



DBE
DEPARTAMENTO
DE BIOENGENHARIA
TÉCNICO LISBOA



DEI
DEPARTAMENTO
DE ENGENHARIA INFORMÁTICA
TÉCNICO LISBOA



Bioinformática

Desafios computacionais e matemáticos no estudo de sistemas biológicos

Verão na Lisboa 2022

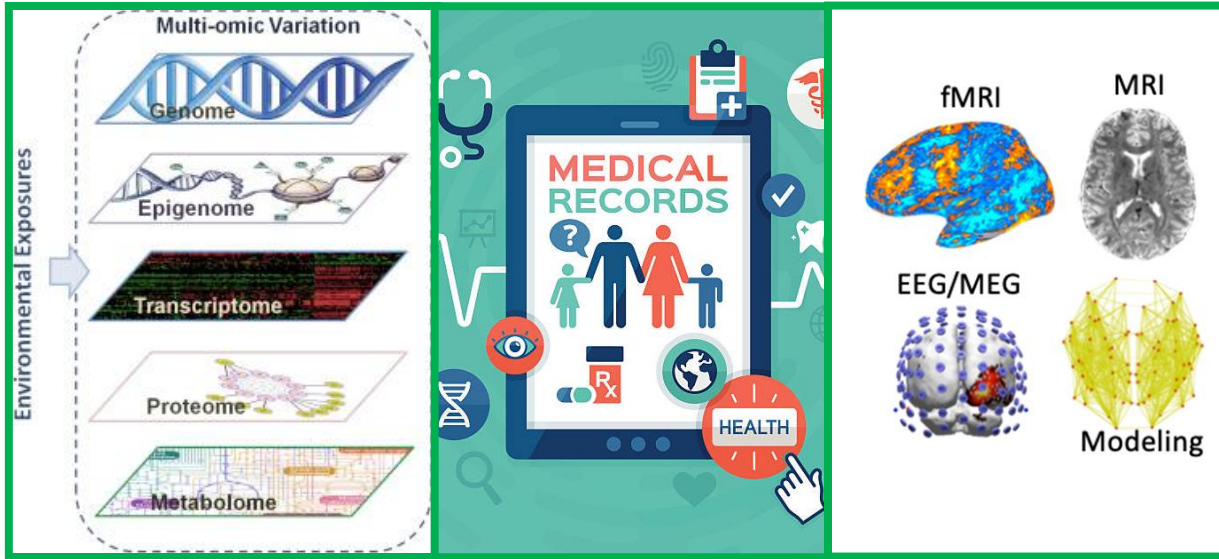
Rui Henriques, rmch@tecnico.ulisboa.pt
Susana Vinga, susana.vinga@tecnico.ulisboa.pt

Tópicos de investigação

Multi-omics

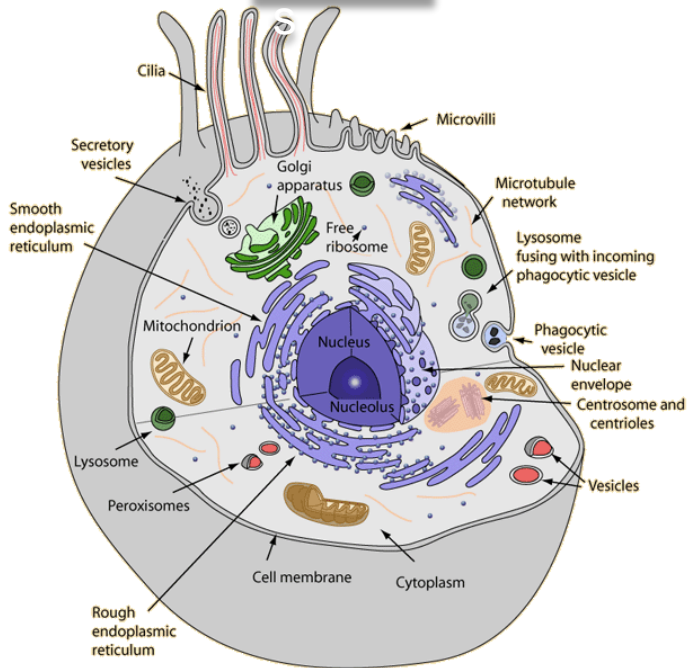
Electronic health records

Brain imaging



Biologia, Matemática e Engenharia

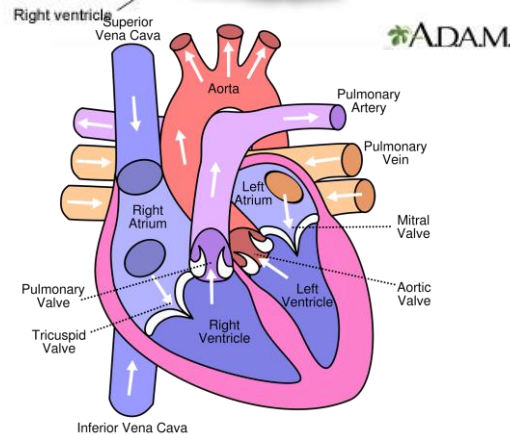
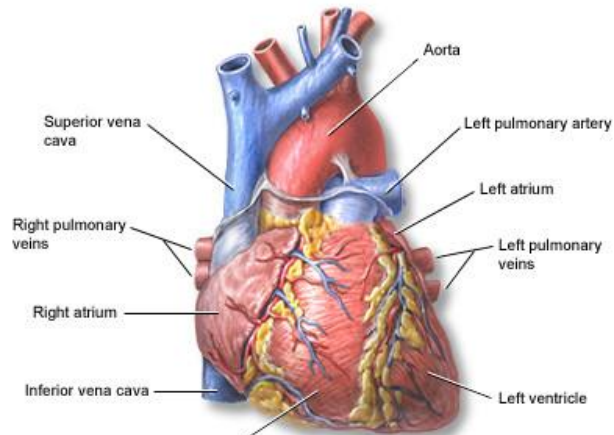
Célula



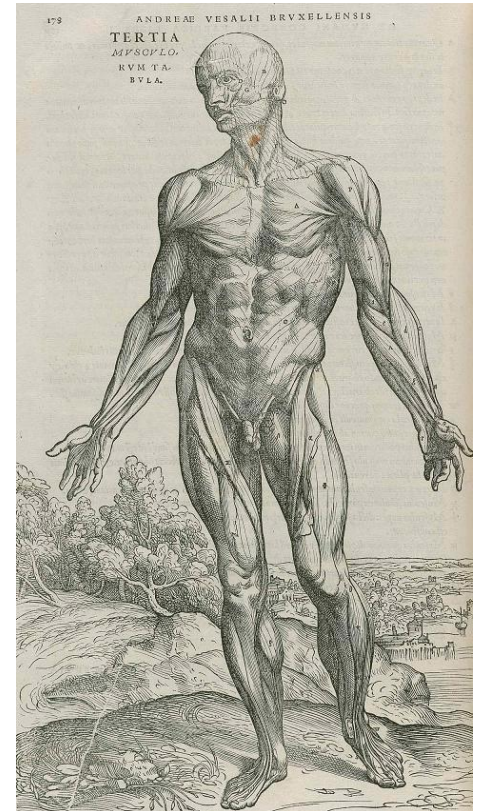
10 μ m, 1ng

<http://hyperphysics.phy-astr.gsu.edu/hbase/biology/cell.html>

Órgãos

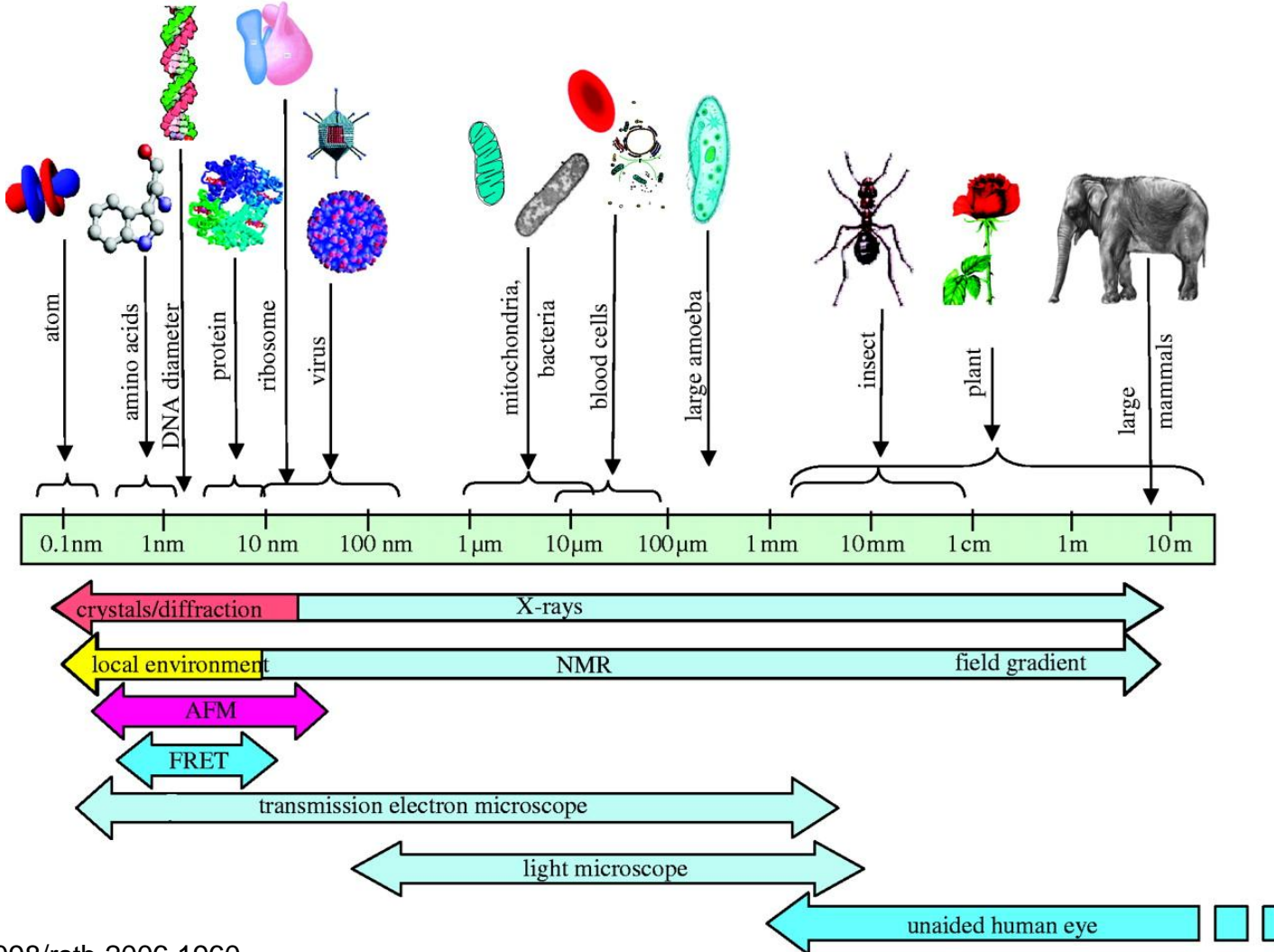


Organismos



Ecosystemas

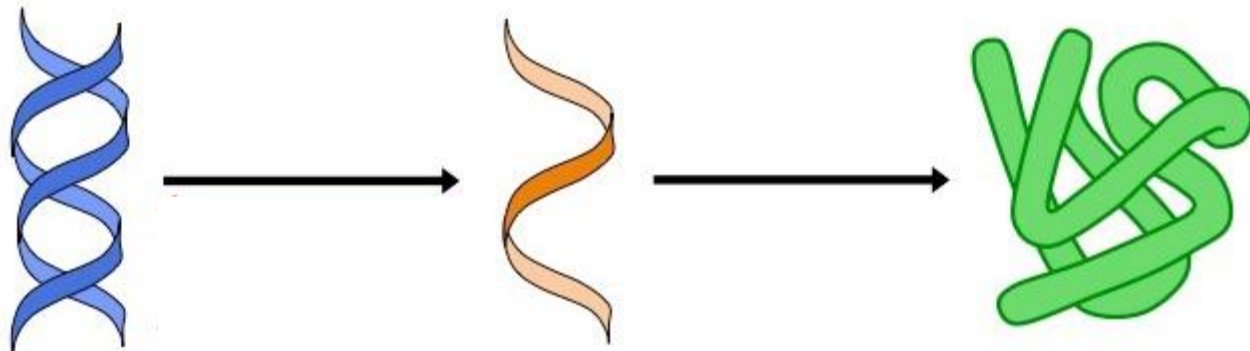
Escalas Biológicas

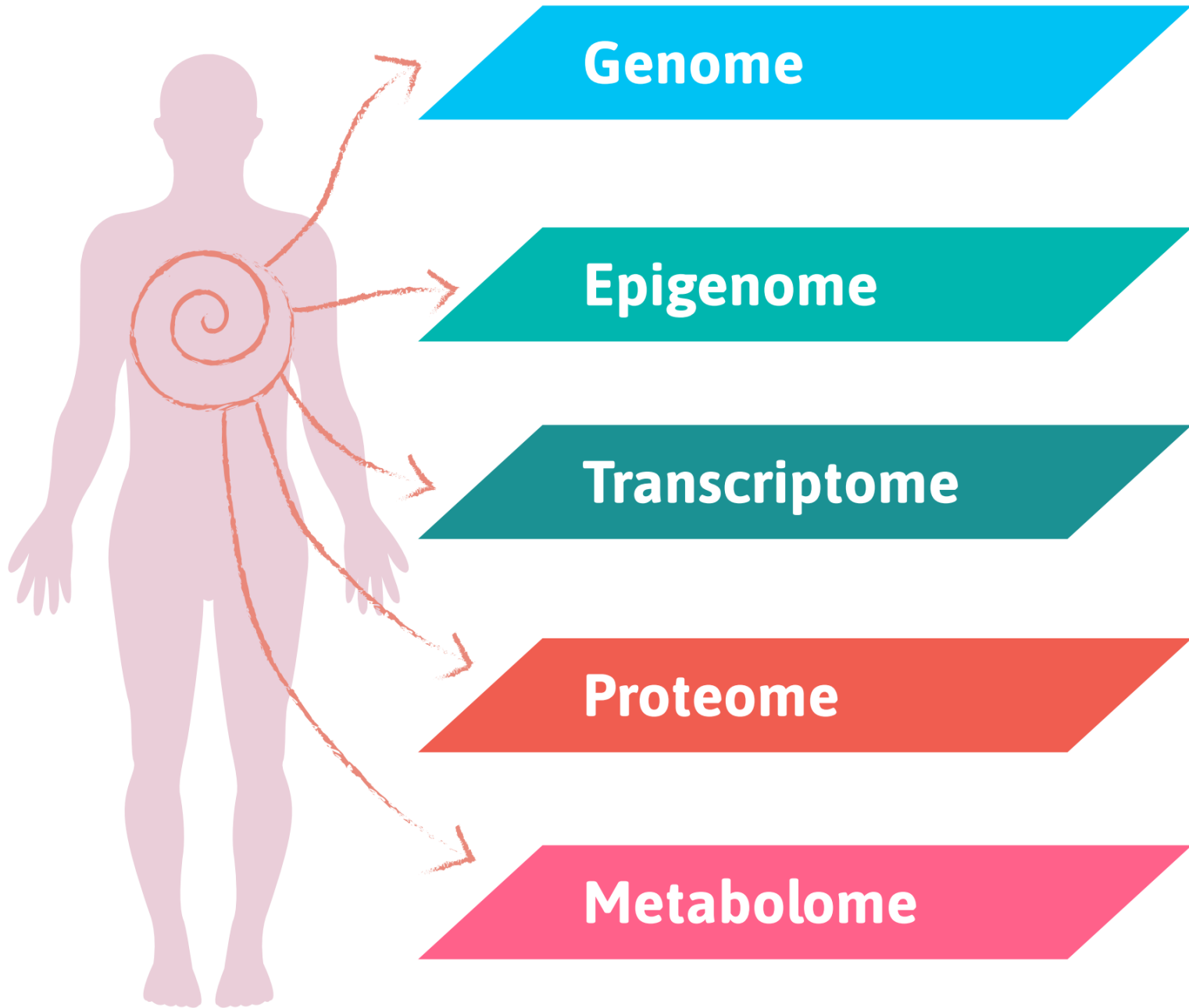


Bioinformática e Biologia Computacional

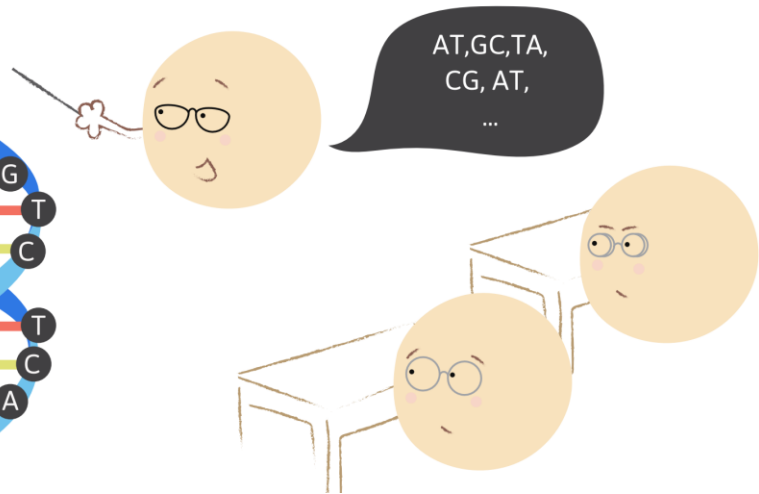
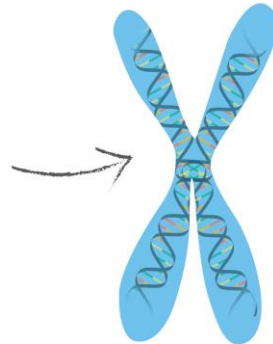
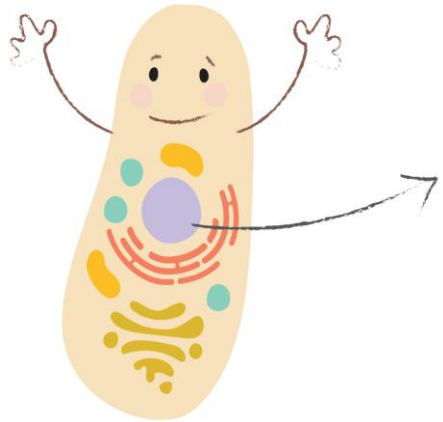
- **Bioinformatics** – application of computational tools to expand the use of biological, medical, behavioral data, including acquisition, consolidation, visualization, analysis
- **Computational biology** – development and application of data-analytical methods, mathematical modeling and computational simulation techniques to study biological, behavioral, and social systems

Fluxo de informação: dogma central

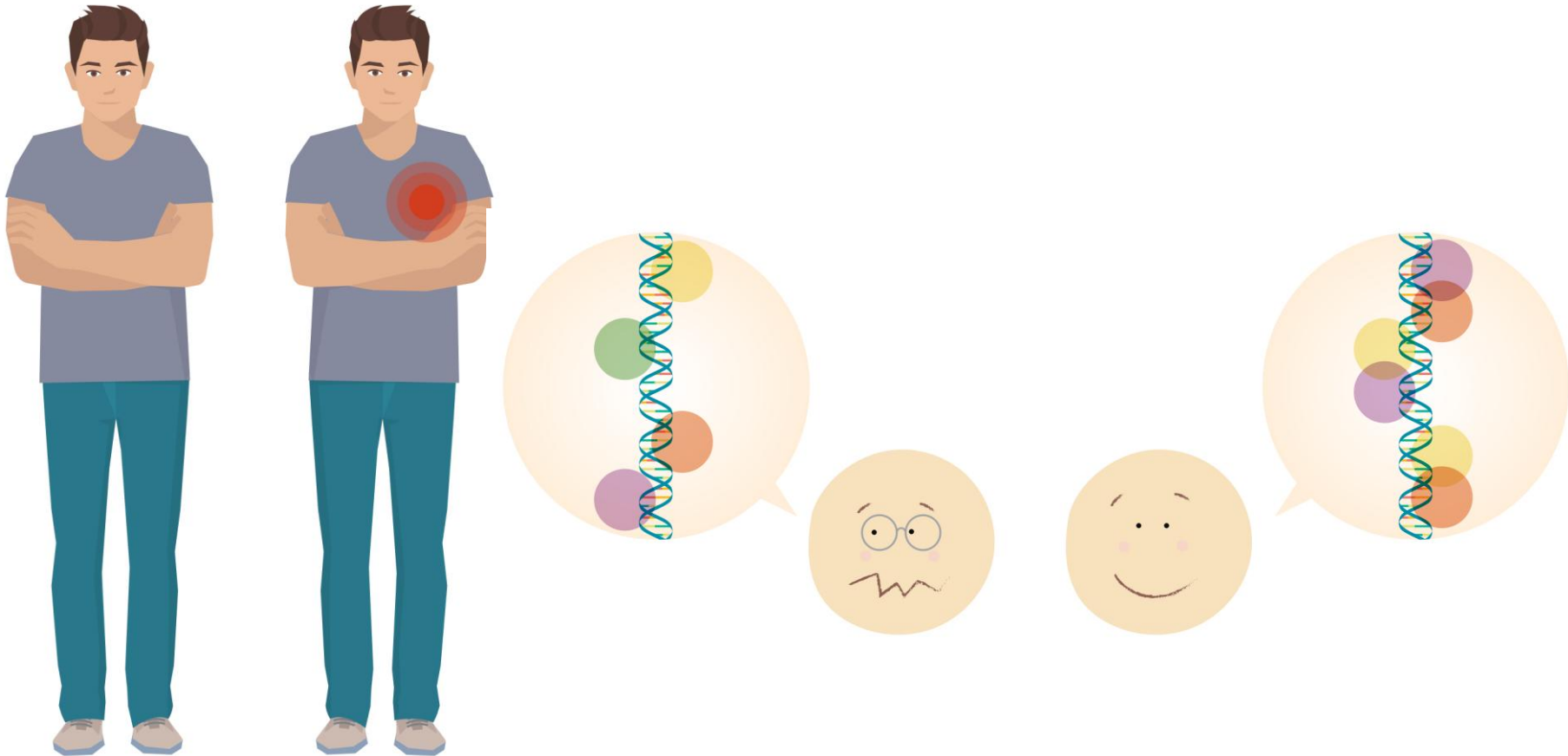




Genética...

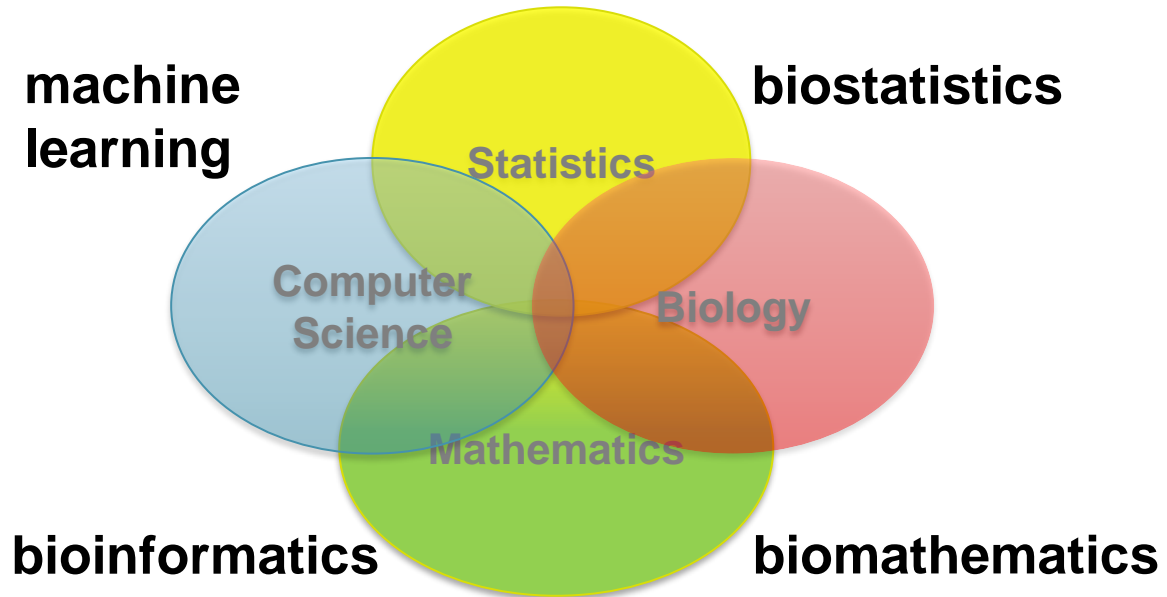


... Epigenética



Bioinformática e Biologia Computacional

Bioinformática e **Biologia Computacional** são áreas científicas nas quais a **informática**, **biologia**, **matemática** e **estatística** se encontram



Tópicos

- Medicina e saúde
- Evolução e filogenia
- Biotecnologia
- Genómica, epigenómica
- Transcriptómica, proteómica
- Biologia estrutural
- Biologia de sistemas
- Microbioma

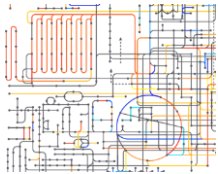


Áreas de investigação



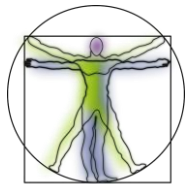
1. Bioinformática

Teoria da Informação: **análise de sequências**



2. Biologia de Sistemas

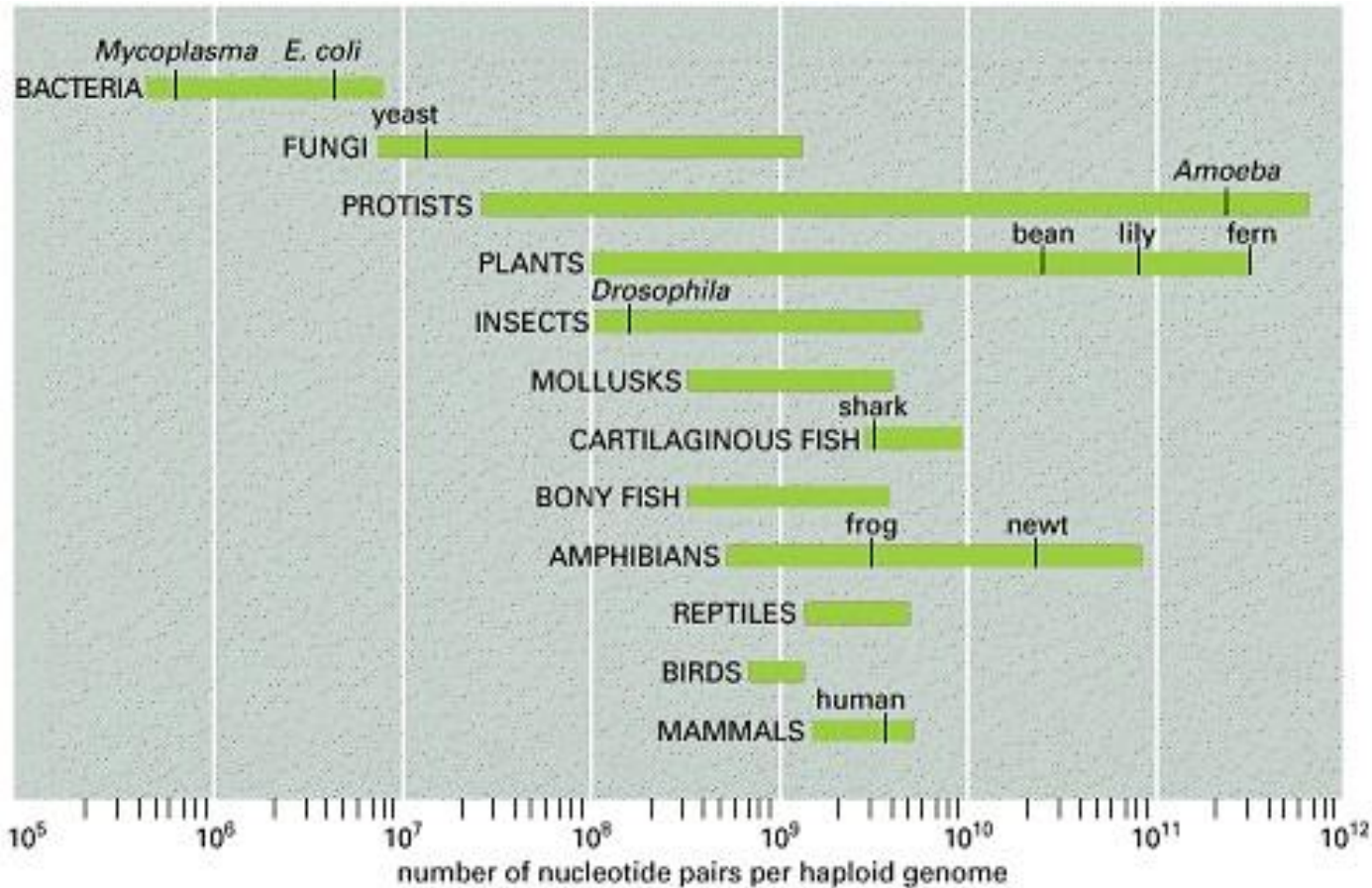
Modelação dinâmica, identificação e optimização de sistemas biológicos: **redes metabólicas**



3. Biomedicina

Bioestatística e aprendizagem automática para análise de dados clínicos, e.g. **infecção e cancro**

Tamanho do Genoma



In: "Molecular Biology of the Cell" – Alberts et al.



A importância crescente da Bioinformática

1990

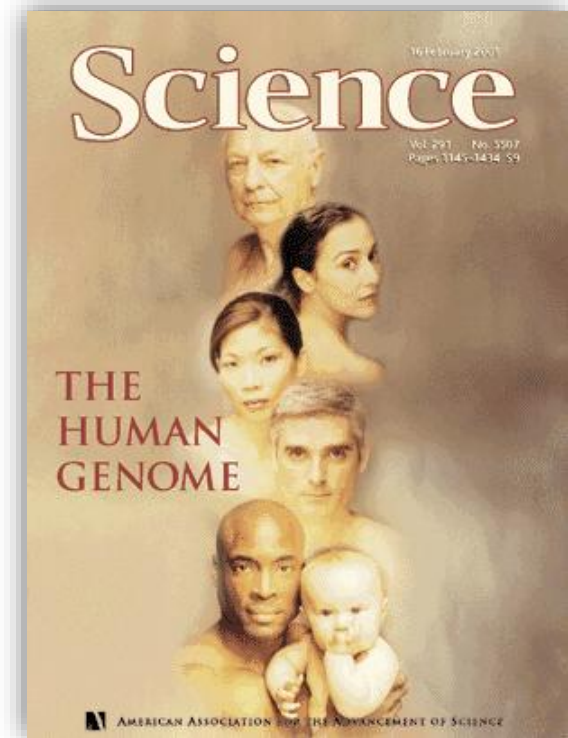
Human Genome Project

Um projecto de 15 anos lançado pelo Congresso Americano



2001

Publicação do primeiro rascunho do genoma Humano



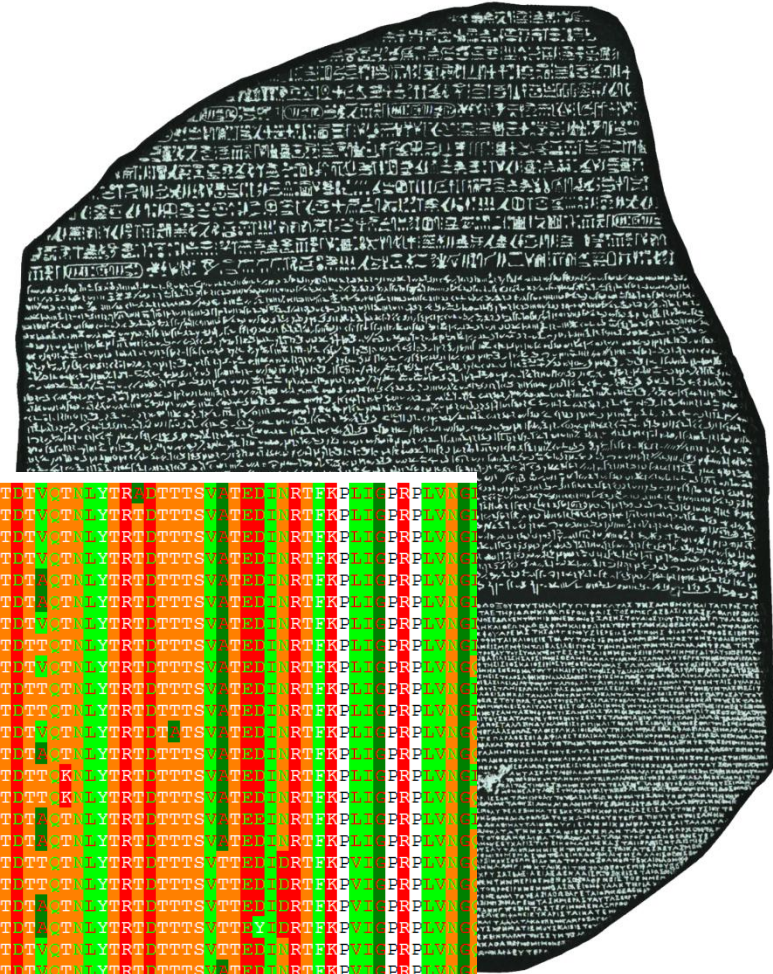
Courtesy: Prof. Ana Teresa Freitas



CTCGAGGGGCTAGACATTGCCCTCCAGAGAGACACCCAACACCCTCCAGGCTTGACCGGCCAGGGTGTCCCTTTCCTACCTTGGAGAGAGCAGC
CCCAGGGCATCCTGCAGGGGGTGTGGGACACCAGCTGGCCTTCAAGGTCTCTGCCTCCCTCCAGCCACCCCACTACACGCTGCTGGGATCCTGGA
TCTCAGCTCCCTGGCCGACAACACTGGCAAACCTCTACTCATCCACGAAGGCCCTCCTGGGCATGGTGGTCTTCCCAGCCTGGCAGTCTGTTCCCT
CACACACCTTGTAGTGCCAGCCCTGAGGTTGCAGCTGGGGGTGTCTCTGAAGGGCTGTGAGCCCCAGGAAGCCCTGGGGAAGTGCCTGCCTT
GCCTCCCCCGGCCCTGCCAGCGCCTGGCTCTGCCCTCCTACCTGGGCTCCCCCATCCAGCCTCCCTCCCTACACACTCCTCTCAAGGAGGCACC
CATGTCTCTCCAGCTGCCGGCCTCAGAGCACTGTGGCGTCTGGGGCAGCCACCGCATGTCTGCTGTGGCATGGCTCAGGGTGGAAAGGGCGG
AAGGGAGGGGTCTGCAGATAGCTGGTGCCACTACCAAACCCGCTCGGGGCAGGAGAGCCAAAGGCTGGGTGTGTGCAGAGCGGCCCCGAGAGGT
TCCGAGGCTGAGGCCAGGGTGGGACATAGGGATGCGAGGGGCGGGGCACAGGATACTCCAACCTGCCTGCCCCATGGTCTCATCCTCCTGCTTC
TGGGACCTCCTGATCCTGCCCTGGTGCTAAGAGGCAGGTAAGGGGCTGCAGGCAGCAGGGCTCGGAGCCCATGCCCCCTACCATGGGTGAGGCT
GGACCTCCAGGTGCCTGTTCTGGGGAGCTGGGAGGGCCGGAGGGGTGTACCCAGGGGCTCAGCCAGATGACACTATGGGGGTGATGGTGTGATG
GGACCTGGCCAGGAGAGGGGAGATGGGCTCCAGAAGAGGAGTGGGGGCTGAGAGGGTGCCTGGGGGGCCAGGACGGAGCTGGGCCAGTGCACAGC
TTCCACACCTGCCACCCCCAGAGTCTGCCGCCACCCCCAGATCACACGGAAGATGAGGTCCGAGTGGCCTGCTGAGGACTTGCTGCTTGTCCC
CAGGTCCCAGGTGATGCCCTCCTTCTGCCACCCTGGGGAGCTGAGGGCCTCAGCTGGGGTGTGTCTAAGGCAGGGTGGGAACTAGGCAGCCA
GCAGGGAGGGGACCCCTCCCTCACTCCCACTCTCCACCCCCACCACCTTGGCCATCCATGGCGGCATCTTGGGCCATCCGGGACTGGGGACAGG
GGTCTTGGGGACAGGGTCCGGGGACAGGGTCTGGGGACAGGGGTGTGGGGACAGGGGTCTGGGGACAGGGGTGTGGGGACAGGGGTGTGGGG
AGGGGTCTGGGGACAGGGGTGTGGGGACAGGGGTCCGGGGACAGGGGTGTGGGGACAGGGGTCTGGGGACAGGGGTGTGGGGACAGGGGTGTGGGG
ACAGGGGTCTGGGGACAGGGGTGTGGGGACAGGGGTCTGGGGACAGGGGTGTGGGGACAGGGGTGTGGGGACAGGGGTGTGGGGACAGGGGTGTG
GGGACAGGGGTCTGGGGATAGGGGTGTGGGGACAGGGGTGTGGGGACAGGGGTCCCGGGACAGGGGTGTGGGGACAGGGGTGTGGGGACAGGGG
TCCTGGGGACAGGGGTCTGAGGACAGGGGTGTGGGCACAGGGGTCTGGGGACAGGGGTCTGGGGACAGGGGTCTGGGGACAGGGGTCTGGGGA
CAGCAGCGCAAAGAGCCCCGCCCTGCAGCCTCCAGTCTCTCCTGGTCTAATGTGGAAAGTGGCCCAGGTGAGGGCTTTGCTCTCCTGGAGACATTTG
CCCCAGCTGTGAGCAGGGACAGGTCTGGCCACCGGGCCCTGGTTAAGACTCTAATGACCCGCTGGTCTGAGGAAGAGGTGCTGACGACCAAGG
AGATCTTCCACAGACCCAGCACCAGGGAAATGGTCCGGAAATTGCAGCCTCAGCCCCAGCCATCTGCCGACCCCCCACCCTAATGGGC
CAGGCGCAGGGGTTGACAGGTAGGGGAGATGGGCTCTGAGACTATAAAGCCAGCGGGGGCCAGCAGCCCTCAGCCCTCCAGGACAGGCTGCATC
AGAAGAGGCCATCAAGCAGGTCTGTTCCAAGGGCCTTTGCGTCAGGTGGGCTCAGGGTTCAGGGTGGCTGGACCCAGGCCCCAGCTCTGCAGCA
GGGAGGACGTGGCTGGGCTCGTGAAGCATGTGGGGGTGAGCCAGGGGCCCAAGGCAGGGCACCTGGCCTTCAGCCTGCCTCAGCCCTGCCTGTC
TCCAGATCACTGTCTTCTGCCATGGCCTGTGGATGCGCCTCTGCCCTGCTGGCGCTGCTGGCCTCTGGGGACCTGACCCAGCCGCAGCCT
TTGTGAACCAACACCTGTGCGGCTCACACCTGGTGAAGCTCTCTACCTAGTGTGCGGGGAACGAGGCTTCTTCTACACACCCAAGACCCGCGGG
AGGCAGAGGACCTGCAGGGTGAAGCAACCGCCATTGCTGCCCTGGCCGCCCCAGCCACCCCTGCTCCTGGCGCTCCACCCAGCATGGGCAG
AAGGGGGCAGGAGGCTGCCACCCAGCAGGGGGTCAAGTGCACCTTTTTTAAAGAAAGTCTCTTGGTCAAGTCTAAAGTGAAGTCCAGGCTCCCTGTG
GCCAGTCAGAATCTCAGCCTGAGGACGGTGTGGCTTCCGGCAGCCCCGAGATACATCAGAGGGTGGGCACGCTCCTCCCTCCACTCGCCCCCTCAA
ACAAATGCCCCGACCCATTTCTCCACCCTCATTGATGACCGCAGATTCAAGTGTTTTGTAAAGTAAAGTCTGGGTGACCTGGGGTACAGGG
TGCCCCACGCTGCCTGCCTCTGGGCGAACACCCCATCACGCCCGGAGGAGGGCGTGGCTGCCTGCCTGAGTGGGCCAGACCCCTGTGCGCAGCCTC
ACGGCAGCTCCATAGTCAGGAGATGGGGAAGATGCTGGGGACAGGCCCTGGGGAGAAGTACTGGGATCACCTGTTCAAGGCTCCCACTGTGACGCTG
CCCCGGGGCGGGGGAAGGAGGTGGGACATGTGGGCGTTGGGGCCTGTAGGTCCACACCCAGTGTGGGTGACCCCTCCCTCTAACCTGGGTCCAGCCC
GGCTGGAGATGGGTGGGAGTGCACCTAGGGCTGGCGGGCAGGCGGGCACTGTGTCTCCCTGACTGTGTCTCCTGTGTCCCTCTGCCTCGCCGCT
GTTCCGGAACCTGCTCTGCGCGGCACGTCTTGGCAGTGGGGCAGGTGGAGCTGGGCGGGGGCCCTGGTGCAGGCAGCCTGCAGCCCTTGGCCCTGG
AGGGGTCCCTGCAGAAGCGTGGCATTGTGGAACAATGCTGTACCAGCATCTGCTCCCTTACCAGCTGGAGAACTACTGCAACTAGACGCAGCCTG

Alinhamento de sequências

A ideia fundamental do alinhamento é que as sequências que **partilham** os mesmos segmentos podem ter a mesma **função** ou estar relacionadas por **homologia**



Hebei_1	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G	
Ningxia*_1	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G	
Beijing_1	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G	
Henan98_1	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G	
Heilong01	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G	
Henan02_1	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G	
Jilin_1	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G	
Guang4/00	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G	
Henan00_1	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G	
Guang10/00	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G	
_1	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G	
1	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G	
01	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G	
09	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	S	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G
/9	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	S	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G
01	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	S	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G
*	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	S	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G
/9	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	S	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G
7/	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	S	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G
6/	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	S	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G
7	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	S	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G
7	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	S	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G
*	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	S	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G
1	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	S	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G
0	:	S	F	Y	R	S	M	R	W	L	T	K	N	A	P	O	A	C	Y	T	N	R	K	S	I	L	F	M	W	G	I	N	H	P	P	T	D	T	O	T	L	Y	T	R	I	T	T	S	V	A	T	I	D	I	R	T	F	E	L	I	E	R	P	L	V	N	G



ISFYRSMRWLTok NaYP O AcyTNNrak ILFMWGINHPPDtd Ot Lyt4dTTtSV Tedi RTFRP6IGRPRLVng

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

```
CCCCTCACCATGGGTCAGGCTGGACCTCCAGGTGCCTGTTCTGGGGAGCTGGGAGGGCC  
GGAGGGGTGTACCCAGGGGCTCAGCCCAGATGACACTATGGGGGTGATGGTGTTCATGGG  
ACCTGGCCAGGAGAGGGGAGATGGGCTCCAGAAGAGGAGTGGGGGCTGAGAGGGTGCC  
TGGGGGCCAGGACGGAGCTGGGCCAGTGCACAGCTTCCACACCTGCCACCCCCAGA  
GTCTGCGCCACCCAGATCACACGAAGATGAGGTCCGAGTGGCCTGCTGAGGACTT
```

Or, upload file

Choose file No file chosen

Job Title

Enter a descriptive title for your BLAST search

 Align two or more sequences

Choose Search Set

Database

 Human genomic + transcript Mouse genomic

Nucleotide collection (nr/nt)

Organism
Optional

Enter organism name or id--completions will be suggested

Enter organism common name, binomial, or tax id. Only 20 taxa will be shown

Exclude
Optional Models (XM/XP) Uncultured/environmental sequencesLimit to
Optional Sequences from type materialEntrez Query
Optional

Enter an Entrez query to limit search

[YouTube](#) [Create custom database](#)

Program Selection

Optimize for

- Highly similar sequences (megablast)
- More dissimilar sequences (discontiguous megablast)
- Somewhat similar sequences (blastn)

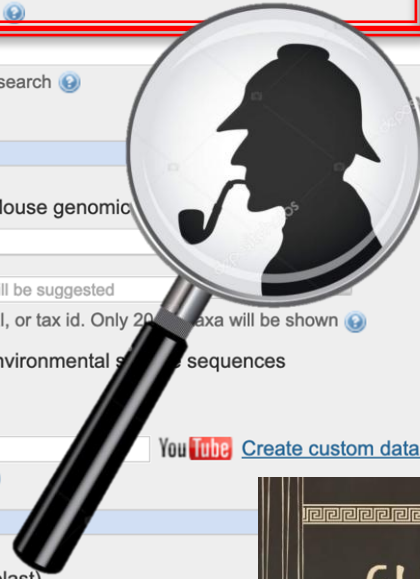
Choose a BLAST algorithm

BLAST

Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)

 Show results in a new window[+ Algorithm parameters](#)

Sequência do slide anterior



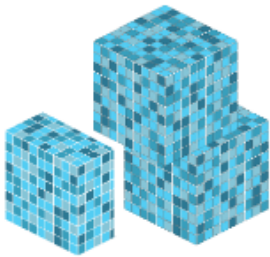
TCGA – The Cancer Genome Atlas

NATIONAL CANCER INSTITUTE THE CANCER GENOME ATLAS

TCGA BY THE NUMBERS

TCGA produced over

2.5
PETABYTES
of data




To put this into perspective, **1 petabyte** of data is equal to

212,000
DVDs



TCGA data describes

 **33**
DIFFERENT
TUMOR TYPES

...including

10
RARE
CANCERS

...based on paired tumor and normal tissue sets collected from

 **11,000**
PATIENTS

...using

7
DIFFERENT
DATA TYPES



CANCER TISSUES COLLECTED FOR STUDY

Last Updated: November 02, 2017

[Expand All](#) | [Collapse All](#)

▶ **Breast**

▶ **Central Nervous System**

▶ **Endocrine**

▶ **Gastrointestinal**

▶ **Gynecologic**

▶ **Head and Neck**

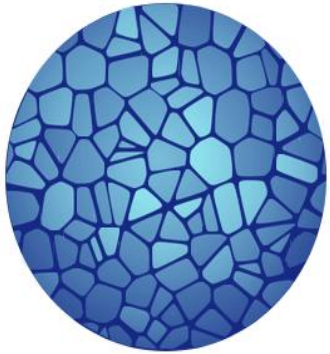
▶ **Hematologic**

▶ **Skin**

▶ **Soft Tissue**

▶ **Thoracic**

▶ **Urologic**



HUMAN CELL ATLAS

MISSION

To create comprehensive reference maps of all human cells—the fundamental units of life—as a basis for both understanding human health and diagnosing, monitoring, and treating disease.

” The Human Cell Atlas will impact almost every aspect of biology and medicine, ultimately leading to a richer understanding of life’s most fundamental units and principles.

<https://www.humancellatlas.org/>




Projecto do Microbioma Humano

Nº células humanas: $\sim 10^{13}$


Nº células microbianas: $\sim 10^{14}$

NIH Human Microbiome Project



Characterization of the microbiomes of healthy human subjects at five major body sites, using 16S and metagenomic shotgun sequencing.

Enter HMP1



Characterization of microbiome and human host from three cohorts of microbiome-associated conditions, using multiple 'omics technologies.

Enter iHMP



Desafios teóricos e computacionais

- Prospecção de dados biomédicos
- Sequências biológicas
- Grafos (estrutura de proteínas, redes biológicas)
- Integração de conhecimento
- Multi-ômica
- Complexidade dos fenómenos biológicos
- Medicina de precisão e personalizada
 - diagnóstico, prognóstico e terapêutica estratificada

Genome Jumper



<https://genome-jumper.sib.swiss/>

Thank You!



rmch@tecnico.ulisboa.pt

