Genome Informatics

The flagship goal of genomic analyses—also known by the more general term 'omics'—is to improve healthcare, and in particular treating cancer, but also to uncover the mysteries of how cells are engineered, how species evolve, or how microbes in our gut help to keep us healthy.

The development of next-generation sequencing has allowed us to investigate genomes in increasing numbers and with increased precision. New techniques go well beyond just reading and assembling DNA; allowing researchers to interrogate many other features such as epigenetics, gene expression, or chromatin contacts, to name just a few. This situation offers exceptional challenges and opportunities for research that leverage a long tradition of cutting edge advances in the broad field of bioinformatics: data-analysis methodologies, databases of prior knowledge, extensive repositories of experimental data, support for enactment of elaborate analytical pipelines, etc.

In this unit we face all the aspects of what is currently known as Data Science or Big Data. Managing our data requires making it accessible and selectively available to foster research while preserving privacy and security. Data must also be pre-processed, validated, linked and augmented by merging data from different sources. In analyzing it we require analytics and statistics, machine learning and data mining, and a good enough understanding of the problem domain. From the technical point of view, we need to think about latency and scalability, for instance to meet the increasing demands forecasted by hospitals embracing new approaches to precision medicine; we need to think about linking and interacting with diverse sources of data with heterogeneous provenance, formats, reliability, or scope; and we need to think about how to make user interfaces for users with different backgrounds and interests, e.g. healthcare professionals, basic researchers, or industry partners.

Objectives

- Identify available software tools, data analysis methods, or data resources for selected tasks and incorporate them into our toolbox
- Adapt tools or develop new ones to address our research needs
- Compile different tools and resources into tailored analytical pipelines

- Generate applications around data analysis tools that support research and enhance discovery

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

---

Source URL (retrieved on *20 abr 2024 - 17:40*): https://www.bsc.es/ca/discover-bsc/organisation/scientific-structure/genome-informatics