
INTRODUCTION TO BIOINFORMATICS

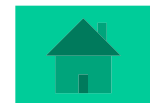


Contents

- [What is bioinformatics?](#)
- [Main thematic of the BIF](#)
- [Big centers and data bases](#)
- [References](#)

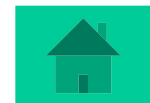


What is bioinformatics?

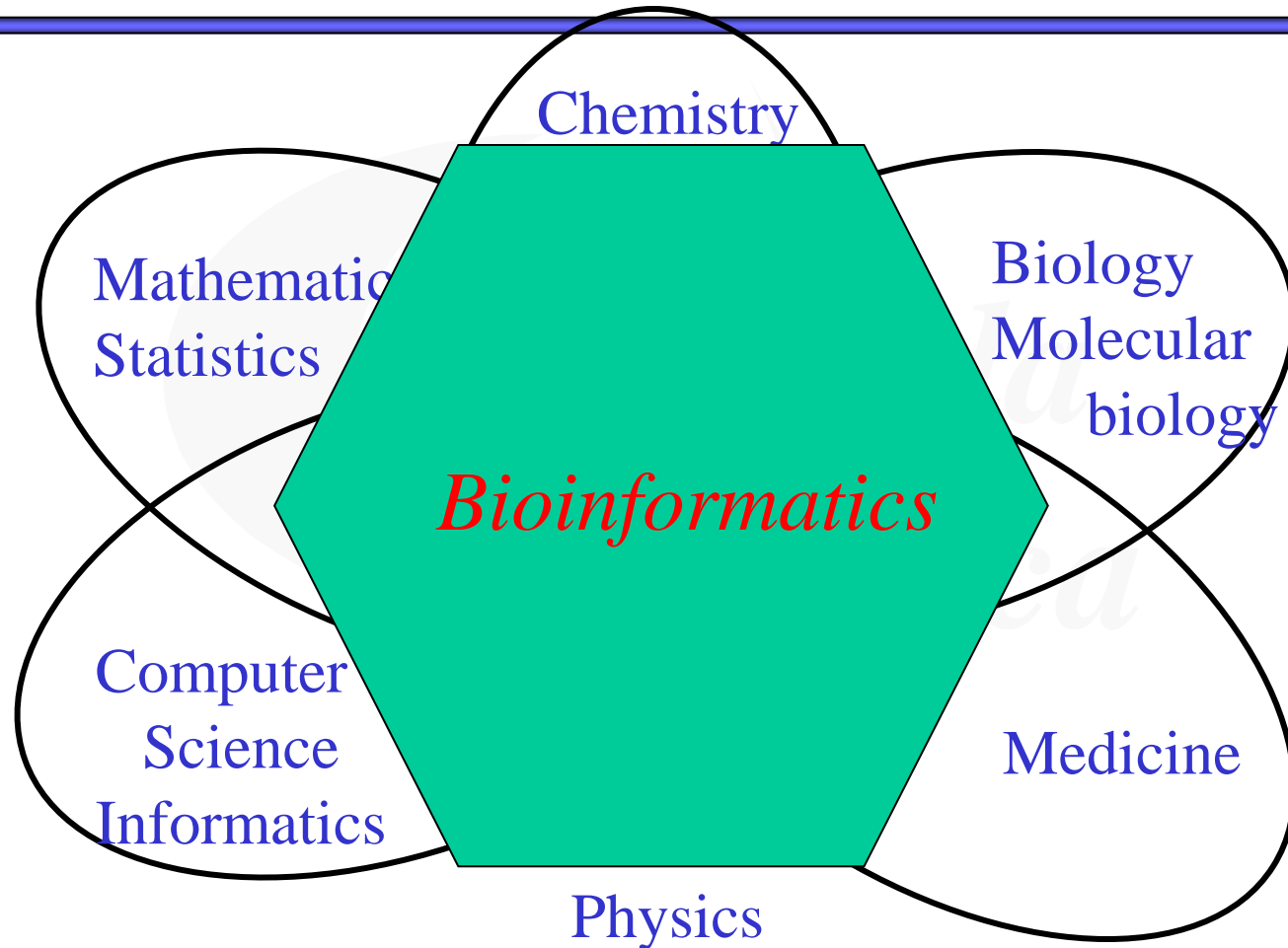


Bioinformatics

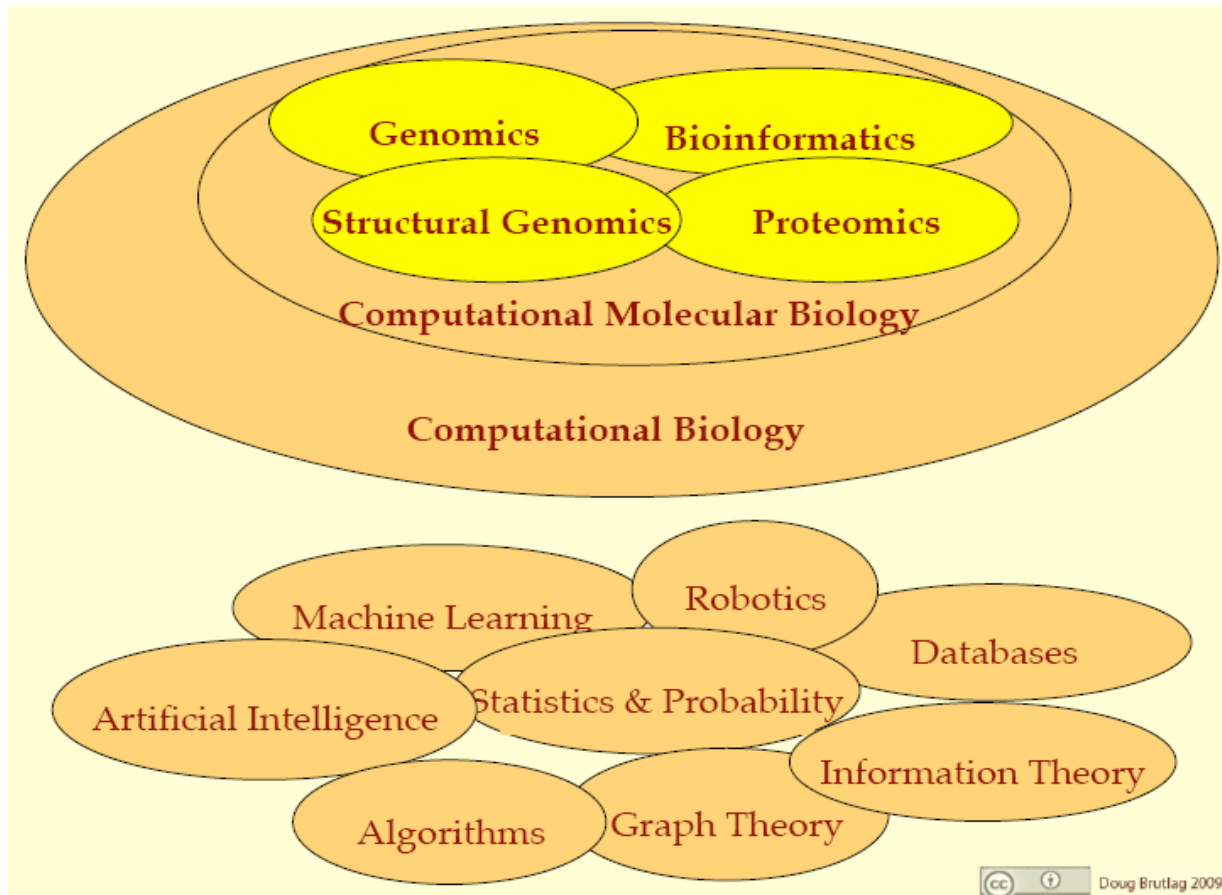
- it was born as consequence
 - development of new technologies and
 - its application to the generation of large amounts of data.
- The scientific discipline that encompasses all aspects of the acquisition, processing, distribution, analysis, interpretation and integration of biological information.



Bioinformatics and Interdisciplinarity



Bioinformatics, Biology and Computacional Biology



Computacional Biology or Bioinformatics

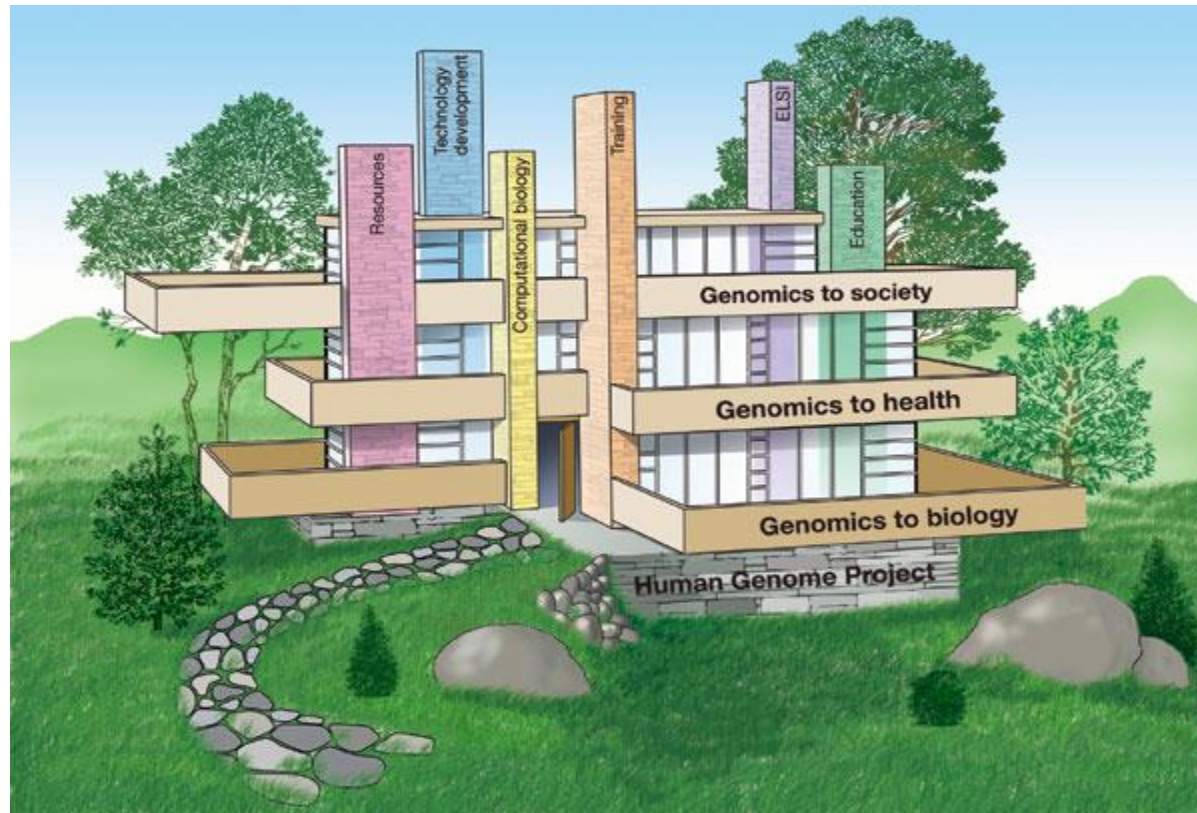
- **Computational biology** applies the techniques of computer science, applied mathematics and statistics to address biological problems.

Bioinformatics is the application of information technology to the field of molecular biology.

It's Biometry, st...



Genomics & Bioinformatics



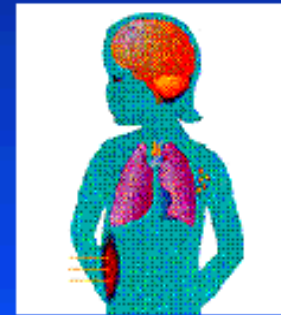
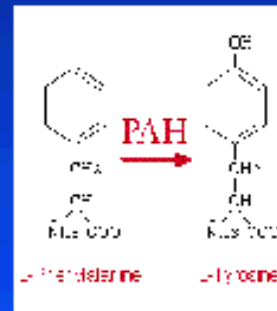
The future of genomics rests on the foundation of the Human Genome Project



Central Paradigm of Bioinformatics

Genetic Information → Molecular Structure → Biochemical Function → Symptoms (Phenotype)

SRAIMKIVA
VSIQTSSAVVN
VSTATVSRALA
GVTTTSSKVIN
SCVSAVSAILN
GVSEMTEDLN
TAYATIHVVE
CSQPTVSRRLA
MSIATITECSN
ISSETVSRILK
FDISRLSGLFR
LSPRLANGLFR
MTVETISRLLC
TLRPHLNLFR



© Doug Brutlag, 1996



The scopes of Bioinformatics

BIOINFORMATICS

Vol. 18 no. 12 2002
Pages 1551-1552

Editorial

BIOINFORMATICS: BIOLOGY BY OTHER MEANS

The success of bioinformatics in its application to genomics and proteomics has complicated the relationship of computation with experimental biology. There is a need to attend to our pressing needs of bioinformatics applications without forgetting other, perhaps less evident but equally important, aspects of computation in biology.

MAINSTREAM BIOINFORMATICS

In the field of genomics and proteomics, bioinformatics provides the key connection between all different forms of data gathered by new high-throughput techniques such as systematic sequencing, proteomics, expression arrays, yeast two-hybrid (y2h), and high throughput screens (Figure 1). We will soon have at our disposal tens of thousands of genomes, thousands of protein structures, protein interactions determined by y2h and tens of thousands of genes with their expression monitored in hundreds of experiments, and hundreds of thousands of single-nucleotide polymorphisms (SNPs).

Handling this massive amount of data requires powerful integrated bioinformatics systems. Issues related to database inter-operability, information representation and data classification (the much abused term 'ontology') are currently being addressed. Also, in fields such as automatic extraction of information from the biological literature, activity has increased greatly since the first papers were published five years ago.

BIOINFORMATICS IN THE STUDY OF SPECIFIC BIOLOGICAL PROBLEMS

BIOINFORMATICS IN THE STUDY OF GENERAL BIOLOGICAL PROBLEMS

A much deeper aspect of bioinformatics extends towards the study of fundamental biological questions, such as gene assembly, protein folding and the nature of functional specificity. Such issues extend beyond the current perception of bioinformatics as a support discipline and address aspects of biological complexity, including the simulation of cellular systems and molecular interaction networks. The contribution of bioinformatics to these areas is related to the development of concepts in theoretical molecular biology, but also to the management and representation of complex biological information.

The study of particular systems is the source of inspiration that guides the formation of general ideas from specific cases to general principles. The study of fundamental problems encourages the interdisciplinary nature of bioinformatics and allows the field to re-invent itself. It may be the interplay of these activities in these areas that defines bioinformatics as the 'biology by other means'.

However, different areas have been developing at different rates. The technical and computational developments are very attractive for newcomers from fields such as computer science, engineering and mathematics. The practical applications of bioinformatics are highly sought after by institutions and companies, and constitute the natural entry point for most molecular biologists and biochemists. Perhaps the work related to the fundamental biological problems is less well-regarded and requires more attention, given its importance for the future of biology as a quantitative science.

Finally, I would like to stress that the training programs in bioinformatics have an essential role in preparing the

molecular
systems
biology

Setting standards in Systems Biology



The scopes of Bioinformatics

- Organization of information
 - Databases
 - Algorithms and exploitation tools
- Analysis and interpretation of experimental results
 - DNA sequencing & genome analysis
 - Comparative Genomics
 - Transcriptomics and gene expression
 - Proteomics, Protein-Protein Interaction networks
 - Models of Biological Systems





NCBI Tools for Data Mining

PubMed	Entrez	BLAST	OMIM	Books	TaxBrowser	Structure
--------	--------	-------	------	-------	------------	-----------

Search for

Nucleotide Sequence Analysis	Protein Sequence Analysis	Structures	Genome Analysis	Gene Expression
------------------------------	---------------------------	------------	-----------------	-----------------

NCBI

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

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 001101011AG 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.

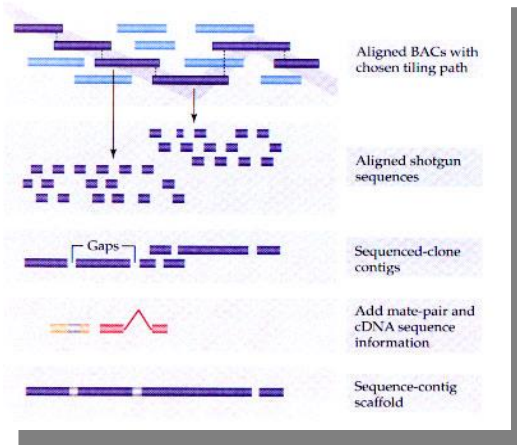
 **Entrez Gene** - each Entrez Gene record encapsulates a wide range of information for a given gene and organism. When possible, the information includes results of analyses that have been done on the sequence data. The amount and type of information presented depend on what is available for a particular 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 CDS, 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.

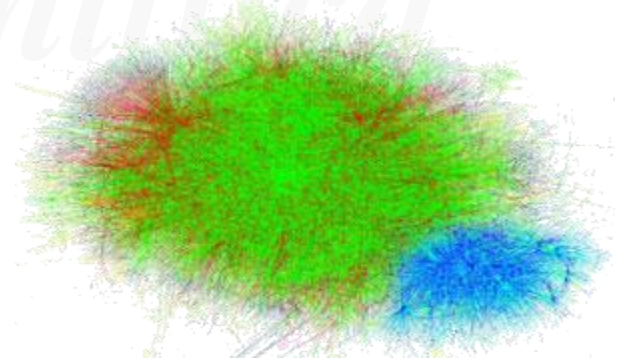
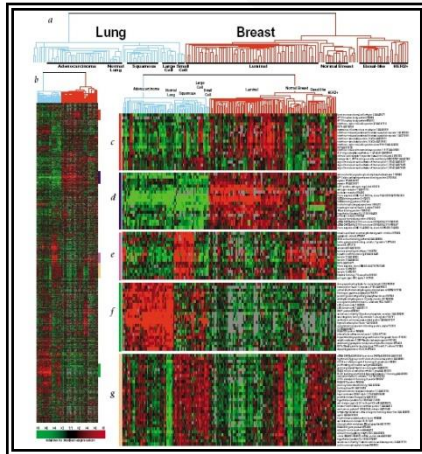
 **ORF Finder** - identifies all possible ORFs in a DNA sequence by locating the standard and alternative stop and start codons. The deduced amino acid sequences can then be used to BLAST against GenBank. ORF finder is also packaged in the sequence submission software **Seqin**.

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

Analysis and interpretation



AGAGTTCTGCTCG
AGGGTTATGCGCG

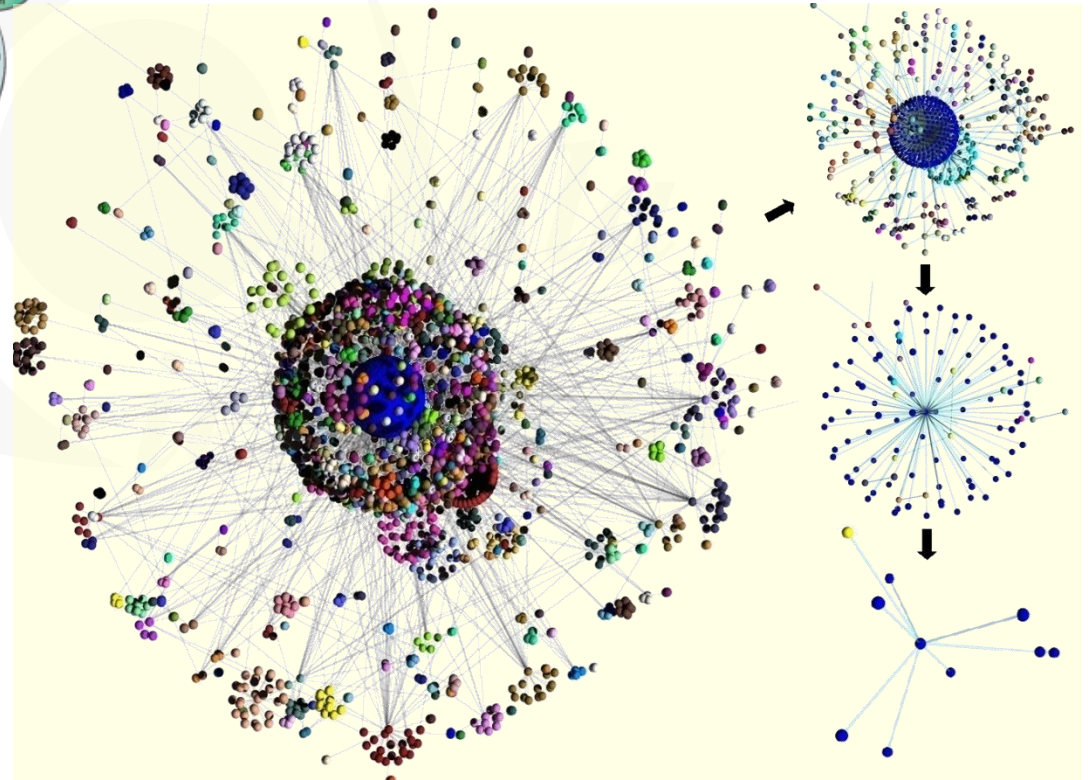
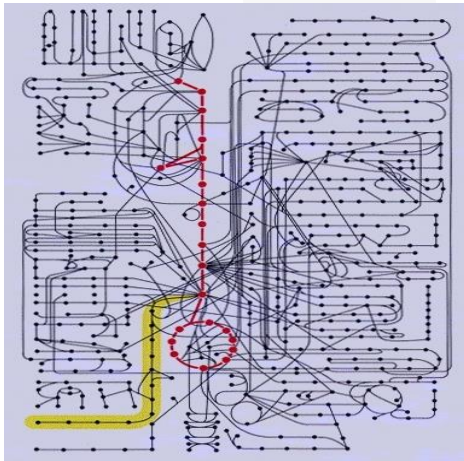
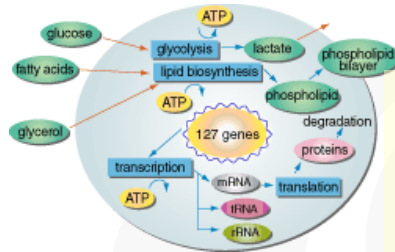


Challenges in bioinformatics data analysis

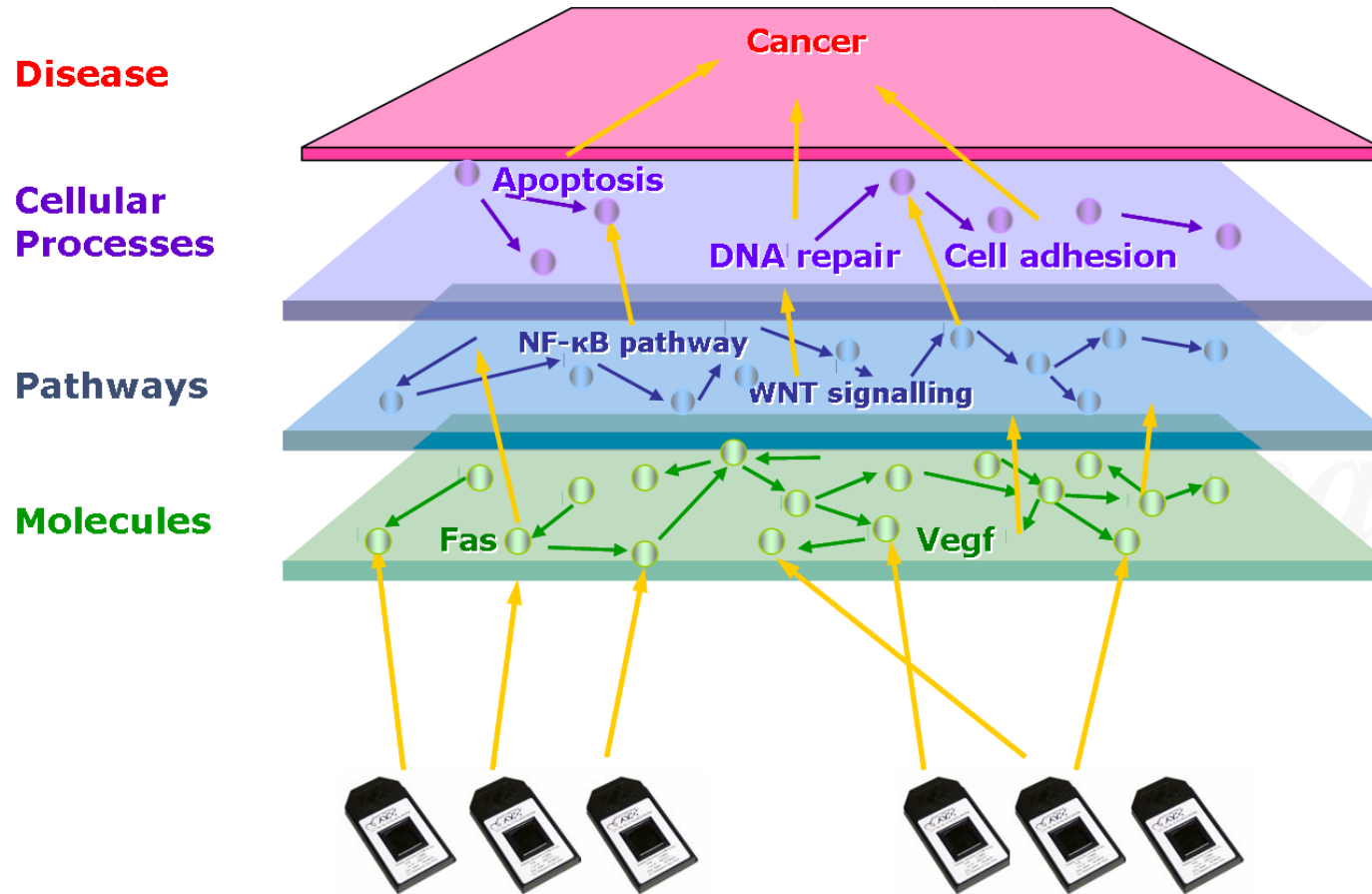
- MULTIPLE-COMPARISONS ISSUE
 - HIGH-DIMENSIONAL BIOLOGICAL DATA
 - SMALL-n AND LARGE-p PROBLEM
 - NOISY HIGH-THROUGHPUT BIOLOGICAL DATA
 - INTEGRATION OF MULTIPLE, HETEROGENEOUS BIOLOGICAL DATA INFORMATION
-
- Textbook: Jae K. Lee. (2010). Statistical Bioinformatics: For Biomedical and Life Science Researchers. Wiley-Blackwell



Models of Biological Systems



Integrative Bioinformatics

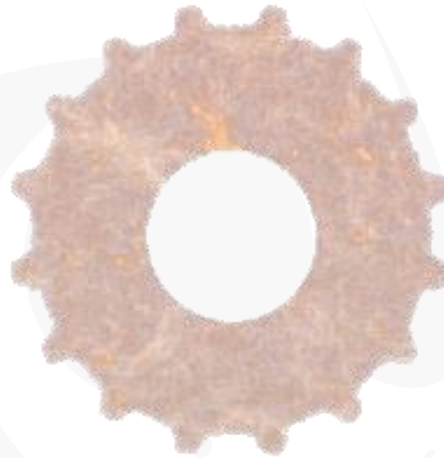


In summary ...

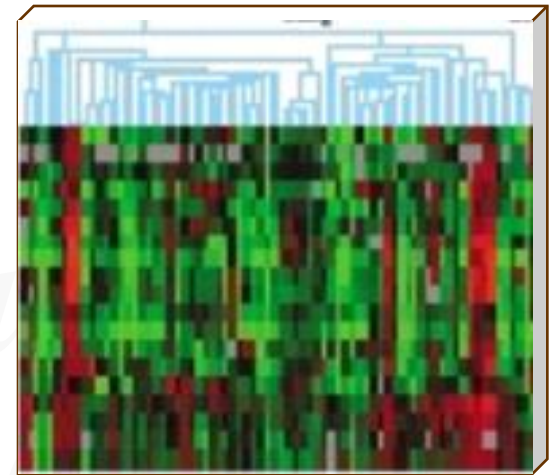


```
ATGTGCAATGCTT
CGTTACGGCTCAA
TATGCCGCAGTAA
GCTGCAGTATCCG
CCGCAGTAACTGG
GCCGCAG.....
```

Data



Biological resources
and tools



Knowledge

- As however defined, from wherever you look, the role of bioinformatics has been, is and will be crucial to the advancement of biology and medicine of the XXI century



What is a **bioinformatic**?

2 – Environment, training, bioinformatics activities



What knowledge have to know a bioinformatics?

- Must have "solid knowledge" in
 - Some biological discipline
 - Biochemistry, Genetics,...
 - Software development environments
 - OS [Linux], Programming languages [Perl, Java, R, Python], Databases[SQL], Web development[PHP, ASP, Ajax...]
 - Some quantitative discipline
 - [Mathematics, Statistics, Physics]

At least two of the above three!!



What does a bioinformatic?

- Information Management
 - Implementation and operation biological databases –offline or online–
 - Installation and maintenance web servers.
- Application Development
 - Creation of local or web programs.
- Operation and data analysis
 - Microarrays, high-performance data



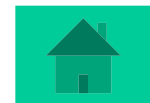
Where it is done bioinformatics?

- Specialized Centers
 - EBI, NCBI, EMBL.
 - INB / Bioinformatics platform of the UAB
- Services bioinformatics of research centers,
 - UEB, UBB, BU
- Universities,
- Pharmaceutical Laboratories,
- ...



How bioinformatics is done?

- Usually, though not necessarily the BIF has "universal" vocation, users access a maximum of:
 - Often sought solutions WEB
 - Often based on projects [more or less] Free open source distribution.
 - This is not any general
 - For example *Ingenuity Pathway Analysis* is not free but it's good.



For more information:

- There are many free resources
 - [Train on line](#) en el EBI
 - [Tutoriales](#) del NCBI
 - “locals” course
 - [Invitacio a la Bioinformatica](#) (Plataforma BIF UAB)
- A variety of books on the subject
 - [List of books on bioinformatics](#)
- Journals and scientific societies
 - [Bioinformatics](#), [Briefings in Bioinformatics](#)
 - [International Society for Computational Biology](#)
 - [List of bioinformatics journals\(wikipedia\)](#)



Bioinformatics centers & Biological Databases



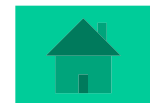
Biological Databases

- Much of the work in bioinformatics involves the construction and / or operation of databases of biological information
- They are used, for example for:
 - Add or seek information ("annotations")
 - Find similarities or patterns
 - Make predictions
 - From protein structure or function
 - From genes to genomes



The access resources

- The WWW has revolutionized the provision of services in bioinformatics
- Many things can be done through internet without local copies of databases or software are needed to exploit
- Despite this globalization there are organizations that centralize resources



Centres and significant resources

- Major centers worldwide
 - EMBL / EBI (www.embl.org / www.ebi.ac.uk)
 - NCBI (www.ncbi.nlm.nih.gov)
 - DDBJ (www.ddbj.nig.ac.jp)
- Biological Databases
 - EMBL DNA sequence database
 - SWISSPROT i TREMBL
 - PIR, PDB
 - Catalog of biological databases
www.infobiogen.fr/services/dbcat

