So far the discussion has covered the basic methodologies of Bioinformatics.
This talk will address the impacts of Bioinformatics, throughout its short existence.

#2

Here is illustrated a time-line showing important advances in Bioinformatics, and the intertwined consequent progress in Genomics and Proteomics.

Blue boxes represent Bioinformatics advances
Brown boxes represent Genomics advances
Green boxes represent Proteomics advances

<click>

#3

In a keynote speech delivered at a conference in the USA in 2003, Lincoln Stein, a respected computational biologist from the other side of the Atlantic, predicted that bioinformatics would become one of a series of core courses taught in undergraduate and graduate biology programs, and that there would be a vanishing market for researchers who focus solely on biological data management.

In 2008 he had to publish a correction as his prediction was essentially wrong.

“...we are witnessing the rise of a new generation of computational biologists who spend part of their time at the bench and part of their time at the computer.”

<click>

#4

The growth of data resources has been accompanied by increasing needs for user services (red curve). This is Bioinformatics in action. Here we look at the NCBI alone but it is sufficient to illustrate the concept.

<click>

#5

Societal and economic pressures have determined the driving forces that led to investments in Bioinformatics.

A simple way of identifying these areas is to look at the buzzwords most frequently used in the calls for grant applications. Funding agencies hold such documents in archives. A text mining analysis on the headlines of newspapers could provide a similar set of directions.

<click>

#6

It is true that coping with the deluge of information that pours from the very many current Omics projects is already demanding. However, the most interesting challenges for Bioinformatics professionals lie ahead.

We need standardised datasets, properly annotated and curated, made available in interoperable public data repositories, so that they can easily be found by researchers.
But we need more:

- We need to be able to better interpret research data, we need to recognise conserved domain architectures, not just domains in proteins. We need to be able to use network and systems biology to generate and curate models to support reasoning so that they can be used to publish scientific results.

- We need to refine the computational methods that we use to make predictions, so that they work well with real data uncertainty.

- We need to be able to look at complex genetic traits and relate phenomes to genomes and transcriptomes, in order to understand how organisms change from a healthy state to conditions that we consider disease.

<click>

#7

Curation and annotation needs to be applied uniformly to datasets. It is now common to build pipelines to achieve this objective in a reproducible way. As a consequence, the analysis of published results can be effectively traced, from raw data acquisition to summary and visualisation.

<click>

#8

We need to develop and use integrative methods extensively so that genotype-phenotype interactions are revealed. Such methods use multidimensional datasets and specific statistical analytical procedures to extract rules from observations. Systematic use of these integrative methods not only lead to substantial gains in knowledge about natural processes, but also about environmental interactions, drug activity, etc.

<click>

#9

As we have seen earlier in this module, Bioinformatics techniques aim at quantifying similarity and its possible consequences.

They often involve alignments. Alignments of multiple genomes can be quite complex, often beyond that which the human brain can imagine.

The potential offered by artificial intelligence (machine learning) is enormous, especially when a fairly recent approach called deep learning is used. Multi-level, hierarchical artificial neural networks can handle models in unprecedentedly complex ways, much more suitable for hard problems like aligning multiple genomes and extracting reasoned results.

<click>

#10

In this example, related to the industrial production of clavulanic acid by biosynthesis, an intermediate component, an enzyme named YgbP, was a serious bottleneck.

Bioinformatics was used to locate genetic signatures of the DNA that can express this enzyme in nature. This finding was crucial in replacing a complex, fail-proned and expensive chemical synthesis with a natural biosynthesis in a microorganism (H. influenza) that can be inexpensively grown in large vessels.

<click>

#11

In this next example, more related to Medicine, Bioinformatics was used to discover the relationship between a number of repeats and the onset and severity Huntington's disease.

<click>

#12

In summary, present day Bioinformatics activities have now to deal with handling large and very large datasets, with tools to generate complex analytical results. The methods often call for specific capabilities from their practitioners, without which serious bottlenecks develop.

<click>

#13

Therefore, we are facing a human resources problem at two levels.
On the one hand the Education system must become ready to equip every single student of Life Sciences with basic Bioinformatics awareness.
On the other hand, those who actually need to handle the datasets need to be provided with practical skills and be comfortable about using those skills autonomously. That role is best fulfilled by training.