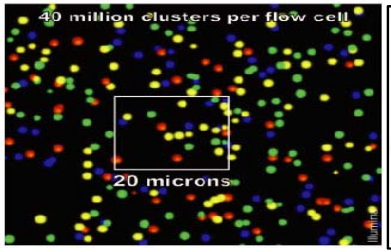
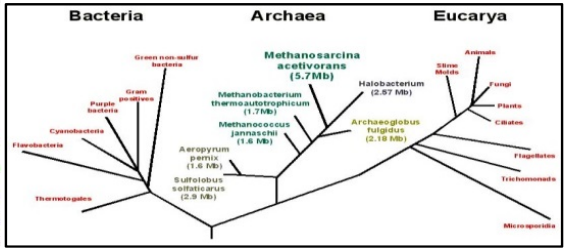
生物信息学：导论与方法
Bioinformatics: Introduction and Methods

Ge Gao 高歌 & Liping Wei 魏丽萍
Center for Bioinformatics, Peking University

TAACCCTAACCCTAACCCTAACCCTAACCCTA
CCTAACCCTAACCCTAACCCTAACCCTAACCC
CCCTAACCCCTAACCCTAACCCTAACCCTAAC
AACCCTAACCCTAACCCTAACCCCTAACCCTA
ACCCTAACCCCAACCCCAACCCCAACCCCAAC
CTACCCTAACCCTAACCCTAACCCTAACCCTA
ACCCTAACCCTAACCCTAACCCCTAACCCTAA



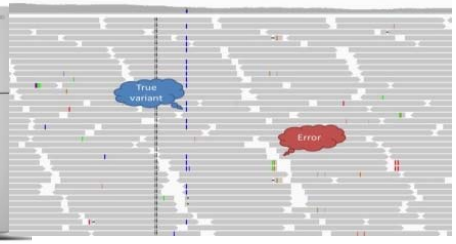


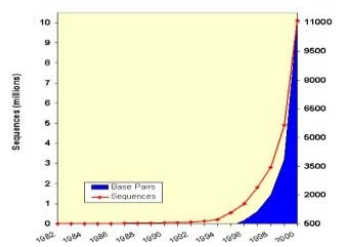
Introduction and History

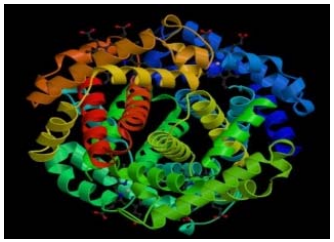
北京大学生物信息学中心 魏丽萍

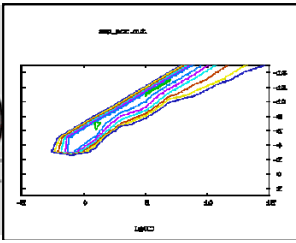
Liping Wei, Ph.D.

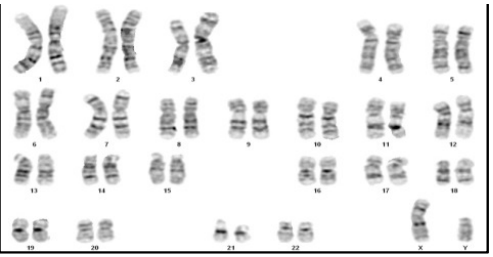
Center for Bioinformatics, Peking University

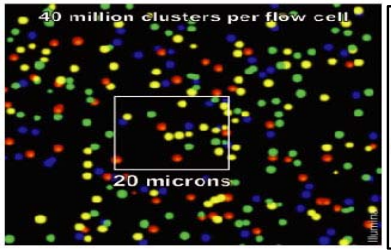


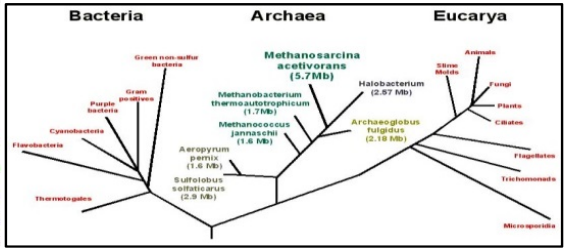










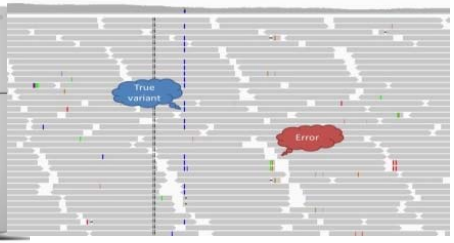


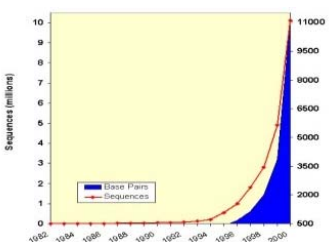
Unit 1:
What is bioinformatics?

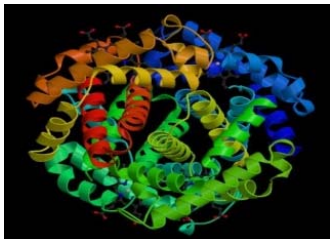
北京大学生物信息学中心 魏丽萍

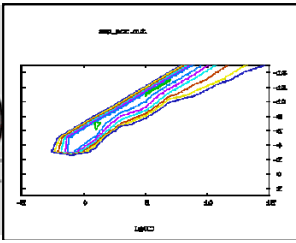
Liping Wei, Ph.D.

Center for Bioinformatics, Peking University

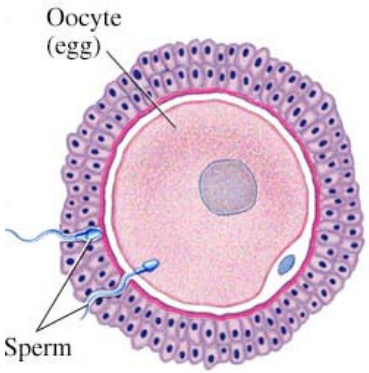






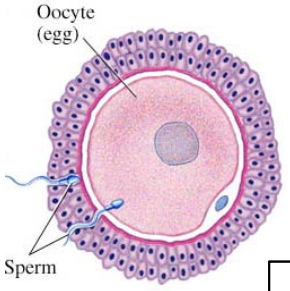


The miracle of life

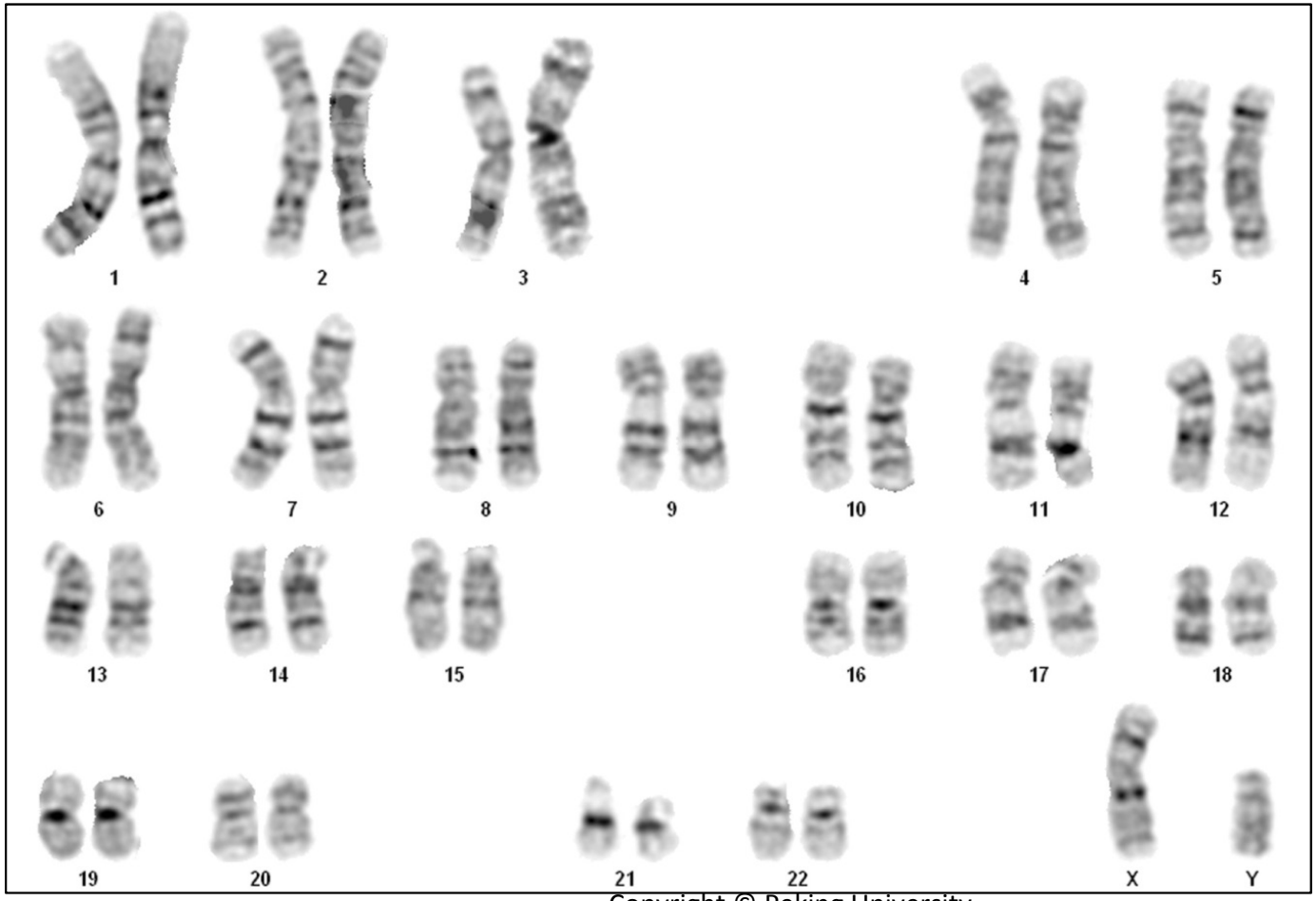


http://img.webmd.com/dtmcms/live/webmd/consumer_assets/site_images/media/medical/hw/n5551221.jpg

http://commons.wikimedia.org/wiki/File:Hospital_newborn_by_Bonnie_Gruenberg2.jpg



Genome: the “manual of life” -- ***Almost***

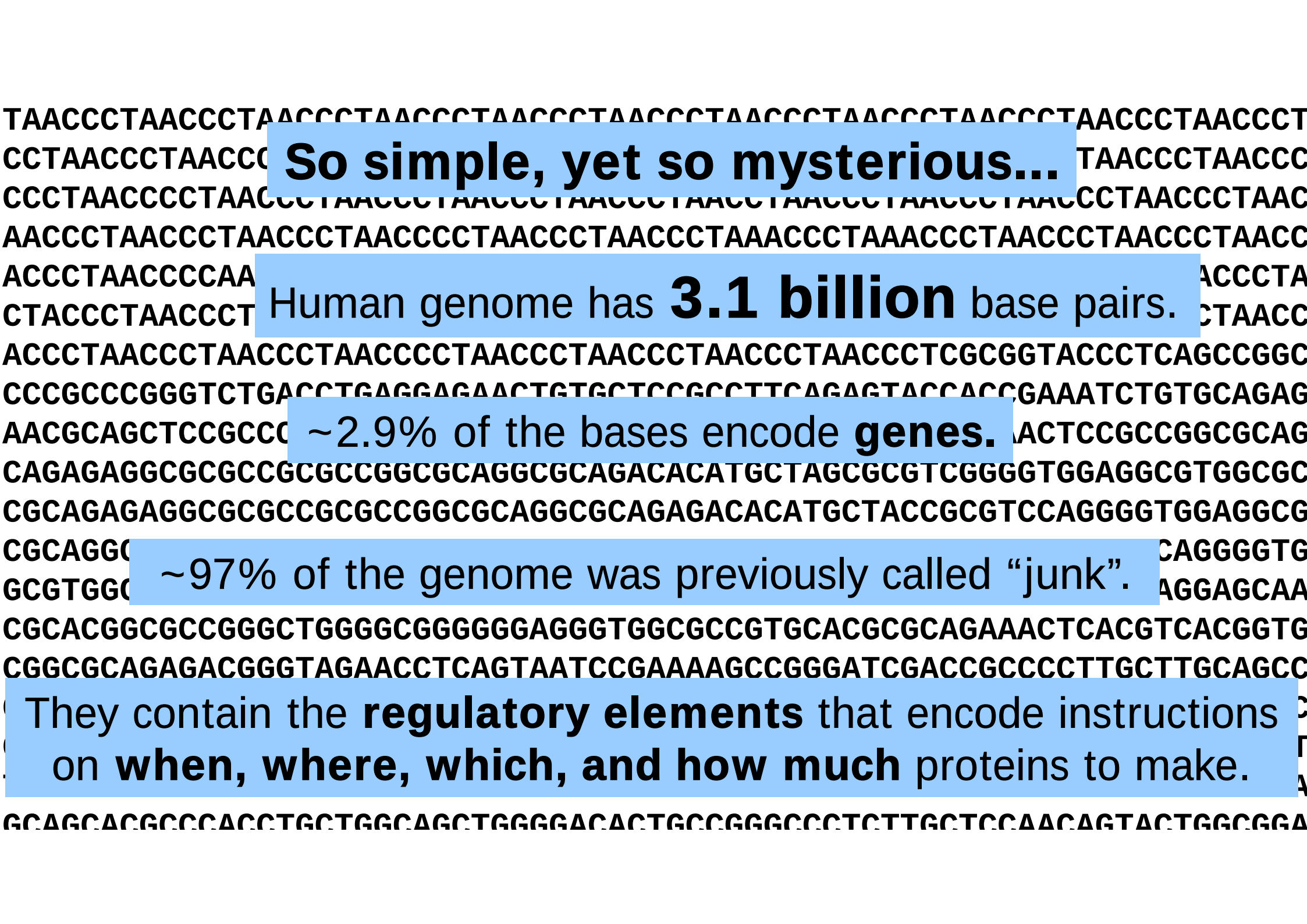


mitochondrial DNA

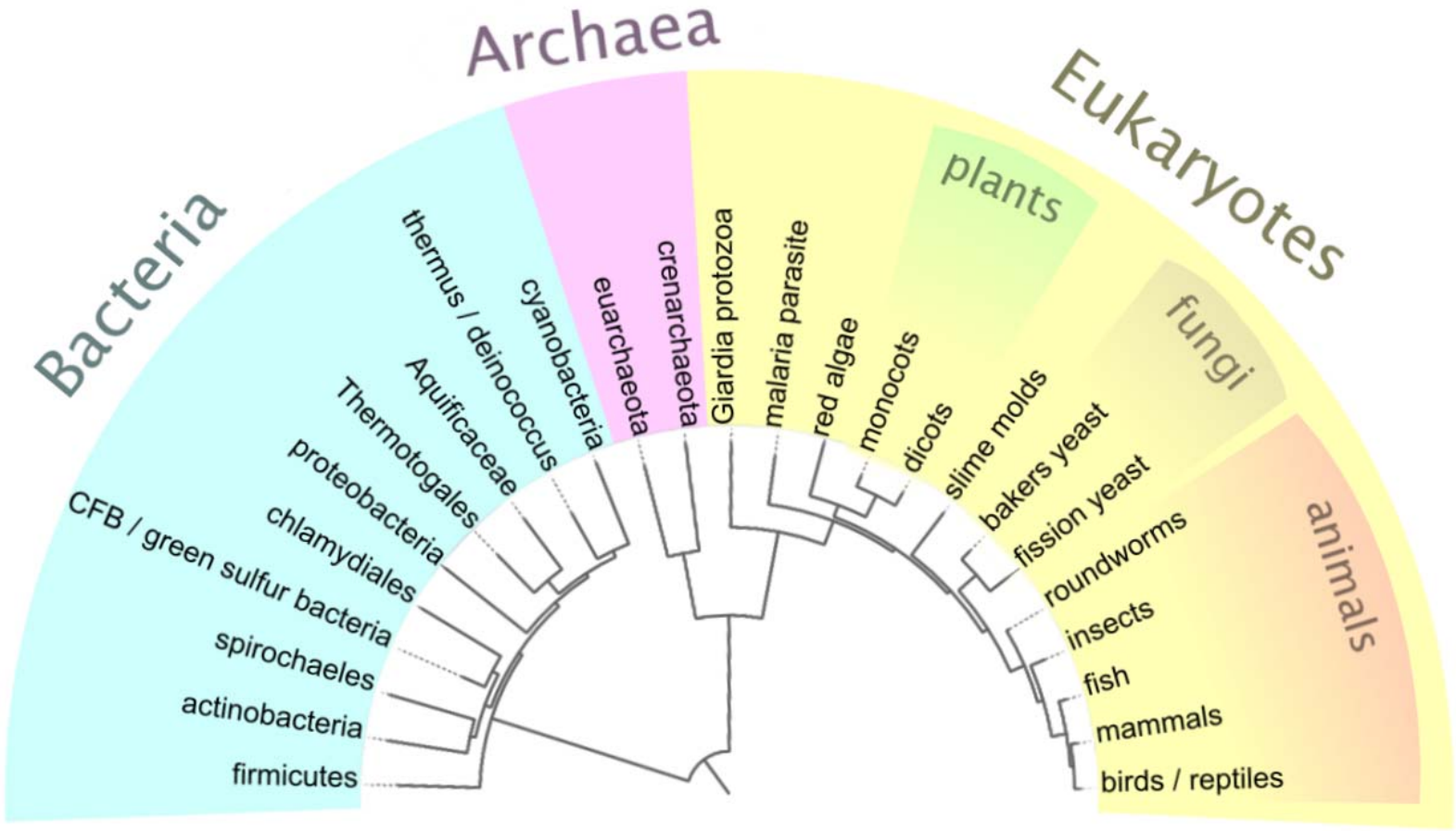
epigenetics

environments/nurture

chance

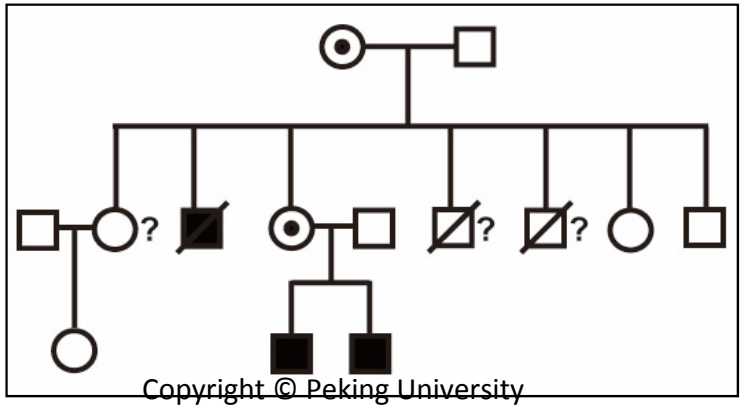

So simple, yet so mysterious...
Human genome has 3.1 billion base pairs.
~2.9% of the bases encode genes.
~97% of the genome was previously called “junk”.
They contain the regulatory elements that encode instructions on when, where, which, and how much proteins to make.

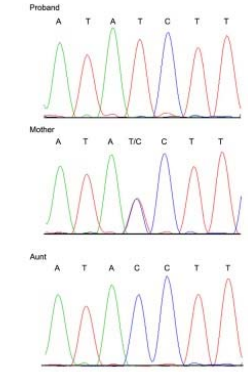
The universal code: Other species' genomes
The Tree of Life

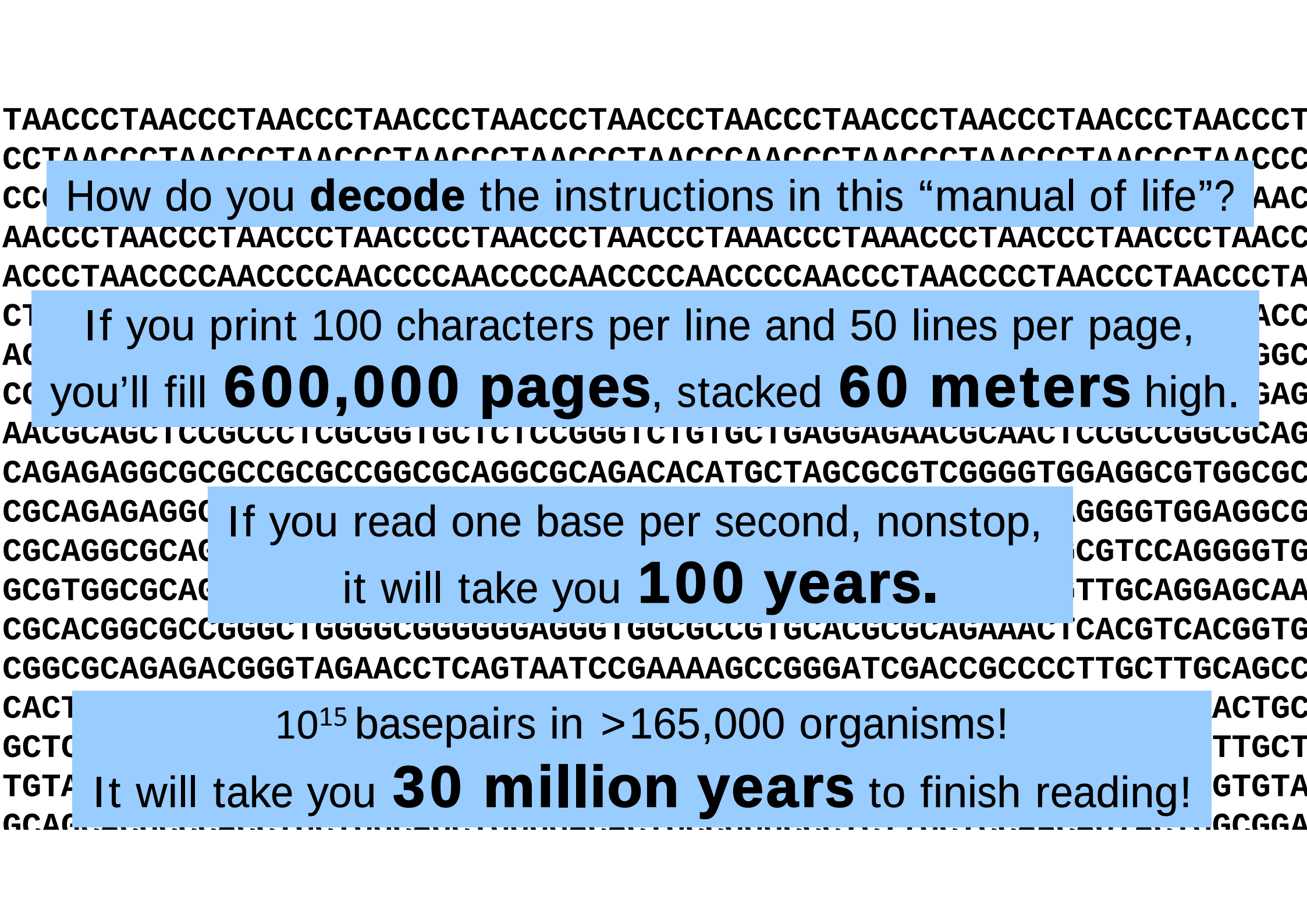


http://commons.wikimedia.org/wiki/File:Simplified_tree.png

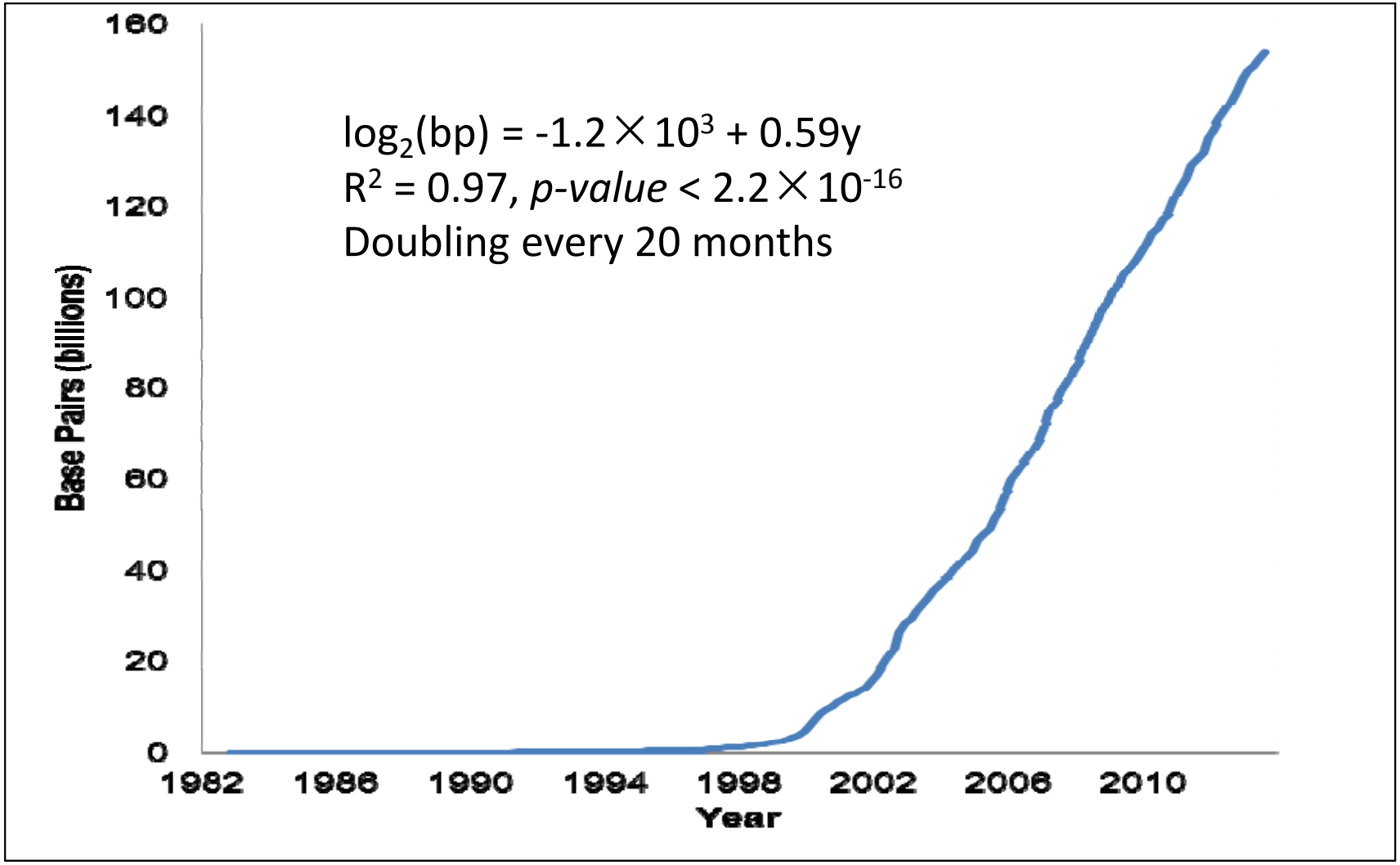
Human Genetic Variations





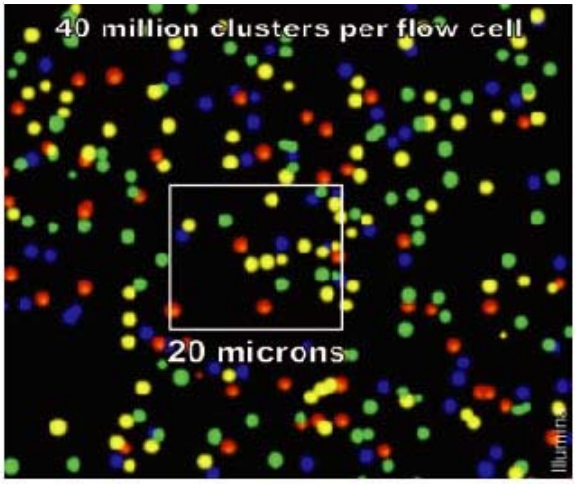

How do you **decode** the instructions in this “manual of life”?
If you print 100 characters per line and 50 lines per page,
you’ll fill **600,000 pages**, stacked **60 meters** high.
If you read one base per second, nonstop,
it will take you **100 years.**
10^{15} basepairs in >165,000 organisms!
It will take you **30 million years** to finish reading!

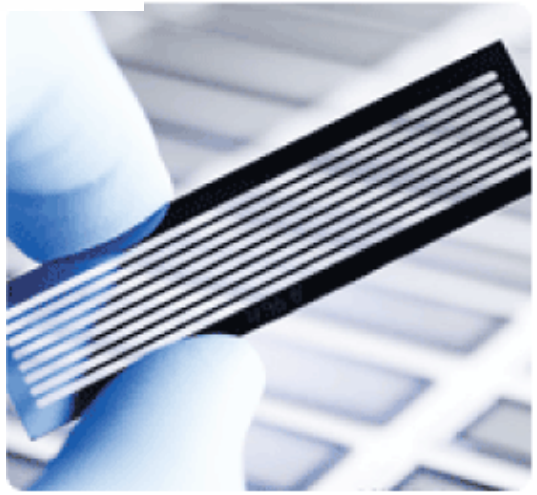
Genbank growth

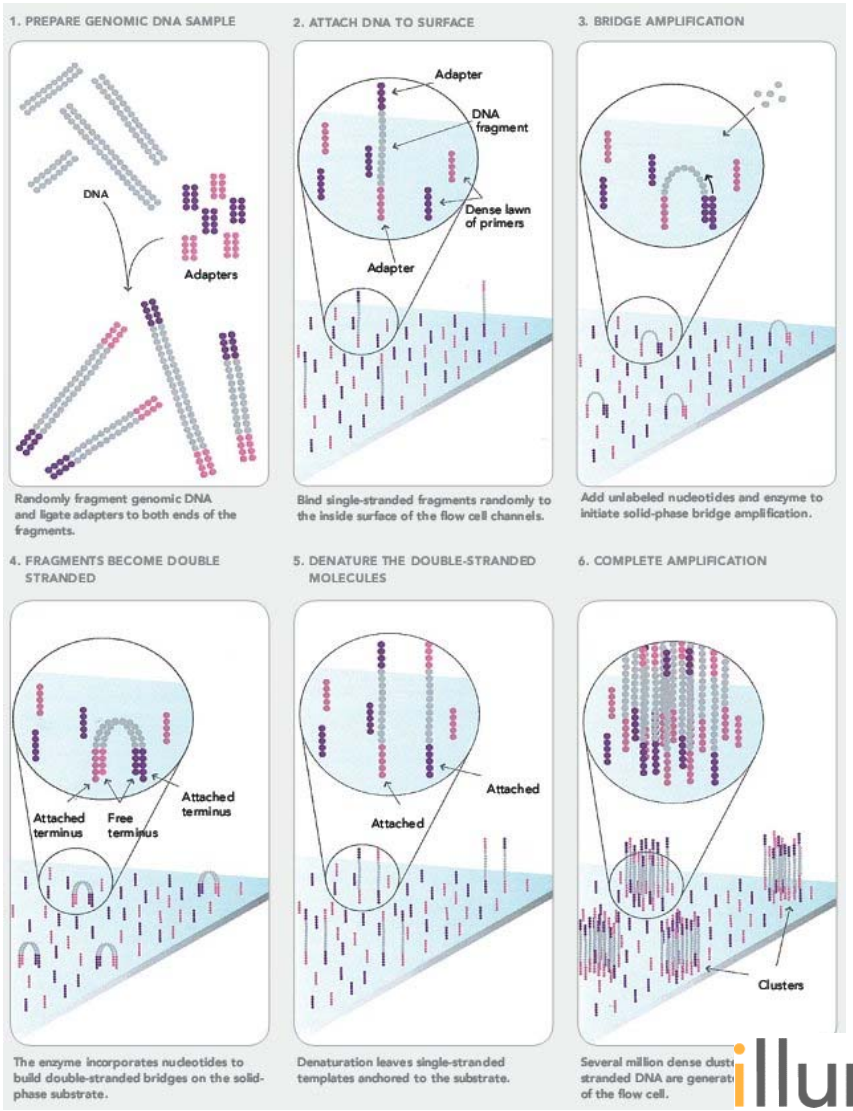


Data Source: ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt

Next-Generation Sequencing: Your genome, one day, $3000!

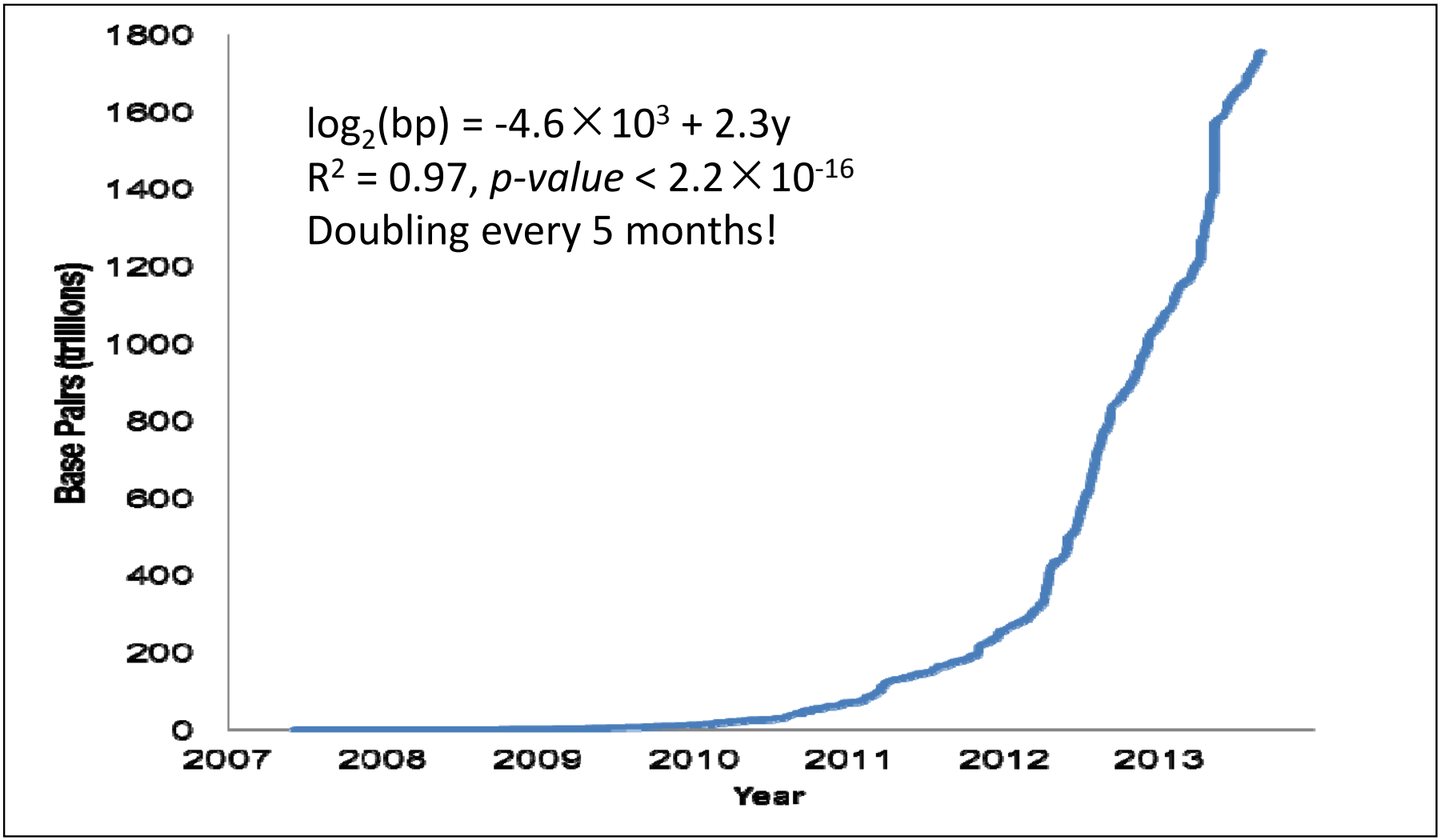




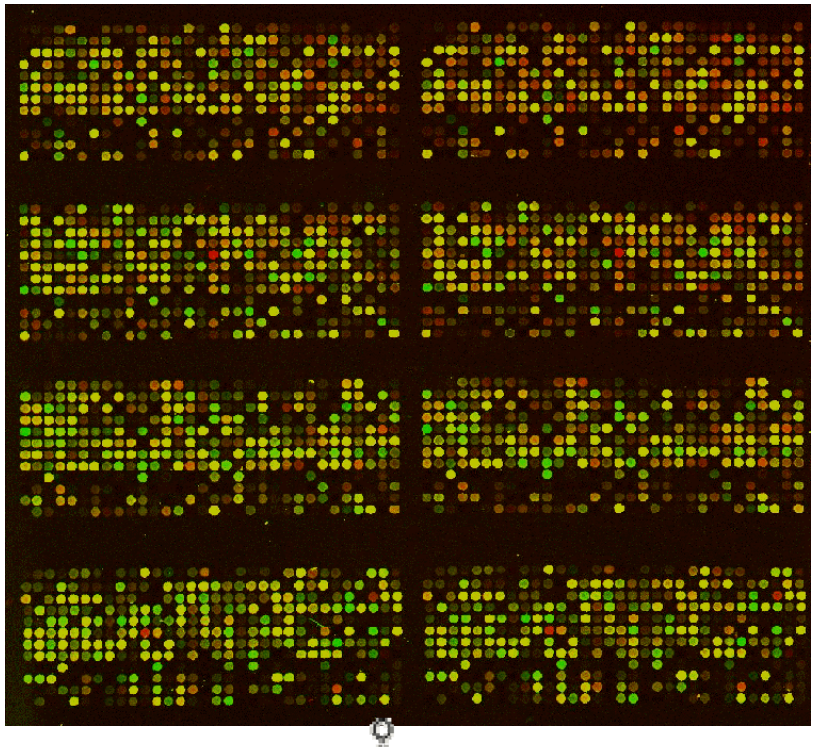


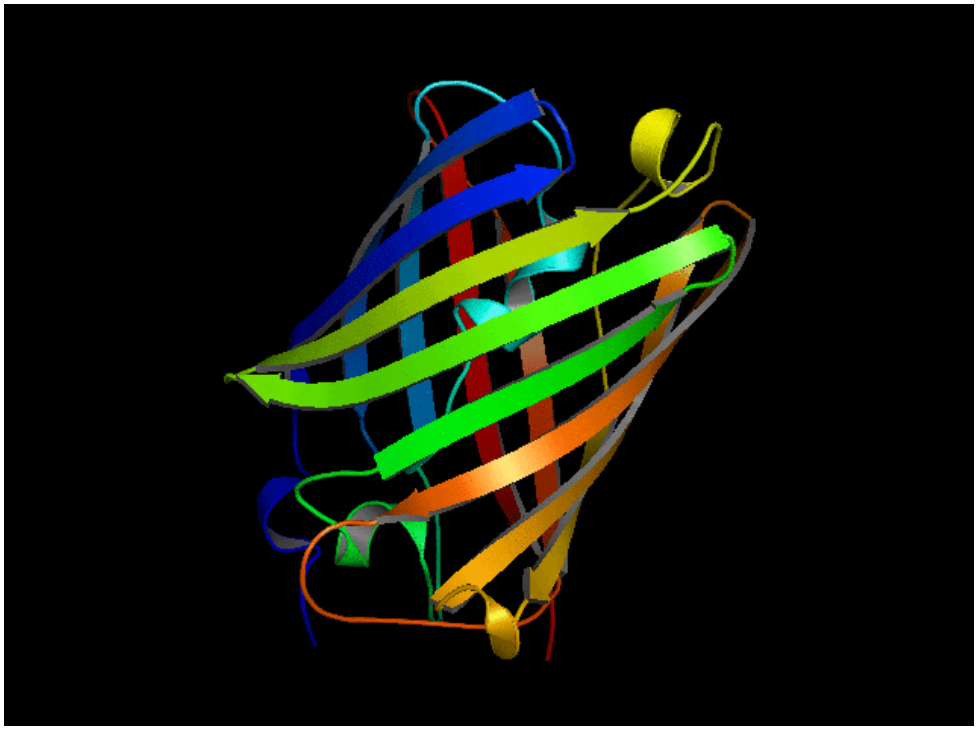
illumina®

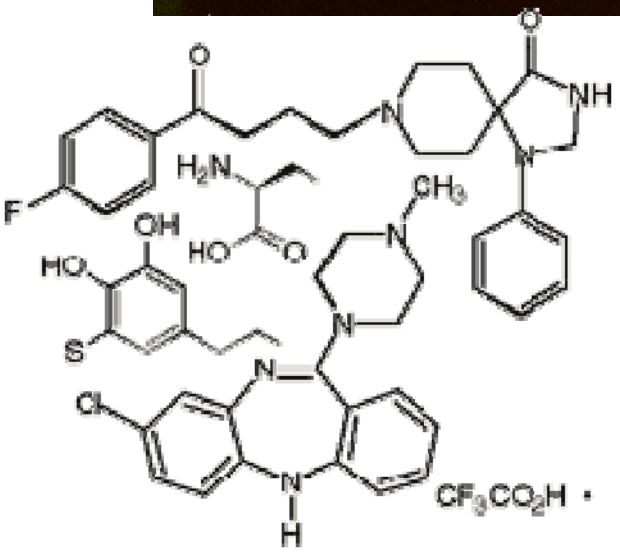
SRA Growth

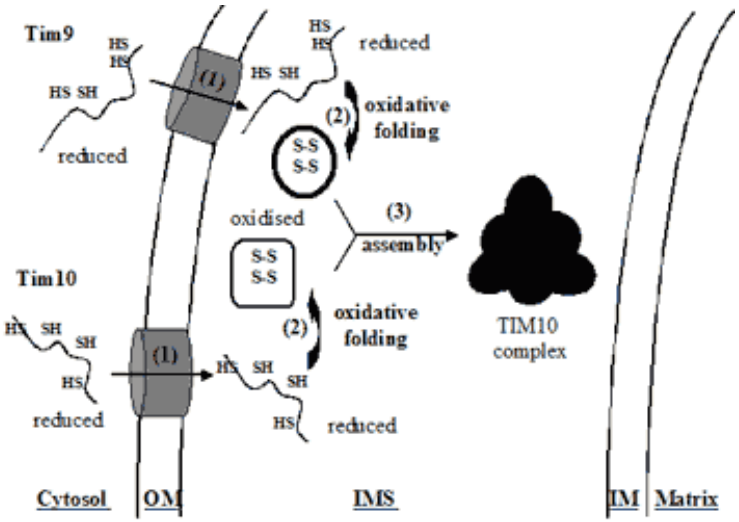


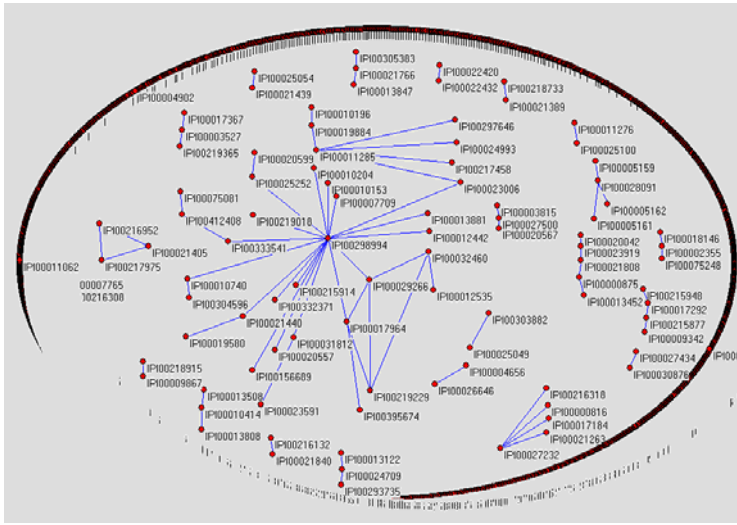
Data source: http://0-www.ncbi.nlm.nih.gov.elis.tmu.edu.tw/Traces/sra/

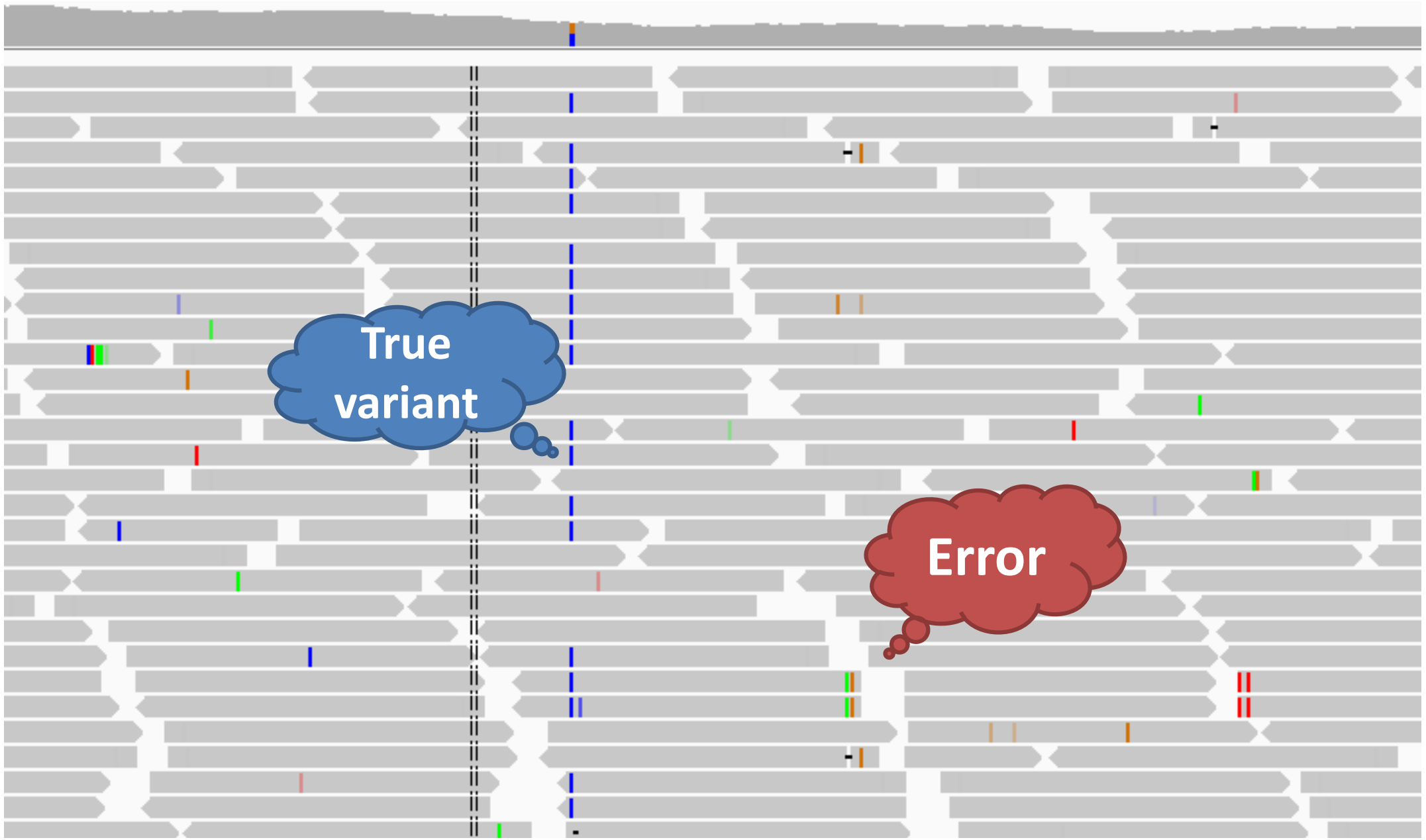





O
NH
F
H_2N
CH_3
OH
HO
HO
O
S
N
Cl
CF_3CO_2H ·

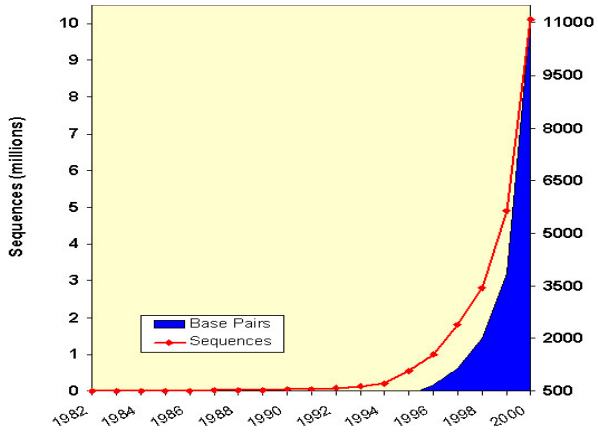

Tim9
HS
HS
HS SH
reduced
(1)
HS SH
reduced
(2)
oxidative
folding
S-S
S-S
oxidised
(3)
assembly
Tim10
HS
SH
SH
HS
reduced
(1)
HS SH
SH
HS
reduced
(2)
oxidative
folding
TIM10
complex
Cytosol
OM
IMS
IM
Matrix




True variant
Error

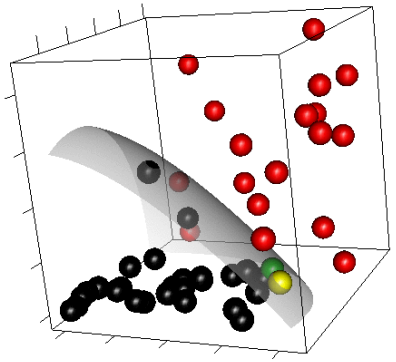
Opportunities and challenges hand-in-hand: the driving forces of bioinformatics

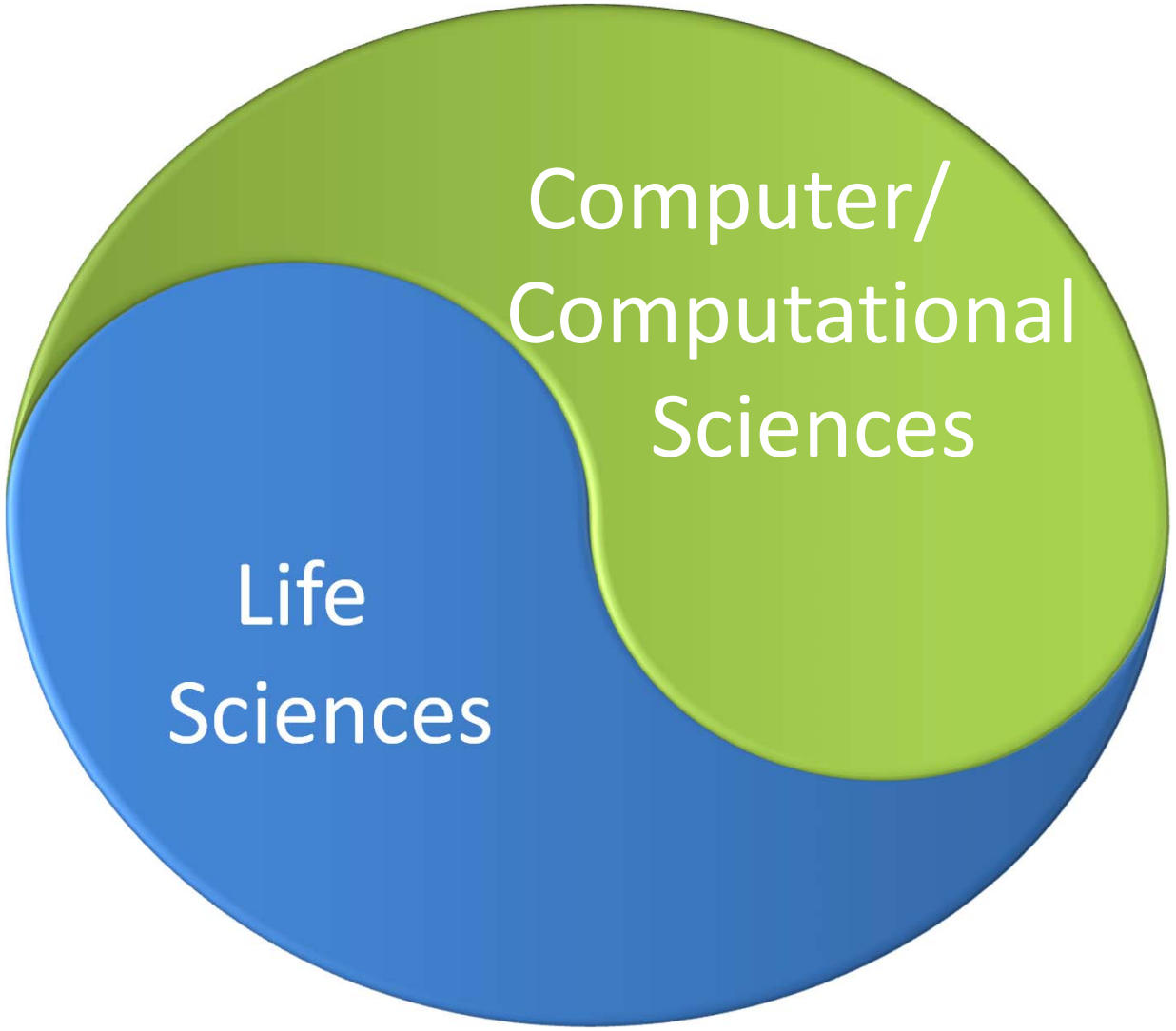
- High-throughput data
  - Huge amount
  - Explosive growth
  - Low signal-to-noise ratio
  - Multiple types



ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt

- Requirements for the methods
  - Data needs to be stored in efficient ontology-based database systems
  - The huge amount of data requires efficient algorithms
  - Exponential growth requires scalable methods
  - The low signal-to-noise ratio requires accurate methods
  - Multiple types of data require data integrative methods





Bioinformatics

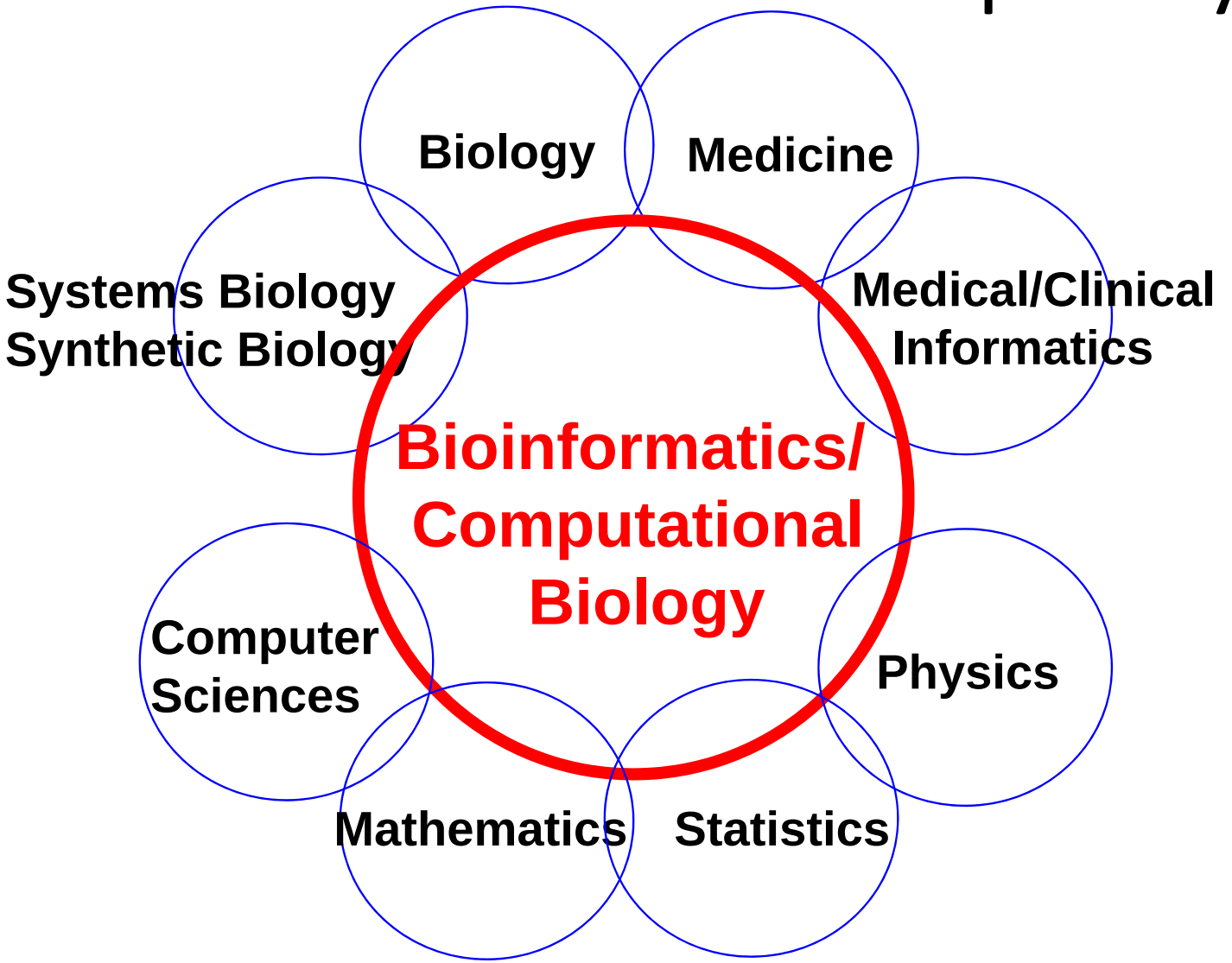
What is bioinformatics?

Bioinformatics: an interdisciplinary field that develops and applies computer and computational technologies to study biomedical questions

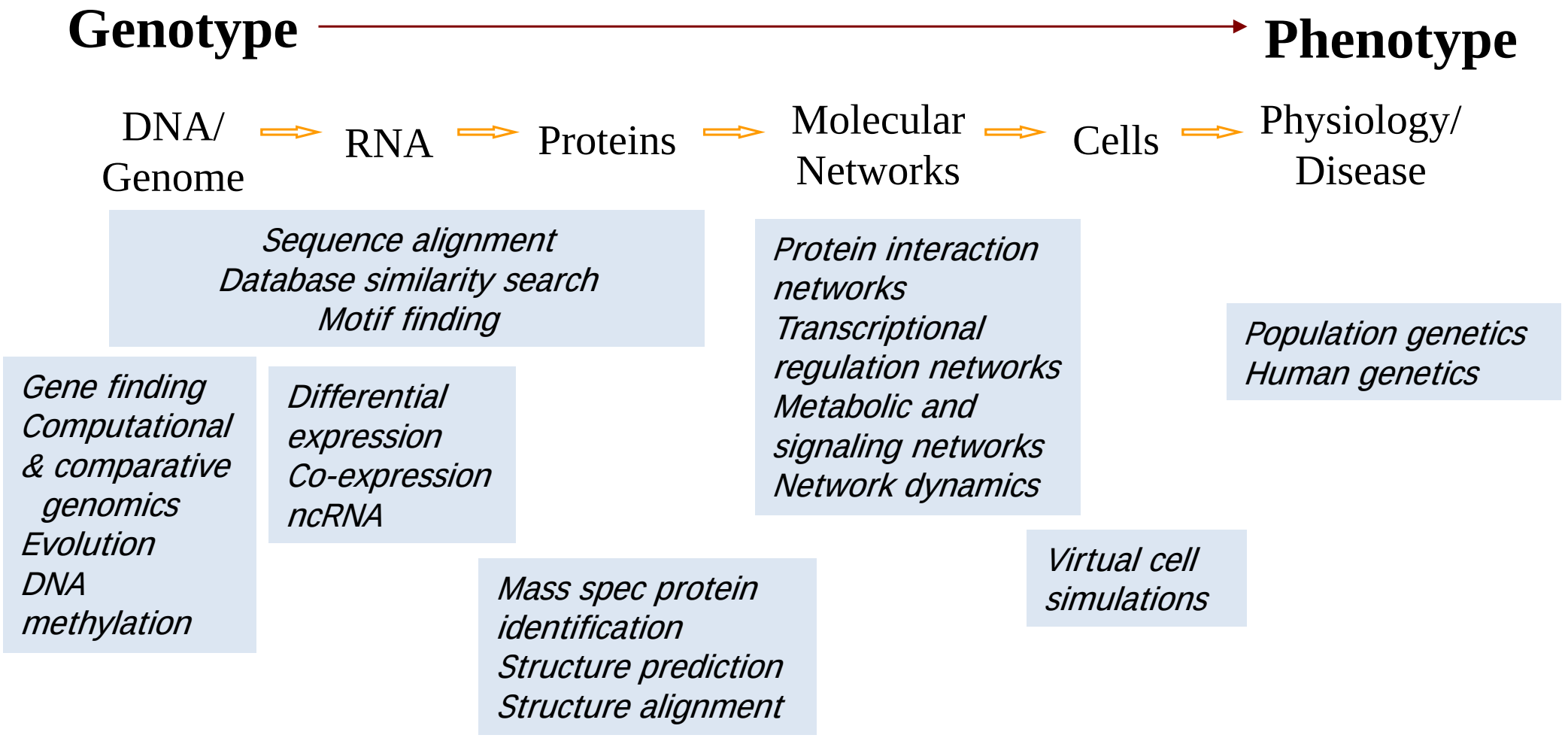
- As a technology, bioinformatics is a powerful technology to manage, query, and analyze big data in life sciences.

- As a methodology, bioinformatics is a top-down, holistic, data-driven, genome-wide, and systems approach that generates new hypotheses, finds new patterns, and discovers new functional elements.

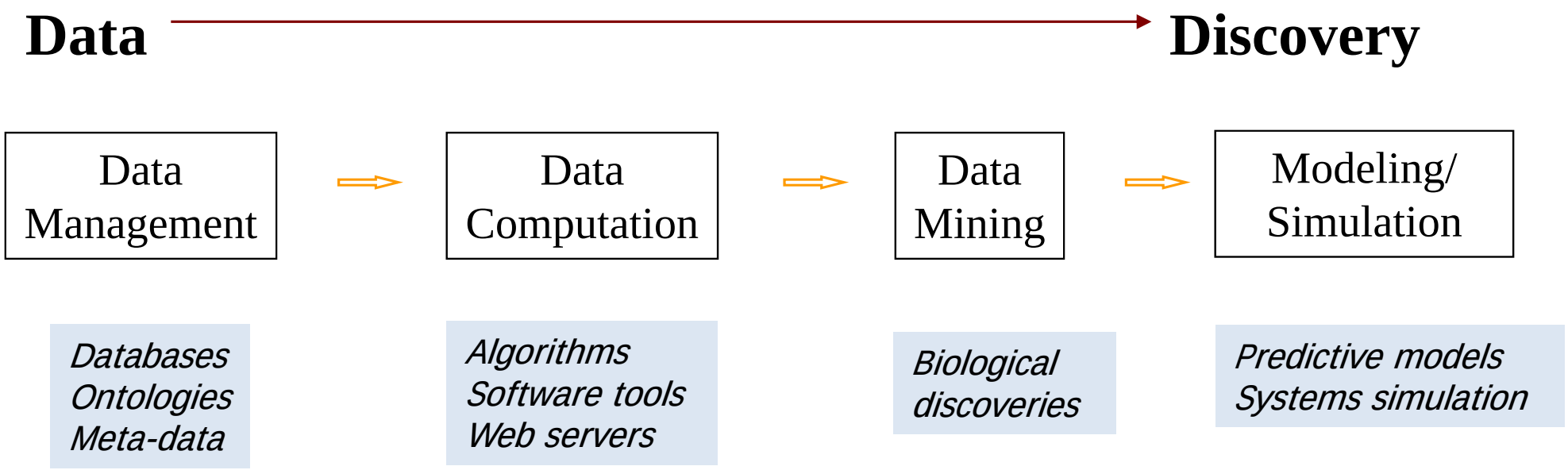
Bioinformatics is an interdisciplinary field



The Bio- in Bioinformatics



The –informatics in Bioinformatics



Summary Questions

- What is bioinformatics?
- Can you name some of the biomedical questions studied in bioinformatics?
- Can you name some of the technical questions studied in bioinformatics?
- Why are you interested in bioinformatics?

生物信息学：导论与方法
Bioinformatics: Introduction and Methods