



БІОЛОГІЧНА ІНФОРМАЦІЯ. БІОІНФОРМАТИКА

*“Bioinformatics is the science of
 managing, mining, and interpreting
 information from biological data”*

Biological Information — What is It?

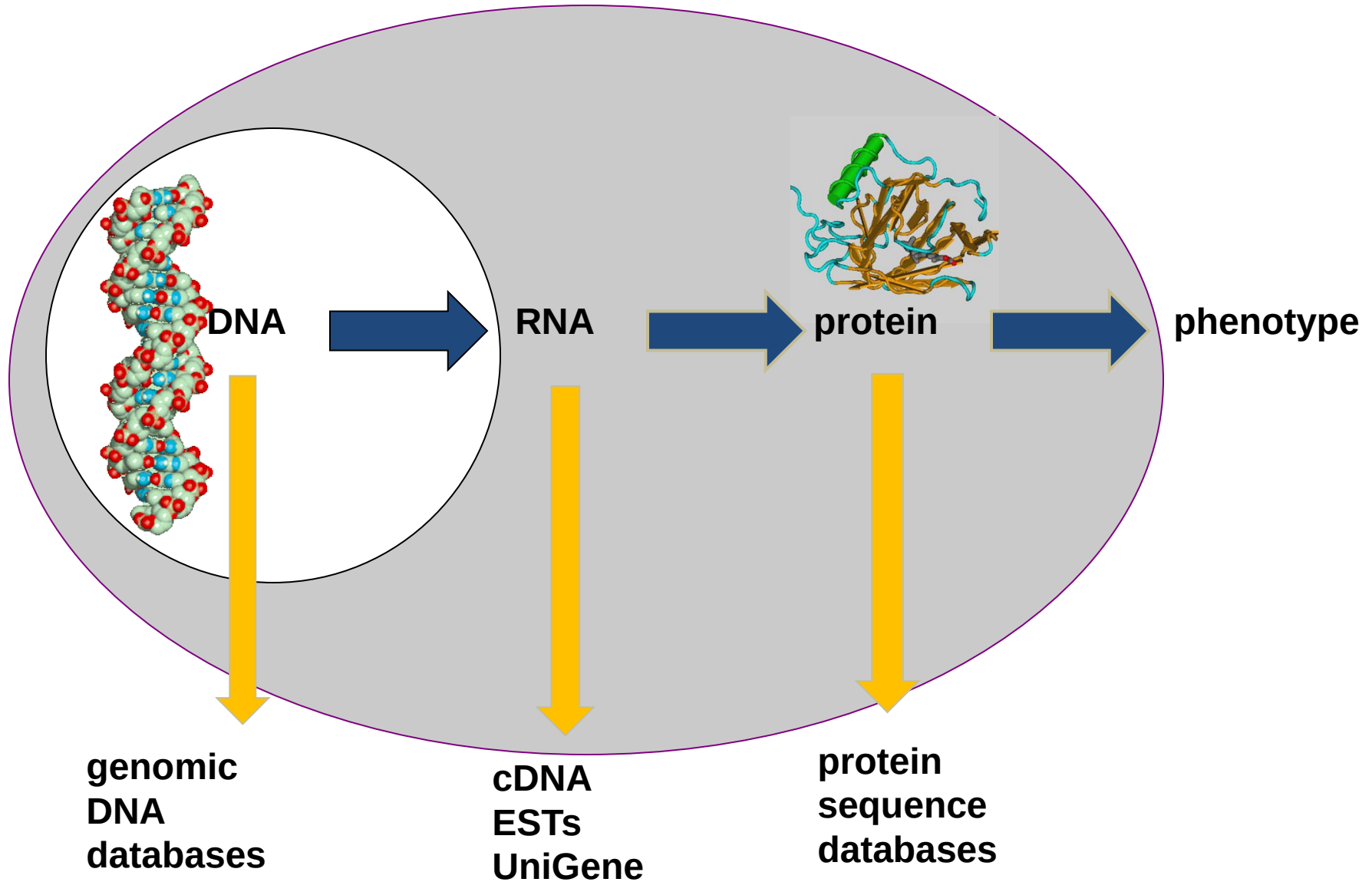
Werner Gitt^{1*}, Robert Compton² and Jorge Fernandez³

¹German Federal Institute of Physics and Technology, Braunschweig, Germany. Former Head of Information Technology. Director and Professor (retired). ²Department of

Інформація в геномі, комунікаційні мережі РНК і білків, міжклітинна комунікація (сигналінг) в регуляції процесів на рівні клітини та організму

The title of this symposium is “Biological Information: New Perspectives”. But what do we mean by the term “biological information”? We suggest that, at present, it cannot be unambiguously defined. Yet, an unambiguous definition would be extremely helpful because multiple levels of communication systems are being researched: from the DNA-coded information in the genome, to the intra-cellular communication networks involving RNA and proteins, to inter-cellular signaling via entities such as hormones, all the way up to and including the nervous system and the brain. Clearly, identifying all of these communication systems and defining the information that is being transferred will be a challenge.

Потік інформації згідно догми молекулярної біології і подання інформації у первинних БД



Послідовність → 3D структура → біологічна функція

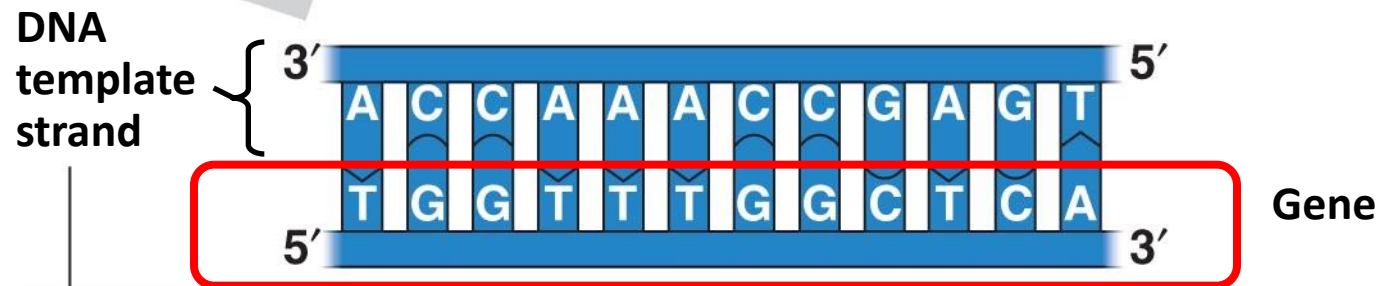
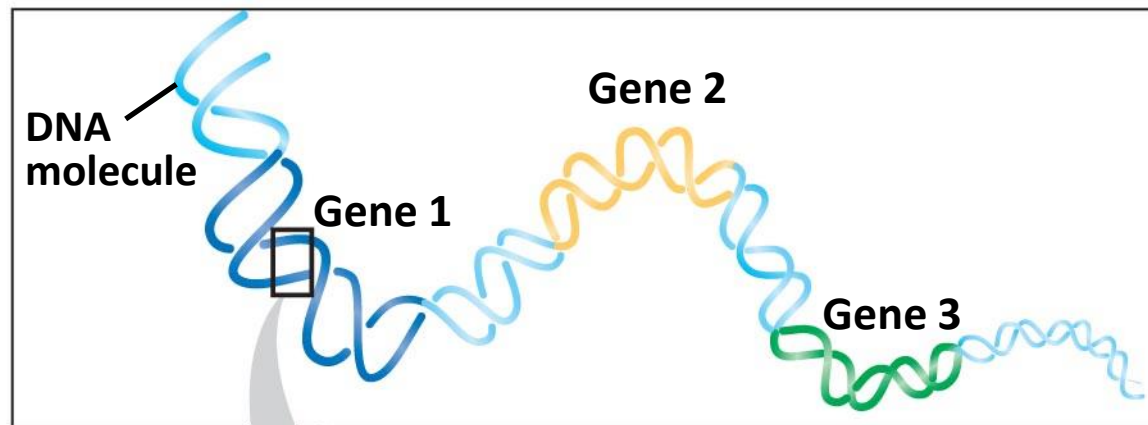
Типи біологічної інформації - 1

Лінійна = послідовності (сіквенци=sequences)

НК і білків – результат секвенування

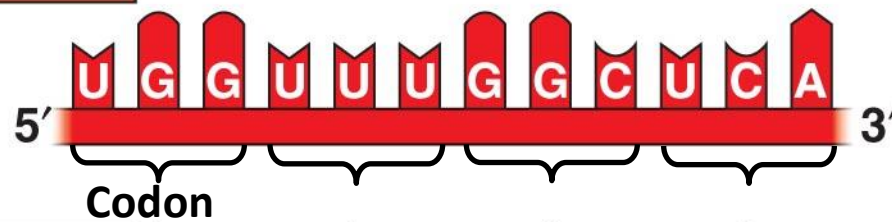
Carriers (носії) of linear biological information

Object	Monomers	Types of monomers	Approx. length (nbases)
Gene	(deoxyribo)-nucleotides	4 (AGCT)	10,000 -100,000
Transcript	(ribo)-nucleotides	4 (AGCU)	?<50 -10,000
Protein	Amino acids	20 + modified	50-1,000 or >



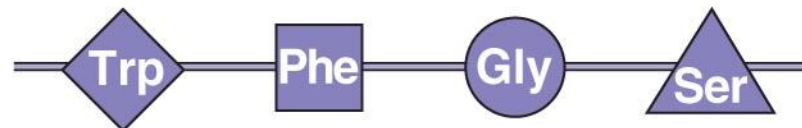
TRANSCRIPTION

mRNA



TRANSLATION

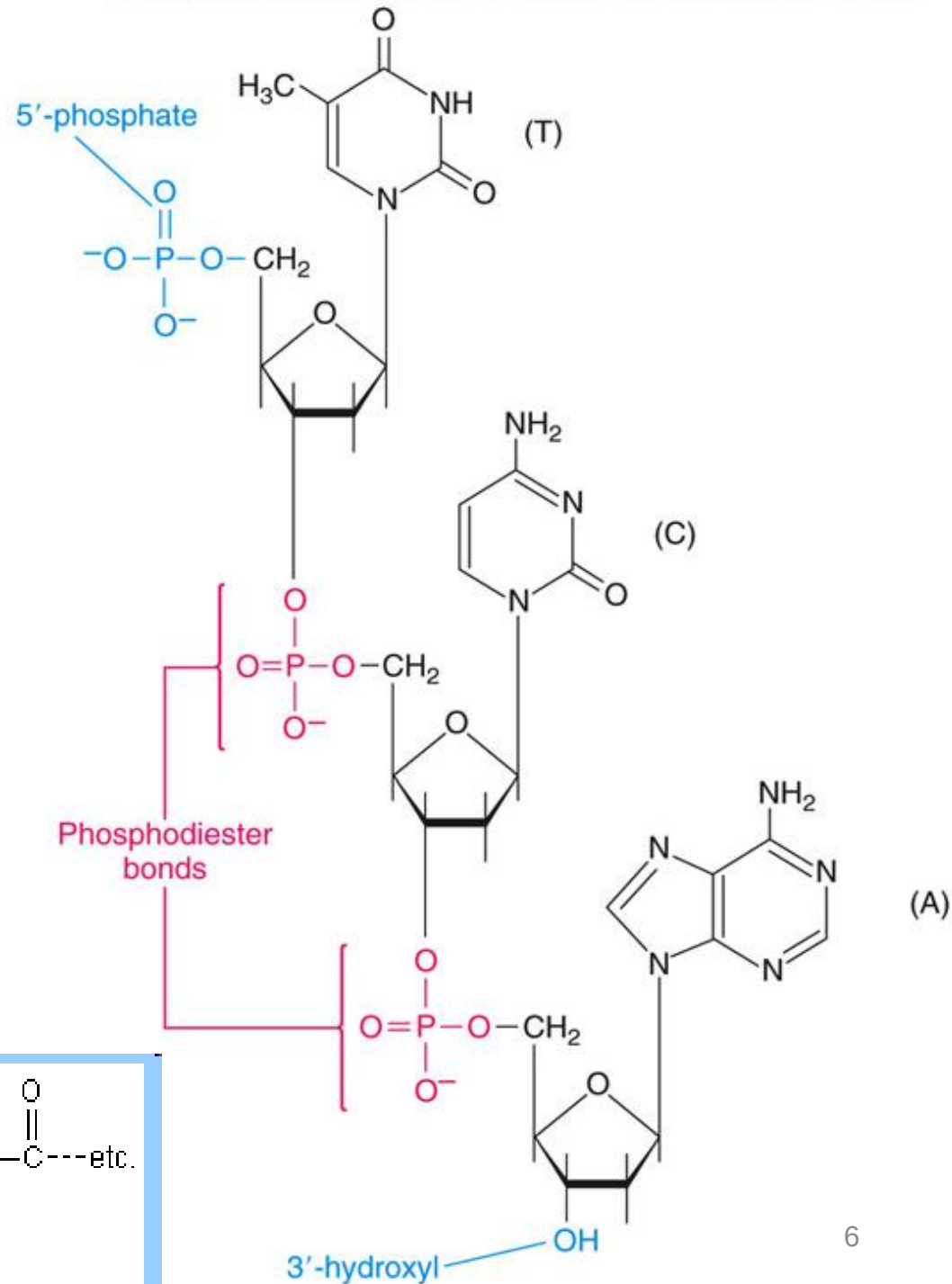
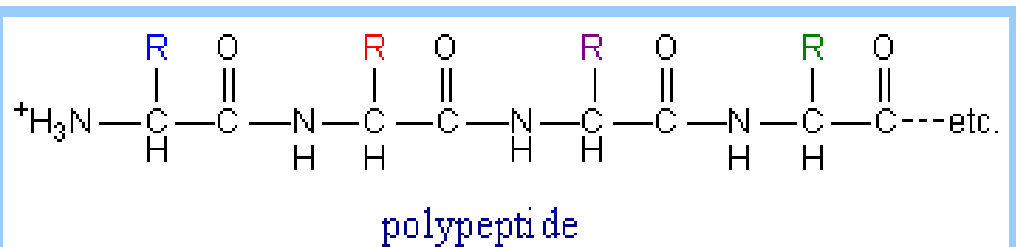
Protein



Amino acid

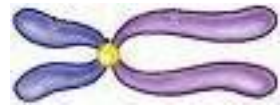
Послідовності (сіквенси=sequences) НК і білків:

- 5'→3';
- N-end → C-end



High throughput sequencing (секвенування з високою пропускнуою здатністю)

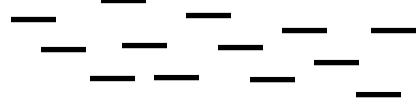
Large-scale sequencing projects



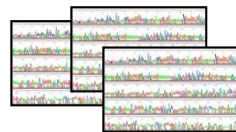
chromosome



fragment



sequencing library

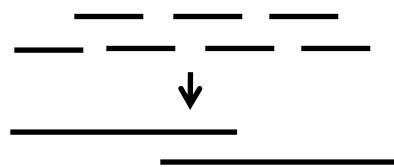


sequence reads

submission <<

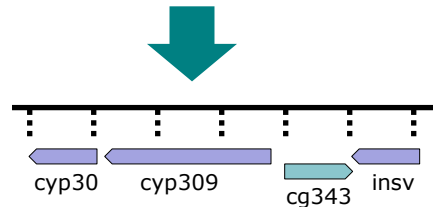
e.g. whole genome shotgun

submission <<



assembly
sequence

submission <<



annotation



ACTGC
TGCTA
GCTAG

>chromosome:GRCh37:11:6462894:-1
CCTCATTACTGTCAACCCCTTCACTTCCAGGCACTTTTTGCAAAGCCCTTTGCCAGTCAGG
GAAGGCGAGAGGCTGGGCATGGGGCTTGGACATTTGACAACAGTGAGACATTATTGTCCC
CAGACTCACTAGCCCAAGGGTAAAGCTGAAGAGGCTTGGGCATGCCCCAGAAAGGCCCT
GATGAAGCTTGGAAAAAGCTGTTCTCTGAGTATTTCTAAGTAAGTTTATCTGTGTGTG
GTTACTAAAAGTAGTAAGTATTGCTGTCTCTAGCTGCCCTTAGAGCAGGGCTTGACACAGT
ACACAGCAATATTAGTTCCCTCCTTTTCTCACCTCCCCCATTTGTGGAGATAAACTCAATC
ACAAAAGGTGATCCTCAGTCTACTCACTTCCCTGACTTATGGATGCCTGGACCCATTGCC
AGTGTGAGAGTCACAGCTGGACGTCAGCAGTGTAGCCAGTTACTGCTTGAAAATTGCTG
AAGGGGGTTGGGGGGCAGCTGCCGGGAAAAAGGAGTCTTGGATTGAGATTTCTGTCCAGA
CCCTGACCTTATTTGCAGTGATGTAATCAGCCAATATTGGCTTAGTCCTGGGAGACAGCA
CATTCCAGTAGAGTTGGAGGTGGGGGTGGTGTCTGTGCCAACTCTATATAGGGAGTTCA
ACTGGTCACCCAGAGCTGTCTGTGGCCTCTGCAGCTCAGCATGGCTAGGGTACTGGGAG
CACCCGTTGCACTGGGGTTGTGGAGCCTATGCTGGTCTCTGGCCATTGCCACCCCTCTTC
CTCCGTGAGTAAAGCTGGGACTAGAAGCGAAGGATTGAGTTCTGGGCTAGGGTAAGGTAG
GGCCAGTTTTTTAGGCCCTCGGTCAAATTTGGGGTCAGGGGCTATGGGAAAGGGATCGGTCC
CAATGGATCAAGATATCTATTTTGTCTCCCTAG**GACTAGTGCCCATGGGAATGTTGCTG**
AAGGCGAGACCAGCCAGACCAGACGTGACTGGTGAGGCCCTGACTCCCTAAGTCTGTC
TTATCTGTCTGGTTGTGTCTCTGCATTTTATCACCTTCTGGTTTTTTTTTTTTTTTTTT
TTTTTTTTTTTTTTTTTACTTTGCCATCTCCCTACCTCCACCCAG**AACGCTGCTCAGAT**
GGCTGGAGCTTTGATGCTACCACCTGGATGACAATGGAACCATGCTGTTTTTAAAGGT
AGGAGGGACTGAGGTTAGGGCGTTTAGGACCTTAGACTTACTCTCCTTCACAAAGGGTGT
CCCTGTCTGTGGGAGGTCTTAGGAATTATCTGATGGTATCACTGACAGCTTCTCTCAAGC
TATCTCAGTAGGTCAAAGGTTTCTCACTGGGCCCCCTCAGTGAGTGTGGGTTTTTTCAG**GG**
GAGTTTGTGTGGAAGAGTCACAAATGGGACGGGAGTTAATCTCAGAGAGATGGAAGAAT
TTCCCCAGCCCTGTGGATGCTGCATTCCGTCAAGGTACAAACAGTGTCTTTCTGATCAAG
GTACTGCTGGGCCAAAATCAGGGCCAGGCTGGAAAGGGCTGGAATCGACACTGGGGACCC
TTCCCCCAAATGGCCTTGGCATGGAGCCCATAGCAATAGGTAGCAGATTTCTTTCCCATG
TGCCCTCCTTTTCTGTAAAAGCTTGGGCTAAGGGAGTGTGCATGCGTGTGGGCCTGGCAG
GTGCACCATCCAGTGGCTGTTCTTCAGTCTAGTCTTAGTTCTACACCGCTCTGCTGTAC
CTCACACTGCTGGCCATCCTTTTTTCTCTGGCAATTGCTTCCCTTGCCCTTCCATGACCC
TGATCAAGTCTCTTCATAGGGCAAGGCAAGTTGTTCCCAACACAATGGCACCTGGCTA
GAAGAGCATGTGGAGCATGAAATCCAGTCTGCT**GTGGTCAACCAAGTCCCATGTGACCCAG**
GCTGTGTCTGCTCAGAGGAAGGGGTGCCTTTTCT

1D biological
Information :
sequences

← Gene of hemopexin
(about 11,000 nb)

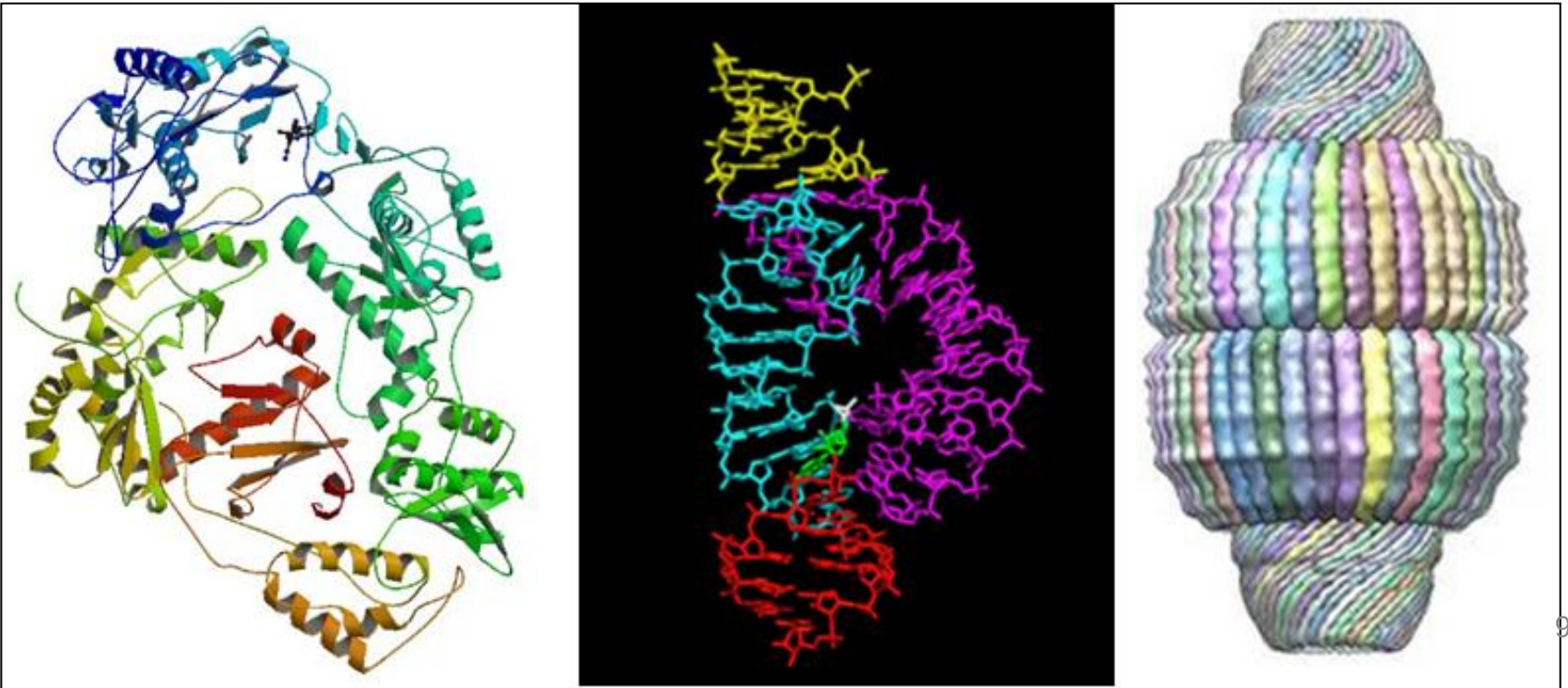
mRNA (1652 nb)

Protein (462 aa)

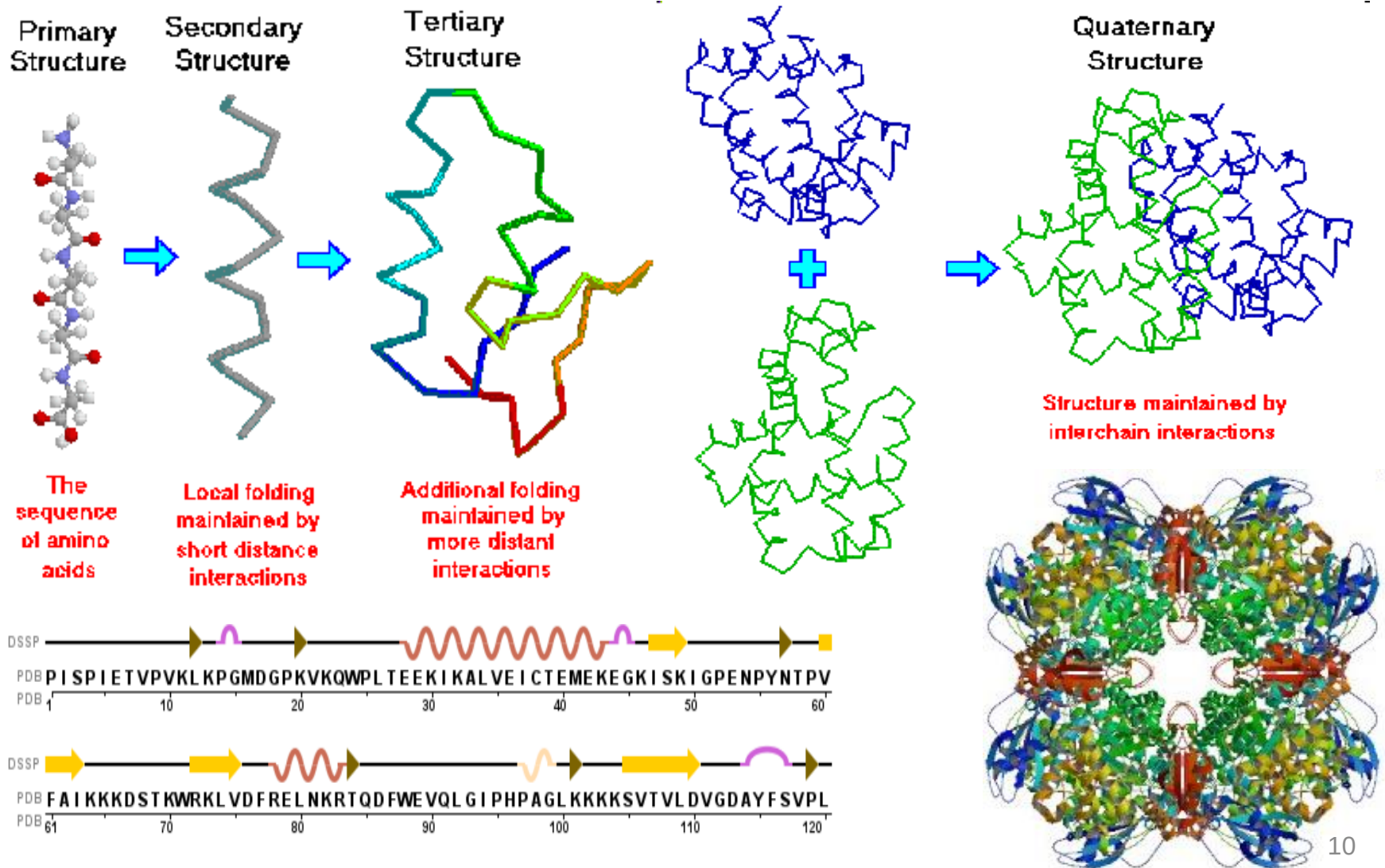
```
>gi|11321561|ref|NP_000604.1| hemopexin precursor [Homo sapiens]
MARVLGAPVALGLWSLCWSLAIATPLPPTSAHGNVAEGETKPDVPTERCSDGWSFDATTLDDNGTMLFF
KGEFVWKSHKWDRELISERWKNFPSPVDAAFRQGHNSVFLIKGDKVWVYPPEKKEKGYPKLLQDEFPGIP
SPLDAAVECHRGECAEGVLFFQGDREWFWDLATGTMKERSWPAVGNCSALRMLGRYYCFQGNQFLRFD
PVRGEVPPPRYPDRDVRDYFMPPCPGRGHGHRNGTGHGNSTHHGPEYMRCSPHLVLSALTSDNHGATYAFSGT
HYWRLDTSRDGWHSWPIAHQWPQGPSAVDAAFSWEEKLYLVQGTQVYVFLTKGGYTLVSGYPKRLEKEVG
TPHGIILDSVDAAFICPGSSRLHIMAGRRLWMLDLKSGAQATWTLPWPHEKVDGALCMEKSLGPNSSCA
NGPGLYL IHGPNLYCYSDVEKLNAAKALPQPQNVTSLLGCTH
```


Типи біологічної інформації - 2

3D = просторова структура (в осн. білків, але також НК) – задована в сиквенсі,
основа для реалізації функцій;

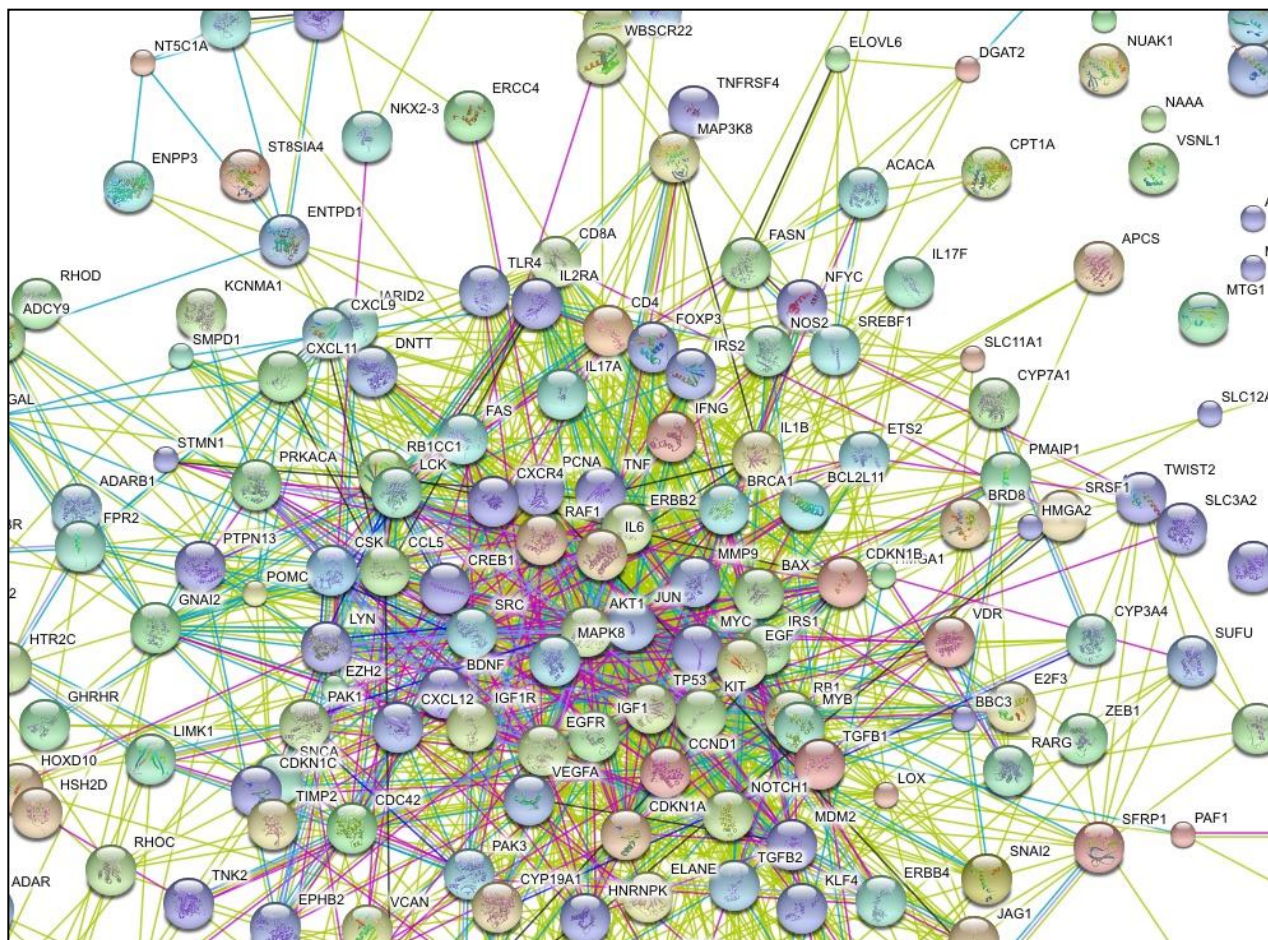


Levels of protein structure: primary (sequence), secondary, tertiary, quaternary



Типи біологічної інформації - 3 :

“4D” а) мережі взаємодії молекул



Б) генна експресія в нормі, при патологіях і стресі

<http://www.ncbi.nlm.nih.gov/geo/>



Gene Expression Omnibus

[GEO Publications](#) | [FAQ](#) | [MIAME](#) | [Email GEO](#)

NCBI » GEO [Login](#)

Gene Expression Omnibus: a public functional genomics data repository supporting [MIAME-compliant](#) data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles. [More information »](#)

GEO navigation

QUERY

DataSets

GO

Gene profiles

GO

GEO accession

GO

GEO BLAST

BROWSE

DataSets

GEO accessions

Platforms

Samples

Series

Site contents

Public data

Platforms	9,835
Samples	701,303
Series 	28,460
DataSets	2,720

Documentation

[Overview](#) | [FAQ](#) | [Find](#)

[Submission guide](#)

[Linking & citing](#)

[Journal citations](#)

[Construct a Query](#)

[Programmatic access](#)

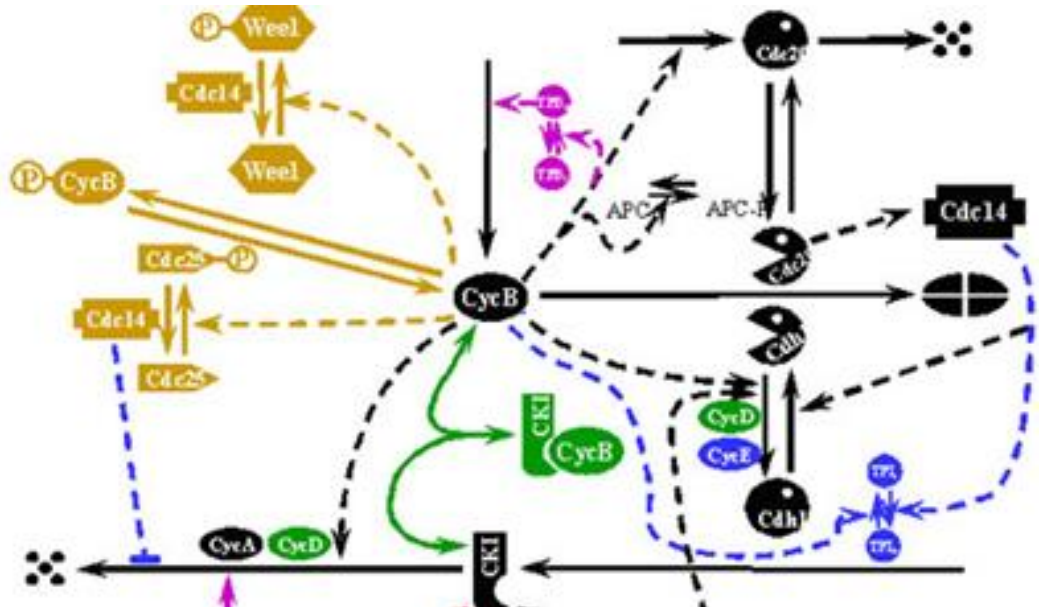
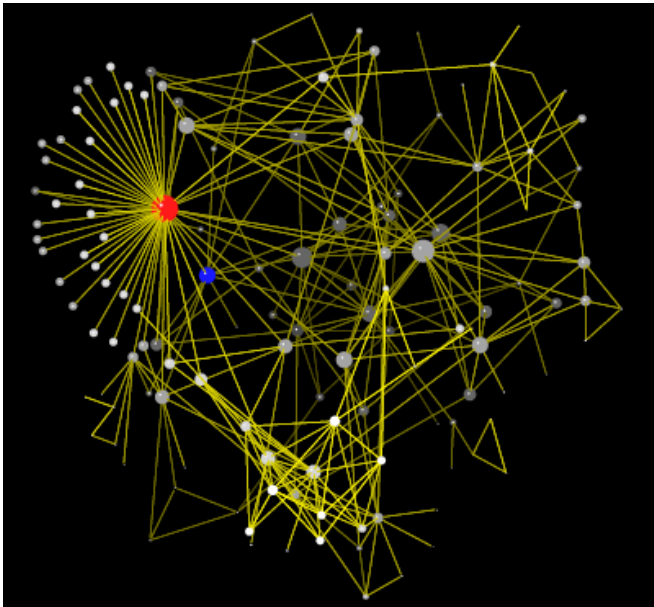
[DataSet clusters](#)

[GEO announce list](#)

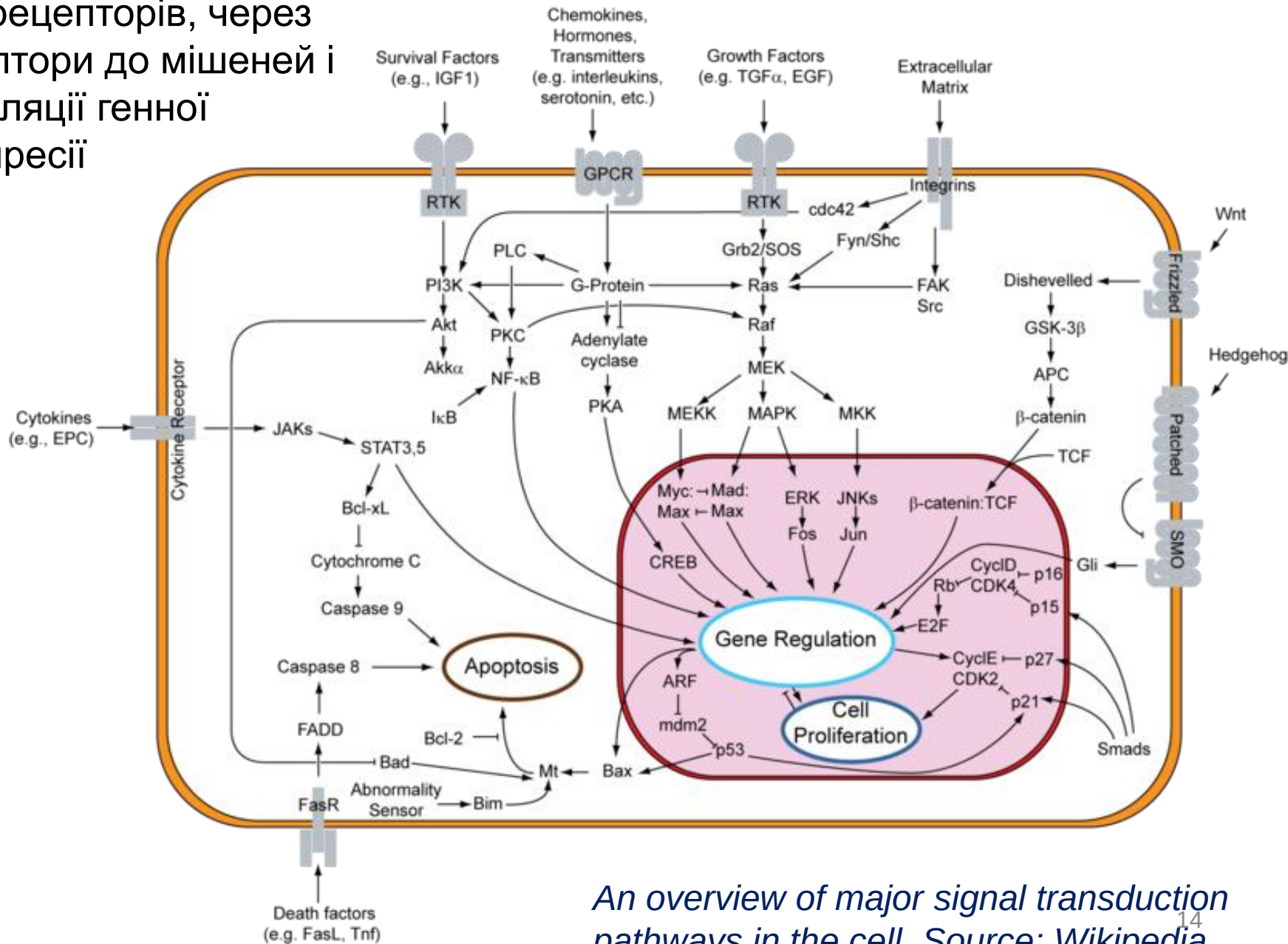
[Data disclaimer](#)

12

В) метаболічні шляхи, потік метаболітів, ферментативні процеси у часі



Г) **сигнальні шляхи** –
від рецепторів, через
адаптори до мішеней і
регуляції генної
експресії



An overview of major signal transduction pathways in the cell. Source: Wikipedia.

Bioinformatics resources

Databases

Annotated
genomic DBs
(on-line only)

DBs
specialized
by object
(organism,
organelle,
molecule),
disease, etc

(on-line &
downloads)

Genome
browsing,
sequences,
annotations
alignments,
references,
cross-links,
phylo-
genetics

Program tools

(on-line & downloads)

**Linear
bio-
informatics**

**Structural
bio-
informatics**

**System
biology tools**

'Omics'
raw data
analysis

**Sequence
analysis:**
alignments,
similarity
search,
phylogene-
tics trees

Structure
(proteins,
RNA)
analysis &
prediction,
modeling,
docking,
drug design

Metabolic
pathways
analysis,
networks
modeling,
virtual cell
simulations

Data
clustering,
identifi-
cation,
normali-
zation

Nucleic Acids Research

Issues Advance Articles Publish ▼ Purchase About ▼



Current Issue
Volume 45, Issue 4
28 February 2017

Impact Factor
9.202
5 year Impact Factor
8.647

Senior Executive
Keith R. Wilson
Barry Stoughton

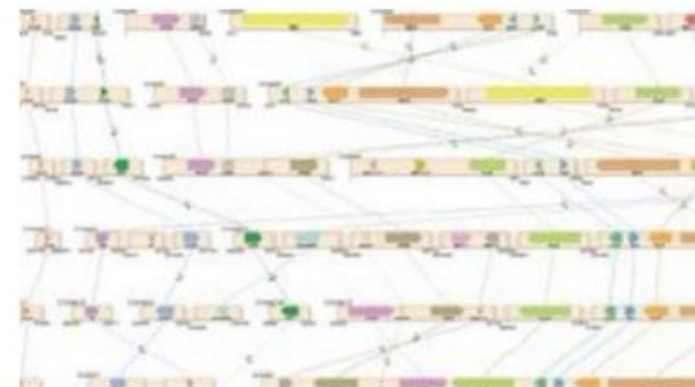
About the journal
Nucleic Acids Research (NAR) is a leading Open Access journal, which has been chosen



Database issue

The issue includes 152 papers, of which 54 describe newly created databases. 82 papers provide updates on databases that have been described in the previous *NAR* Database Issues, and 16 contain updates on databases whose descriptions have previously been published in other journals.

[View the issue](#)

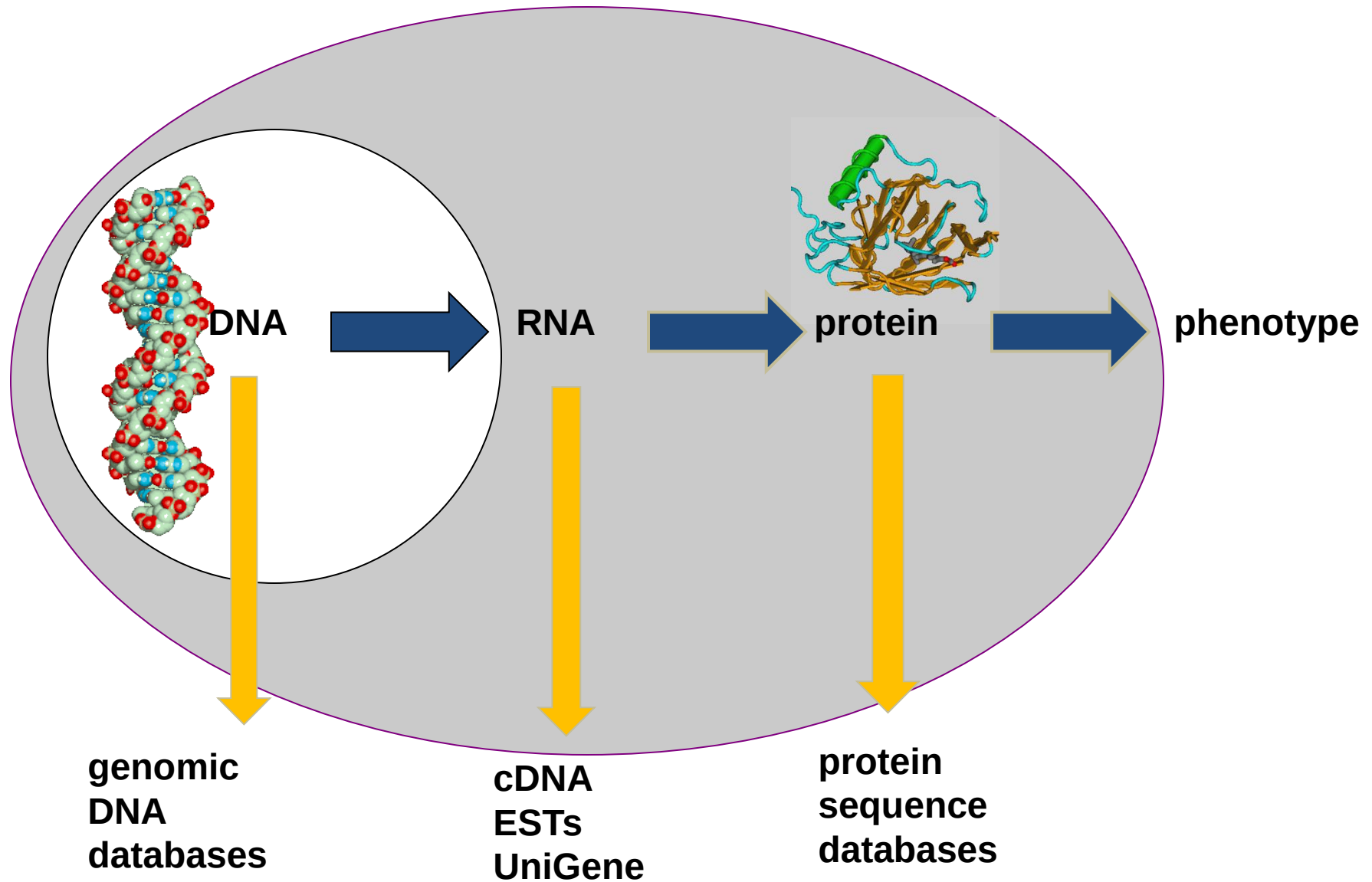


2016 Web Server issue

The papers in this issue reflect current and emerging trends in bioinformatics and computational biology. Of particular interest are recent works which move beyond an understanding of bio-molecular phenomena toward an emphasis on the development of engineering tools to exploit that understanding.

[View the issue](#)

Потік інформації згідно догми молекулярної біології і подання інформації у первинних БД



Послідовність → 3D структура → біологічна функція

Типи біологічної інформації - 1

Лінійна = послідовності (сіквенци=sequences)
інформаційних біополімерів –
НК і білків – вивчаються через
секвенування

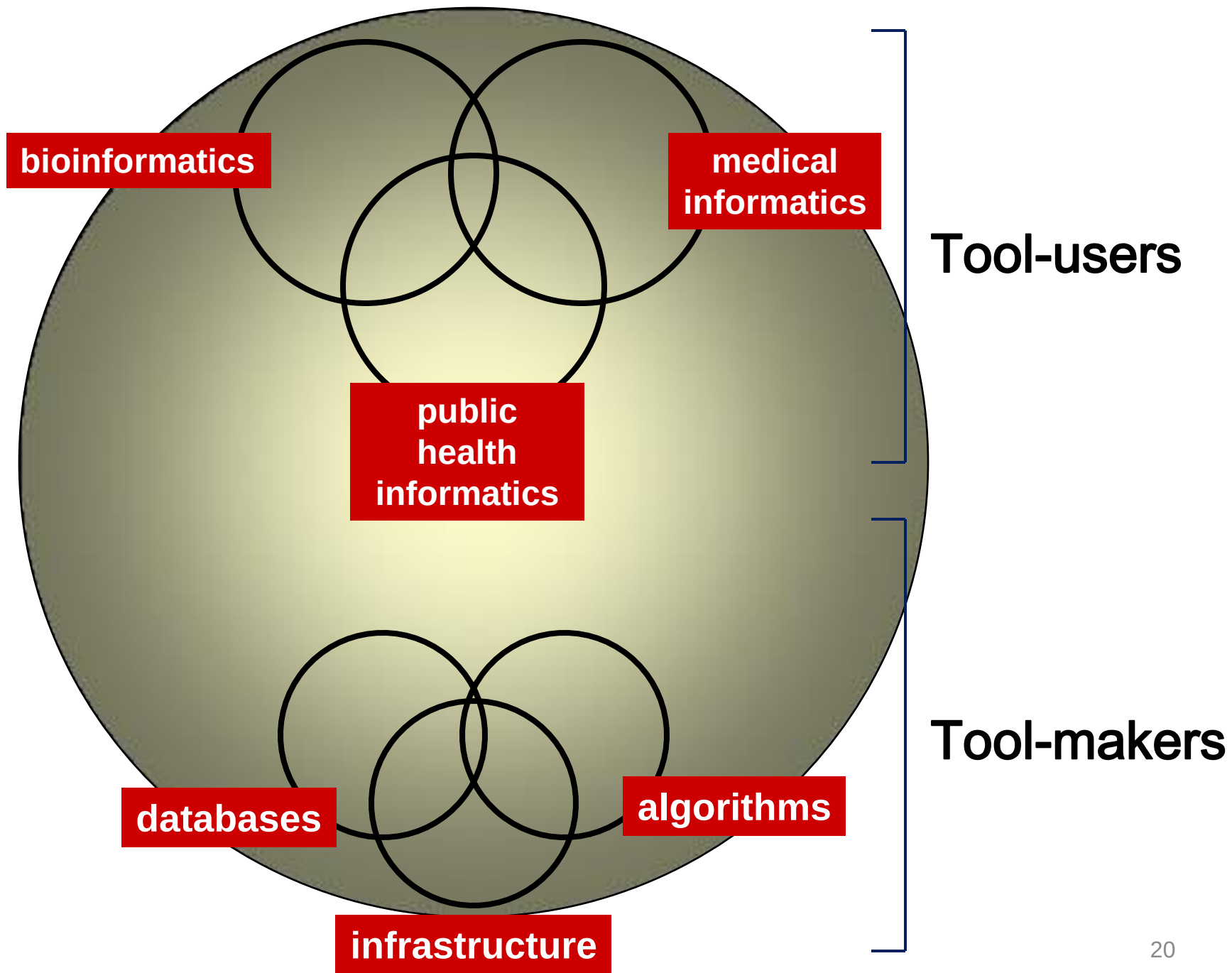
Carriers (носії) of linear biological information

Object	Monomers	Types of monomers	Approx. length (nbases)
Gene	(deoxyribo)-nucleotides	4 (AGCT)	10,000 -100,000
Transcript	(ribo)-nucleotides	4 (AGCU)	?<50 -10,000
Protein	Amino acids	20 + modified	50-1,000 or >

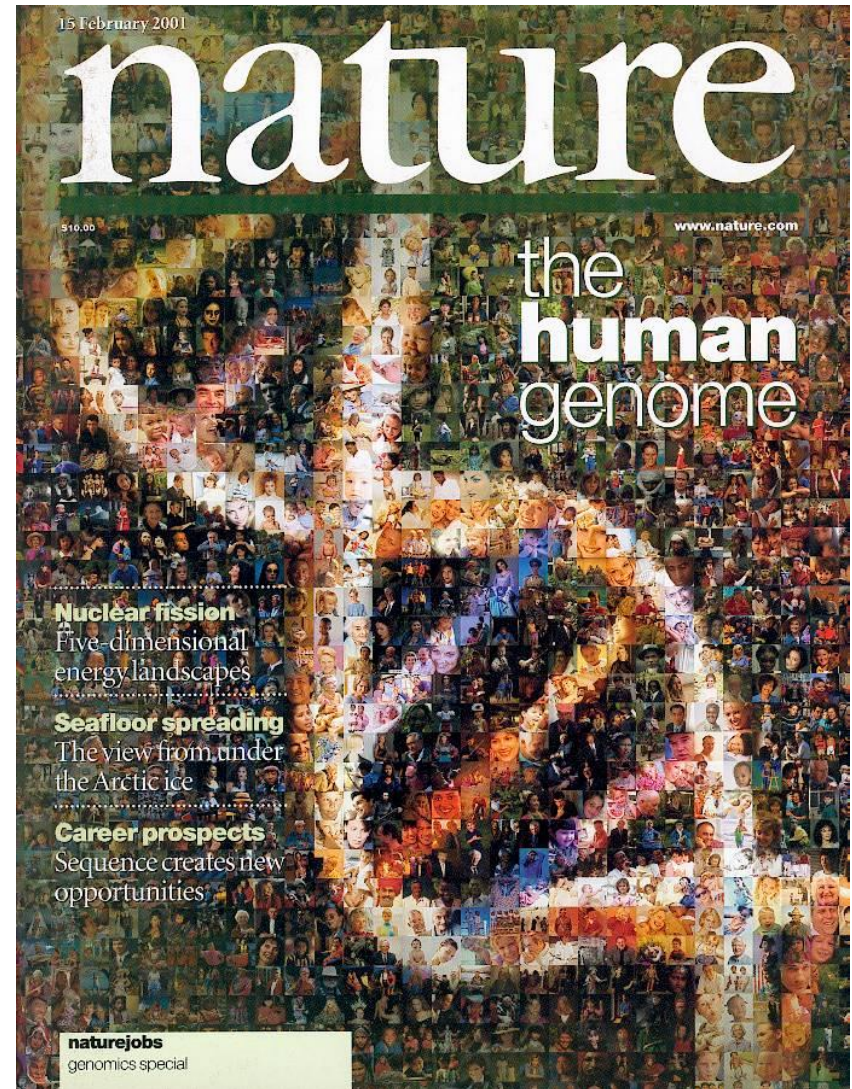
Предумови появи біоінформатики

- 1) розвиток біохімічних і біофізичних методів дослідження (розділення і очистка молекул, секвенування, мікрочипи, кристалографія і т.і.)
- 2) Розвиток комп'ютерної техніки і методів програмування
- 3) Розвиток мережевих технологій (інтернет, ГРИД-технології)





Pre-genomic → Post-genomic era



Bioinformatics resources

Databases

Annotated
genomic DBs
(on-line only)

DBs
specialized
by object
(organism,
organelle,
molecule),
disease, etc

(on-line &
downloads)

Genome
browsing,
sequences,
annotations
alignments,
references,
cross-links,
phylo-
genetics

Program tools

(on-line & downloads)

**Linear
bio-
informatics**

**Structural
bio-
informatics**

**System
biology tools**

**'Omics'
raw data
analysis**

**Sequence
analysis:**
alignments,
similarity
search,
phylogene-
tics trees)

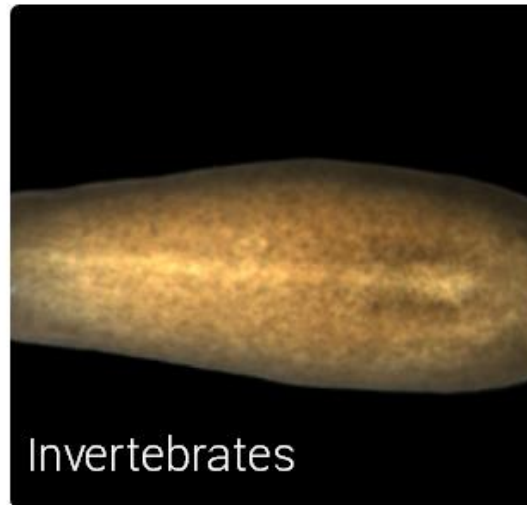
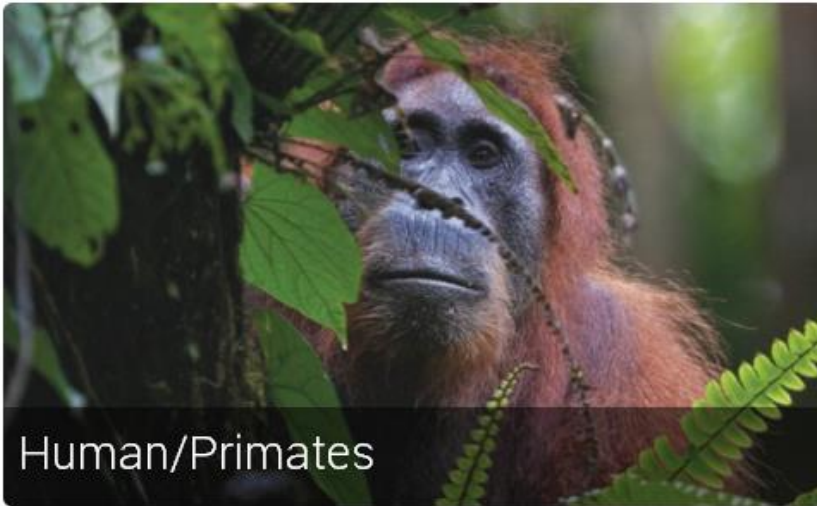
Structure
(proteins,
RNA)
analysis &
prediction,
modeling,
docking,
drug design

Metabolic
pathways
analysis,
networks
modeling,
virtual cell
simulations

Data
clustering,
identifi-
cation,
normali-
zation

**Результати біохімічних, фізико-хімічних досліджень –
основний шлях до наповнення первинних БД**

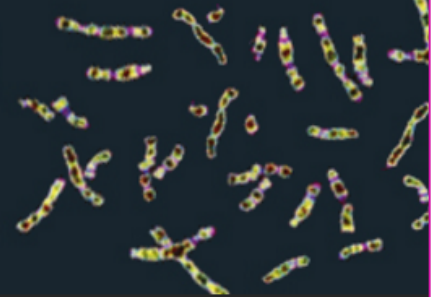
Genomes



<http://www.1000genomes.org/about>

IGSR: The International Genome Sample Resource

Providing ongoing support for the 1000 Genomes Project data



[Home](#) [About](#) [Data](#) [Analysis](#) [Participants](#) [Contact](#) [Browser](#) [FAQ](#)

About IGSR and the 1000 Genomes Project

The [1000 Genomes Project](#) ran between 2008 and 2015, creating the largest public catalogue of human variation and genotype data. As the project ended, the Data Coordination Centre at [EMBL-EBI](#) has received continued funding from [the Wellcome Trust](#) to maintain and expand the resource. The International Genome Sample Resource (IGSR) was set up to do this and has the following aims:

1. [Ensure the future access to and usability of the 1000 Genomes reference data](#)
2. [Incorporate additional published genomic data on the 1000 Genomes samples](#)
3. [Expand the data collection to include new populations not represented in the 1000 Genomes Project](#)

The 1000 Genomes Project

Overview of the 1000 Genomes Project

The goal of the 1000 Genomes Project was to find most genetic variants with frequencies of at least 1% in the populations studied.

<http://www.sanger.ac.uk/genetics/CGP/>



InformationProjectsOther Services

Cancer Genome Project
Genomics & Genetics
Overview
CGP
Faculty
Stratton
Futreal
Projects
Cancer Gene Census
COSMIC
CGP Resequencing Studies
Copy Number Mapping
NCI-60
Planned studies
Genomics of Drug

The Cancer Genome Project

Summary

All cancers occur due to abnormalities in DNA sequence. Throughout life, the genome within cells of the human body is exposed to mutagens and suffers mistakes in replication. These corrosive influences result in progressive, subtle divergence of the DNA sequence in each cell from that originally constituted in the fertilised egg. Occasionally, one of these somatic mutations alters the function of a critical gene, providing growth advantage to the cell in which it has occurred and resulting in the emergence of an expanded clone derived from this cell. Acquisition of additional mutations, and consequent waves of clonal expansion result in the evolution of the mutinous cells that invade surrounding tissues and metastasise. One in

Data Resources

 **Cancer Gene Census:**
Mutated genes causally implicated in human cancer.

 **COSMIC:**
Catalogue Of Somatic Mutations In Cancer

 **CGP Resequencing Studies:**
Somatic mutations from systematic large scale resequencing of genes in human cancers.

Our Science

[Home](#) > [Our Science](#)

<http://jgi.doe.gov/our-science/science-programs/>

Science Programs

[Plant Program](#)[Fungal Program](#)[Metagenome Program](#)[Microbial Program](#)[DNA Synthesis Science Program](#)[Emerging Science Program](#)

Fungal Genomics



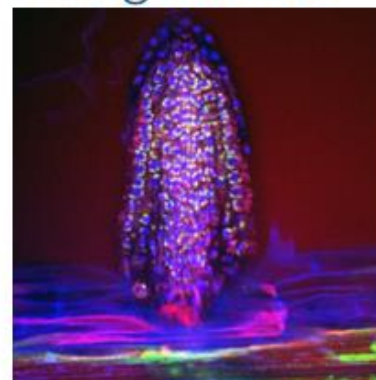
Plant Genomics



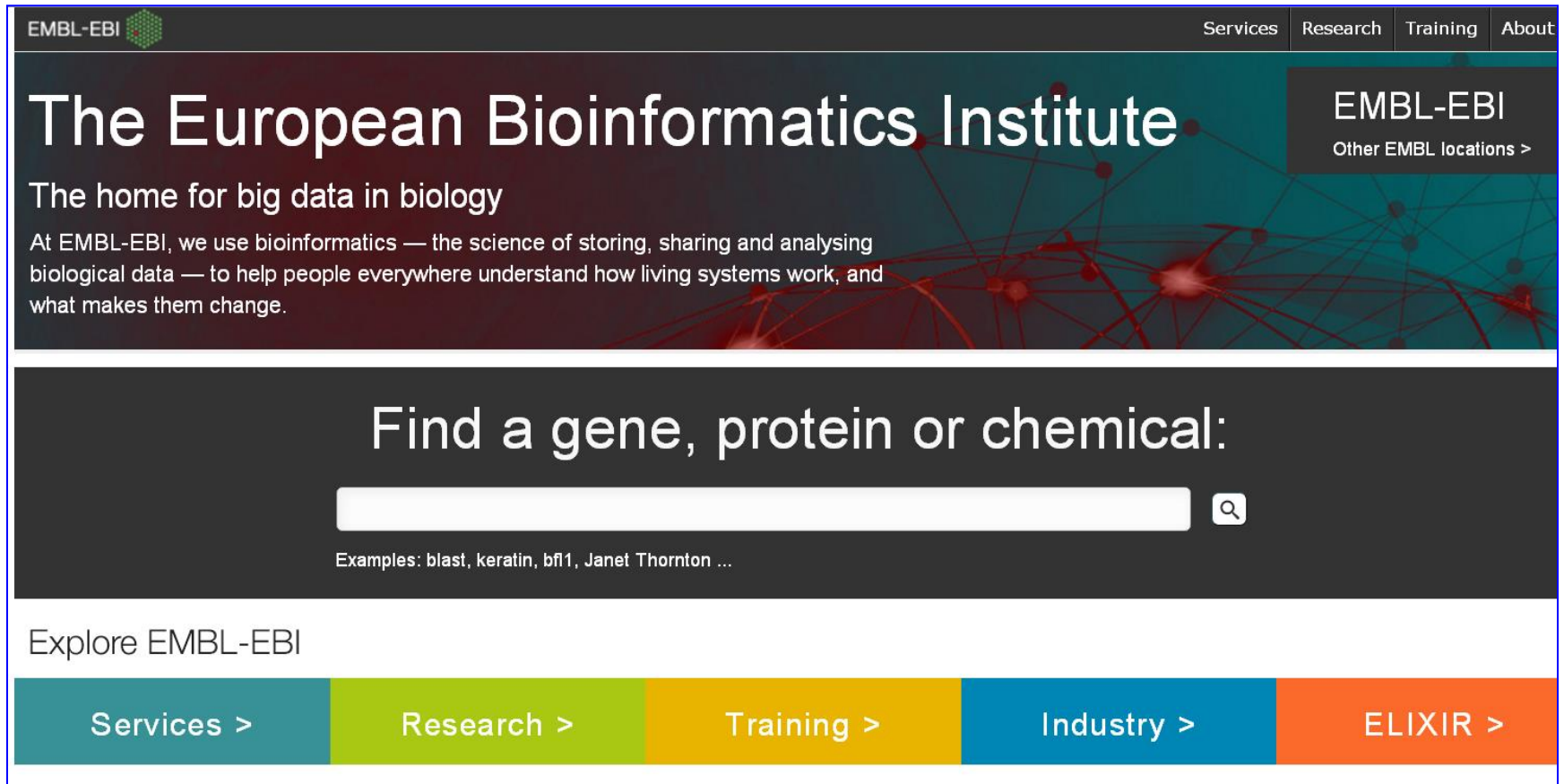
Microbial Genomics



Metagenomics



<http://www.ebi.ac.uk>



The screenshot shows the EMBL-EBI website homepage. At the top left is the EMBL-EBI logo. To the right is a navigation bar with links for Services, Research, Training, and About. The main header features the text 'The European Bioinformatics Institute' in large white font, followed by 'The home for big data in biology'. Below this is a paragraph explaining the institute's mission: 'At EMBL-EBI, we use bioinformatics — the science of storing, sharing and analysing biological data — to help people everywhere understand how living systems work, and what makes them change.' To the right of the header is a box with 'EMBL-EBI' and a link for 'Other EMBL locations >'. The background of the header has a network diagram with red nodes and lines. Below the header is a dark grey section with the text 'Find a gene, protein or chemical:' and a search bar. Below the search bar is a text input field with a magnifying glass icon. Below the input field are examples: 'Examples: blast, keratin, bfl1, Janet Thornton ...'. At the bottom is a section titled 'Explore EMBL-EBI' with five colored buttons: Services > (teal), Research > (green), Training > (yellow), Industry > (blue), and ELIXIR > (orange).

EMBL-EBI

Services Research Training About

The European Bioinformatics Institute

The home for big data in biology

At EMBL-EBI, we use bioinformatics — the science of storing, sharing and analysing biological data — to help people everywhere understand how living systems work, and what makes them change.

EMBL-EBI
Other EMBL locations >

Find a gene, protein or chemical:

Examples: blast, keratin, bfl1, Janet Thornton ...

Explore EMBL-EBI

Services > Research > Training > Industry > ELIXIR >

**Сервіси
Європейського
Біоінформаційного
Інституту:
основні категорії**

Browse by type



DNA & RNA



Gene
Expression



Proteins



Structures



Systems



Chemical
biology



Ontologies

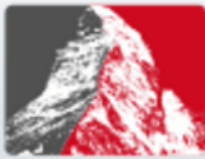


Literature



Cross
domain

Other European resources <http://www.expasy.org>



ExPASy
Bioinformatics Resource Portal

Query all databases



Visual Guidance

Categories

proteomics

genomics

structural bioinformatics

systems biology

phylogeny/evolution

population genetics

transcriptomics

biophysics

imaging

IT infrastructure

drug design

ExPASy is the **SIB Bioinformatics Resource Portal** which provides access to scientific databases and software tools (i.e., *resources*) in different areas of life sciences including proteomics, genomics, phylogeny, systems biology, population genetics, transcriptomics etc. (see **Categories** in the left menu). On this page you can find resources from many different SIB groups as well as external institutions.

Featuring today

Protein Spotlight

Monthly review articles describing a specific protein or family of proteins on an informal tone
[\[details\]](#)



<https://www.france-bioinformatique.fr>



French Institute of
Bioinformatics

Services and cloud ▾

Trainings ▾

Platforms ▾

Events ▾

About us ▾

Formations en Bioinformatique

L'IFB propose des formations en bioinformatique pour la recherche biomédicale et les sciences du vivant.

EN SAVOIR PLUS



CALL OF PROJECTS



SUPPORT REQUEST



WORK GROUP



JOB OFFERS



JOIN IFB



IFB CLOUD

<https://www.helmholtz-muenchen.de/ibis/>

IBIS Institute of Bioinformatics and Systems Biology

HelmholtzZentrum münchen
German Research Center for Environmental Health

[Institute](#) [News](#) [Resources](#) [Publications](#) [Staff](#) [Service](#)

search term 

Institute of Bioinformatics and Systems Biology

Welcome to the Institute of Bioinformatics and Systems Biology

Welcome to the Institute of Bioinformatics and Systems Biology

We analyse and interpret biological data to capture information on the etiology and progression of human diseases. Our focus is on qualitative and quantitative modeling, on the interconnection of metabolic profiles with genetic variance, on the systematic interpretation of high throughput data and on the systems biology of small molecules.

Groups and research topics

 Curation of Human Disease Networks >

 Metabolomics >

News

all News

**30.03.2017**
SNiPA v3.2 (released in March 2017) >

**15.03.2017**
Prof. Meuser retires after 16

Contact

Helmholtz Zentrum München
German Research Center
for Environmental Health (GmbH)
Institute of Bioinformatics and Systems Biology
Ingolstädter Landstraße 1

**Dr. Gabi Kastenmüller**
 [Email](#)
Phone +49 89 3187 3578

<https://www.molgen.mpg.de/en/bioinformatics>



**MAX PLANCK INSTITUTE
FOR MOLECULAR GENETICS**

ABOUT US | **PUBLIC RELATION** | **RESEARCH** | **YOUNG SCIENTISTS** | **PUBLICATION OVERVIEW** | **STAFF** | **CENTRAL FACILITIES**

JOB OFFERS

► Home ► Research ► Department Computational Molecular Biology(Martin Vingron)

**Department
Computational
Molecular Biology
(Martin Vingron)**

- Martin Vingron
- Groups
- Team
- PuRe Publications

COMPUTATIONAL MOLECULAR BIOLOGY

Department Computational Molecular Biology

Computational biology studies biological questions with mathematical and computational methods. In the area of molecular biology and genomics, the possibility to apply such formal methods, of course, comes from the availability not only of genome sequences, but also of large amounts of functional data about cellular processes. Computational molecular biology encompasses both development and adaptation of methods in the areas of mathematics, statistics, and computer science, as well as pursuing biological questions applying these tools and close collaborations with experimentalists. The research interest of the Computational Molecular Biology Department lies in understanding gene regulatory mechanisms in the context of structure and evolution of the eukaryotic genome. To this end, mathematical,

<http://computing.bio.cam.ac.uk>



UNIVERSITY OF
CAMBRIDGE

Study at Cambridge

About the University

Research at Cambridge

Quick links



Se

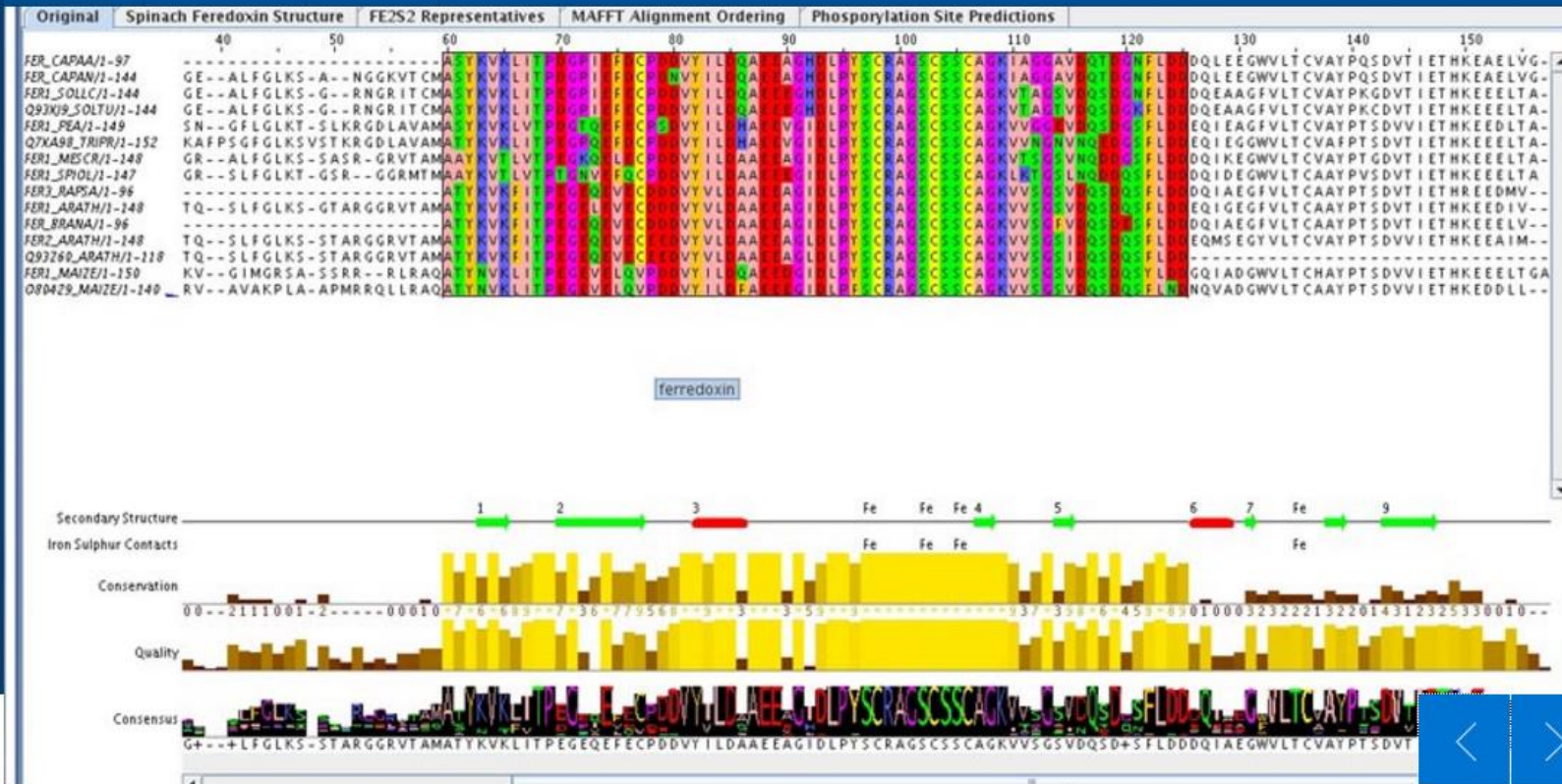
[Home](#) / [School of the Biological Sciences](#) / [Bioinformatics and Computational Biology Services](#)

Bioinformatics and Computational Biology Services

[Home](#)

[Training](#)

[News](#)



Top Links

[Book courses](#)

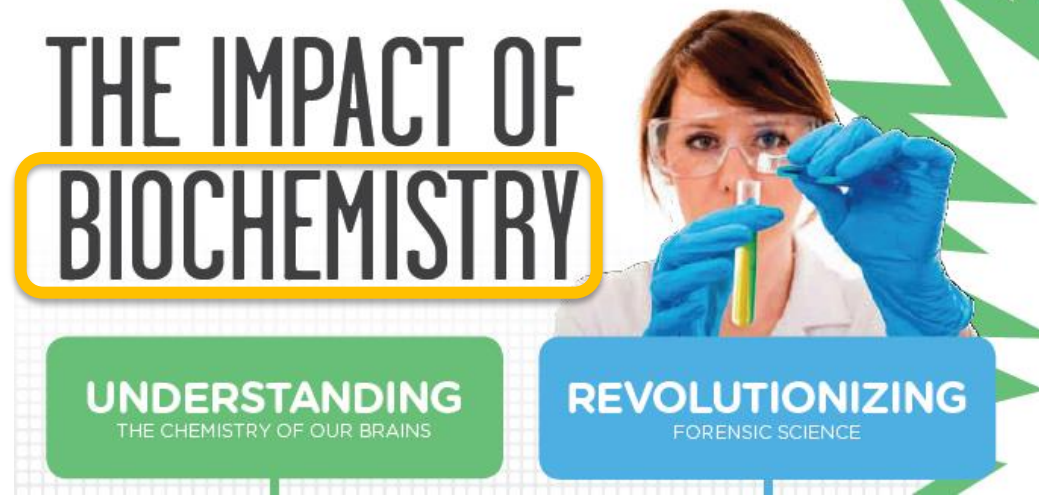
Omics Technologies як результат розвитку біохімічних методів, ІТ-технологій і біоінформатики

- Методи очищення і розподілення молекул
- Секвенування
- Ідентифікація молекул (полімерів і малих молекул)
- Аналіз структури
- Аналіз взаємодії молекул

XX – сторіччя фізики
XXI – сторіччя біології



- Геноміка і метагеноміка
- Транскриптоміка
- Протеоміка
- Метаболоміка
- Інтерактоміка, ін.



Future

