

# Introduction to Bioinformatics

**Mireia Ferrer<sup>1</sup>, Álex Sánchez<sup>1,2</sup>, Esther Camacho<sup>1</sup>, Angel Blanco<sup>1,2</sup>, Berta Miró<sup>1</sup>**

<sup>1</sup> Unitat d'Estadística i Bioinformàtica (UEB) VHIR

<sup>2</sup> Departament de Genètica Microbiologia i Estadística, UB

*We are drowning in information  
and starved for knowledge*

John Naisbitt

*Who on efficient work is bent,  
Must choose the fittest instrument.*

Goethe (Fausto)

# What is Bioinformatics?

# A (first) definition

*Bioinformatics is the application of computer technology to the management of biological information.*

*Computers are used to gather, store, analyze and integrate biological information.*

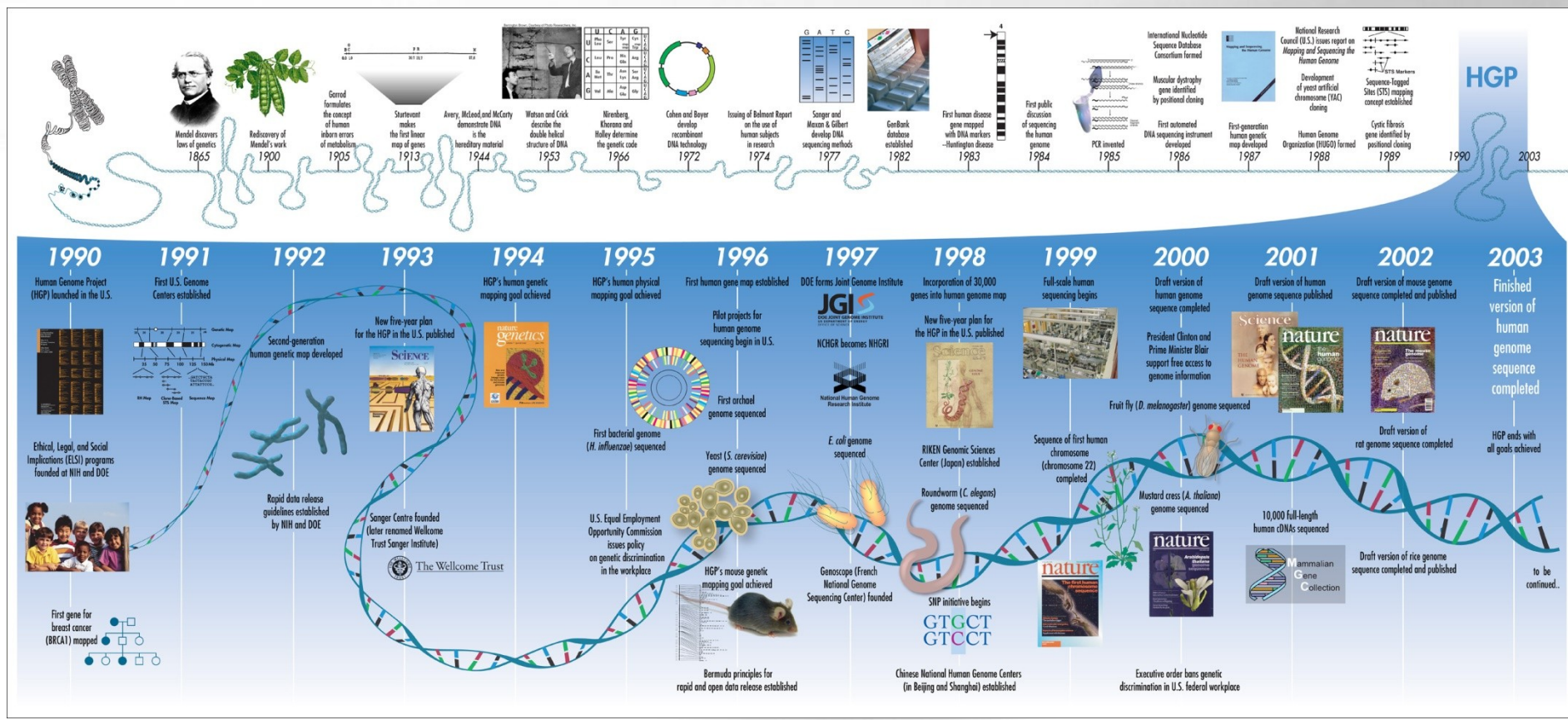


# A historical approach

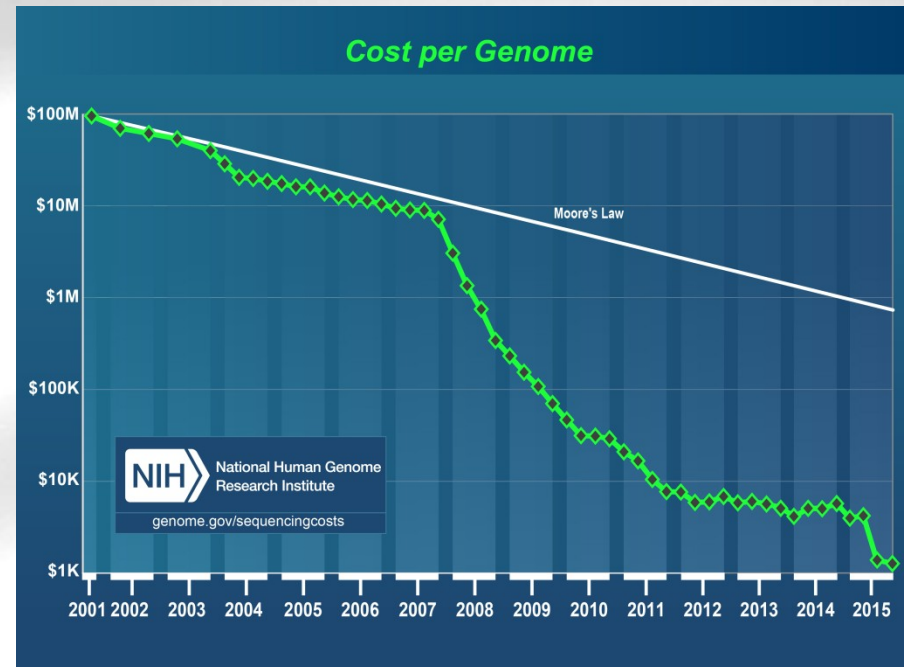
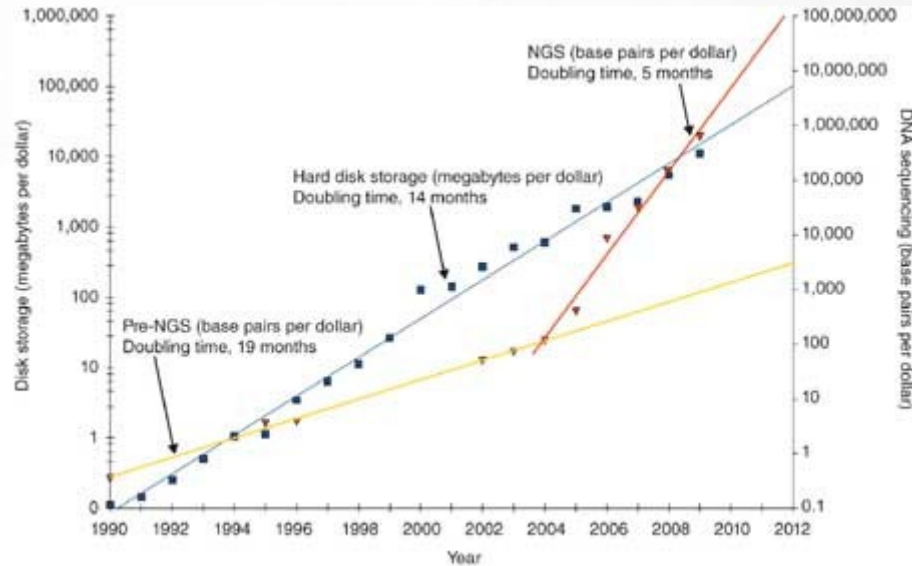
- The term appeared in the 70's
- It became popular/important with the development of the human genome project
- Bioinformatics is entering a big data era that will foster new possibilities.



# The Human Genome Project



# “Nextgen” sequencing revolution



# New paradigm shift: *Precision Medicine*

## 100,000 Genomes Project

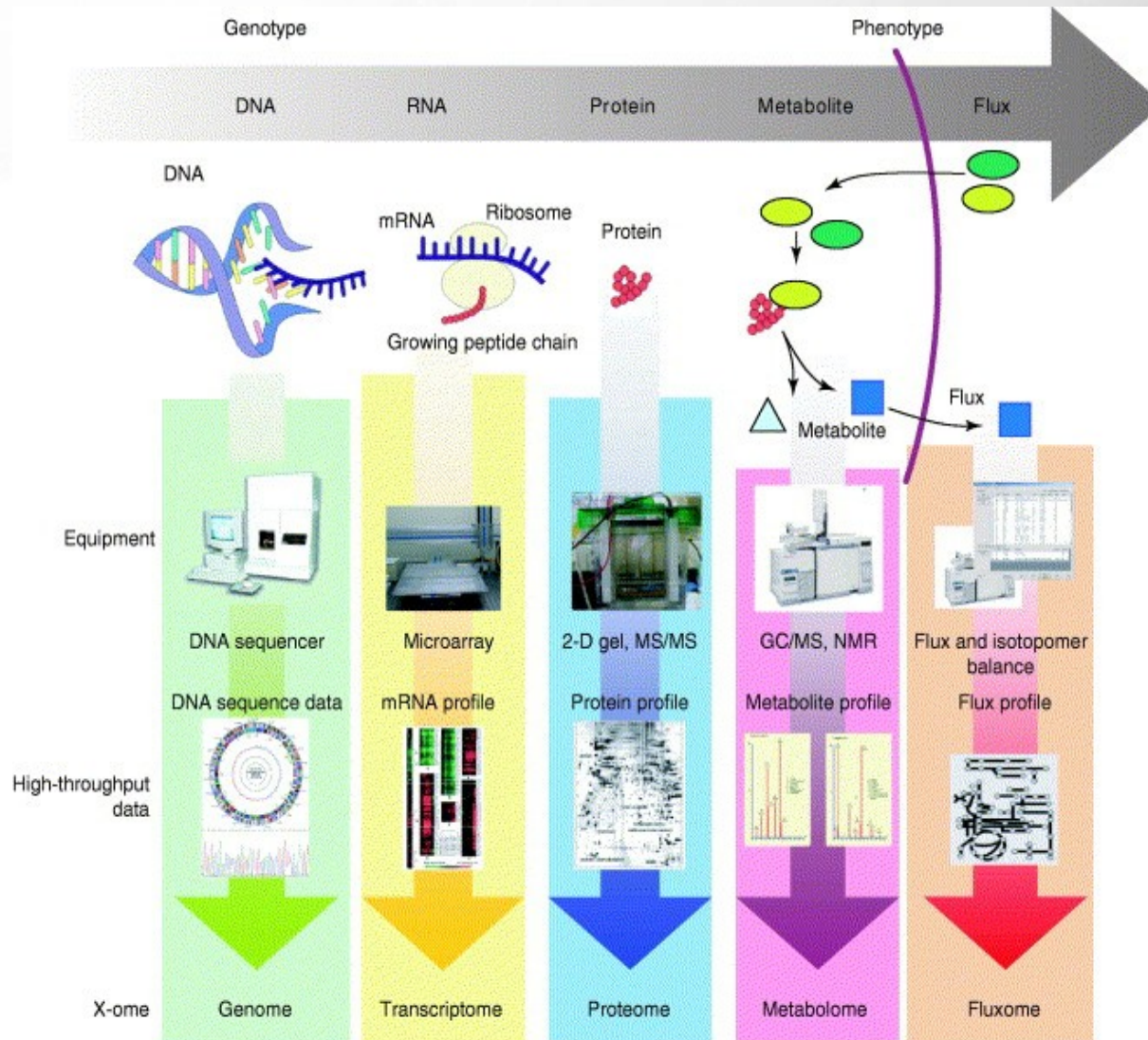


*"It is crucial that we continue to push the boundaries and this new plan will mean we are the first country in the world to use DNA codes in the mainstream of the health service"*

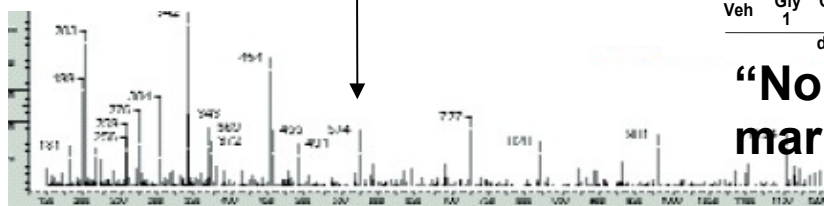
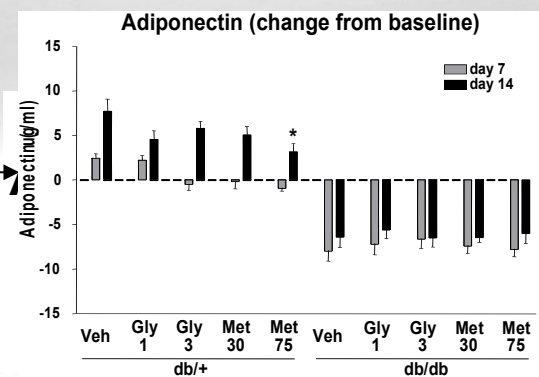
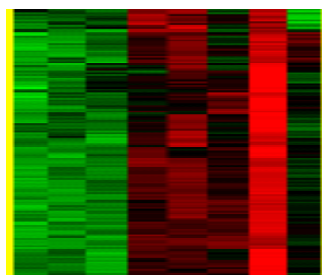
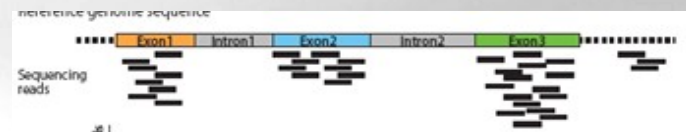
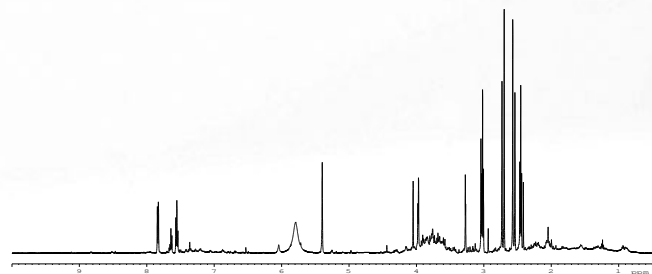
**The Rt Hon David Cameron MP**  
**The Prime Minister**  
10 December 2012



# Omics technologies



# New paradigm shift: *Integromics*

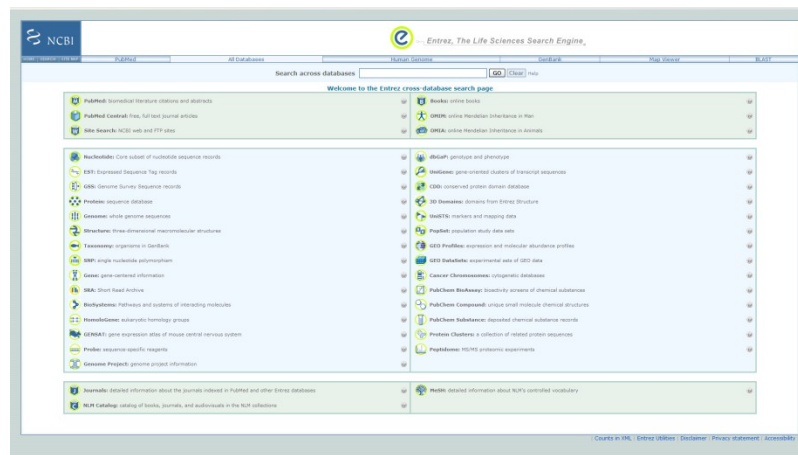


“Non-omic”  
markers

# Bioinformatics domains

- Information management
  - Databases, databanks
  - Algorithms and tools for database querying and searching
- Information Modelling
  - Protein structure characterization
- Analysis & Interpretation of results
  - Genome sequencing and analysis
  - Comparative genomics
  - Transcriptomics and gene expression
  - Metagenomics
  - Proteomics, metabolomics, ...
- Biological system modeling

# Information management





## Tools for Data Mining

PubMed   Entrez   BLAST   OMIM   Books   TaxBrowser   Structure

Search  for

Nucleotide Sequence Analysis   Protein Sequence Analysis   Structures   Genome Analysis   Gene Expression

### Tools - Nucleotide Sequence Analysis

**BLAST** The Basic Local Alignment Search Tool (BLAST) for comparing gene and protein sequences against others in public databases, now comes in several types including PSI-BLAST, PHI-BLAST, and BLAST 2 sequences. Specialized BLASTs are also available for human, microbial, malaria, and other genomes, as well as for vector contamination, immunoglobulins, and tentative human consensus sequences.

**electronic PCR** **Electronic PCR** - allows you to search your DNA sequence for sequence tagged sites (STSs) that have been used as landmarks in various types of genomic maps. It compares the query sequence against data in NCBI's **UniSTS**, a unified, non-redundant view of STSs from a wide range of sources.

**Entrez Gene** - each Entrez Gene record encapsulates a wide range of information for a given gene and organism. When possible, the information includes results of analyses that have been done on the sequence data. The amount and type of information presented depend on what is available for a particular gene and organism and can include: (1) graphic summary of the genomic context, intron/exon structure, and flanking genes, (2) link to a graphic view of the mRNA sequence, which in turn shows biological features such as CD: SNPs, etc., (3) links to gene ontology and phenotypic information, (4) links to corresponding protein sequence data and conserved domains, (5) links to related resources, such as mutation databases. Entrez Gene is a successor to LocusLink.

**Model Maker** - allows you to view the evidence (mRNAs, ESTs, and gene predictions) that was aligned to assembled genomic sequence to build gene model and to edit the model by selecting or removing putative exons. You can then view the mRNA sequence and potential ORFs for the edited model and save the mRNA sequence data for use in other programs. Model Maker is accessible from sequence maps that were analyzed at NCBI and displayed in Map Viewer.

**ORF Finder** - identifies all possible ORFs in a DNA sequence by locating the standard and alternative stop and start codons. The deduced amino acid sequences can then be used to BLAST against GenBank. ORF finder is also packaged in the sequence submission software **Sequin**.

**Organism Specific Resources** - Bee, Cat, Chicken, Cow, etc.

**SAGEman** - provides a tool for performing statistical tests designed

### Site Map

Guide to NCBI resources

### Tools for Programmers

**BLAST**  
Standard tool for sequence analysis

**BLink**  
BLAST Link

**CDART**  
Conserved Domain Architecture Retrieval Tool

**CD search**  
Conserved Domain Database search

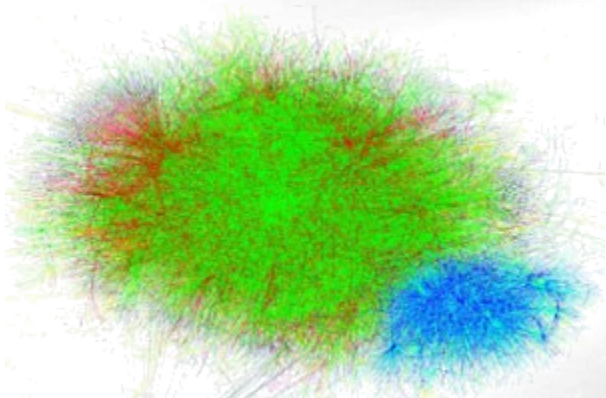
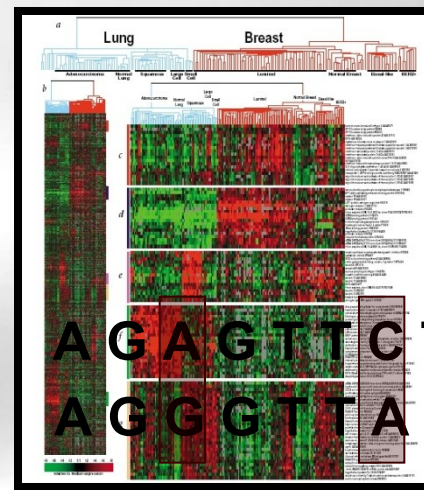
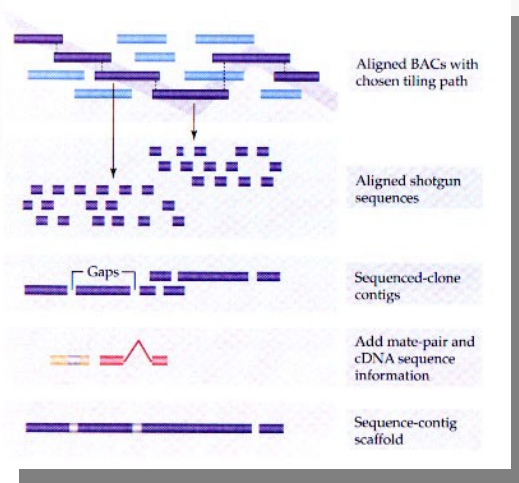
**CGAP**  
Cancer Gene Anatomy Project

**Cn3D**  
View 3-dimensional structures

**COGs**  
Clusters of Orthologous Groups

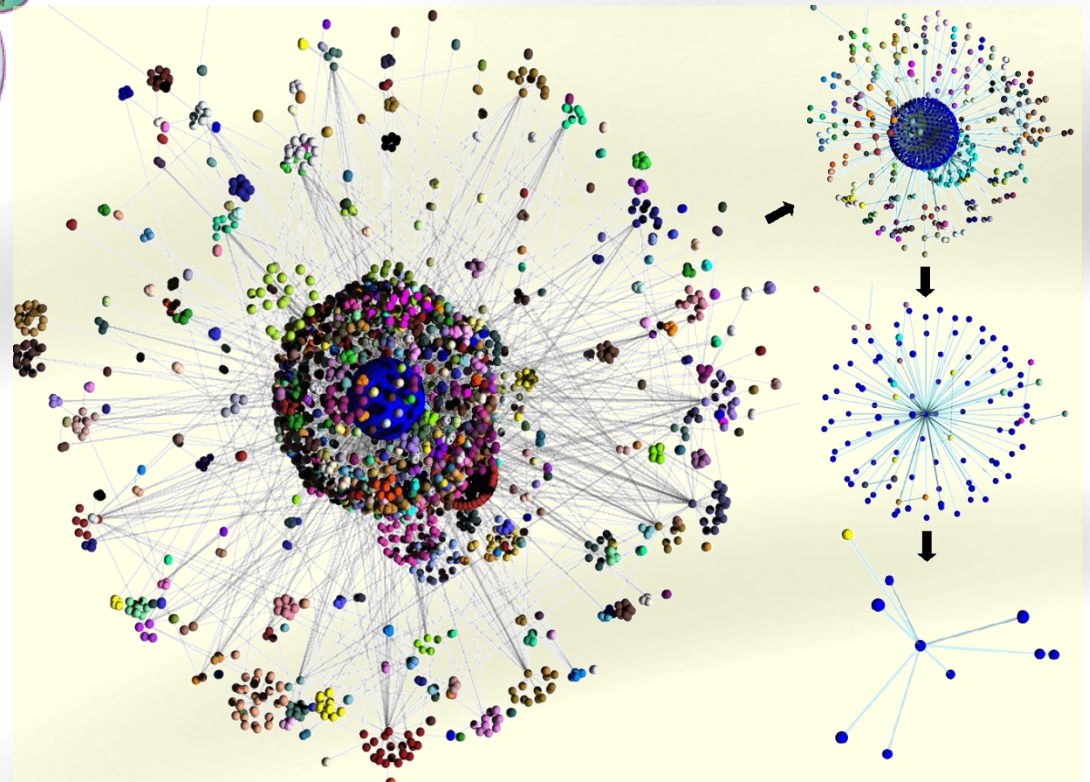
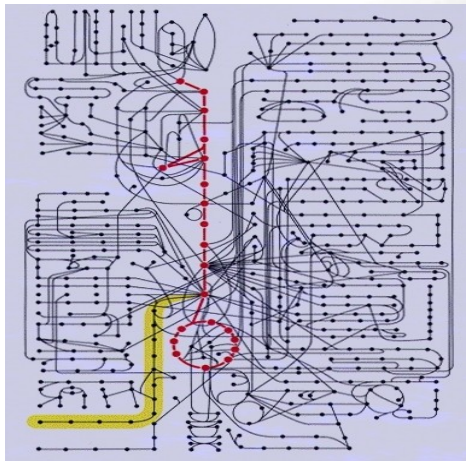
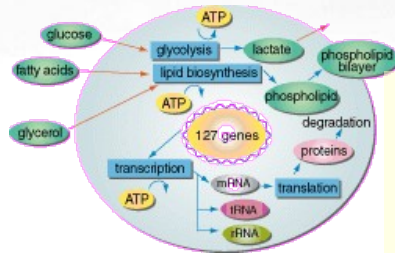


# Analysis and interpretation

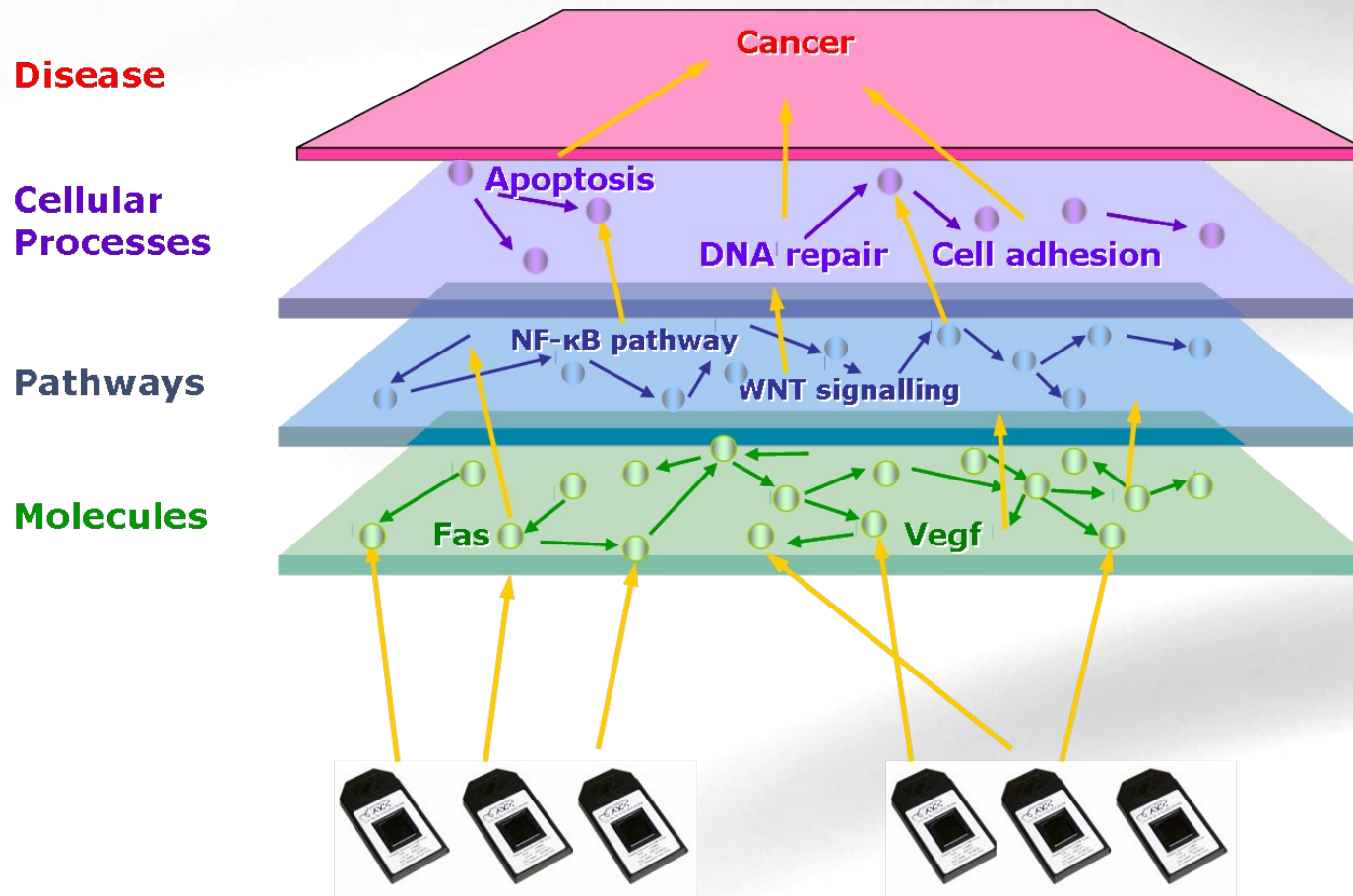




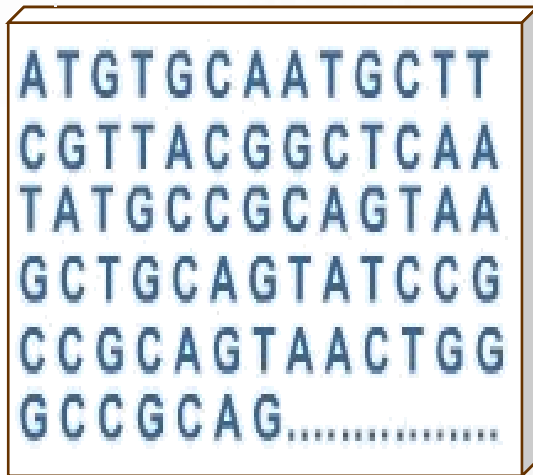
# Biological system modeling



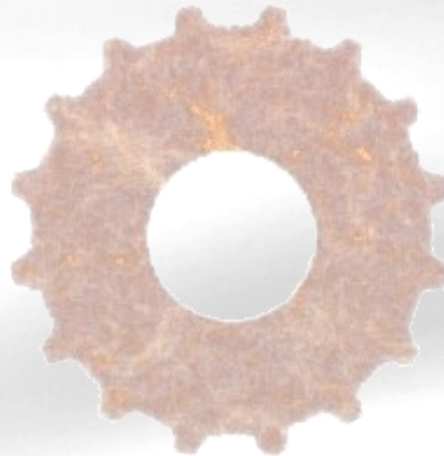
# Integrative bioinformatics



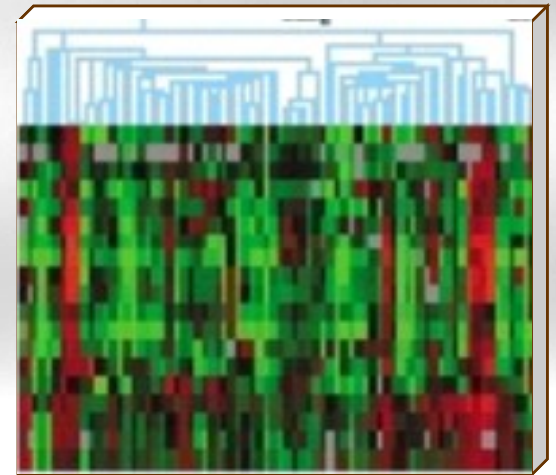
# In summary...



**Data**



**Bioinformatics  
methods and  
resources**



**Knowledge**

# In summary, bioinformatics ...

- Was born with
  - Development of new technologies
  - Its application for generating –increasingly huge– of big masses of biological data.
- *Has become now an interdisciplinary science encompassing all aspects of the Acquisition, Processing, Distribution, Analysis, Integration and Interpretation of biological information.*

# “Doing bioinformatics”

- Bioinformatics analyses
  - Database searching/querying
  - Sequence analysis, Omics data analysis
  - Systems biology
- Can be done differently
  - From console-based systems
    - Using scripts (perl/python/R) for automating processes
    - Doing data analysis with R
  - Or working with graphical/web interfaces to do (almost) the same things
- Each user has a different preferred approach