

Life Science Lattices for Biomedicine and Bioinformatics

Tony Breton *

Department of Science and Technology, Nanjing Agricultural University, Nanjing, China

*Corresponding author: Tony Breton, Department of Science and Technology, Nanjing Agricultural University, Nanjing, China; E-mail:

Breton_T@Hed.cn

Received date: March 28, 2022, Manuscript No. IPCHI-22- 13564; **Editor assigned date:** March 30, 2022, PreQC No. IPCHI-22- 13564 (PQ); **Reviewed date:** April 11, 2022, QC No. IPCHI-22- 13564; **Revised date:** April 21, 2022, Manuscript No. IPCHI-22- 13564 (R); **Published date:** April 28, 2022, DOI:10.36648/2470-6973.8.2.119

Citation: Breton T (2022) Life Science Lattices for Biomedicine and Bioinformatics. Chem inform Vol.8 No.2: 119.

Description

Bioinformatics is an interdisciplinary field that develops methods and software tools for understanding biological data, in particular when the data sets are large and complex. As an interdisciplinary field of science, bioinformatics combines biology, chemistry, physics, computer science, information engineering, mathematics and statistics to analyse and interpret the biological data. Bioinformatics has been used for in silico analyses of biological queries using mathematical and statistical techniques. Bioinformatics includes biological studies that use computer programming as part of their methodology, as well as specific analysis "pipelines" that are repeatedly used, particularly in the field of genomics. Common uses of bioinformatics include the identification of candidate genes and Single Nucleotide Polymorphisms (SNPs). Often, such identification is made with the aim to better understand the genetic basis of disease, unique adaptations, desirable properties or differences between populations. In a less formal way, bioinformatics also tries to understand the organizational principles within nucleic acid and protein sequences, called proteomics.

Protein Structure Alignment

The primary goal of bioinformatics is to increase the understanding of biological processes. What sets it apart from other approaches, however, is its focus on developing and applying computationally intensive techniques to achieve this goal. Examples include: Pattern recognition, data mining, machine learning algorithms, and visualization. Major research efforts in the field include sequence alignment, gene finding, genome assembly, drug design, drug discovery, protein structure alignment, protein structure prediction, prediction of gene expression and protein-protein interactions, genome-wide association studies, the modeling of evolution and cell division/mitosis. Bioinformatics now entails the creation and advancement of databases, algorithms, computational and statistical techniques, and theory to solve formal and practical problems arising from the management and analysis of biological data. Over the past few decades, rapid developments in genomic and other molecular research technologies and developments in information technologies have combined to produce a tremendous amount of information related to molecular biology. Bioinformatics is the name given to these

mathematical and computing approaches used to glean understanding of biological processes. Common activities in bioinformatics include mapping and analyzing DNA and protein sequences, aligning DNA and protein sequences to compare them and creating and viewing 3-D models of protein structures.

Image and signal processing allow extraction of useful results from large amounts of raw data. In the field of genetics, it aids in sequencing and annotating genomes and their observed mutations. It plays a role in the text mining of biological literature and the development of biological and gene ontologies to organize and query biological data. It also plays a role in the analysis of gene and protein expression and regulation. Bioinformatics tools aid in comparing, analyzing and interpreting genetic and genomic data and more generally in the understanding of evolutionary aspects of molecular biology. At a more integrative level, it helps analyze and catalogue the biological pathways and networks that are an important part of systems biology. In structural biology, it aids in the simulation and modeling of DNA, as well as bimolecular interactions.

Bimolecular Interactions

Bioinformatics is a science field that is similar to but distinct from biological computation, while it is often considered synonymous to computational biology. Biological computation uses bioengineering and biology to build biological computers, whereas bioinformatics uses computation to better understand biology. Bioinformatics and computational biology involve the analysis of biological data, particularly DNA, RNA, and protein sequences. The field of bioinformatics experienced explosive growth starting in the mid-1990s, driven largely by the human genome project and by rapid advances in DNA sequencing technology. Analyzing biological data to produce meaningful information involves writing and running software programs that use algorithms from graph theory, artificial intelligence, soft computing, data mining, image processing and computer simulation. The algorithms in turn depend on theoretical foundations such as discrete mathematics, control theory, system theory, information theory, and statistics. In the context of genomics, annotation is the process of marking the genes and other biological features in a DNA sequence. This process needs to be automated because most genomes are too large to annotate by hand, not to mention the desire to annotate as many genomes as possible, as the rate of sequencing has ceased

to pose a bottleneck. Annotation is made possible by the fact that genes have recognizable start and stop regions, although the exact sequence found in these regions can vary between genes. Genome annotation can be classified into three levels: The nucleotide, protein, and process levels.

Gene finding is a chief aspect of nucleotide-level annotation. For complex genomes, the most successful methods use a combination of initio gene prediction and sequence comparison with expressed sequence databases and other organisms. Nucleotide-level annotation also allows the integration of genome sequence with other genetic and physical maps of the genome. The principal aim of protein-level annotation is to assign function to the products of the genome. Databases of protein sequences and functional domains and motifs are powerful resources for this type of annotation. Nevertheless, half of the predicted proteins in a new genome sequence tend to have no obvious function. Understanding the function of genes and their products in the context of cellular and organismal physiology is the goal of process-level annotation. One of the obstacles to this level of annotation has been the inconsistency of terms used by different model systems. The Gene Ontology Consortium is helping to solve this problem. Most current genome annotation systems work similarly, but the programs available for analysis of genomic DNA.

In cancer, the genomes of affected cells are rearranged in complex or even unpredictable ways. Massive sequencing efforts are used to identify previously unknown point mutations in a variety of genes in cancer. Bioinformaticians continue to produce specialized automated systems to manage the sheer volume of sequence data produced, and they create new algorithms and software to compare the sequencing results to the growing collection of human genome sequences and germline polymorphisms. New physical detection technologies are employed, such as oligonucleotide microarrays to identify chromosomal gains and losses (called comparative genomic hybridization), and single-nucleotide polymorphism arrays to detect known point mutations. These detection methods simultaneously measure several hundred thousand sites throughout the genome, and when used in high-throughput to measure thousands of samples, generate terabytes of data per experiment. Again the massive amounts and new types of data generate new opportunities for bioinformaticians. The data is often found to contain considerable variability, or noise. Two important principles can be used in the analysis of cancer genomes bioinformatically pertaining to the identification of mutations in the exome. First, cancer is a disease of accumulated somatic mutations in genes. Second cancer contains driver mutations which need to be distinguished from passengers.