

## Lecture 13. Bioinformatics.

### Learning outcomes:

1. Explain the terms “bioinformatics”, “computational biology” and “system biology”.
2. Analyze the bioinformatical methods used in different “Omics” technologies, give the specific examples.
3. Explain the differences between structural, functional and evolutionary bioinformatics.
4. Give and describe the examples of bioinformatical computer programs used for different tasks.
5. Classify and describe the main bioinformatical databases, give the specific examples.

**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, computer science, information engineering, mathematics and statistics to analyze 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 a 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 of better understanding the genetic basis of disease, unique adaptations, desirable properties (esp. in agricultural species), 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**.

**Systems biology** is the **computational and mathematical analysis and modeling of complex biological systems**. It is a biology-based interdisciplinary field of study that focuses on complex interactions within biological systems, using a **holistic approach** (holism instead of the more traditional **reductionism**) to biological research.

Particularly from year 2000 onwards, the concept has been used widely in biology in a variety of contexts. The **Human Genome Project** is an example of applied systems thinking in biology which has led to new, collaborative ways of working on problems in the biological field of genetics. One of the aims of systems biology is to model and discover emergent properties, properties of cells, tissues and organisms functioning as a system whose theoretical description is only possible using techniques of systems biology. These typically involve metabolic networks or cell signaling networks.

**Computational biology** involves the development and application of data-analytical and theoretical methods, mathematical modelling and computational simulation techniques to the study of biological, ecological, behavioural, and social systems. The field is broadly defined and includes foundations in biology, applied mathematics, statistics, biochemistry, chemistry, biophysics, molecular biology, genetics, genomics, computer science, and evolution.

Computational biology is different from **biological computing**, which is a subfield of computer engineering using bioengineering and biology to build computers.

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**.

**Structural bioinformatics** is the branch of bioinformatics that is related to the analysis and prediction of the three-dimensional structure of biological macromolecules such as proteins,

RNA, and DNA. It deals with generalizations about macromolecular **3D structures** such as comparisons of overall folds and local motifs, principles of molecular folding, evolution, binding interactions, and structure/function relationships, working both from experimentally solved structures and from computational models. The term structural has the same meaning as in structural biology, and structural bioinformatics can be seen as a part of computational structural biology. The main objective of structural bioinformatics is the creation of new methods of analysing and manipulating biological macromolecular data in order to solve problems in biology and generate new knowledge.

**Evolutionary biology** is the study of the origin and descent of species, as well as their change over time. **Informatics** has assisted evolutionary biologists by enabling researchers to:

- trace the evolution of a large number of organisms by measuring changes in their DNA, rather than through physical taxonomy or physiological observations alone,
- compare entire genomes, which permits the study of more complex evolutionary events, such as gene duplication, horizontal gene transfer, and the prediction of factors important in bacterial speciation,
- build complex computational population genetics models to predict the outcome of the system over time,
- track and share information on an increasingly large number of species and organisms.

Future work endeavours to reconstruct the now more complex **tree of life**.

The area of research within computer science that uses genetic algorithms is sometimes confused with computational evolutionary biology, but the two areas are not necessarily related.

With the advent of **next-generation sequencing** we are obtaining enough sequence data to map the genes of complex diseases **infertility, breast cancer** or **Alzheimer's disease**. **Genome-wide association studies** are a useful approach to pinpoint the mutations responsible for such complex diseases. Through these studies, thousands of DNA variants have been identified that are associated with similar diseases and traits. Furthermore, the possibility for genes to be used at prognosis, diagnosis or treatment is one of the most essential applications. Many studies are discussing both the promising ways to choose the genes to be used and the problems and pitfalls of using genes to predict disease presence or prognosis.

#### **The questions for self - control:**

1. What are the “bioinformatics”, “computational biology” and “system biology”?
2. Structural, functional and evolutionary bioinformatics.
3. Methods, computer programs and databases of bioinformatics.

#### **Recommended readings:**

1. Lesk, A. M. (26 July 2013). "Bioinformatics". Encyclopaedia Britannica.
2. Tavassoly, Iman; Goldfarb, Joseph; Iyengar, Ravi (2018-10-04). "Systems biology primer: the basic methods and approaches". *Essays in Biochemistry*. 62 (4): 487–500. doi:10.1042/EBC20180003. ISSN 0071-1365. PMID 30287586.
3. Zewail, Ahmed (2008). *Physical Biology: From Atoms to Medicine*. Imperial College Press. p. 339.
4. Longo, Giuseppe; Montévil, Maël (2014). *Perspectives on Organisms* - Springer. *Lecture Notes in Morphogenesis*. doi:10.1007/978-3-642-35938-5. ISBN 978-3-642-35937-8. S2CID 27653540.
5. Bu Z, Callaway DJ (2011). "Proteins MOVE! Protein dynamics and long-range allostery in cell signaling". *Protein Structure and Diseases. Advances in Protein Chemistry and Structural Biology*. 83. pp. 163–221. doi:10.1016/B978-0-12-381262-9.00005-7. ISBN 978-0-123-81262-9. PMID 21570668.
6. "NIH working definition of bioinformatics and computational biology" (PDF). Biomedical Information Science and Technology Initiative. 17 July 2000. Archived from the original (PDF) on 5 September 2012. Retrieved 18 August 2012.

7. "About the CCMB". Center for Computational Molecular Biology. Retrieved 18 August 2012.
8. Gu J, Bourne PE (2011). *Structural Bioinformatics* (2nd ed.). Hoboken: John Wiley & Sons. ISBN 978-1-118-21056-7. OCLC 778339075.
9. Xiong, Jin (2006). *Essential Bioinformatics*. Cambridge, United Kingdom: Cambridge University Press. pp. 4. ISBN 978-0-511-16815-4 – via Internet Archive.
10. Carvajal-Rodríguez A (2012). "Simulation of Genes and Genomes Forward in Time". *Current Genomics*. 11 (1): 58–61. doi:10.2174/138920210790218007. PMC 2851118. PMID 20808525.
11. Aston KI (2014). "Genetic susceptibility to male infertility: News from genome-wide association studies". *Andrology*. 2 (3): 315–21. doi:10.1111/j.2047-2927.2014.00188.x. PMID 24574159. S2CID 206007180.
12. Véron A, Blein S, Cox DG (2014). "Genome-wide association studies and the clinic: A focus on breast cancer". *Biomarkers in Medicine*. 8 (2): 287–96. doi:10.2217/bmm.13.121. PMID 24521025.
13. Tosto G, Reitz C (2013). "Genome-wide association studies in Alzheimer's disease: A review". *Current Neurology and Neuroscience Reports*. 13 (10): 381. doi:10.1007/s11910-013-0381-0. PMC 3809844. PMID 23954969.
14. Londin E, Yadav P, Surrey S, Kricka LJ, Fortina P (2013). "Use of linkage analysis, genome-wide association studies, and next-generation sequencing in the identification of disease-causing mutations". *Pharmacogenomics. Methods in Molecular Biology*. 1015. pp. 127–46. doi:10.1007/978-1-62703-435-7\_8. ISBN 978-1-62703-434-0. PMID 23824853. Missing or empty |title= (help)
15. Hindorff, L.A.; et al. (2009). "Potential etiologic and functional implications of genome-wide association loci for human diseases and traits". *Proc. Natl. Acad. Sci. USA*. 106 (23): 9362–67. Bibcode:2009PNAS..106.9362H. doi:10.1073/pnas.0903103106. PMC 2687147. PMID 19474294.
16. Hall, L.O. (2010). "Finding the right genes for disease and prognosis prediction". 2010 International Conference on System Science and Engineering. *System Science and Engineering (ICSSE), 2010 International Conference*. pp. 1–2. doi:10.1109/ICSSE.2010.5551766. ISBN 978-1-4244-6472-2. S2CID 21622726.