

IBT 432 Aplikasi Bioinformatika

Kontrak belajar dan pengenalan bioinformatika aplikatif

Riza Arief Putranto

Pertemuan I

1. Pendahuluan
2. Pengenalan era “omics”
3. Pengenalan bioinformatika
4. Refreshing “NCBI”

Pertemuan I

1. Pendahuluan
2. Pengenalan era “omics”
3. Pengenalan bioinformatika
4. Refreshing “NCBI”

Rencana Perkuliahan

1. Kontrak belajar dan pengenalan bioinformatika aplikatif
2. Database sekuen dan analisis genomika
3. Anotasi sekuen ke genom - Praktik
4. Analisis komparasi genomika I
5. Analisis komparasi genomika II
6. Analisis komparasi genomika III
7. Analisis komparasi genomika – Praktik
8. Protein modelling I
9. Protein modelling II
10. Protein modelling III
11. Protein modelling - Praktik
12. Visualisasi protein modelling
13. Visualisasi protein modelling - Praktik
14. Presentasi mahasiswa

Sistematika Pembelajaran

- ✓ **Tatap muka**
- ✓ **Tanya jawab/diskusi**
- ✓ **Hands-on/Simulasi dan Praktikum**
- ✓ **Tugas kelompok: Makalah dan Presentasi**
- ✓ **Evaluasi: UTS, UAS, Kuis**

Apakah ada diantara mahasiswa yang **tidak** menggunakan laptop?

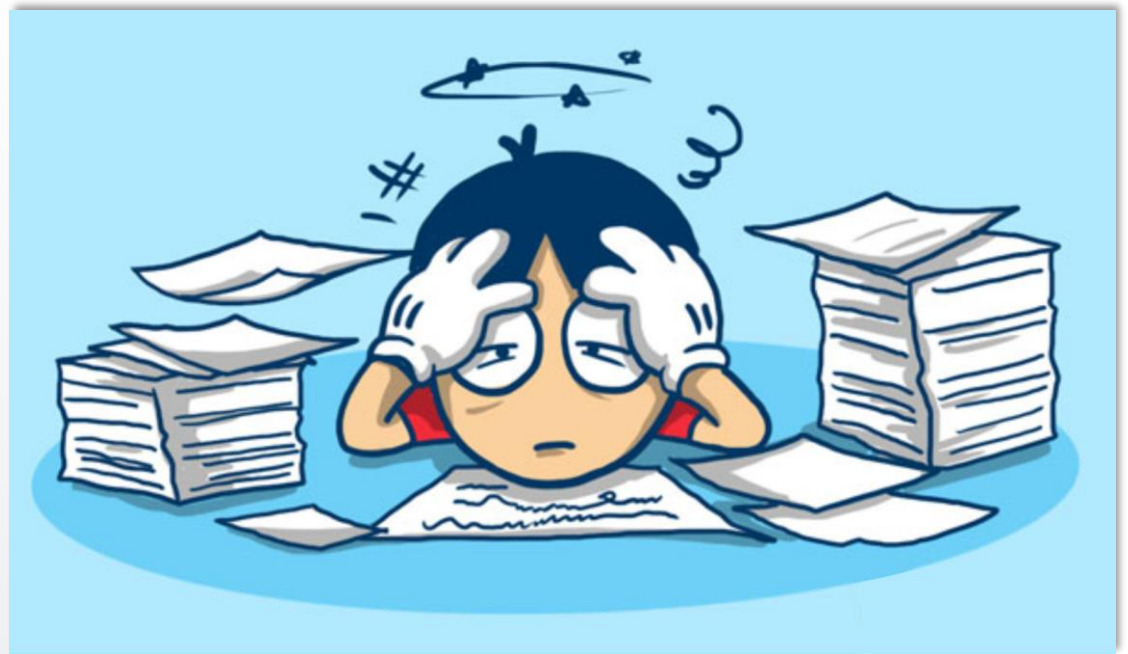
Pemanfaatan teknologi IT – **Materi Kuliah di dalam Google Drive** goo.gl/BbeR9o

Tujuan Perkuliahan

- ❑ Mahasiswa **mengetahui konsep** aplikasi bioinformatika
- ❑ Mahasiswa **memahami beberapa contoh** aplikasi bioinformatika
- ❑ Mahasiswa **memahami dan melakukan kombinasi teknik** aplikasi bioinformatika
- ❑ Mahasiswa **melakukan analisis data** penelitian berbasis bioinformatika

Komponen Penilaian

- ☐ Kehadiran = 10%
- ☐ Tugas = 10%
- ☐ **Praktikum = 20%**
- ☐ **UTS = 25%**
- ☐ **UAS = 35%**



Referensi Perkuliahan

❑ Materi kuliah

- ❑ Claverie, J.-M. & Notredame, C. (2011) **Bioinformatics for dummies**. John Wiley & Sons.
- ❑ Zvelebil, M. J. & Baum, J. O. (2008) **Understanding bioinformatics**, 1. London, UK: Garland Science.
- ❑ DeLano, W. L. & Bromberg, S. (2004) **PyMOL User Guide**. California, USA: DeLano Scientific LLC.

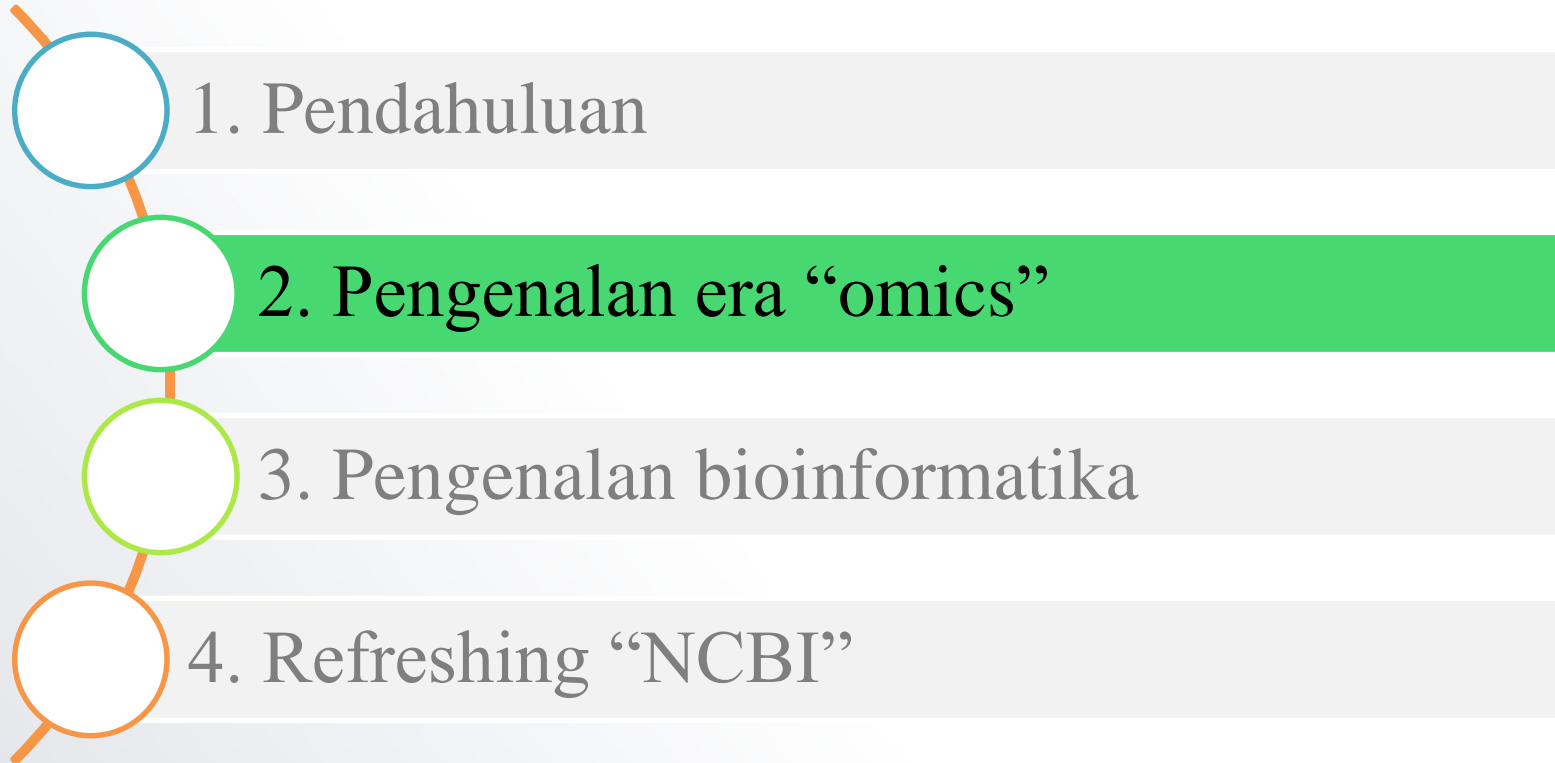
Kontrak Belajar

- ❑ Dosen dan mahasiswa **wajib datang tepat waktu**
- ❑ Diberikan **toleransi kedatangan 15 menit**, setelah itu mahasiswa boleh masuk tapi tidak diperkenankan untuk absen
- ❑ **Kehadiran mahasiswa minimal 75%** (Jadi tidak hadir maksimal 3x perkuliahan) untuk mengikuti UAS
- ❑ Wajib mengenakan **pakaian sopan** (tidak menggunakan kaos oblong dan/atau sandal)
- ❑ Apabila kuliah tidak bisa dilakukan sesuai jadwal akan dikenakan **kelas pengganti**
- ❑ **Tidak diperkenankan mencontek** setiap UTS dan UAS
- ❑ **Apabila diketahui mencontek, nilai UTS atau UAS menjadi E**

Partisipasi Mahasiswa

- Mengikuti perkuliahan Instrumentasi Bioteknologi dengan baik
- Rajin/selalu hadir dalam perkuliahan/selalu berangkat kuliah
- Belajar dengan baik/belajar dengan sungguh-sungguh
- Mengembangkan materi yang telah diberikan
- Disiplin/tidak terlambat/datang tepat waktu
- Aktif bertanya
- Mengerjakan tugas dengan baik
- Mematuhi aturan perkuliahan yang sudah disepakati bersama

Pertemuan I

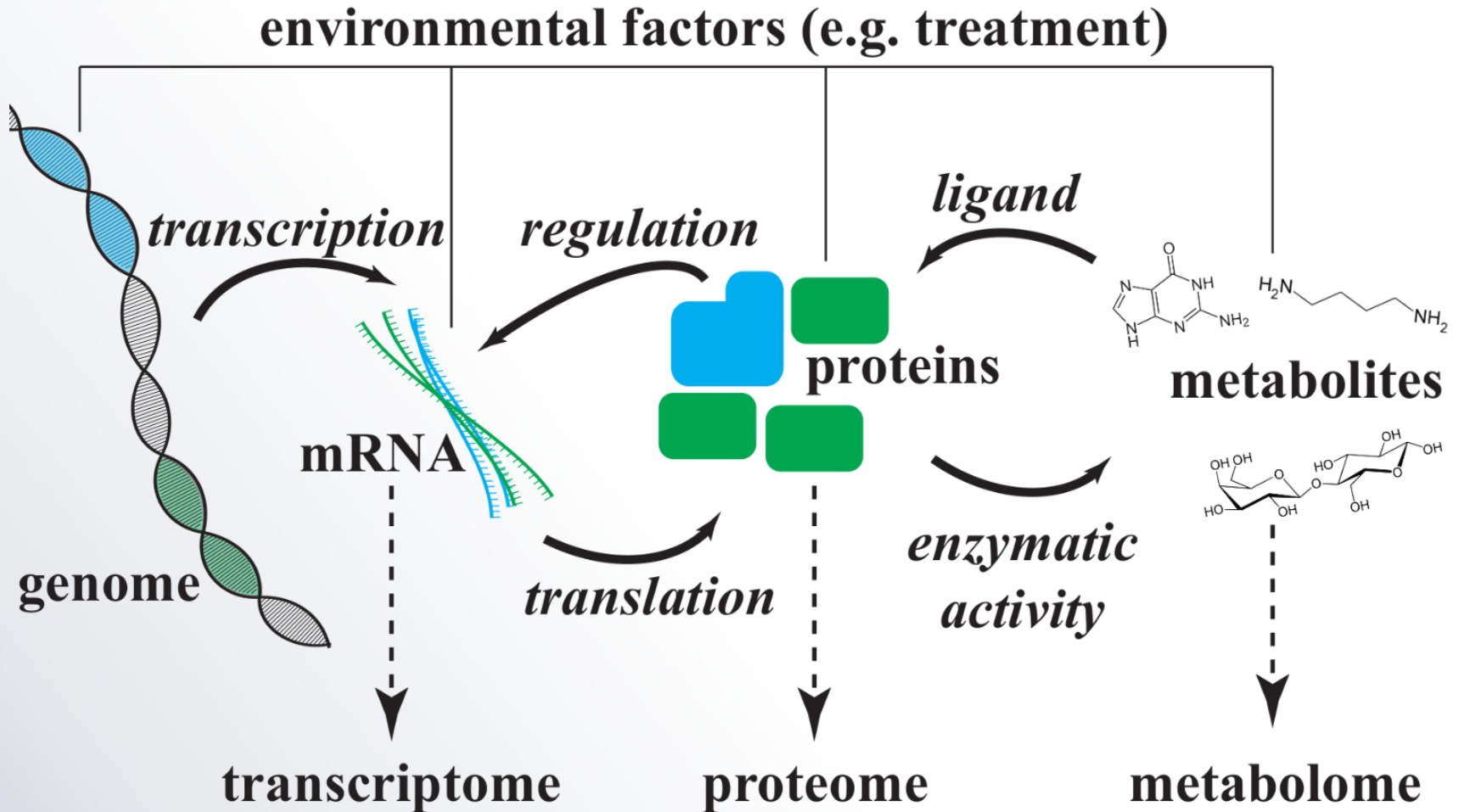
- 
1. Pendahuluan
 2. Pengenalan era “omics”
 3. Pengenalan bioinformatika
 4. Refreshing “NCBI”

The era of Omics



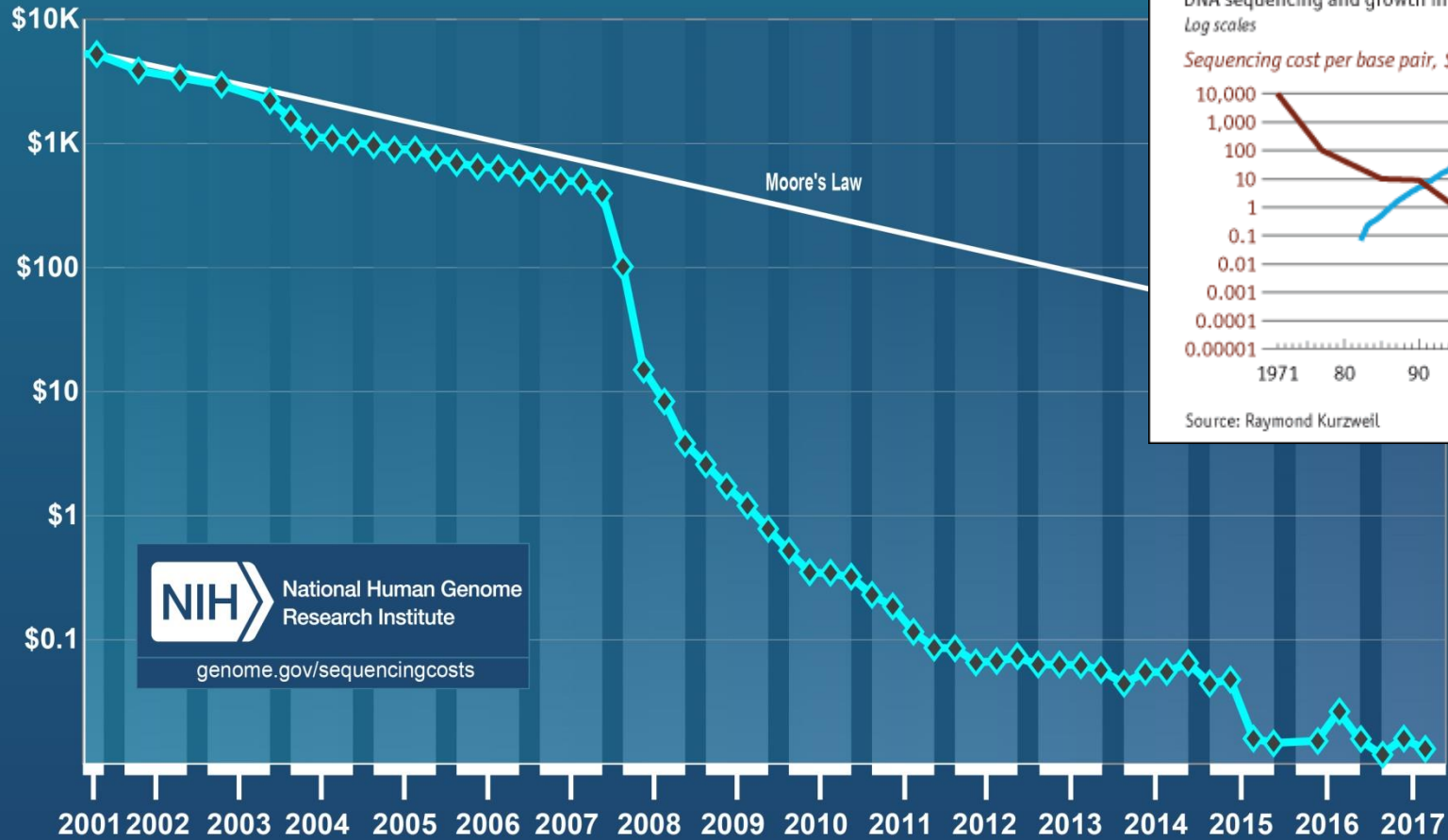
What is Omics?

The era of Omics



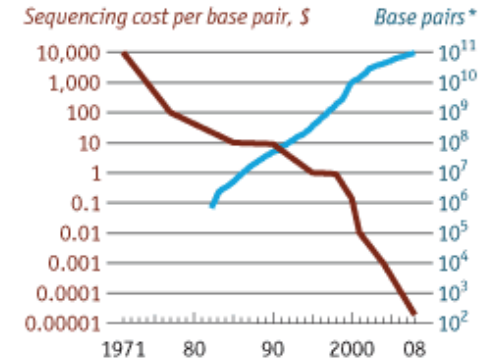
The era of Omics needs Next Gen Seq

Cost per Raw Megabase of DNA Sequence



The road to commoditisation

DNA sequencing and growth in DNA data
Log scales



Source: Raymond Kurzweil

*Growth in Genbank DNA sequence data

DNA and RNA sequencing

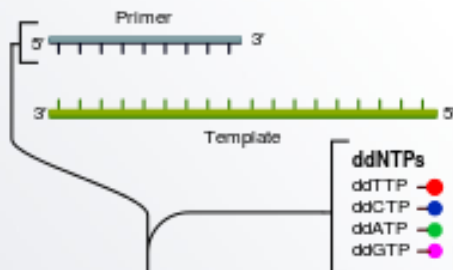
- **DNA sequencing** is the process of **determining** the **precise order** of **nucleotides** (bases—adenine, guanine, cytosine, and thymine) within a DNA molecule.
- **RNA-seq** (RNA sequencing), also called **Whole Transcriptome Shotgun Sequencing (WTSS)**, is a technology that uses the capabilities of **Next-Generation Sequencing** to reveal a snapshot of RNA presence and quantity from a genome at a given moment in time ([Chu & Corey 2012 Nucl Ac Ther 22: 271-274](#)).
- The **transcriptome** is the **complete set of transcripts** in a cell, and their quantity, for a specific developmental stage or physiological condition ([Wang et al. 2009 Nat Rev Gen 10: 57-63](#)).

First Generation Sequencing

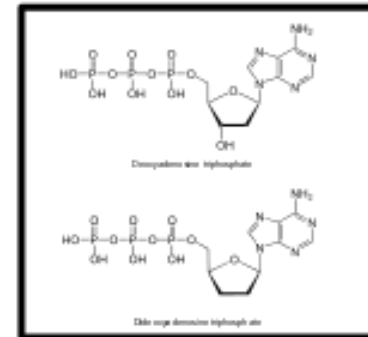
❑ Sanger: 1976

① Reaction mixture

- ▶ Primer and DNA template
- ▶ DNA polymerase
- ▶ ddNTPs with flouorchromes
- ▶ dNTPs (dATP, dCTP, dGTP, and dTTP)



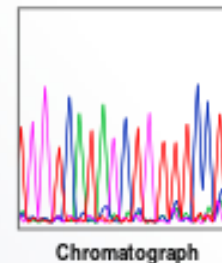
② Primer elongation and chain termination



③ Capillary gel electrophoresis separation of DNA fragments



④ Laser detection of flouorchromes and computational sequence analysis



Next Generation Sequencing

- ❑ Roche/454 FLX: **2004**
- ❑ Illumina Solexa Genome Analyzer: **2006**
- ❑ Applied Biosystems SOLiD™ System: **2007**
- ❑ Helicos Heliscope™ : **2008**
- ❑ Pacific Biosciences SMRT: **2010**

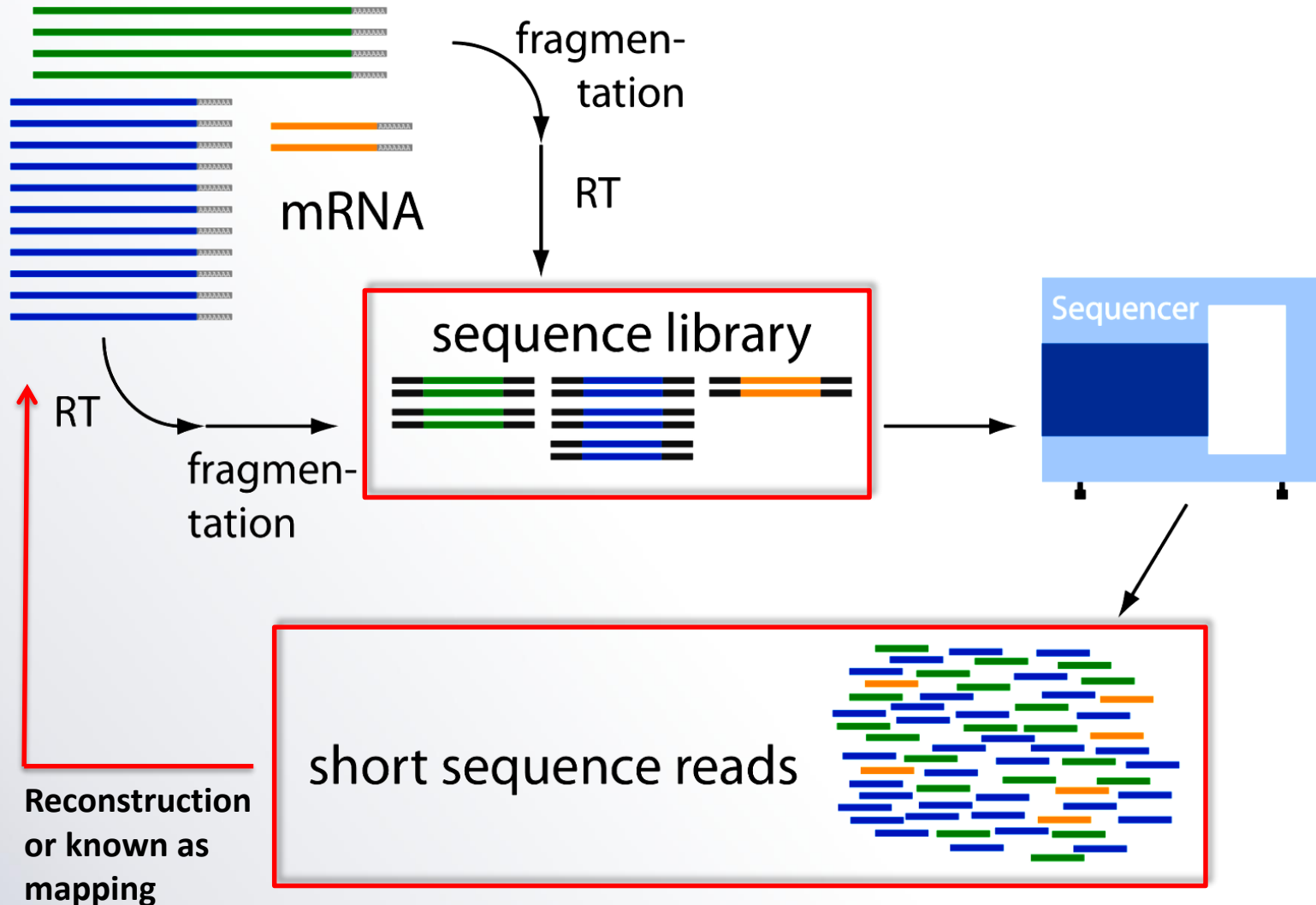


Next Generation Sequencing

High-end sequencing- Platform†	Sequencing chemistry	Read lengths/through put	Run time	Template prep	Application
Roche 454 -Titanium FLX	Pyrosequencing	400 bp 400 Mb/run	10 hours	Emulsion PCR	Denovo WGS of microbes, pathogen discovery, Exome seq
Illumina/Solexa -HiSeq 2000	Reversible terminator chemistry	2×100bp 600 GB/run (dual cell)	11.5 days	Solid-phase	Human WGS, exome seq, RNA-seq, Methylation
ABI/LifeTechnology-SOLiD 5550XL	Sequencing by ligation	2×60bp 15 GB/day	8 days	Emulsion PCR	Human WGS, exome seq, RNA-seq, Methylation
HelicosBiotechnologies	Reversible Terminator chemistry	25-55 bp 28 GB/run (avg)	>1 GB/hour	Single molecule	Human WGS, exome seq, RNA-seq, Methylation
Roche 454- GS Junior	Pyrosequencing	400 bp 50 Mb/run	10 hours	Emulsion PCR	Denovo WGS of microbes, pathogen discovery, Exome seq
Illumina/Solexa- MiSeq	Reversible terminator chemistry	2×150bp 1.0-1.4 Gb	26 hours	Solid-phase	Microbial discovery, Exome seq, Targeted capture
ABI/ Lifetechnology- Iontorrent	H+ Ion sensitive transistor	320 Mb/run	8 hours*	Emulsion PCR	Microbial discovery, Exome seq, Targeted capture

*Sample preparation – 6 hours, sequencing time – 2 hours, †Data shown here represent the highest figures currently available on the company website and is highly likely to change by the time this article is published

Principal workflow of NGS



The trend of sequencing will not yet stop

Biologists propose to sequence the DNA of all life on Earth

By Elizabeth Pennisi | Feb. 24, 2017, 1:15 PM

WASHINGTON, D.C.—When it comes to genome sequencing, visionaries like to throw around big numbers: There's the UK Biobank, for example, **which promises to decipher the genomes of 500,000 individuals**, or Iceland's effort to **study the genomes of its entire human population**. Yesterday, at a meeting here organized by the Smithsonian Initiative on Biodiversity Genomics and the Shenzhen, China-based sequencing powerhouse BGI, a small group of researchers upped the ante even more, announcing their intent to, eventually, sequence "all life on Earth."

Deep sequencing of 10,000 human genomes



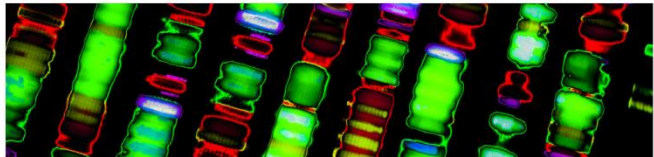
Amalio Telenti, Levi C. T. Pierce, William H. Biggs, Julia di Iulio, Emily H. M. Wong, Martin M. Fabani, Ewen F. Kirkness, Ahmed Moustafa, Naisha Shah, Chao Xie, Suzanne C. Brewerton, Nadeem Bulsara, Chad Garner, Gary Metzker, Efrén Sandoval, Brad A. Perkins, Franz J. Och, Yaron Turpaz and J. Craig Venter

PNAS October 18, 2016. 113 (42) 11901-11906; published ahead of print October 4, 2016.
<https://doi.org/10.1073/pnas.1613365113>

Contributed by J. Craig Venter, August 18, 2016 (sent for review July 1, 2016; reviewed by David B. Goldstein and Stephen W. Scherer)

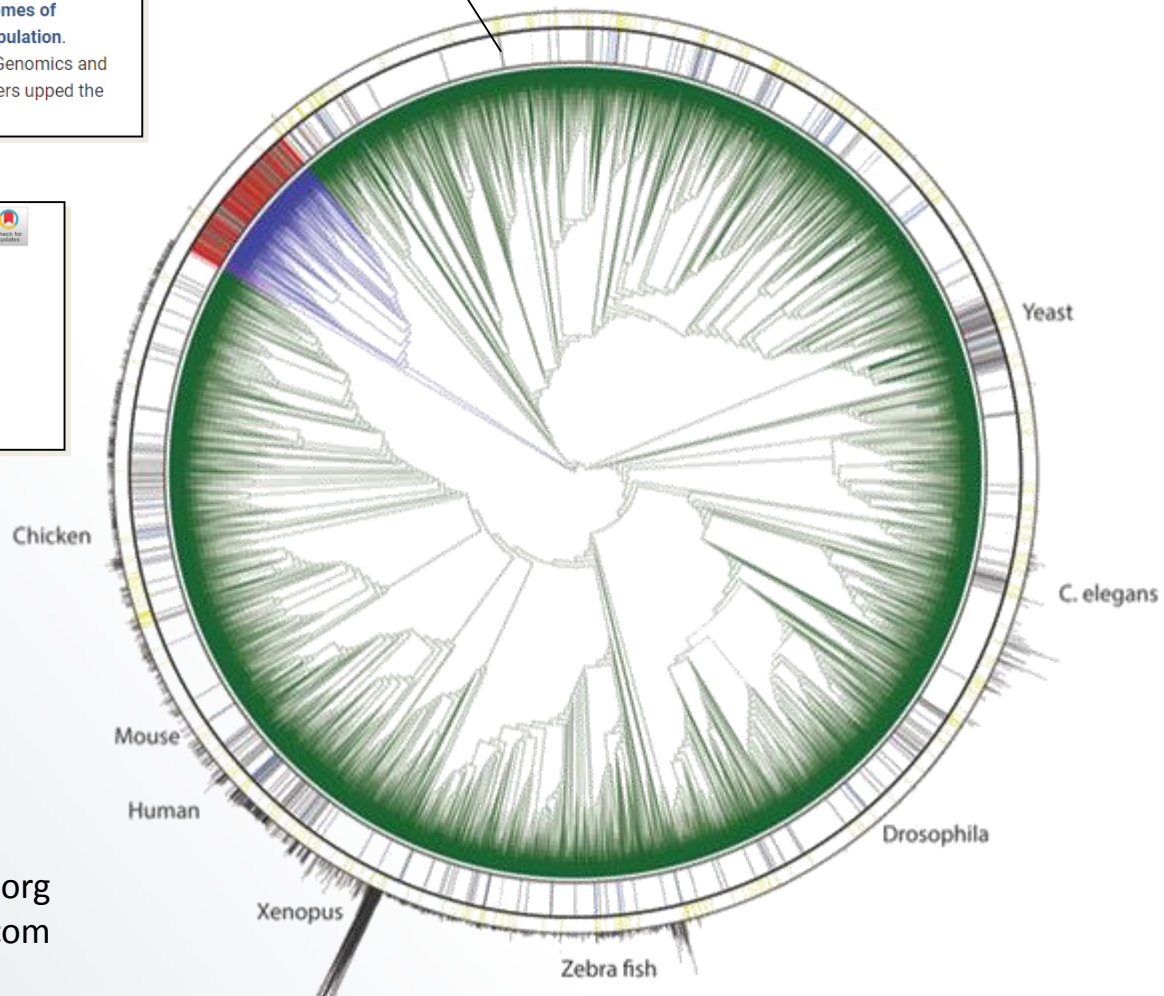
ILLUMINA STAT+ Illumina says it can deliver a \$100 genome — soon

By MEGHANA KESHAVAN @meghesh / JANUARY 9, 2017



Genome sequenced

"The tree of life"



<http://www.sciencemag.org>
<https://www.statnews.com>

The future generation sequencing

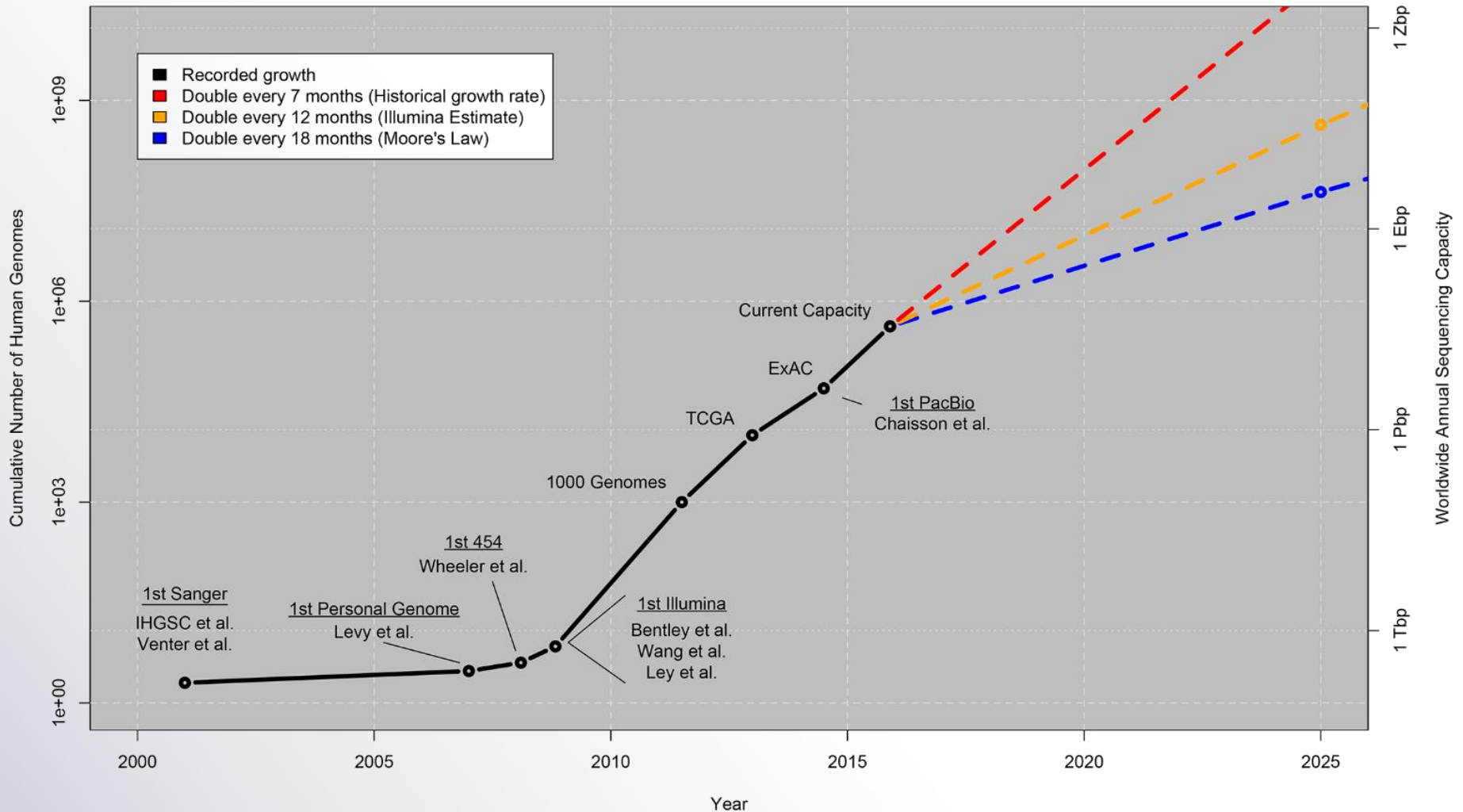


“The MinION has been used to successfully read the genome of a lambda bacteriophage, which has **48,500-ish base pairs, twice during one pass**. That's impressive, because reading 100,000 base pairs during a single DNA capture has never been managed before using traditional sequencing techniques.

The operational life of the [MinION](#) is only about **six hours**, but during that time it can read more than **150 million base pairs**. That's somewhat short of the larger human chromosomes (which contain up to 250 million base pairs), but Oxford Nanopore has also introduced GridION -- a platform where multiple cartridges can be clustered together. The company reckon that a 20-node GridION setup can sequence a complete human genome in just 15 minutes.”

The trend of data "tsunami" (big data)

Growth of DNA Sequencing



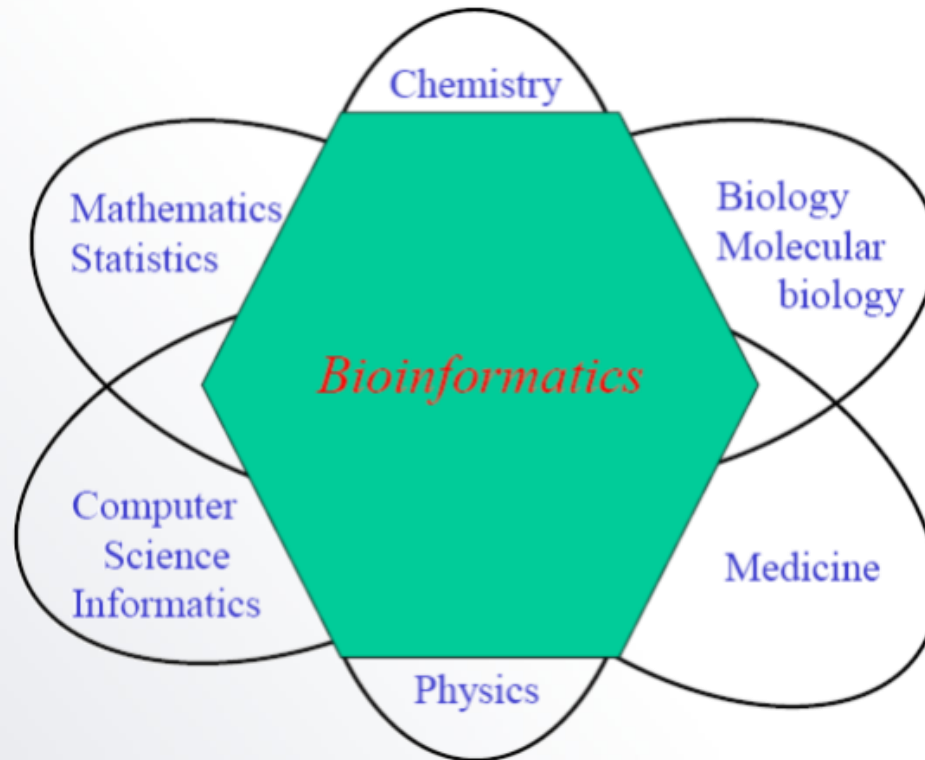
How to handle the big data?



Drew Sheneman New Jersey – The Newark star Ledger 2002

The fact

All analysis of NGS, genomics, transcriptomics, proteomics, metabolomics, **NEED** bioinformatics



Pertemuan I

1. Pendahuluan
2. Pengenalan era “omics”
- 3. Pengenalan bioinformatika**
4. Refreshing “NCBI”

History of bioinformatics

Important factors facilitating the emergence of **Computational Biology** in early 1960

Idea that **macromolecules** carry information

Expanding **collection of amino acid sequences**

High-speed digital computers developed from weapon research during WW-II

History of bioinformatics

- The birth of bioinformatics: 1960s
- Margaret Oakley Dayhoff created
 - ✓ The first protein database
 - ✓ The first program for sequence assembly
- The need of computers and algorithms for accessing, processing, storing, sharing, retrieving, visualizing, annotating...



Definitions of bioinformatics

- The **body of tools, algorithms** needed to handle **large and complex biological information**.
- A new scientific discipline created from **the interaction of biology and computer**.
- The NCBI defines bioinformatics as: "Bioinformatics is **the field of science** in which **biology, computer science**, and **information technology** merge into **a single discipline**".
- The use of **computer technology** for **storage, retrieval, manipulation**, and **distribution** of information related to **biological macromolecules** such as DNA, RNA, and proteins (Luscombe *et al* 2001).

Bioinformatics vs Computational biology

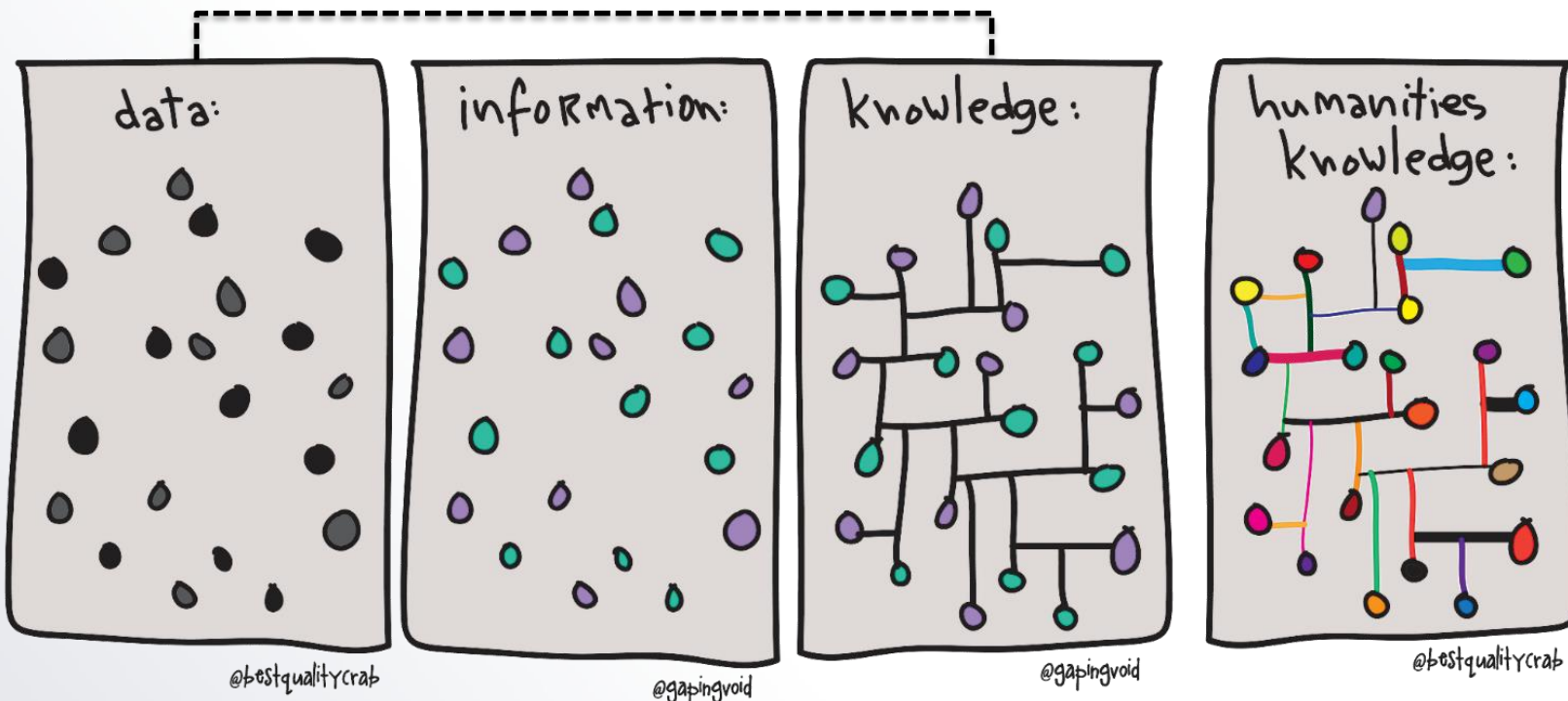
Bioinformatics is **limited** to sequence, structural, and functional analysis of genes and genomes and their corresponding products

Often considered as **computational molecular biology**

Computational biology encompasses **all biological areas that involve computation** (example: mathematical modeling of ecosystems, population dynamics, application of the game theory in behavioral studies)

The Data and Knowledge wisdom

Bioinformatics



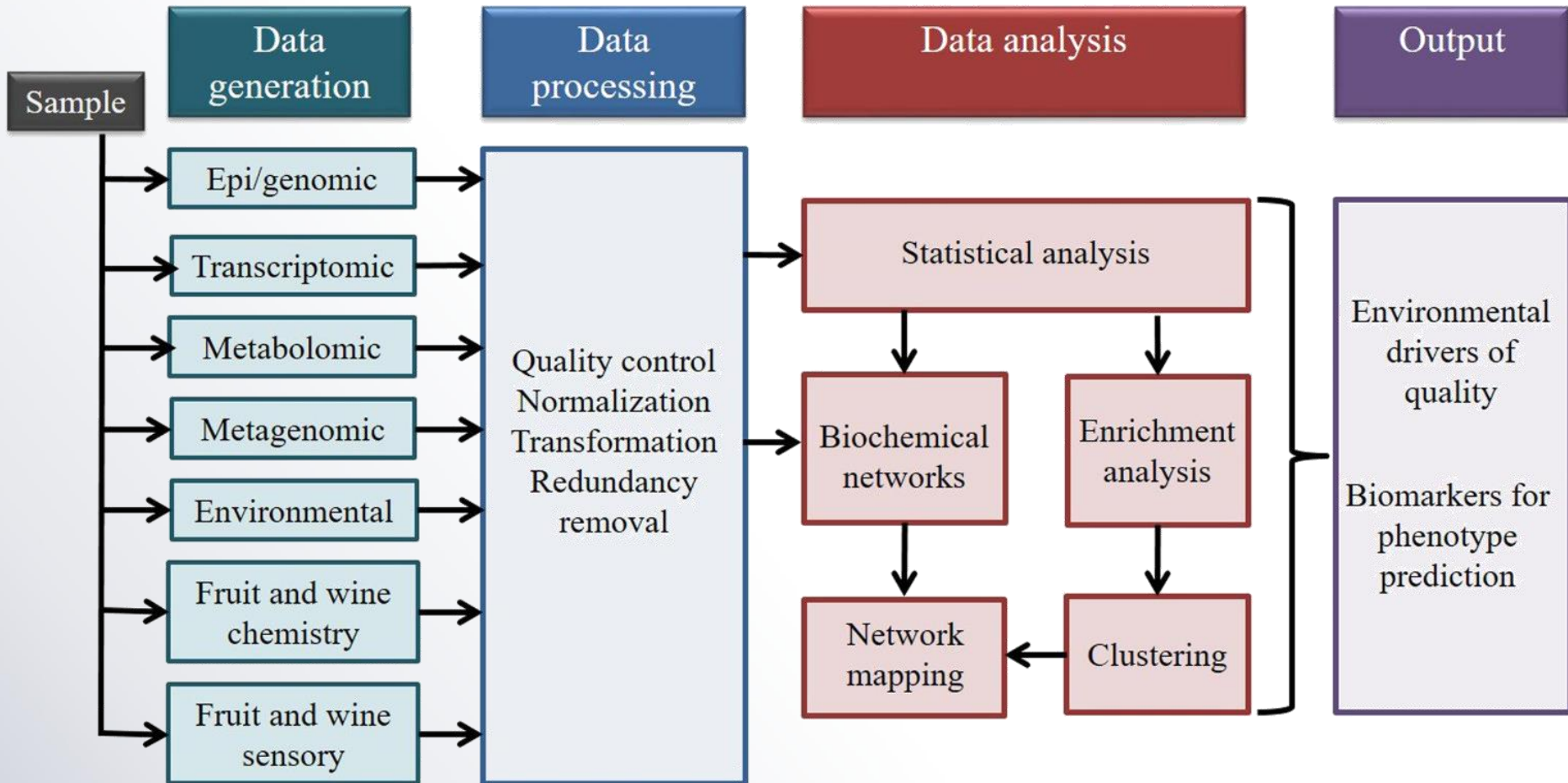
Genome
Transcriptome
Proteome
Metabolome

Interaction

Meta-concept A
Meta-concept B

Applications
Medicine
Agriculture

Example of Data to Knowledge of *Vitis*

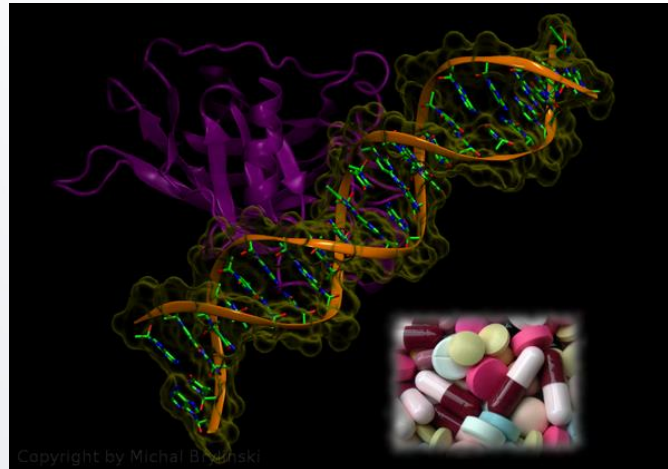


Multi-Omics Data Integration for Terroir Analysis in *Vitis vinifera*

Fabres et al 2017 *Frontiers in Plant Science* 8

Goal of bioinformatics

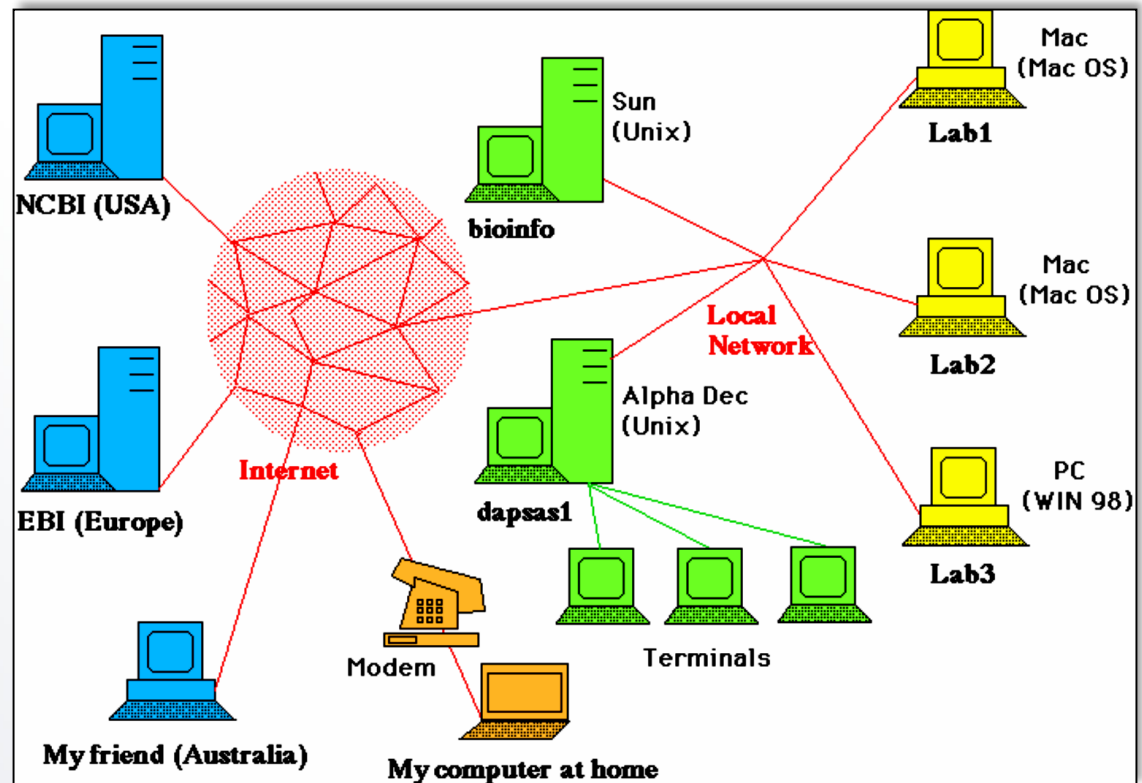
The goal of bioinformatics is to create **a better data management** in order to **uncover the wealth of biological information** hidden in the **mass of data** and to obtain a **clearer insight** into the fundamental **biology of organisms**



Copyright by Michal B. Jankowski

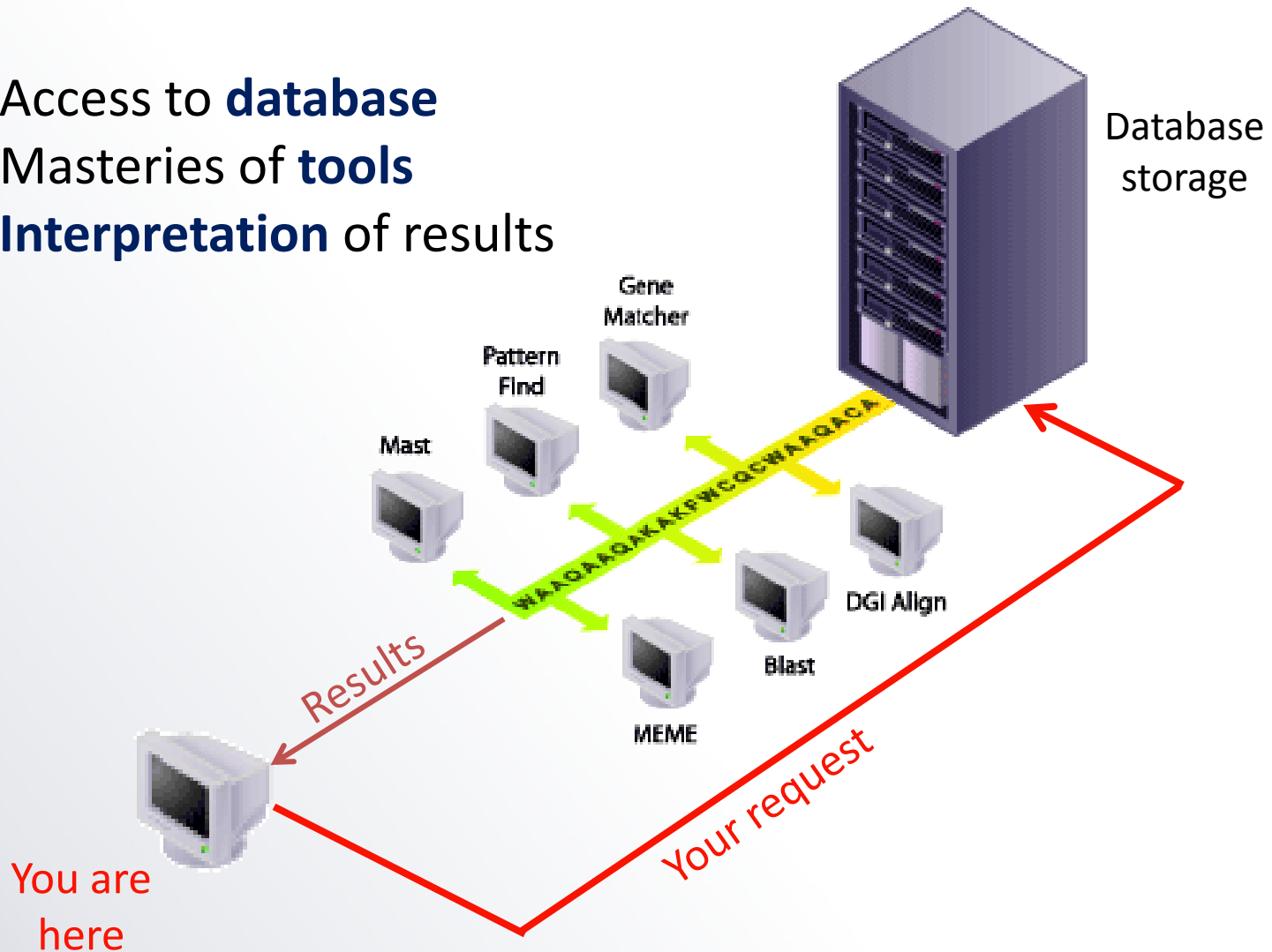
Who keep the data

- In bioinformatics, **everyone participates** in maintaining the data, thanks to the internet.
- Huge amount of data stored on computers all over the world.
- **Impossible to maintain up-to-date copies** of all relevant databases within the lab.



Principals of bioinformatics research

1. Access to **database**
2. Masteries of **tools**
3. **Interpretation** of results



Two subfields of bioinformatics

Development of computational **tools** and **databases**



Writing software for sequence, structural, and functional analysis.
Construction and curating of biological databases

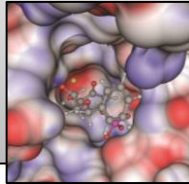
Application of tools and databases in **generating biological knowledge**



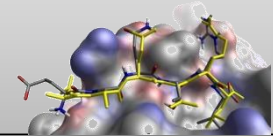
Sequence analysis
Protein structural analysis
Gene and protein functional analysis

Applications of bioinformatics

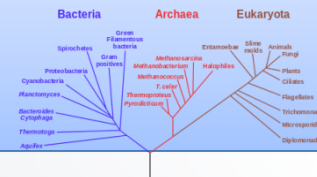
Protein – Ligand Interactions
Modelling



Molecular docking for drug design



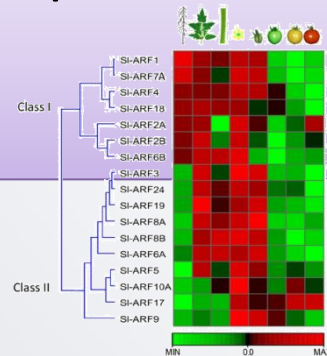
Molecular and phylogenetic analyses



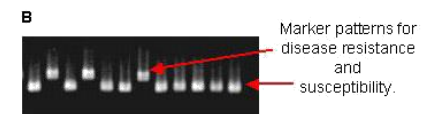
Molecular forensics,
biodiversity & conservation



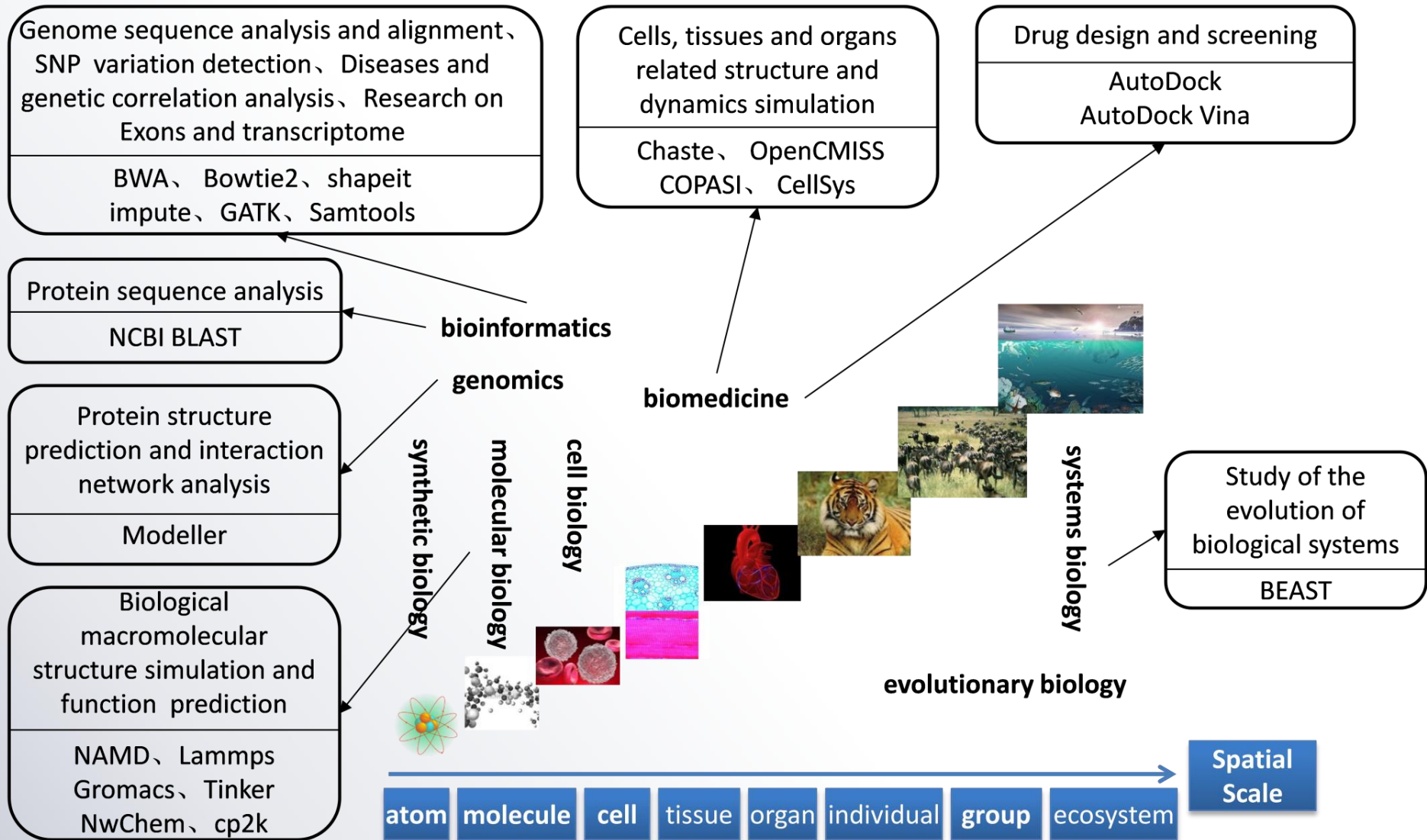
Gene expression profile
analyses



Comparative NGS analysis



These are the applied bioinformatics

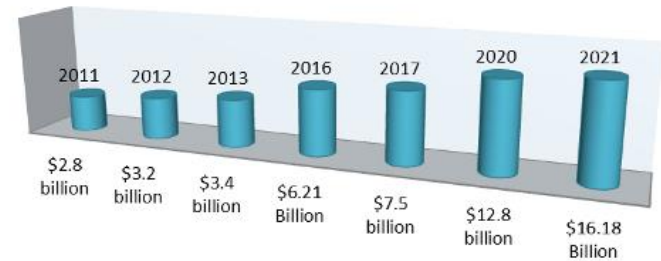


Excellent market growth of bioinformatics

GLOBAL BIOINFORMATICS MARKET 2017-2025



Bioinformatics Research Market Growth



~50 Bioinformatics companies:

Genomatrix Software, Genaissance Pharmaceuticals, Lynx, Lexicon Genetics, DeCode Genetics, CuraGen, AlphaGene, Bionavigation, Pangene, InforMax, TimeLogic, GeneCodes, LabOnWeb.com, Darwin, Celera, Incyte, BioResearch Online, BioTools, Oxford Molecular, Genomica, NetGenics, Rosetta, Lion BioScience, DoubleTwist, eBioinformatics, Prospect Genomics, Neomorphic, Molecular Mining, GeneLogic, GeneFormatics, Molecular Simulations, Bioinformatics Solutions....

MARKET BY



APPLICATION

MARKET BY



PRODUCTS

● 2016

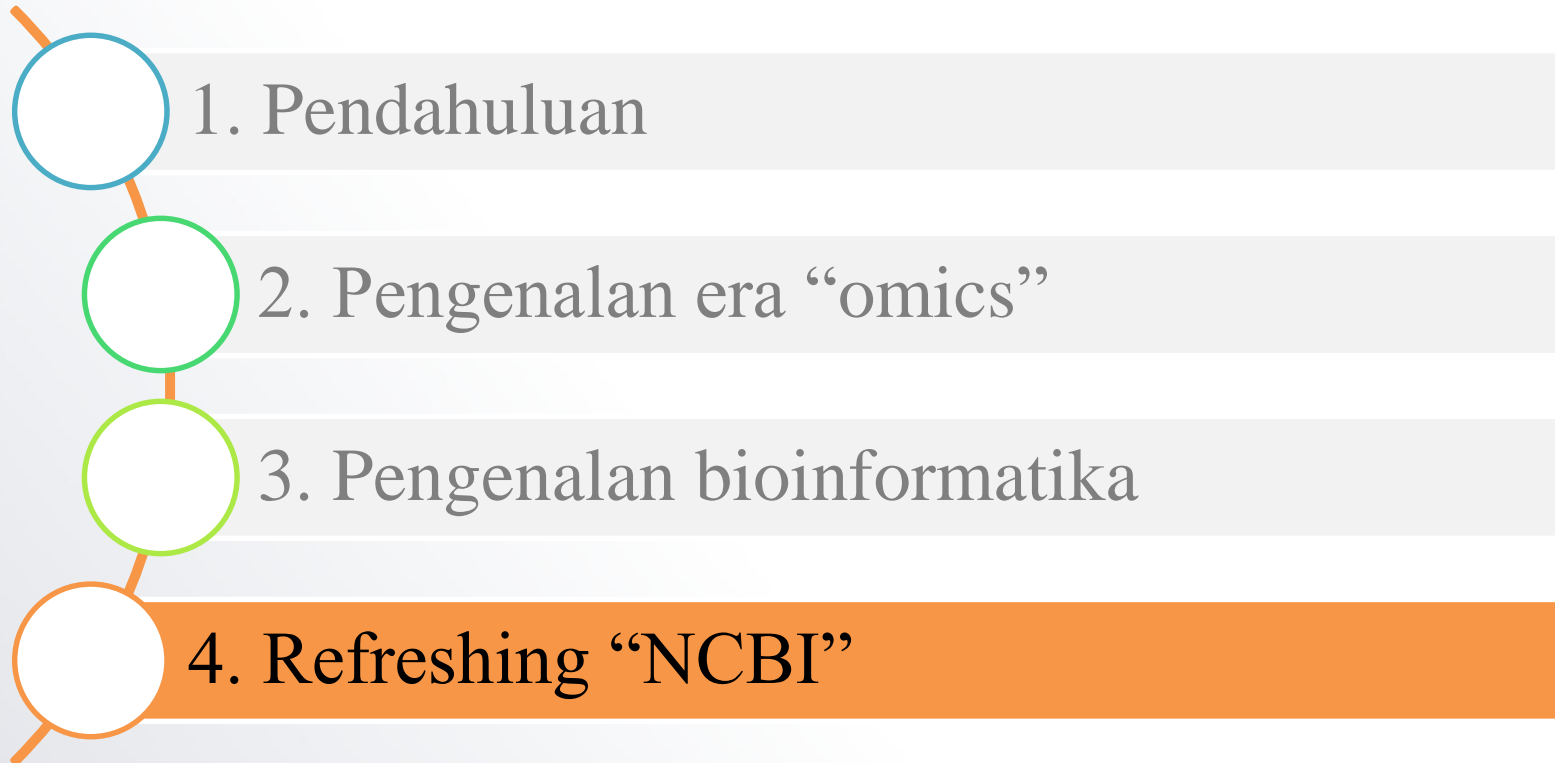
● 2025

www.inkwoodresearch.com

● 2016

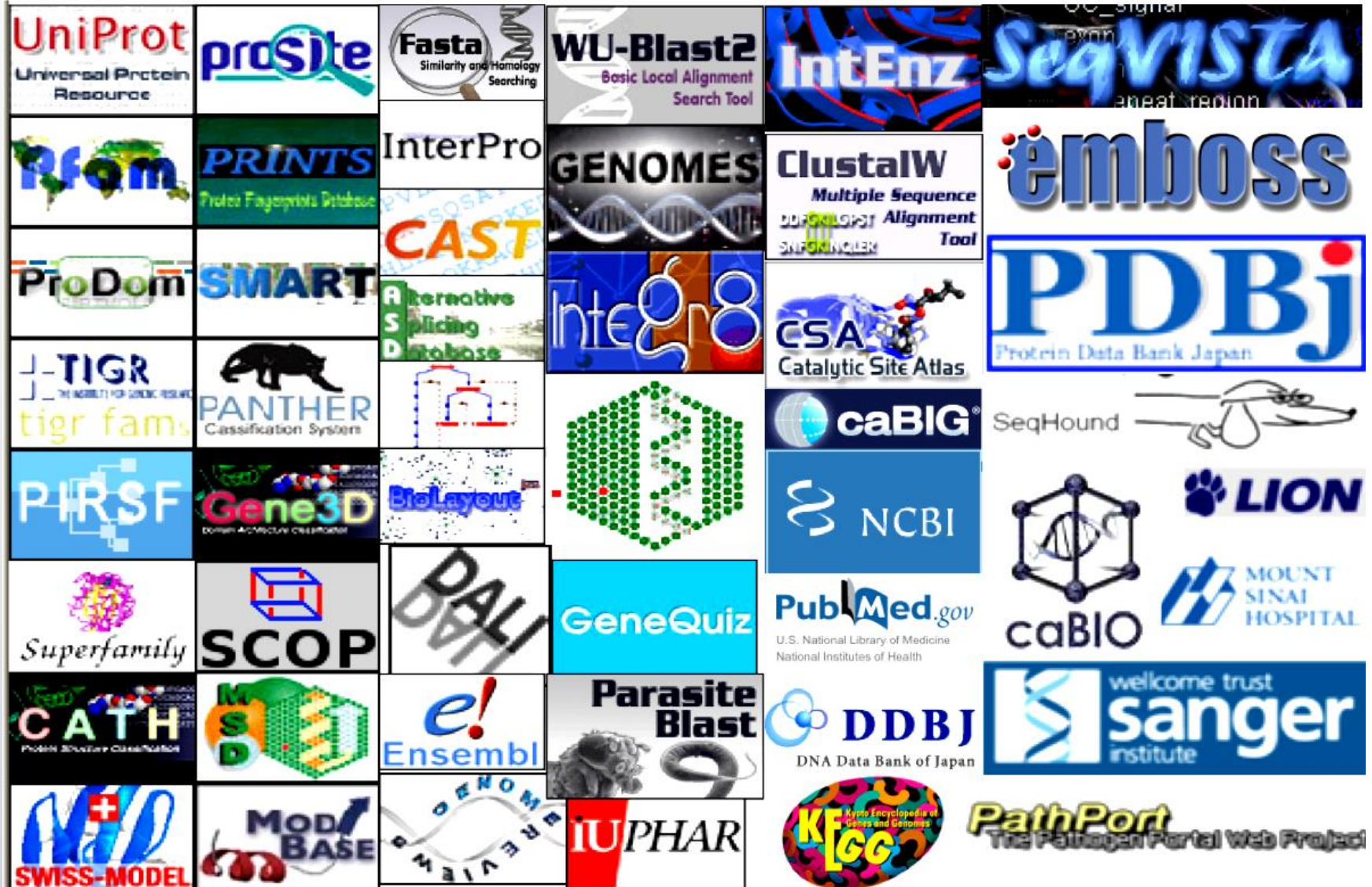
● 2025

Pertemuan I

- 
1. Pendahuluan
 2. Pengenalan era “omics”
 3. Pengenalan bioinformatika
 4. Refreshing “NCBI”

How many Bioinformatics Resources?

The 2016 issue has a list of about 180 databases



Refreshing of previous semester

>sequence

```
GGCCCCAGAGCAAGGAGCTGTTTGTGGGCTTACCACTGCTGTTCCCATATGCCCCAACTGCCTCCCACT
TCTTTCCCCACAGCCTGGTCAGACATGGCACTACCACTAATGGAATCTTTCTTGCCATCTTTTTCTTGCC
GCTTAACAGTGGCAGTGACACTTTGACTCCTGATTTAAGCCTGATTCTGCTTAACTTTTTCCCTTGACTT
TGGCATTTCCTTTGACATGTTCCCTGAGAGCCTGGGGGGTGGGGAACAGCTCCAGCTGGTGACGTTTG
GGGCCGGCCCAGGCCTAGGGTGTGGAGGAGCCTTGCCATCGGGCTTCCTGTCTCTCTTCATTTAAGCACG
ACTCTGCAGAAGGAACAAAGCACCCCTCCCACTGGCTCCTGGTTGCAGAGCTCCAAGTCCTCACACAGAT
ACGCCTGTTTGAGAAGCAGCGGGCAAGAAAGACGCAAGCCCAGAGGTAAGGTGGTCAGACTCGGCTTCCT
TCCCCGGAGCTGAGAGGGAGGGGAACGTGGGGCAGATGCACAGGAATGTGCTCTGCCAGTTGTCTGCC
ACAGCTCTGGCCACCTTCTCTTGCATTTCTCTTGGAACTGGTCATGAGCAGCGATTTCCCACTGGAAGT
TGAGCTTCCAGAGGTCAGAGACTGTGCTAGACTCCTCTCTGCAGCCCCAGCGTGCACAGCTCAGTGTCCA
```

Link: goo.gl/BbeR9o

Please do a search on NCBI

What tool do you use?

Refreshing of previous semester

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information Sign in to NCBI

BLAST » blastn suite Home Recent Results Saved Strategies Help

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

Enter Query Sequence BLASTN programs search nucleotide databases using a nucleotide query. [more...](#) [Reset page](#) [Bookmark](#)

Enter accession number(s), gi(s), or FASTA sequence(s) Query subrange

```
ACGCCTGTTTGAGAAGCAGCGGGCAAGAAAGACGCAAGCCAGAGGTAAGGTGGTCAGACTCGGCTTCCT
TCCCCGGAGCTGAGAGGGAGGGGAACGTGGGGCAGATGCACAGGAATGTGCTTGCCCAAGTTGTCTGCC
ACAGCTGGCCACCTTCTCTTGCAATTTCTTTGGAACGGTCATGAGCAGCGATTTCCCACTGGAAGT
TGAGCTTCCAGAGGTGAGAGACTGTGCTAGACTCCTCTGTCAGCCCCAGCGTGACAGCTCAGTGTCCA
```

From

To

Or, upload file Aucun fichier choisi

Job Title

Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database Human genomic + transcript Mouse genomic + transcript Others (nr etc.):
Nucleotide collection (nr/nt)

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

Exclude Models (XM/XP) Uncultured/environmental sample sequences

Limit to Sequences from type material

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

**Copy paste sequence to a normal
BLAST nucleotide**

Refreshing of previous semester

BLAST Results

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#)

[YouTube](#) [How to read this page](#) [Blast report description](#)

Job title: Nucleotide Sequence (700 letters)

RID [T55WJMAD01R](#) (Expires on 09-08 23:21 pm)

Query ID Id|Query_119253
Description None
Molecule type nucleic acid
Query Length 700

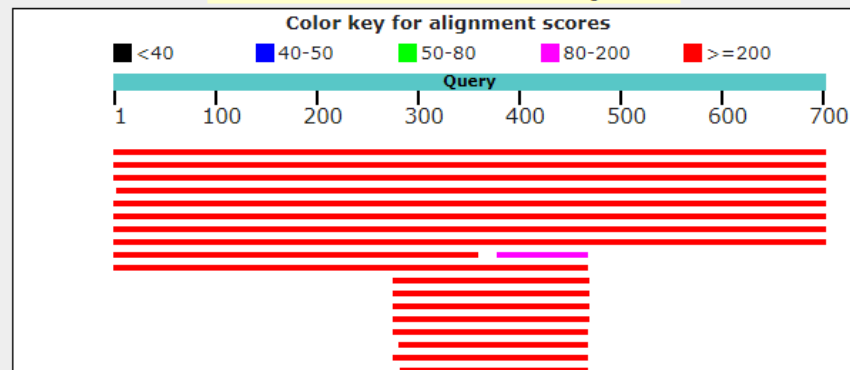
Database Name nr
Description Nucleotide collection (nt)
Program BLASTN 2.8.0+ [Citation](#)

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [MSA viewer](#)

Graphic Summary

Distribution of the top 89 Blast Hits on 89 subject sequences

Mouse over to see the title, click to show alignments



What do you get?

Refreshing of previous semester

Homo sapiens CD4 gene

GenBank: X87579.1

[FASTA](#) [Graphics](#)

[Go to:](#)

LOCUS X87579 13133 bp DNA linear PRI 26-JUL-2016
DEFINITION Homo sapiens CD4 gene.
ACCESSION X87579
VERSION X87579.1
KEYWORDS Alu repeat; CD4 gene.
SOURCE Homo sapiens (human)
ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Zverev,V.V., Blinov,V.M. and Nedospasov,S.A.
TITLE Splice-mediated insertion of antisense and sense Alu repeats in
human CD4 gene: identification of three exons of CD4 mRNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 13133)
AUTHORS Zverev,V.V.
TITLE Direct Submission
JOURNAL Submitted (30-APR-1993) V.V. Zverev, Institute for Viral
Preparation, 109088 Moscow 15, 1st Dubrovskaya, RUSSIA
FEATURES
source Location/Qualifiers
1..13133
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
[gene](#) 455..>11266
/gene="CD4"
[mRNA](#) join(455..465,10869..10983,11102..11266)
/gene="CD4"
[exon](#) 455..465
/gene="CD4"

Customize view

Analyze this sequence

[Run BLAST](#)

[Pick Primers](#)

[Highlight Sequence Features](#)

[Find in this Sequence](#)

Articles about the CD4 gene

[HIV-1 gp41 Residues Modulate CD4-Induced Conformational Changes in the Env \[J Virol. 2018\]](#)

[Redox exchange of the disulfides of human two-domain C \[Biochem Biophys Res Commun. 2018\]](#)

[Host MicroRNAs-221 and -222 Inhibit HIV-1 Entry in Macrophages by Targeti \[Cell Rep. 2017\]](#)

[See all...](#)

Reference sequence information

[RefSeq alternative splicing](#)

See 6 reference mRNA sequence splice variants for the CD4 gene.

More about the CD4 gene

This gene encodes a membrane glycoprotein of

What is a Homo sapiens CD4 gene?

The Human CD4 gene

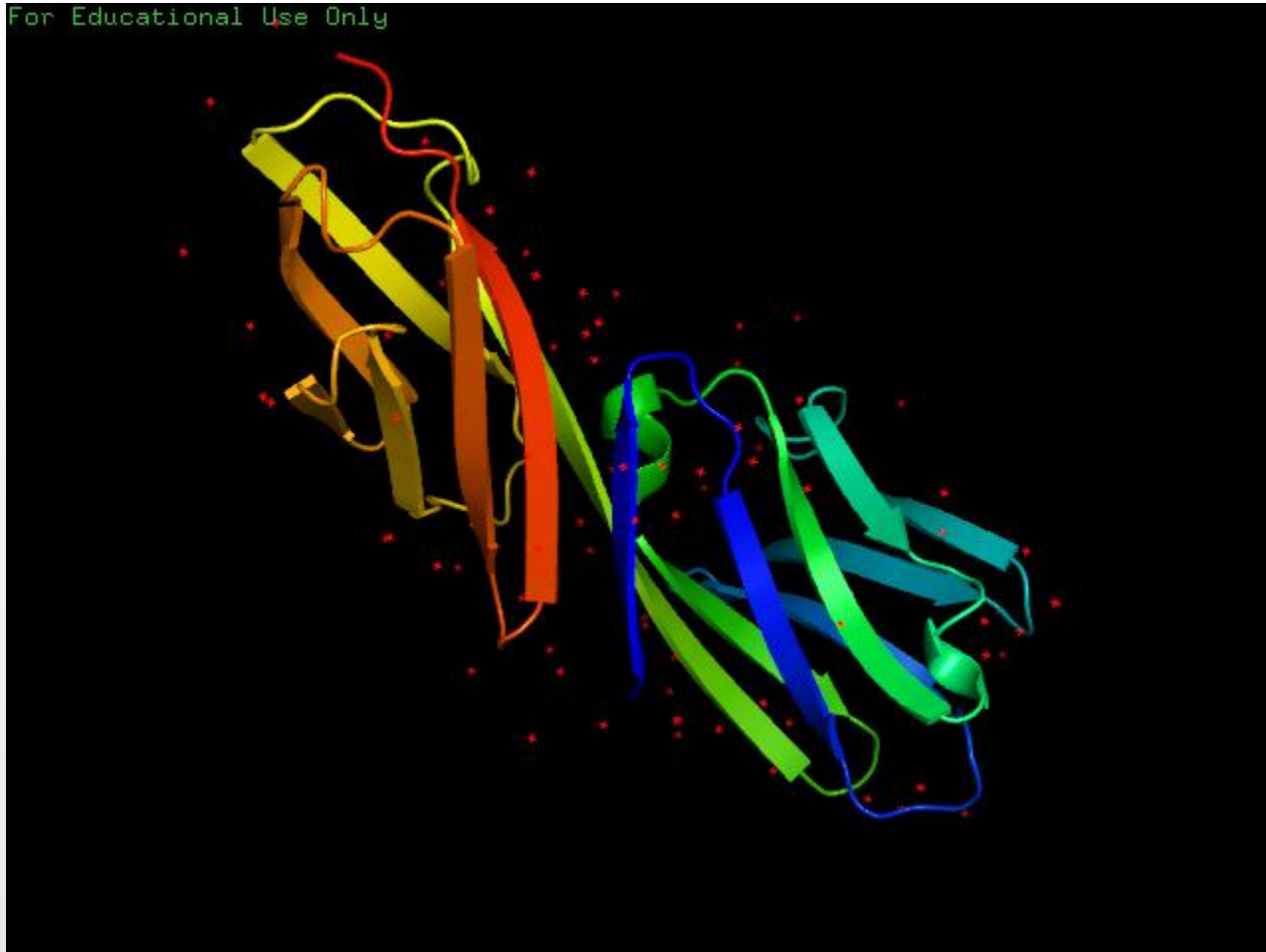
>Function

Integral membrane glycoprotein that plays an essential role in the immune response and serves multiple functions in responses against both external and internal offenses. In T-cells, functions primarily as a coreceptor for MHC class II molecule:peptide complex. The antigens presented by class II peptides are derived from extracellular proteins while class I peptides are derived from cytosolic proteins. Interacts simultaneously with the T-cell receptor (TCR) and the MHC class II presented by antigen presenting cells (APCs). In turn, recruits the Src kinase LCK to the vicinity of the TCR-CD3 complex. LCK then initiates different intracellular signaling pathways by phosphorylating various substrates ultimately leading to lymphokine production, motility, adhesion and activation of T-helper cells. In other cells such as macrophages or NK cells, plays a role in differentiation/activation, cytokine expression and cell migration in a TCR/LCK-independent pathway. Participates in the development of T-helper cells in the thymus and triggers the differentiation of monocytes into functional mature macrophages.

>Function

Primary receptor for human immunodeficiency virus-1 (HIV-1)

The Human CD4 gene



<https://www.uniprot.org/uniprot/P01730>

Thank you for your attention

