of the Human Genome Sequence. Later after 13

years of research and work from 1990 up to spring 2003, in which the official announcement of the Human Genome Sequence Project took place.

In 1988, the Human Genome organization (HUGO) was founded. The first complete genome map was published of bacteria *Haemophilus Influenza*. In 1990, the Human Genome Project was started. By 1991, a total of 1879 human genes had been mapped. In France, in 1993, Genethon, a human genome research center produced a physical map of the human genome. Three years later, Genethon published the final version of the human genetic map.

### **III. BIOINFORMATICS AIMS**

There are five main aims of Bioinformatics [3]:

1. To organize the biological data in an easy manner that helps biologists and researchers to store and access exiting information.
2. To develop and design software tools that help in the analysis and management of data.
3. To use these biological data in the analysis and interpretation of the results in a biological meaningful manner.
4. To assist researchers in the pharmaceutical industry to understand the protein structures that lead and help in the drugs industry development.
5. To help and assist physicians in the medical fields to understand gene structures that will help in detecting and diagnosing disease like cancer.

#### **Bioinformatics databases:**

Bioinformatics is the application of Information technology to store, organize and analyze the vast amount of biological data which is available in the form of sequences and structures of proteins (the building blocks of organisms) and nucleic acids (the information carrier). The biological information of nucleic acids is available as sequences while the data of proteins is available as sequences and structures. Sequences are represented in single dimension where as the structure contains the three dimensional data of sequences. At some time during the course of any bioinformatics project, a researcher must go to a database that houses biological data. Whether it is a local database that records internal data from that laboratory's experiments or a public database accessed through the Internet, such as NCBI's GenBank or EBI's EMBL, researchers use biological databases for multiple reasons

#### **Type of data:**

- nucleotide sequences
- protein sequences
- 3D structures
- gene expression data
- metabolic pathways.

Human diseases, including both communicable and non-communicable diseases, occur through a complex interplay between environmental and host (and pathogen, in the case of infectious diseases) genetic factors. Traditional methods for studying diseases looked at single factors, but now technologies exist for studying multiple factors simultaneously, each with possibly thousands of variables. Integrative, interdisciplinary research using these technologies is being used to unravel the underlying the molecular biology of the disease and move towards personalized medicine. This talk outlines some of the key factors involved in human disease research and the application of bioinformatics methods to understand these diseases and move towards new, more appropriate treatment regimes.

### **IV. BIOINFORMATICS IN CANCER RESEARCH:**

Cancer is a disease determined by several genetic and epigenetic alterations. Due to technological advances in the omics disciplines, cancer research is going through a revolution. The technological advances that lead to the post-genome era have allowed molecular biologists to make meticulous studies on the DNA (genome), the mRNA (transcriptome) and the protein sequences (proteome). Development and use of bioinformatics is essential for the future of cancer therapeutics.

Most cancer treatments work for only a subset of patients and this is likely to remain true for many molecularly-targeted drugs. This results in a large proportion of patients receiving ineffective treatments and is a huge financial burden on our health care system. It is essential that we develop accurate tools for delivering the right treatment to the right patient based on biological characterization of each patient's tumor [4] Gene-expression profiling of tumors using DNA microarrays is a powerful tool for pharmacogenomic targeting of treatments. A good example is the Oncotype DX™ assay (Genomic Health) recently described for identifying the subset of node-negative estrogen-receptor-positive breast cancer patients who do not require adjuvant chemotherapy.

Development of genomic tests that are sufficiently validated for broad clinical application requires the sustained effort of a team that includes clinical investigators, biologic scientists and bio statisticians. Microarray studies of gene expression usually analyze hundreds to tens of thousands of genes. Typical questions to be asked involve the statistical significance of an observed differential expression pattern between samples, or the function of a set of genes with a different expression pattern.

The area of cancer research is not an exception. A typical scenario of cancer research using bioinformatics tools is analysis of global profiles of gene expression in cancer [5], [6]. Gene expression patterns of cancer cells are compared

with those of normal cells or those of other subtypes of the cancer, and genes over/under-expressed in the cancer tissue are identified and clustered (identifying cancer signatures).

The immensely popular fields of cancer research and bioinformatics overlap in many different areas, e.g. large data repositories that allow for users to analyze data from many experiments (data handling, databases), pattern mining, microarray data analysis, and interpretation of proteomics data. There are many newly available resources in these areas that may be unfamiliar to most cancer researchers wanting to incorporate bioinformatics tools and analyses into their work, and also to bioinformaticians looking for real data to develop and test algorithms.[7]

The area of cancer research is not an exception. Rhodes *et al* proposed a statistical model for performing meta-analysis of gene expression data across independent studies, and applied it to expression profiles of prostate cancer [8] . Statistical analysis of characteristic patterns of gene expression are practically very powerful in distinguishing cancer from normal tissue and distinguishing between subtypes of the cancer [9]. A recent excellent review by Rhodes in 2005 discusses the use of external functional information for interpreting and summarizing large cancer signatures [10]. In a functional analysis of a set of genes, it is desired that the employed method can assign accurate function to as large a number of genes as possible in the dataset. However, conventional homology search algorithms, such as BLAST [11] or FASTA [12], can typically cover only 50% or less of the genes in a genome.

## V. CONCLUSIONS

The paper tried to give an overview of this multidisciplinary field, by forming a unique clear definition that is introduced by the reaction of Biology and Computer Science in addition to some assessment factors like statistics and mathematics to result into the newly born field “Bioinformatics”. Taking advantage of genomic technologies to develop drugs effectively and target them to the right patients depends on the use of bio-informatics, in its broadest sense. Bioinformatics is conceptualizing biology in terms of molecules (in the sense of physical-chemistry) and then applying “informatics ” techniques (derived from disciplines such as applied math, CS, and statistics) to understand and organize the information associated with these molecules, on a large-scale.

## REFERENCES

- [1] Yearbook of Medical Informatics 2001.
- [2] Jacques Cohen, *Computer science and bioinformatics*, Communications of the ACM, v.48 n.3, p.72-78, March ,2005
- [3] Jana, R., Aqel, M., Srivastava, P., and Mahanti, P. K., *Soft Computing Methodologies in Bioinformatics, European Journal of Scientific Research*, Vol. 26, No. 2, pp. 189-203, 2009.
- [4] [www.nature.com/clinicalpractice/onc](http://www.nature.com/clinicalpractice/onc)
- [5] Dressman MA . Gene expression profiling detects gene amplification and differentiates tumor types in breast cancer. *Cancer Res*, 63:2194-2199, 2003.
- [6] Subramanian S . Gastrointestinal stromal tumors (GISTs) with KIT and PDGFRA mutations have distinct gene expression profiles. *Oncogene*, 23:7780-7790, 2004.
- [7] Daisuke Kihara, Yifeng David Yang, and Troy Hawkins. Bioinformatics resources for cancer research with an emphasis on gene function and structure prediction tools. *Cancer Inform.*2: 25–35 , 2006.
- [8] Rhodes DR, Barrette TR, Rubin MA, Ghosh D, Chinnaiyan AM, Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res.*; 62(15):4427-33, 2002 .
- [9] Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lønning PE, Brown PO, Børresen-Dale AL, Botstein D, Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A.*; 100(14):8418-23, 2003.
- [10] Rhodes DR, Chinnaiyan AM ' Integrative analysis of the cancer transcriptome. *Nat Gene*; 37 Suppl:S31-7, 2005.
- [11] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, Basic local alignment search tool. *J Mol Biol.*; 215(3):403-10, 1990 .
- [12] Pearson WR, Lipman DJ , Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A.*; 85(8):2444-8, 1988 .