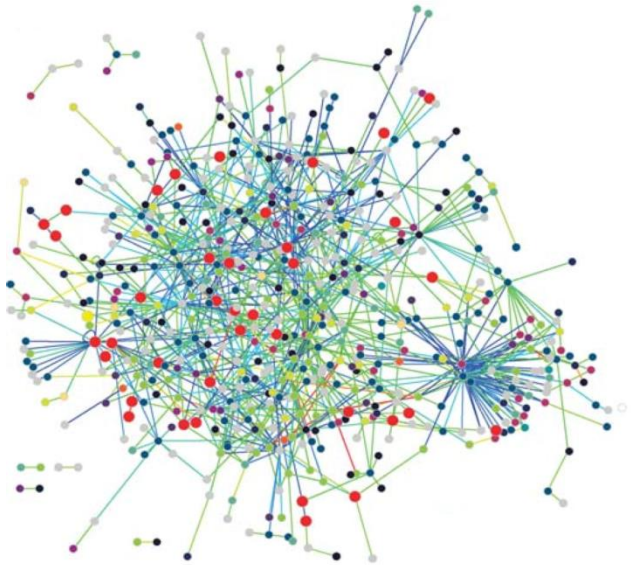
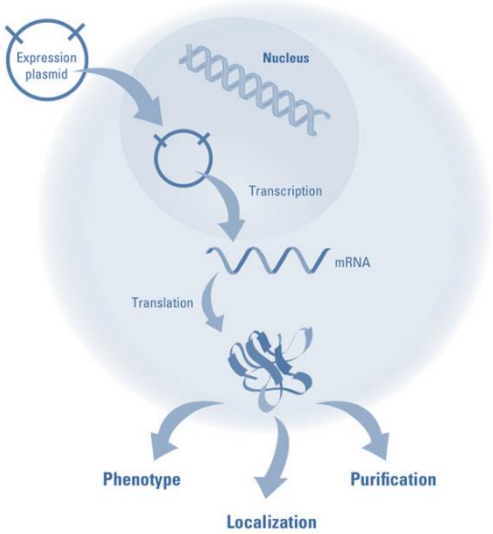




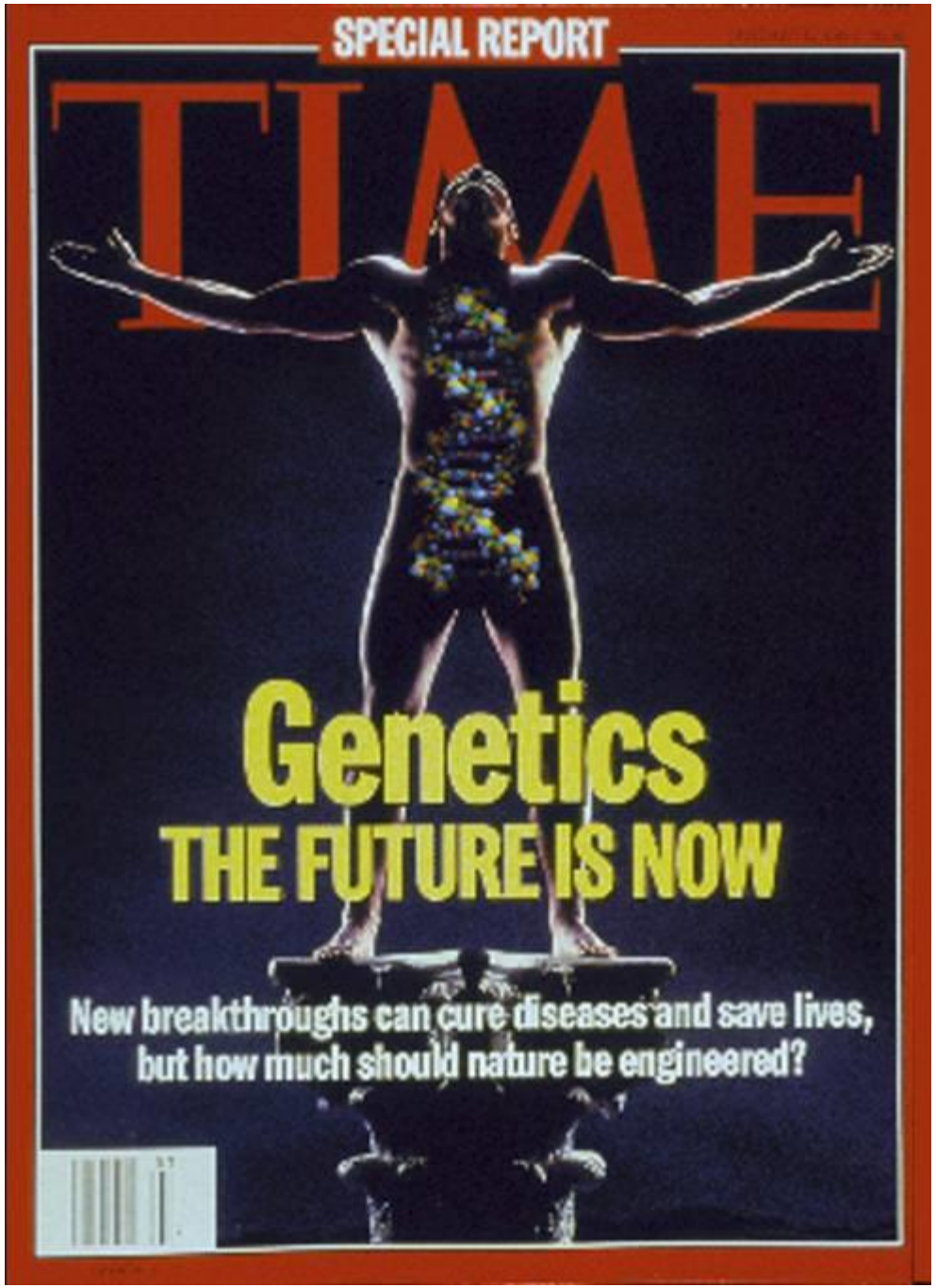
Introductory of Bioinformatics?

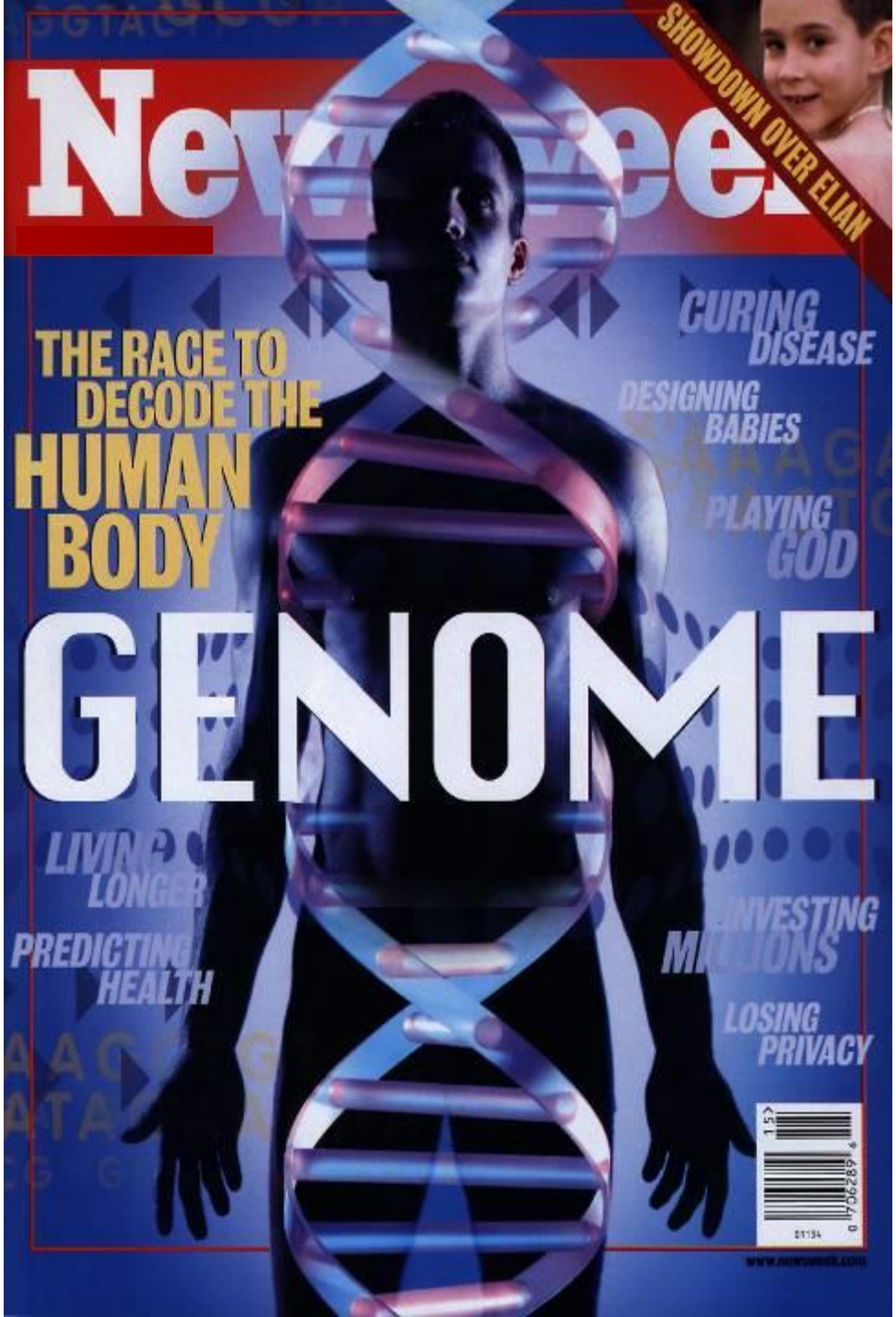
Dr . Amr El Kelish

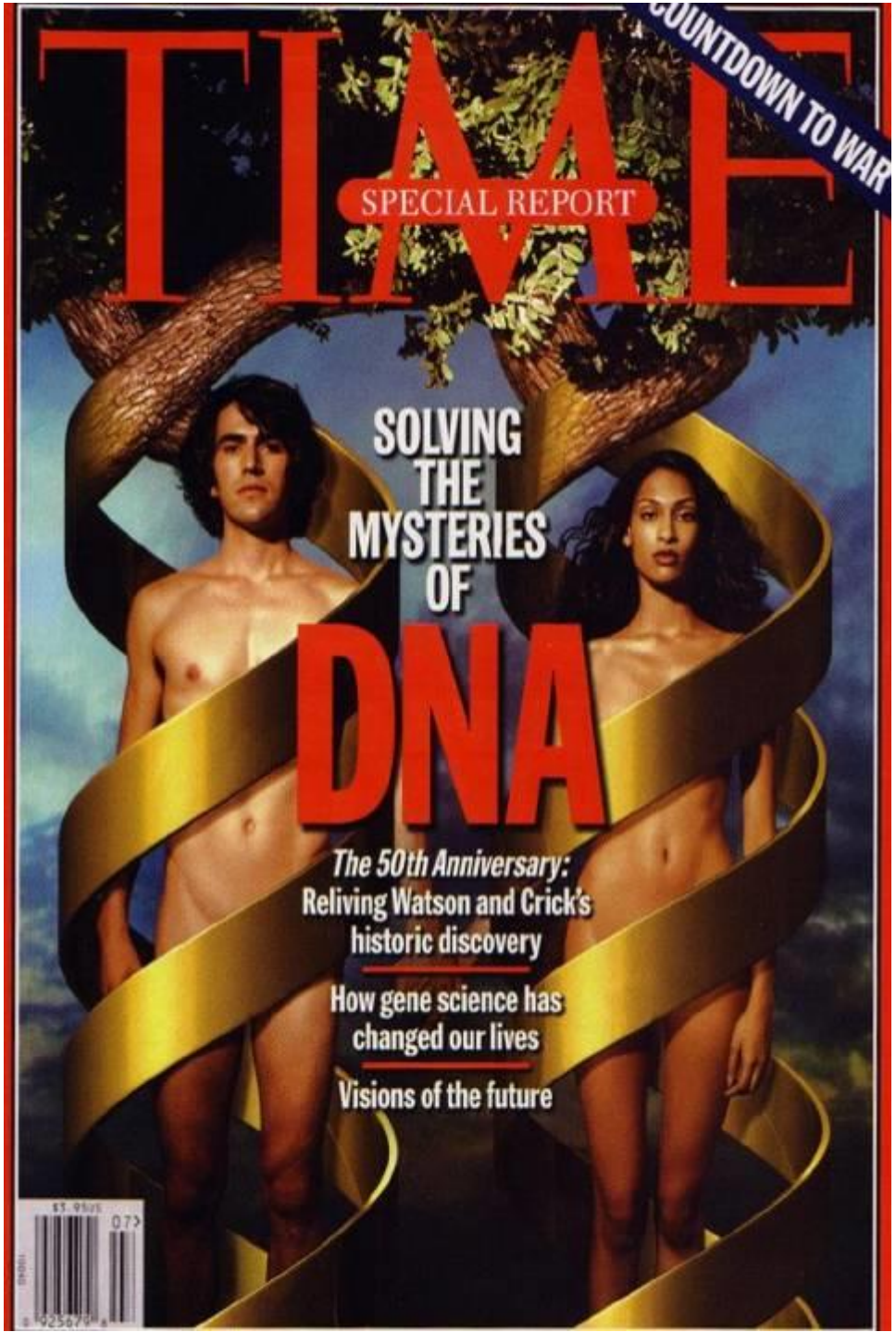
Lecture of Plant Physiology, Faculty Of Science, Suez Canal University

PhD. Technical University Of Munich, Germany

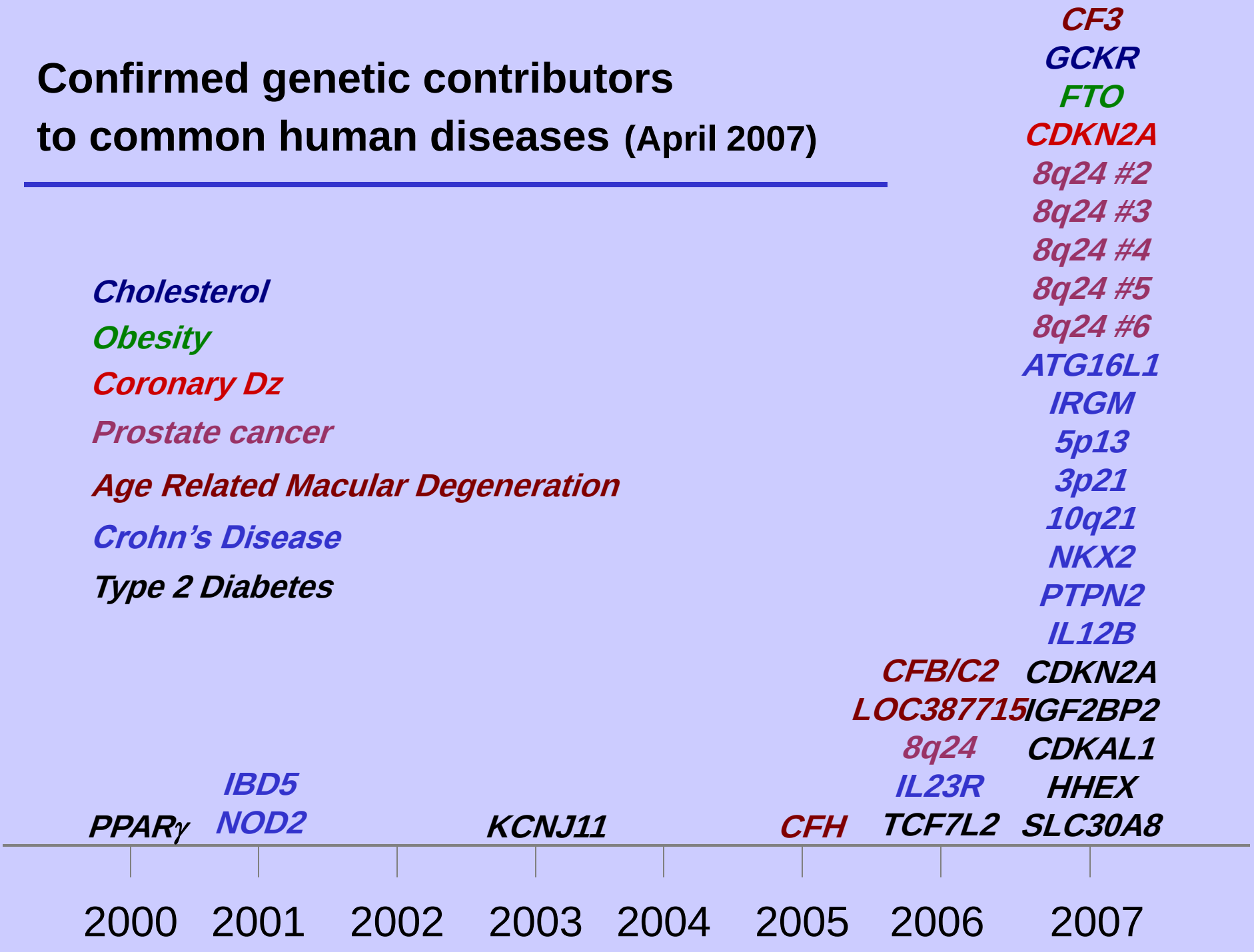
16-08-2015


SPECIAL REPORT
TIME
Genetics
THE FUTURE IS NOW
New breakthroughs can cure diseases and save lives,
but how much should nature be engineered?

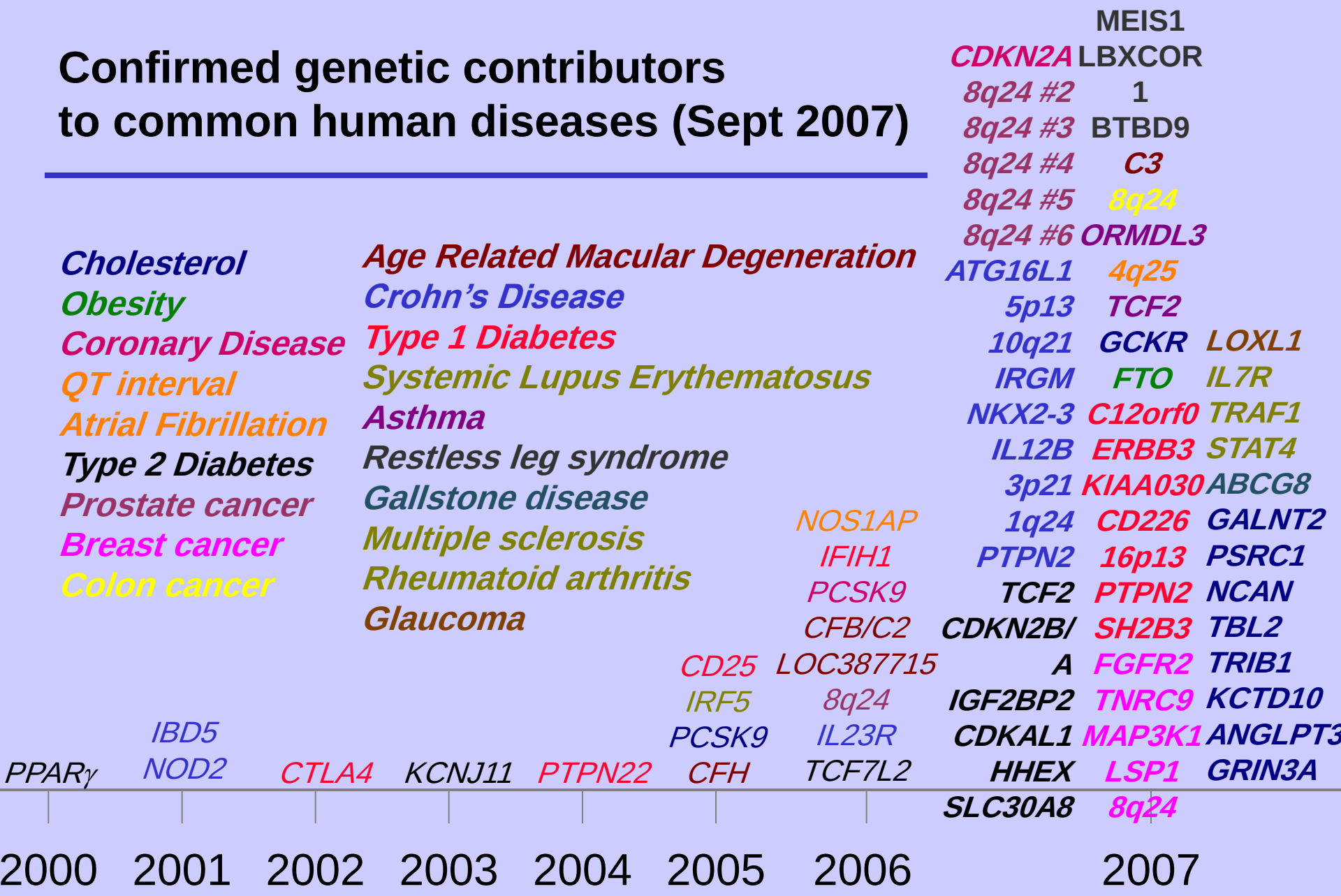

Newsweek
SHOWDOWN OVER ELIAN
THE RACE TO DECODE THE HUMAN BODY
GENOME
CURING DISEASE
DESIGNING BABIES
PLAYING GOD
LIVING LONGER
PREDICTING HEALTH
INVESTING
LOSING PRIVACY
www.newsweek.com


TIME
SPECIAL REPORT
COUNTDOWN TO WAR
SOLVING THE MYSTERIES OF
DNA
The 50th Anniversary:
Reliving Watson and Crick's historic discovery
How gene science has changed our lives
Visions of the future

Confirmed genetic contributors to common human diseases (April 2007)



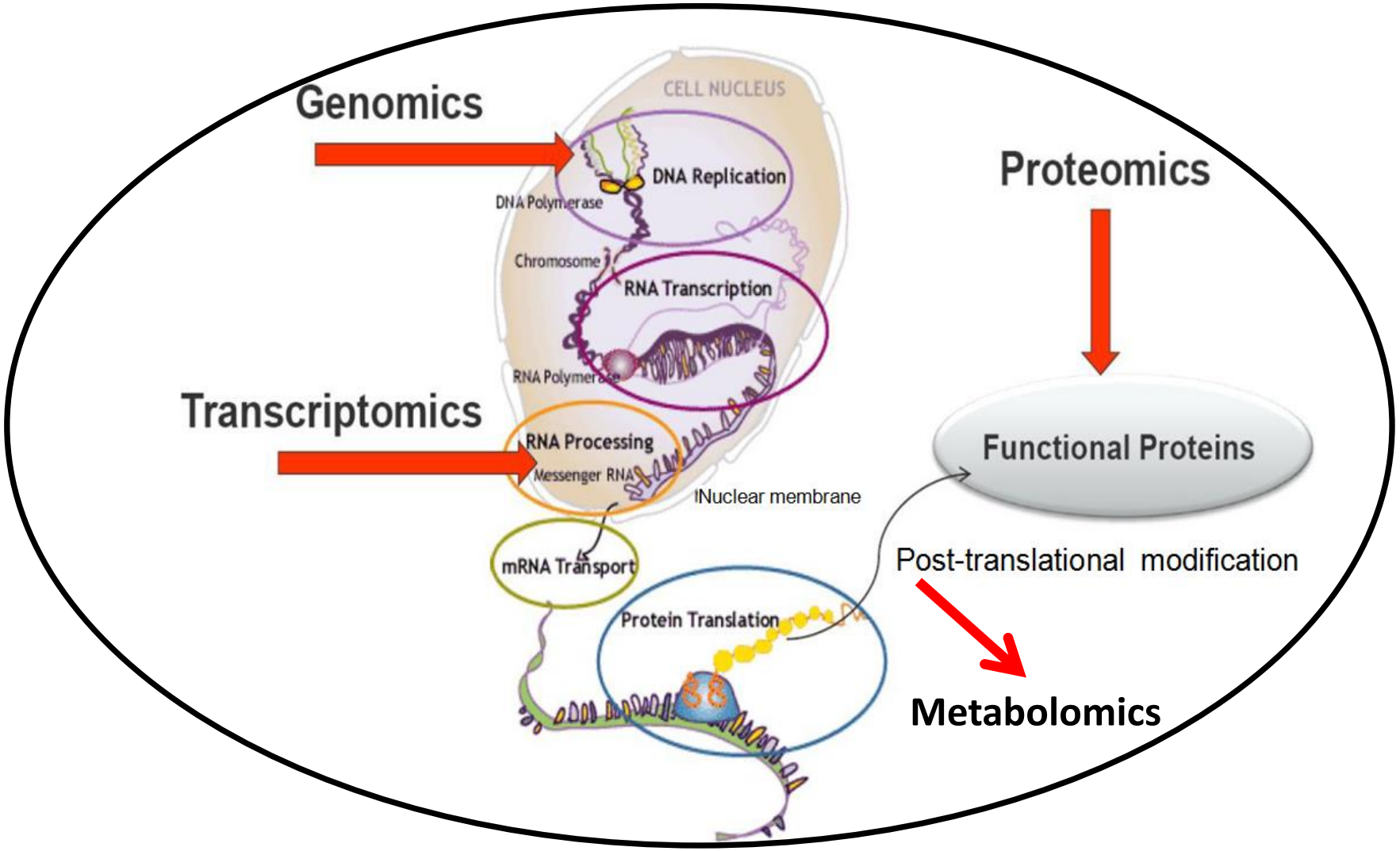
Confirmed genetic contributors to common human diseases (Sept 2007)

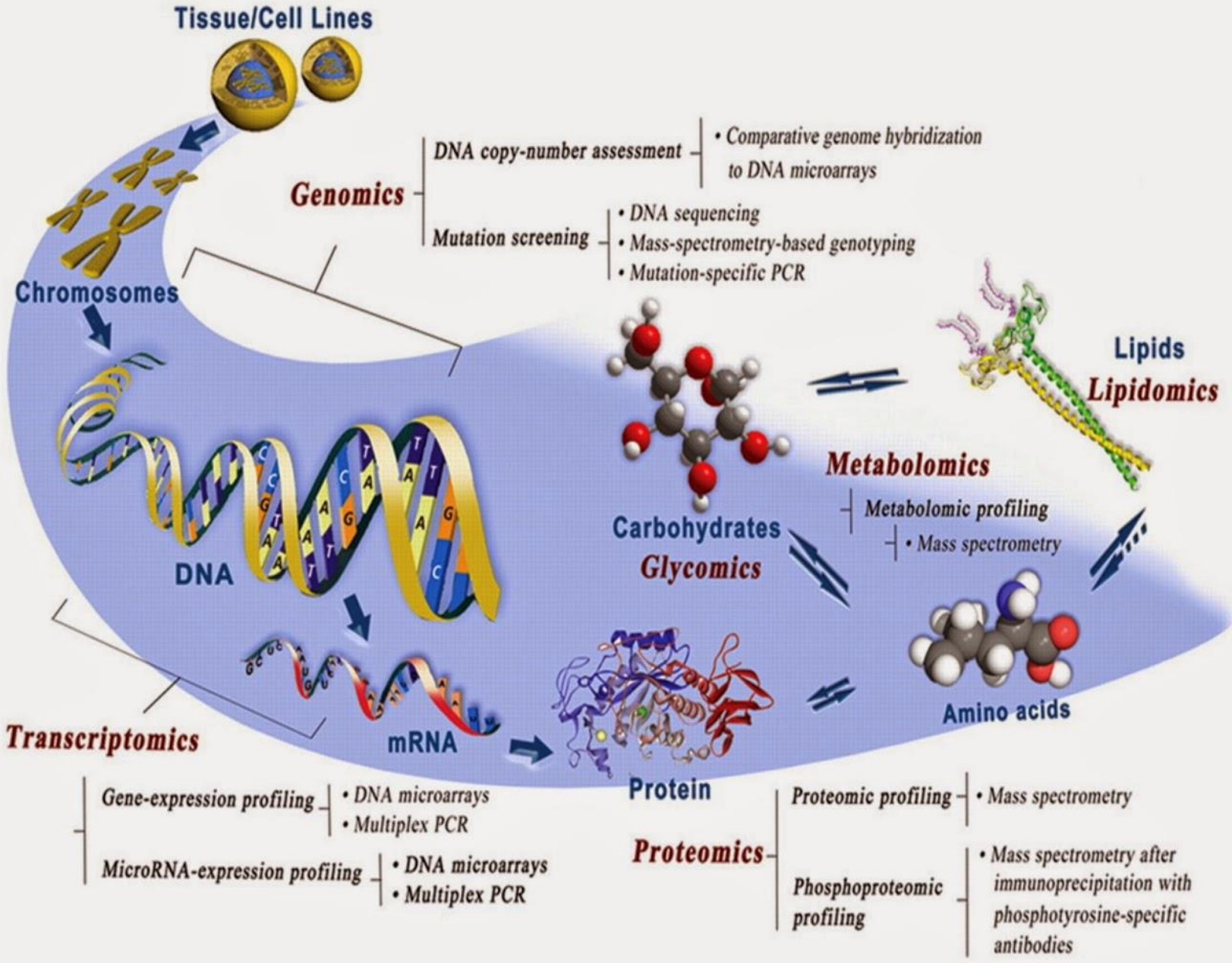


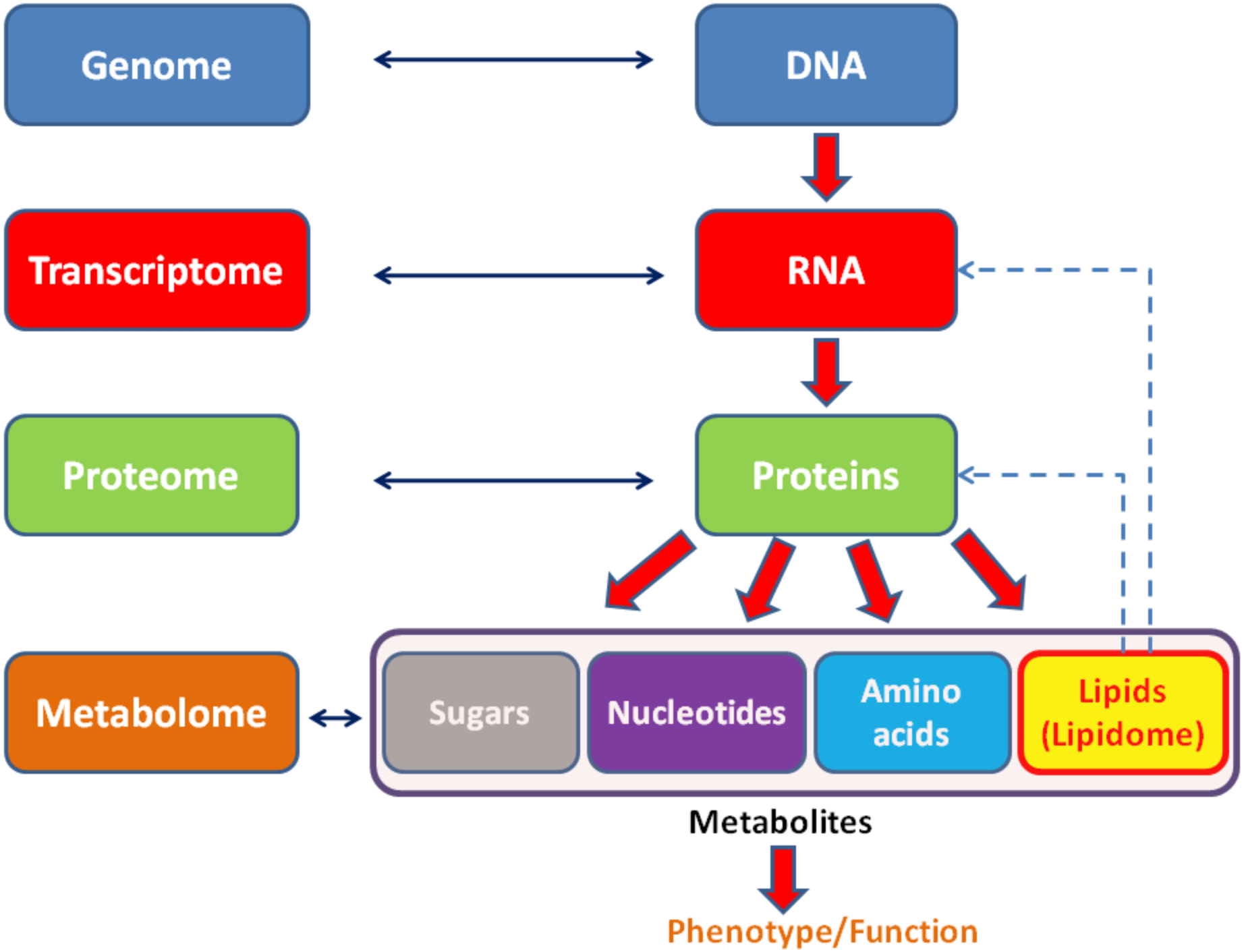
Omic' Technologies

- Holistic view of the molecules that make up a cell, tissue or organism.
- They include the universal detection of **genes** (**genomics**), **mRNA** (**transcriptomics**), **proteins** (**proteomics**) and **metabolites** (**metabolomics**) in a specific biological sample in a non-targeted and non-biased manner.
- This can also be referred to as high-dimensional biology; the integration of these techniques is called **Systems Biology**

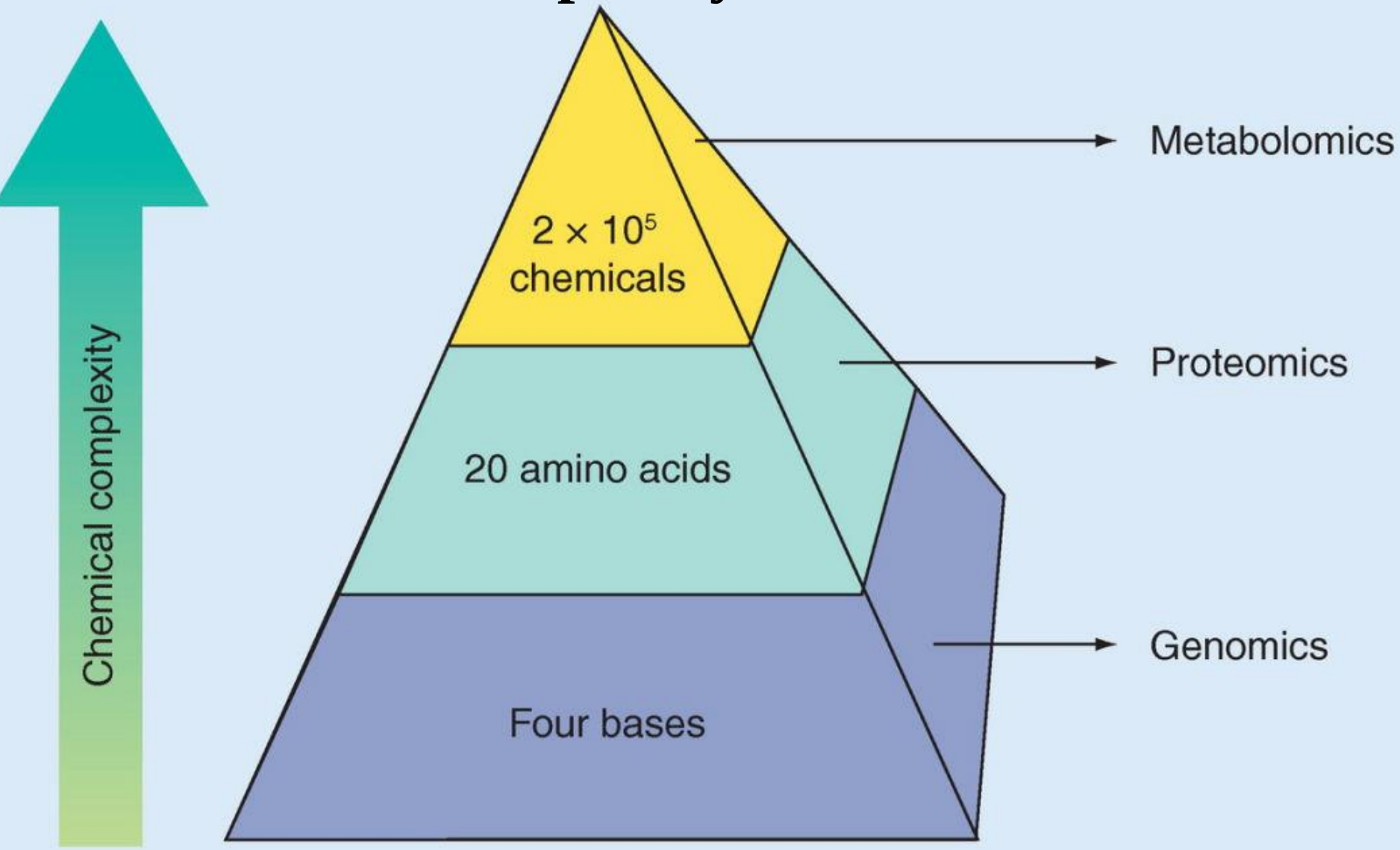
Central Dogma of life




Tissue/Cell Lines
Chromosomes
DNA
mRNA
Protein
Genomics
DNA copy-number assessment
• Comparative genome hybridization to DNA microarrays
Mutation screening
• DNA sequencing
• Mass-spectrometry-based genotyping
• Mutation-specific PCR
Transcriptomics
Gene-expression profiling
• DNA microarrays
• Multiplex PCR
MicroRNA-expression profiling
• DNA microarrays
• Multiplex PCR
Proteomics
Proteomic profiling
• Mass spectrometry
Phosphoproteomic profiling
• Mass spectrometry after immunoprecipitation with phosphotyrosine-specific antibodies
Carbohydrates
Glycomics
Lipids
Lipidomics
Metabolomics
Metabolomic profiling
• Mass spectrometry
Amino acids


Genome
DNA
Transcriptome
RNA
Proteome
Proteins
Metabolome
Sugars
Nucleotides
Amino acids
Lipids (Lipidome)
Metabolites
Phenotype/Function

Complexity level



Genomics

- It is the systematic study of an organism's genome.
- The genome is the total DNA of a cell or organism.
- The human genome contains **3.2 billion bases**
- They can reveal **abnormalities** such as chromosomal **insertions** and **deletions** or **abnormal** chromosomal numbers in a process called **comparative genomic hybridisation.**

Research area of Genomics

- **Functional genomics**

  Describe gene (and protein) functions and interactions**)**

- **Structural genomics**

  Describe the 3-dimensional structure of every protein encoded by a given genome**).**

- **Epigenomics**

  Chemical changes to **DNA** and **histone** proteins of an organism that can be passed down to an offspring

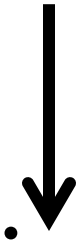
- **Metagenomics**

  Study of genetic material recovered directly from environment samples

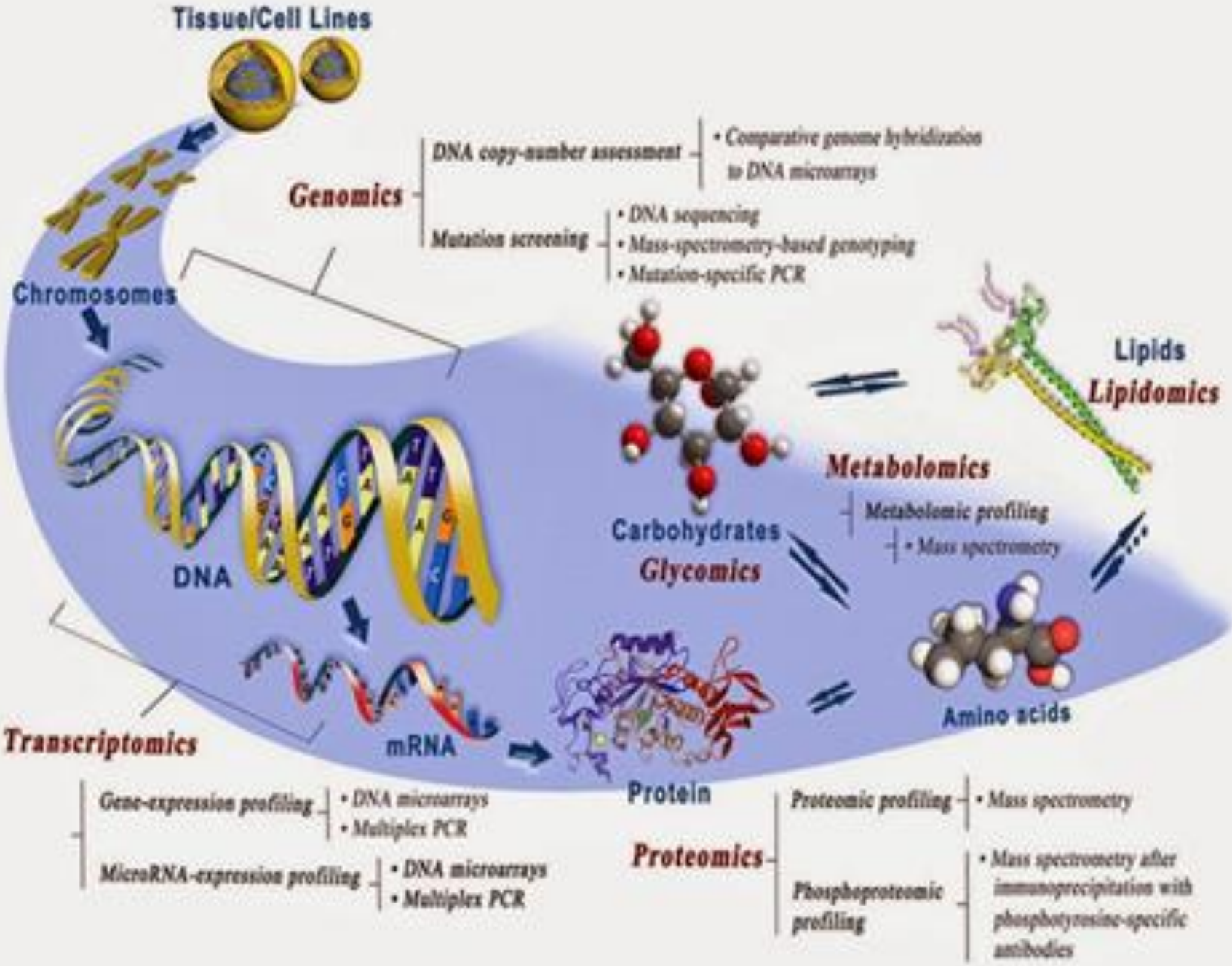
Importance of Genomics for Human

- All humans have 99.9% identical genetic makeup.
- The remaining 0.1% may provide useful information about diseases.
- **Goal** :

To show why some people get sick from certain infections and environmental changes while others do not.



Genetic Disorders


Tissue/Cell Lines
Chromosomes
DNA
mRNA
Protein
Carbohydrates
Glycomics
Lipids
Lipidomics
Metabolomics
Metabolomic profiling
• Mass spectrometry
Amino acids
Genomics
DNA copy-number assessment
• Comparative genome hybridization to DNA microarrays
Mutation screening
• DNA sequencing
• Mass-spectrometry-based genotyping
• Mutation-specific PCR
Transcriptomics
Gene-expression profiling
• DNA microarrays
• Multiplex PCR
MicroRNA-expression profiling
• DNA microarrays
• Multiplex PCR
Proteomics
Proteomic profiling
• Mass spectrometry
Phosphoproteomic profiling
• Mass spectrometry after immunoprecipitation with phosphotyrosine-specific antibodies

Transcriptomics

Specific gene expression

Genes and signatures determined by particular genetic, epigenetic regulatory factors, environmental exposures

Exploratory approaches

E.g global gene expression in tumors versus healthy tissues, differential responses to distinct environmental exposures

Disease etiology and classification

Patterns/signatures rather than single markers can improve knowledge about etiology and diagnosis

Biology

- Development and Morphology (juveniles vs adults)
- Interactions between organisms (antagonistic, mutualistic)
- Interactions between organisms and their environments

(temperature, radiation, draught, toxins and heavy metals)

- Evolution (within- and between species variation)
- Functional analyses (wild type vs mutant)

Medicine

- Disease-associated expression patterns (diagnosis)
- Cell-cycle monitoring (cancer research)
- Treatment-induced expression pattern (drug development and response)

Proteomics

- Set of all expressed proteins in a cell, tissue or organism.
- Proteomics aims to characterize information flow within the cell and the organism, through protein pathways and networks.
- The proteome is a dynamic reflection of both **genes** and the **environment** and is thought to hold special promise for **biomarker** discovery.

Tumor markers

- **Diagnostic markers:** **detection of malignant disease**

  Sensitivity = **detected positives / real positives**

  Specificity = **detected negatives / real negatives**

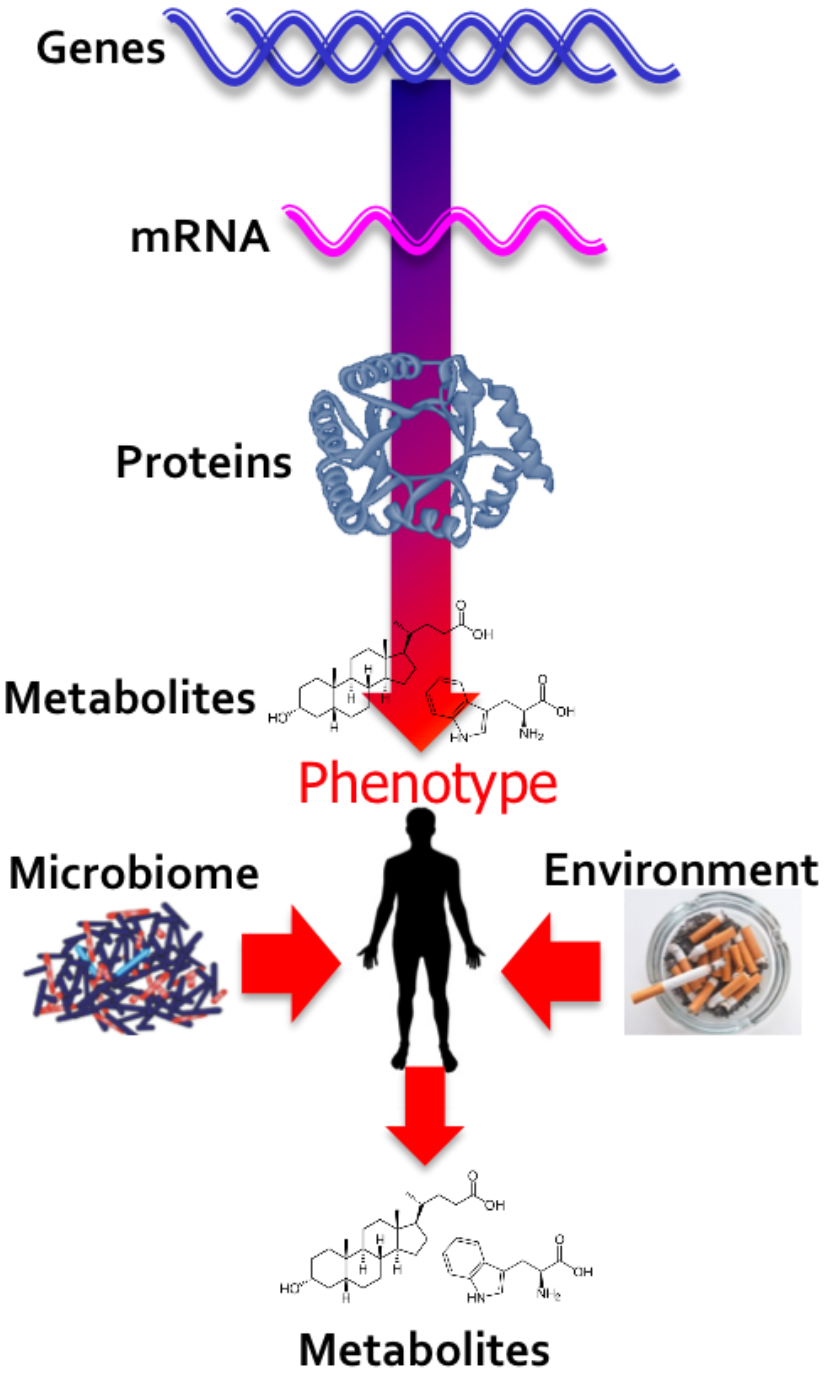
- **Prognostic markers:** **malignant potential, disease recurrence**

- **Predictive markers:** **response to different therapies**

- **Positioning markers:** **positional information (e.g. for surgery)**

Metabolomics

- Study of global metabolite profiles in a system (cell, tissue or organism) under a given set of conditions.
- Additionally, as the downstream product, the metabolome is closest to the **phenotype** of the biological system studied.
- **Metabolom** is more **physically** and **chemically complex** than the other ‘omes’.


Genes
mRNA
Proteins
Metabolites
Phenotype
Microbiome
Environment
Metabolites

Systems Biology

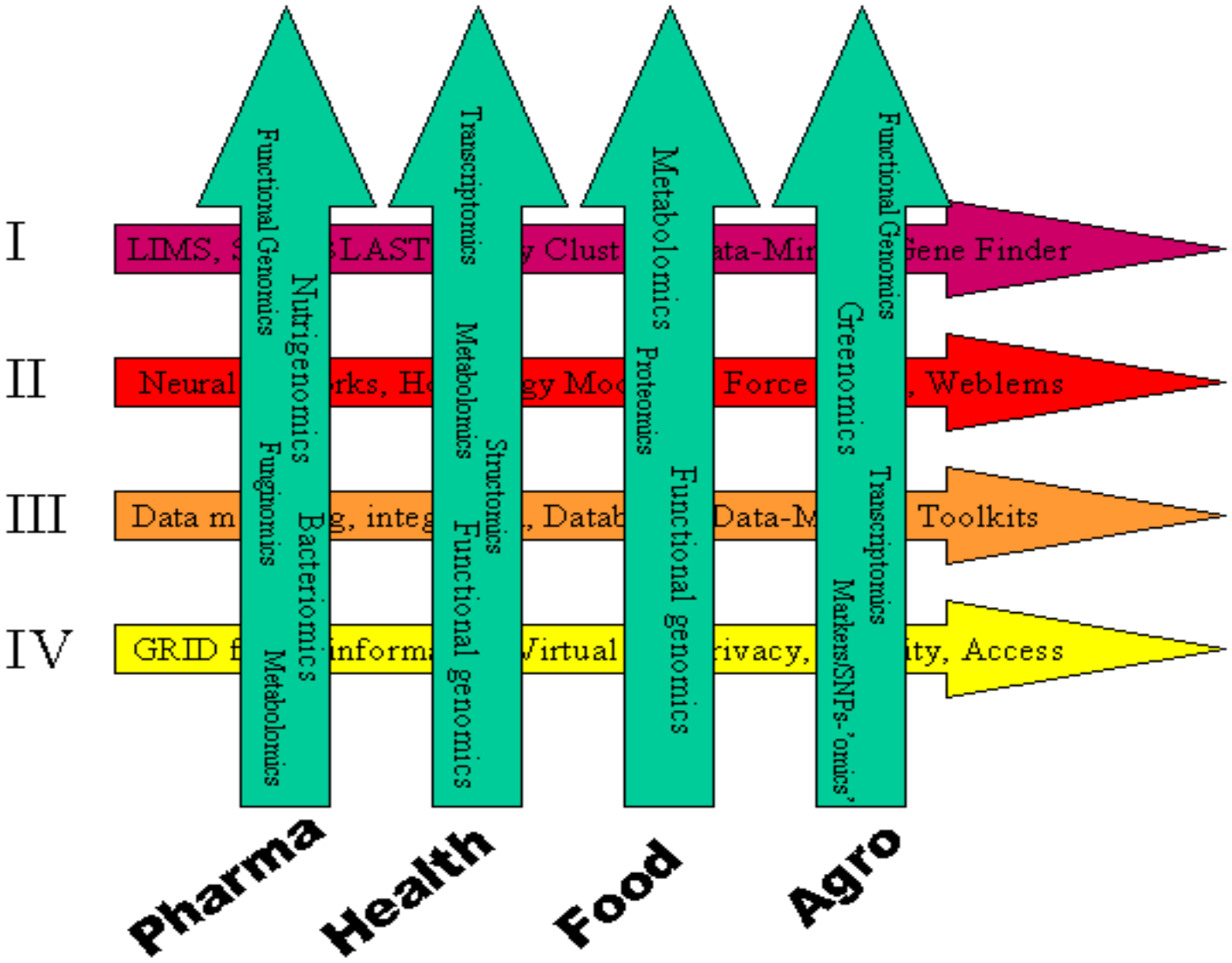
- Biological research focusing on the systematic study of complex interactions in biological systems using integration models.

- The ultimate aim is to understand whole systems, e.g. complex **cellular pathways**, by studying the effect of altered external factors on the **genome**, **transcriptome**, **proteome** and **metabolome** simultaneously

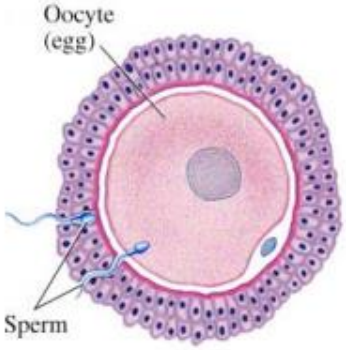
Why Omics ?

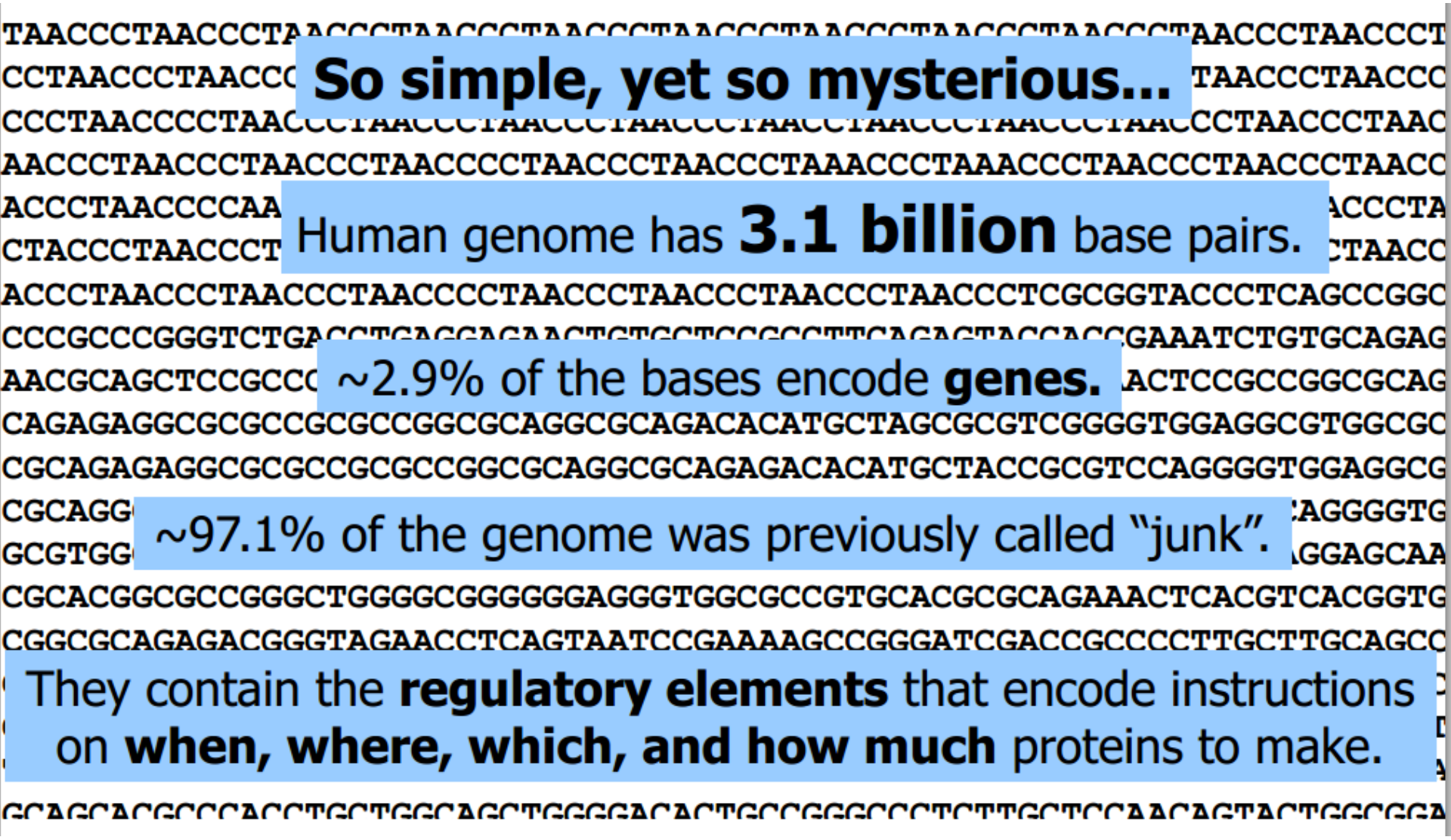
- The basic aspect of these approaches is that a **complex system** can be understood more thoroughly if considered as a whole.
- Omic technology can be applied not only for the greater understanding of **normal physiological** processes but also in disease processes where they play a role in **screening**, **diagnosis** and **prognosis** as well as aiding our understanding of the **aetiology** of diseases.

Omics impact

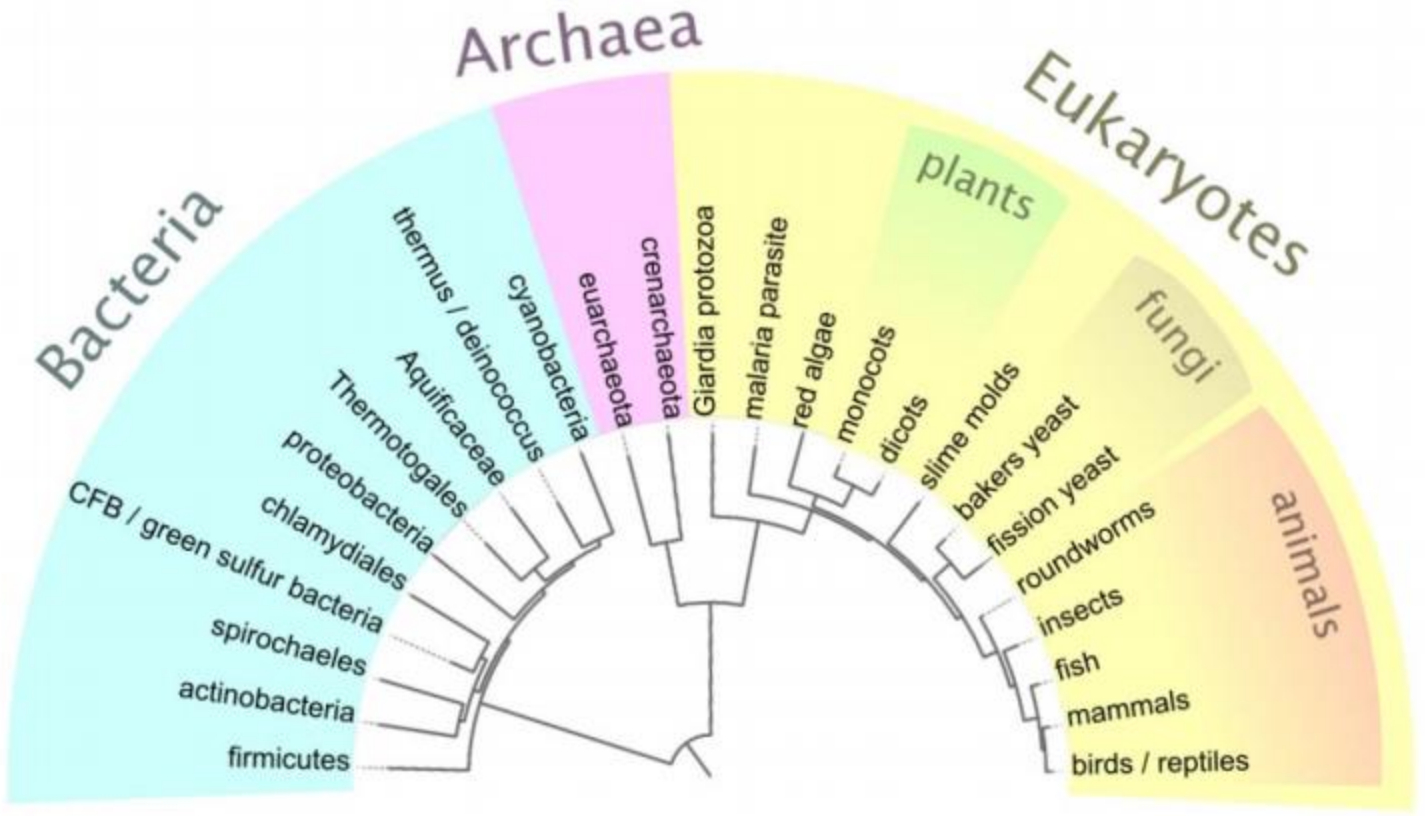


The miracle of life



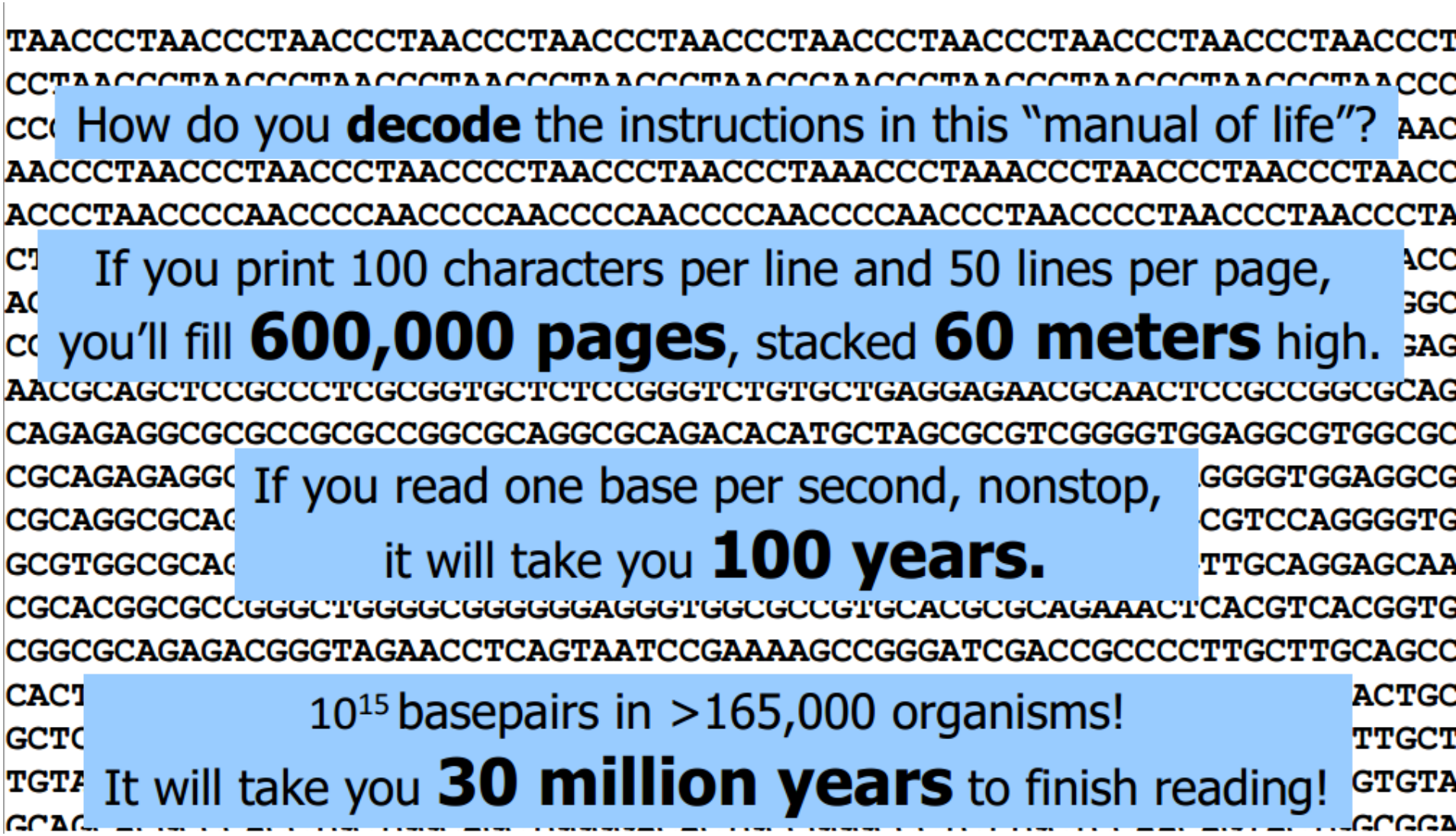

So simple, yet so mysterious...
Human genome has 3.1 billion base pairs.
~2.9% of the bases encode genes.
~97.1% of the genome was previously called "junk".
They contain the regulatory elements that encode instructions on when, where, which, and how much proteins to make.

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

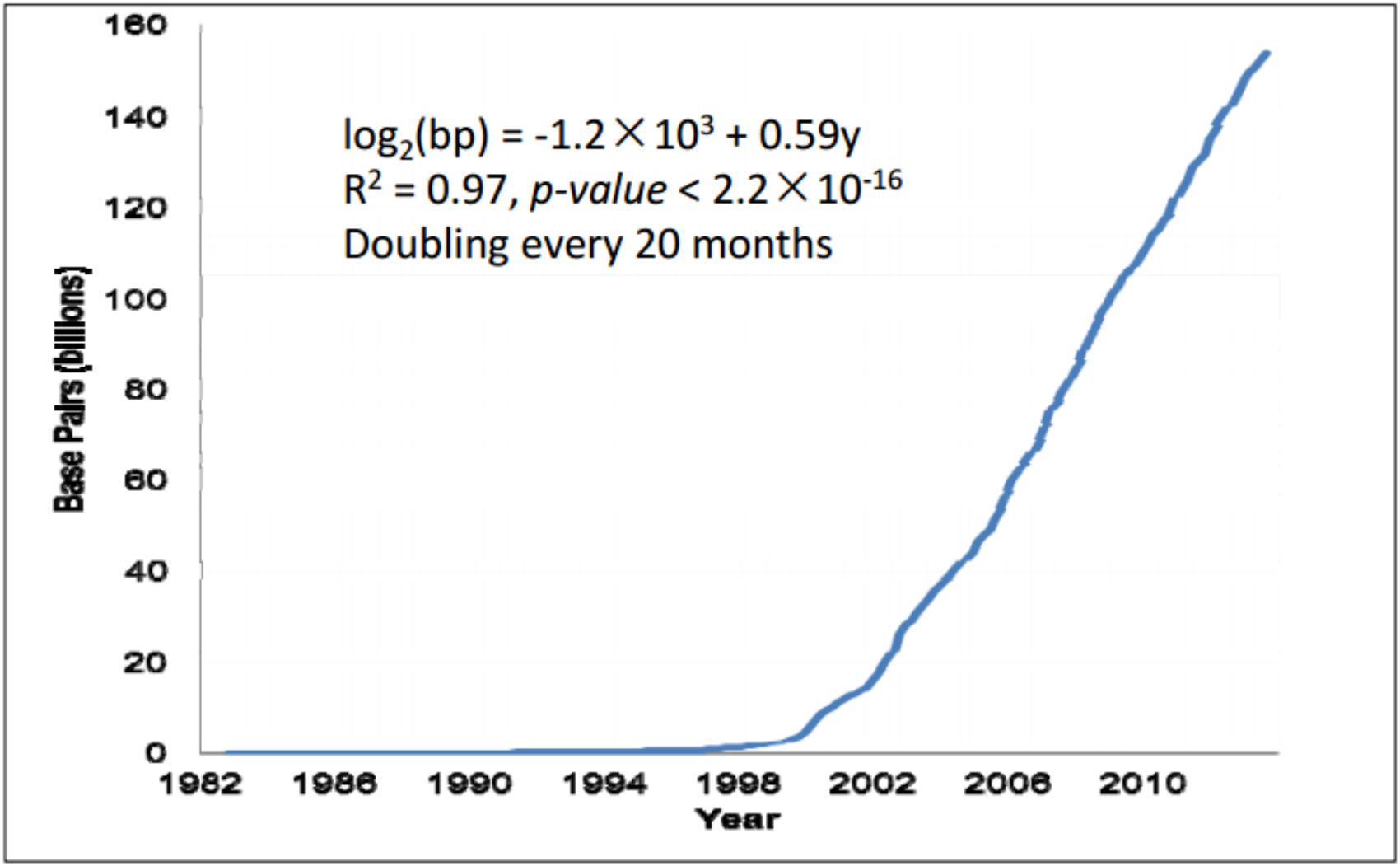


http://commons.wikimedia.org/wiki/File:Simplified_tree.png

Human genetic variation

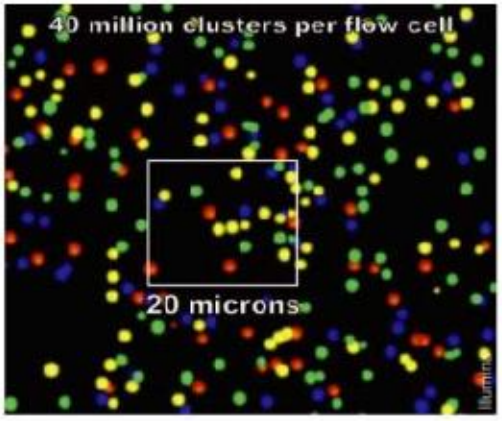

How do you **decode** the instructions in this "manual of life"?
If you print 100 characters per line and 50 lines per page,
you'll fill **600,000 pages**, stacked **60 meters** high.
If you read one base per second, nonstop,
it will take you **100 years.**
10^{15} basepairs in >165,000 organisms!
It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

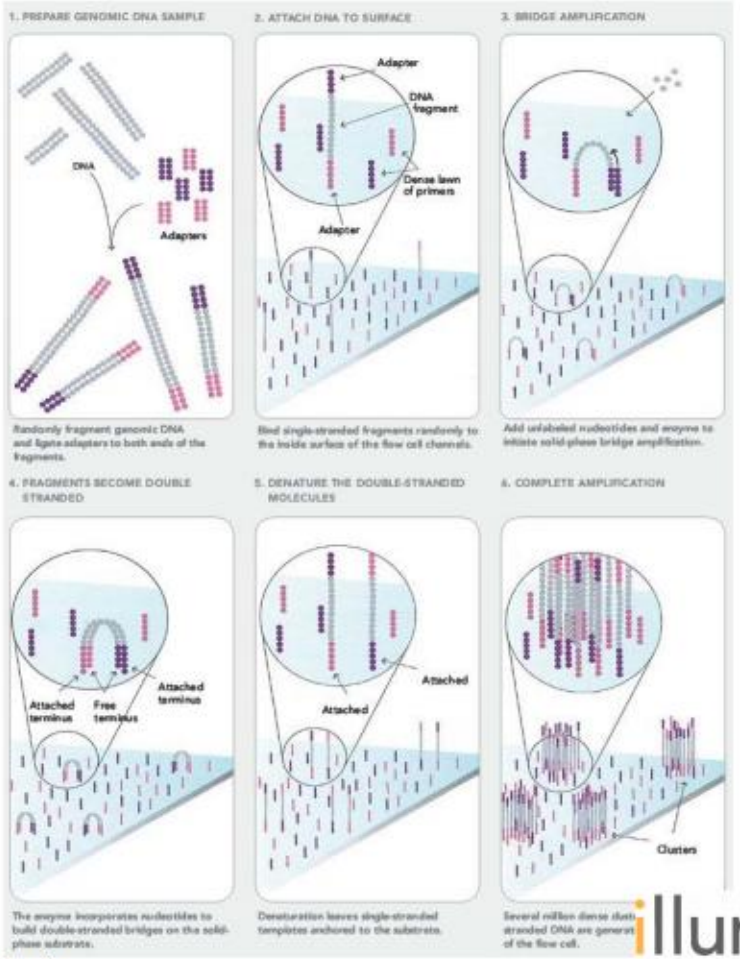
Genbank growth

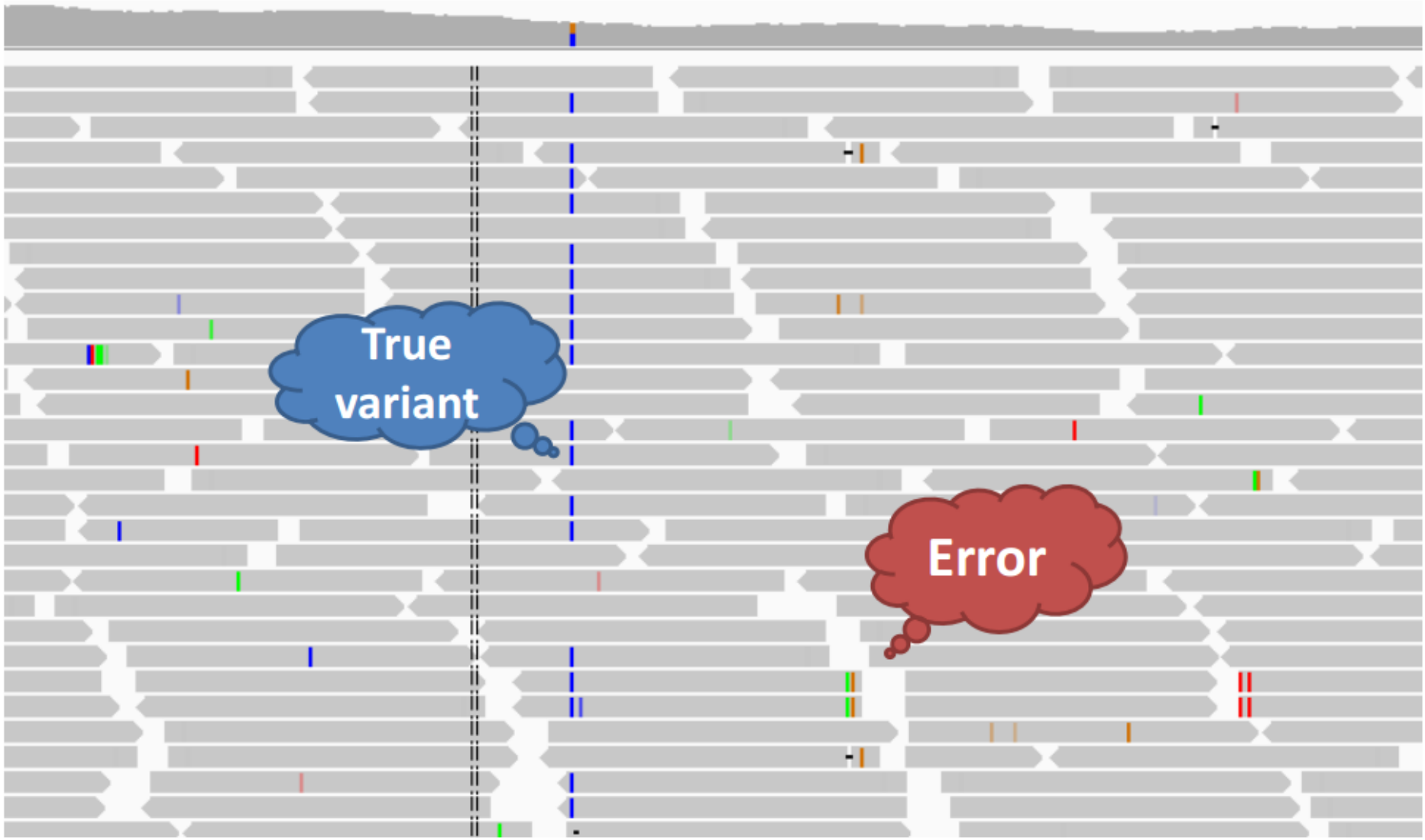


Data Source: ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt

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



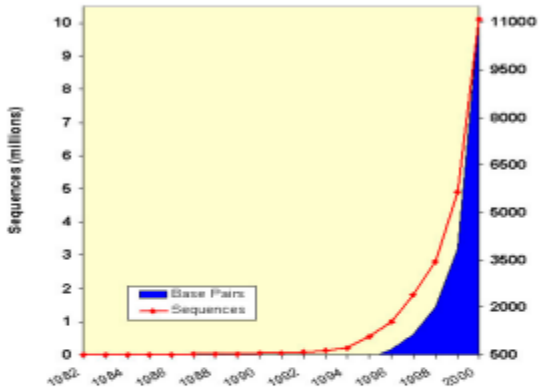



True variant
Error

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

High-throughput data

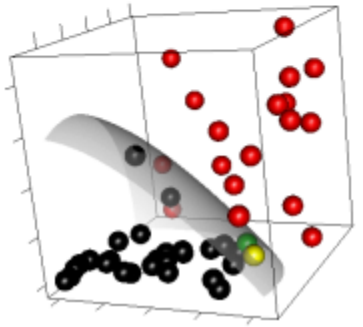
- Huge amount
- Explosive growth
- Low signal-to-noise ratio
- Multiple types



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

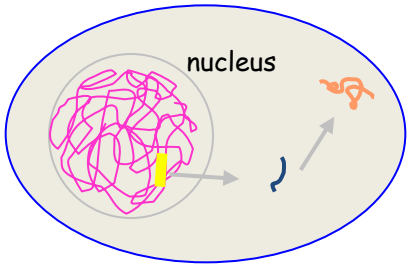


Results in Paradigm shift in Life sciences

- Past experiments where hypothesis driven
  - Evaluate hypothesis
  - Complement existing knowledge

- Present experiments are data driven
  - Discover knowledge from large amounts of data

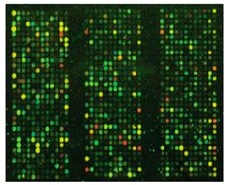
Life sciences research: from gene to function



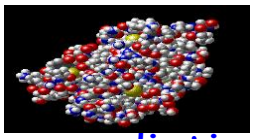
cell

Whole-genome sequence projects

Genome-wide micro-array analysis



"High-throughput" protein-analysis

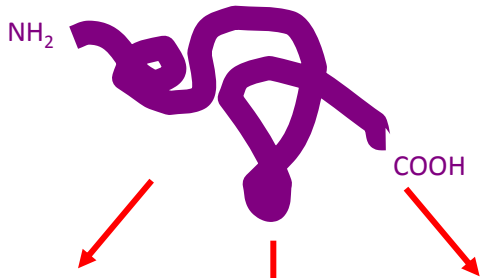


Protein function:
-prediction by bioinformatics
-proof by laboratory research

Gene expression by
RNA synthesis

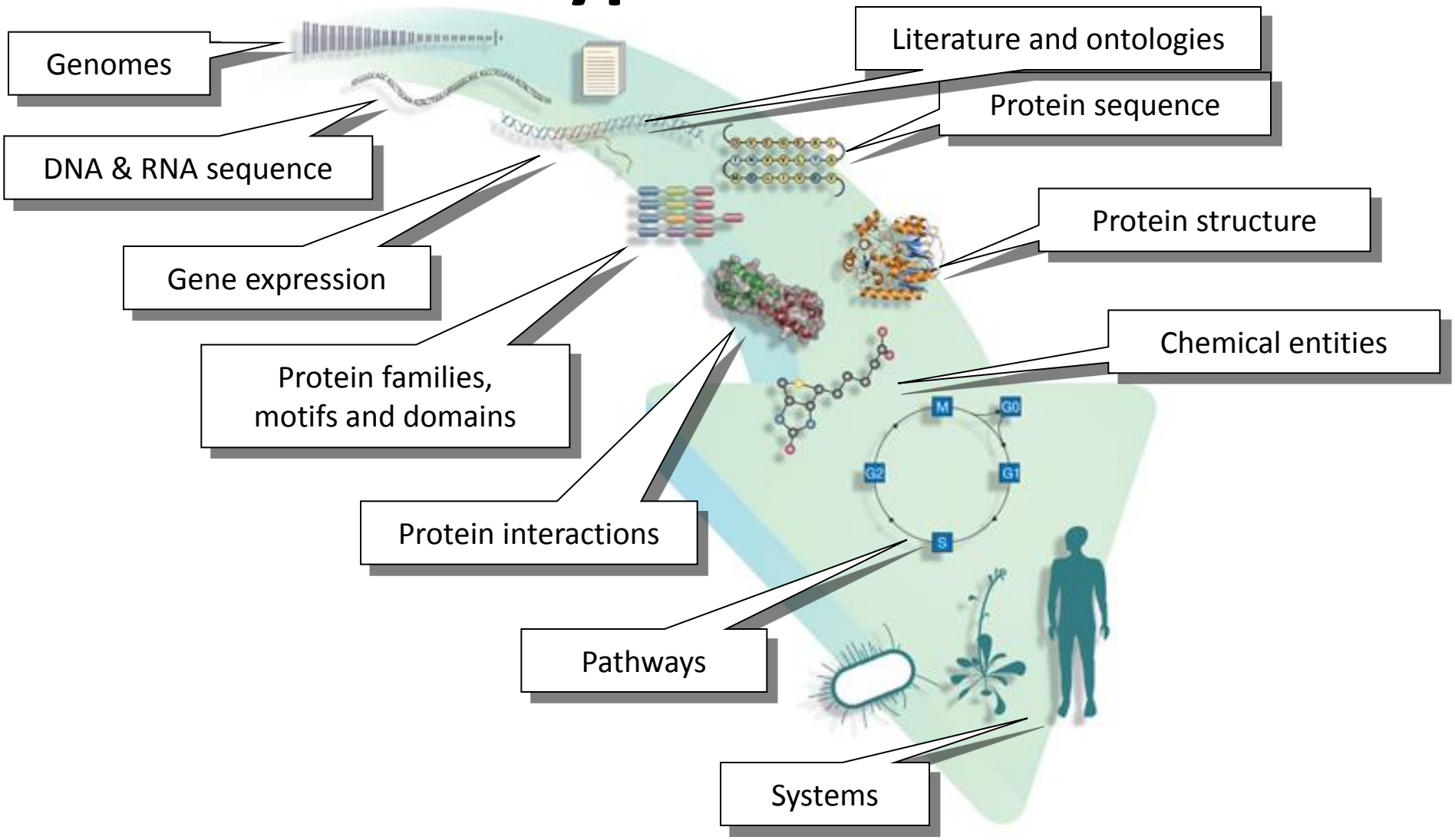
mRNA

mRNA translation by
protein synthesis

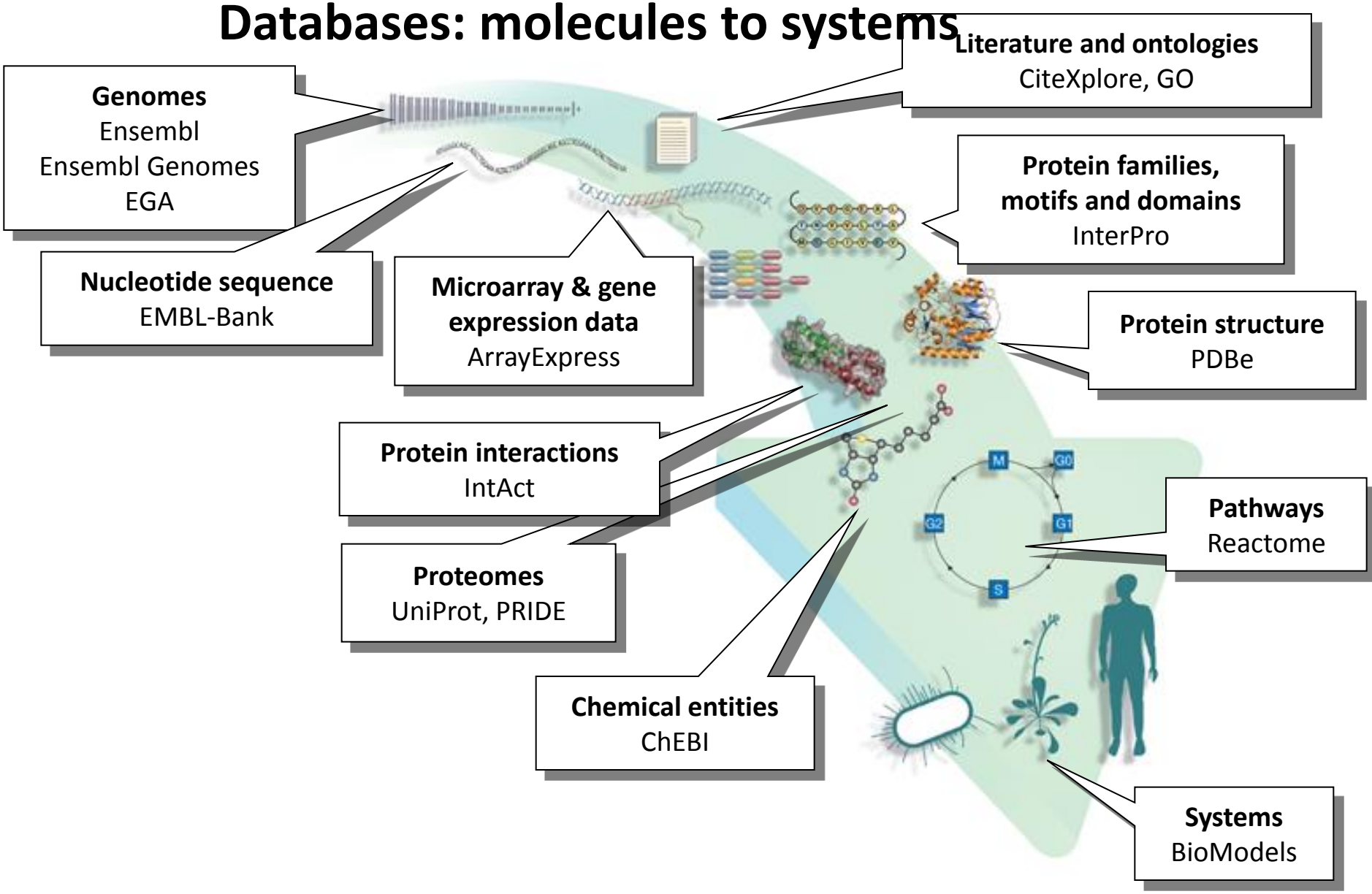


Protein

New types of data



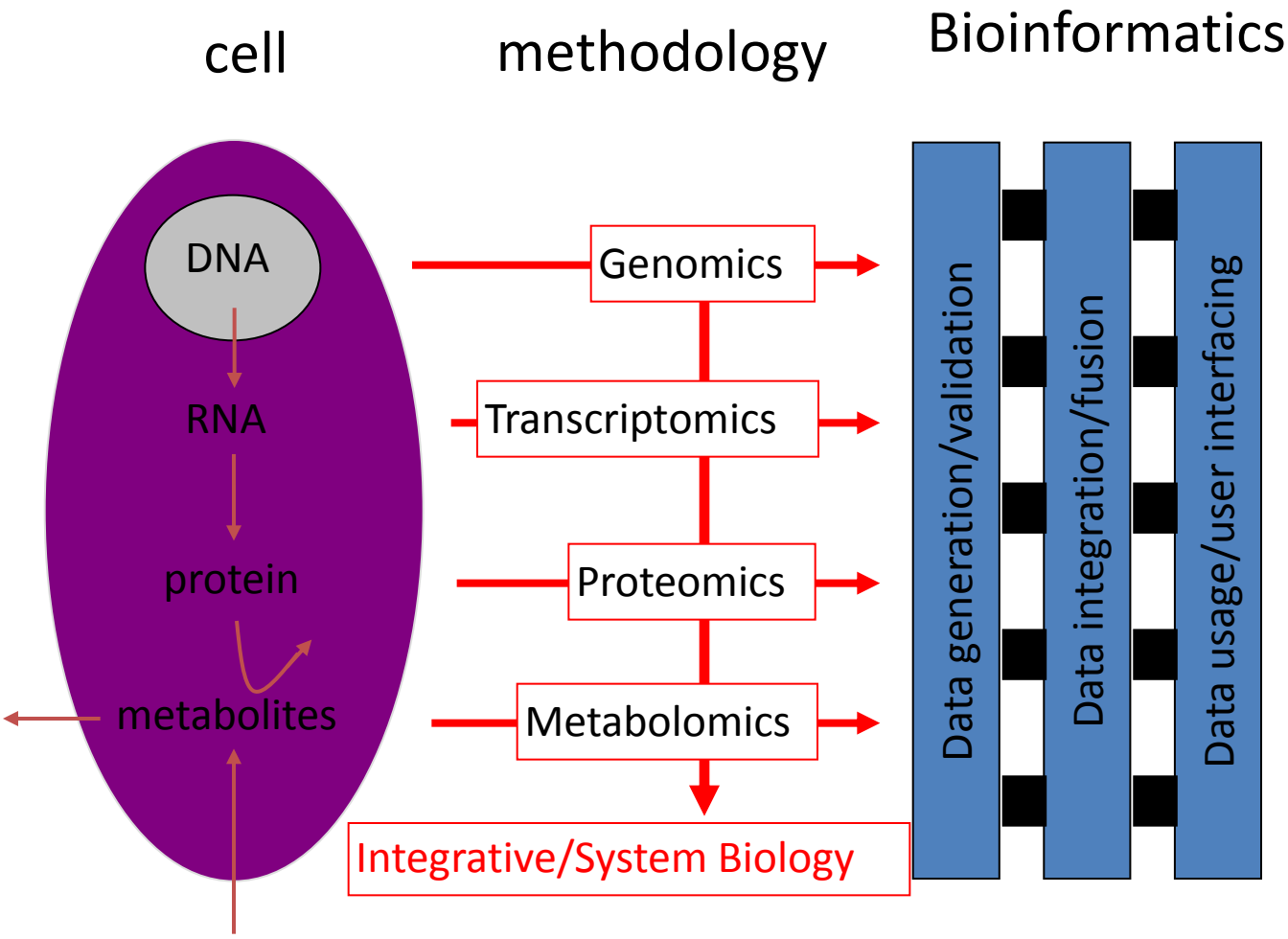
Databases: molecules to systems



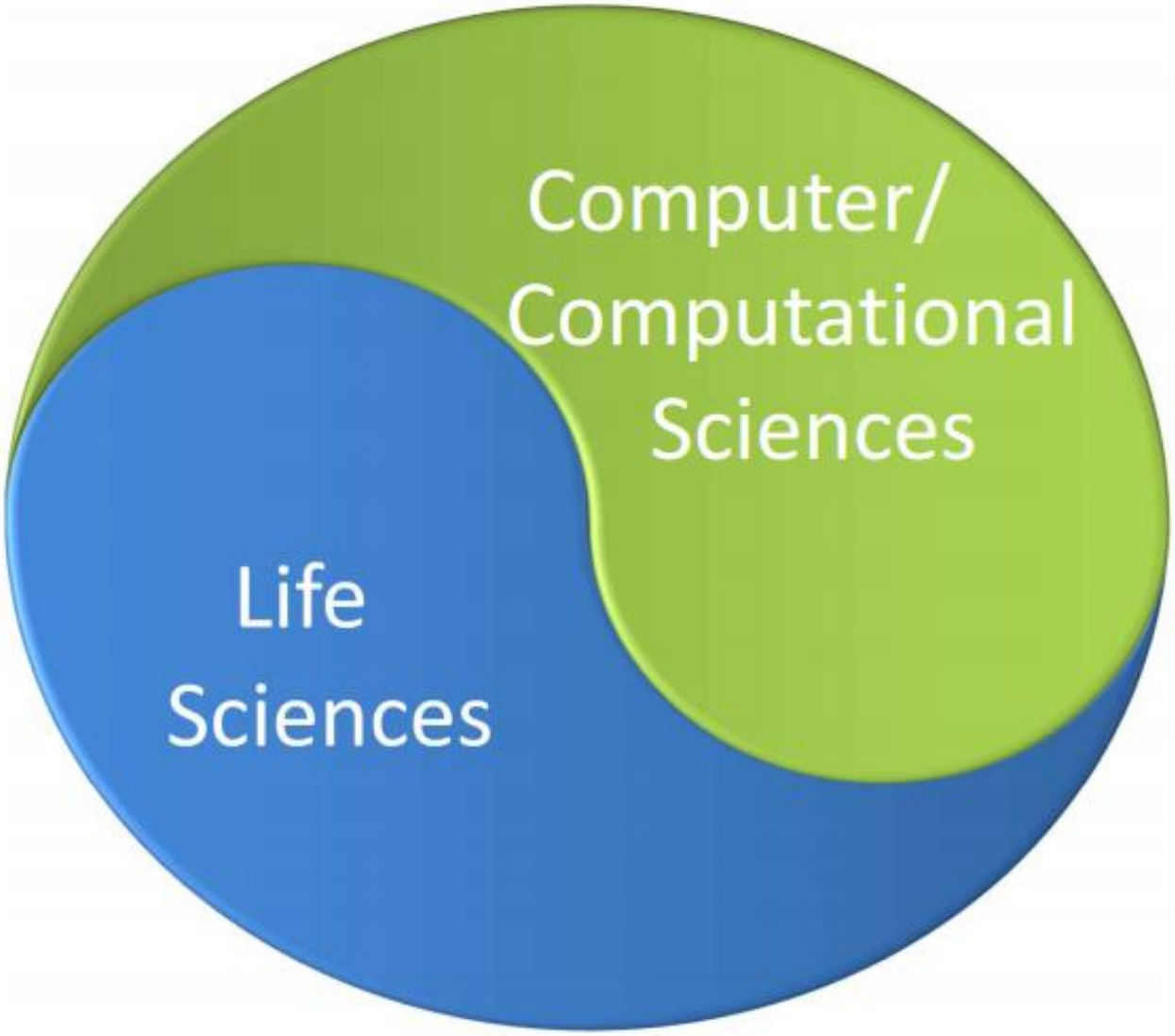
Developments towards Bio-informatics & e-Science

- Experiments become increasingly more complex
- Driven by increase of detector developments
- Results in an increase in amount and complexity of data
- Something has to be done to harness this development
  - Bio-informatics to translate data into useful biological, medical, pharmaceutical & agricultural knowledge

Role of bioinformatics



Bioinformatics

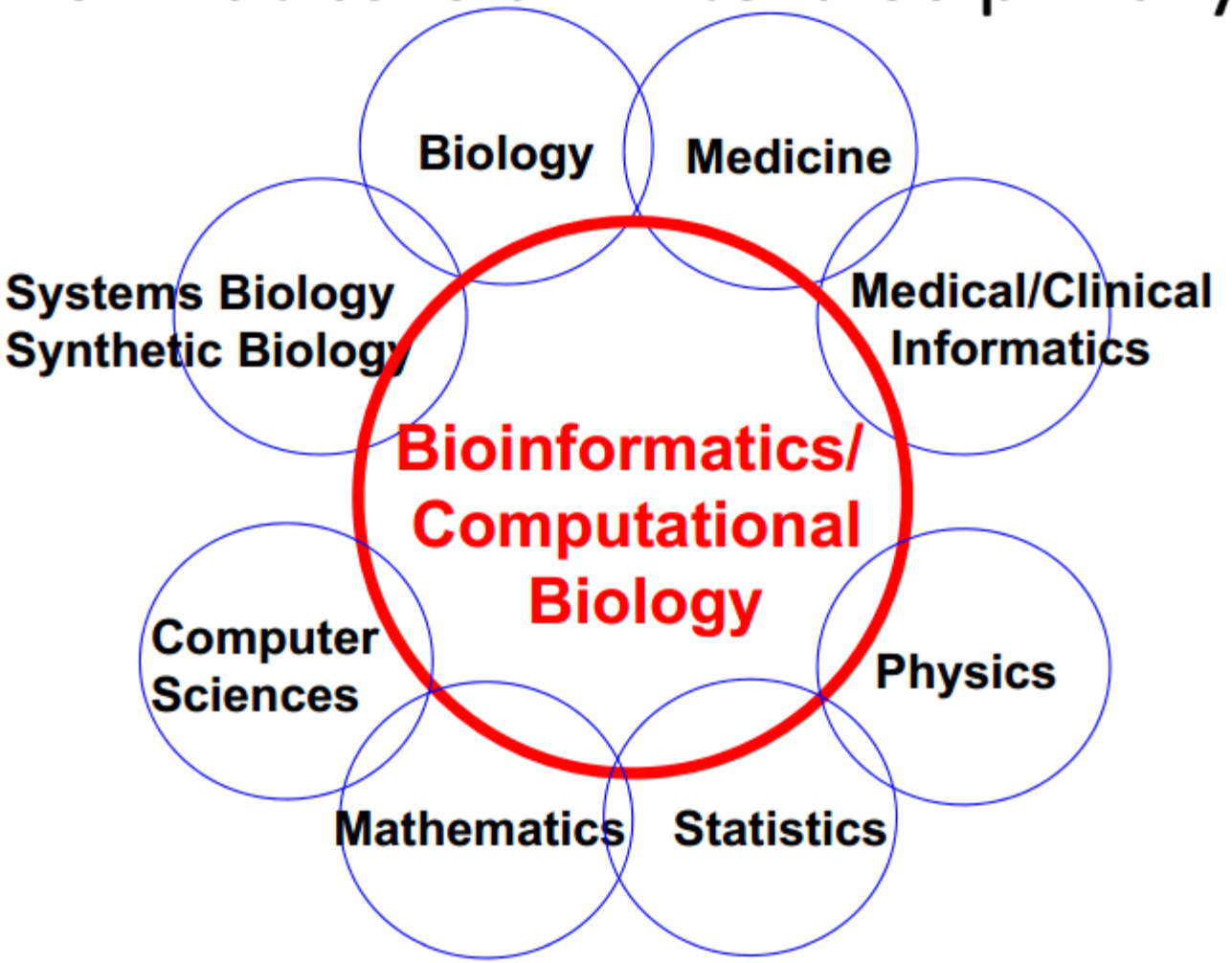


Bioinformatics: an interdisciplinary field that develops and applies computer and computational technologies to study biomedical questions

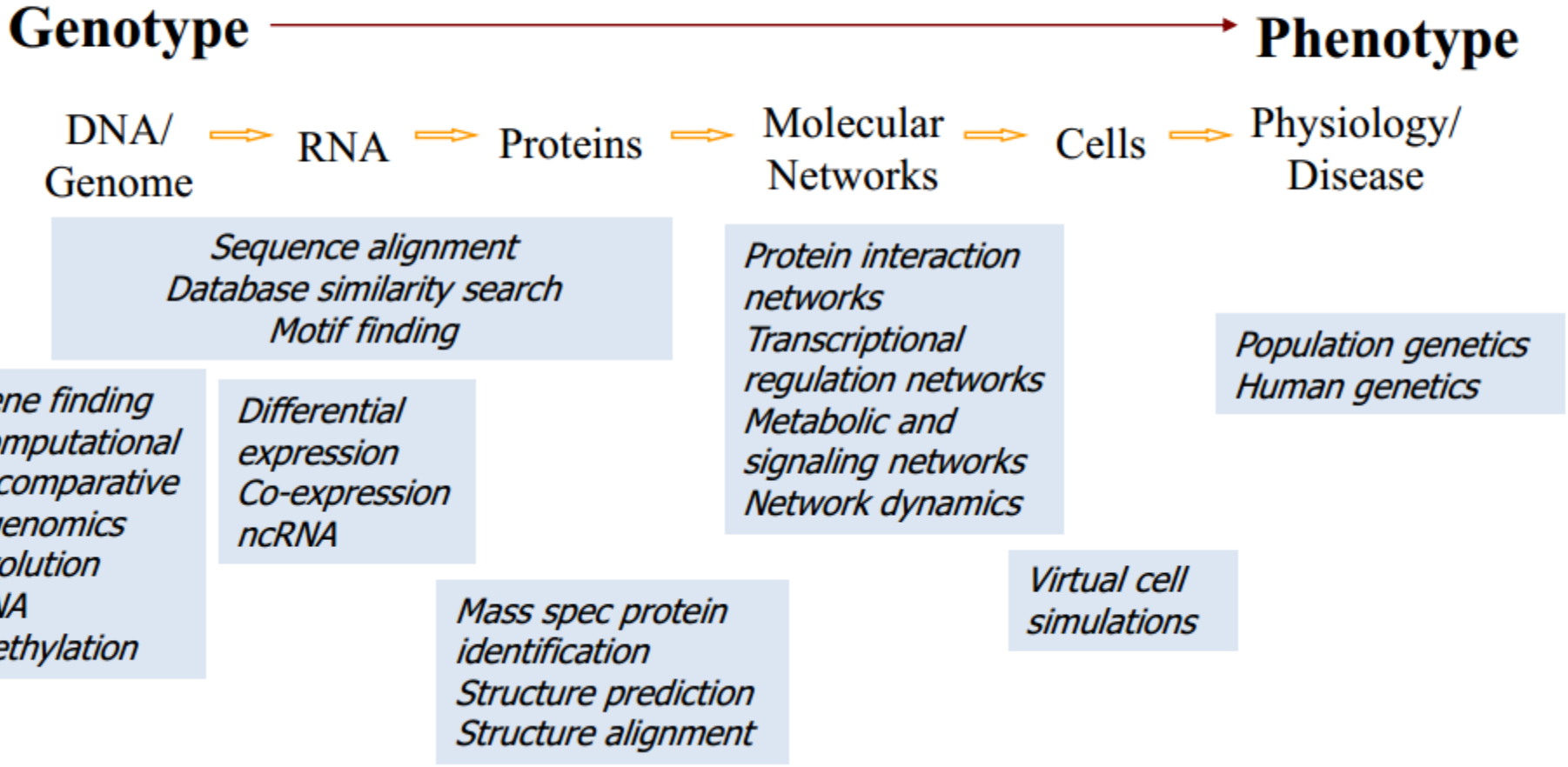
- As a technology, bioinformatics is a powerful technology to manage, query, and analyze big data in life sciences.

- 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.

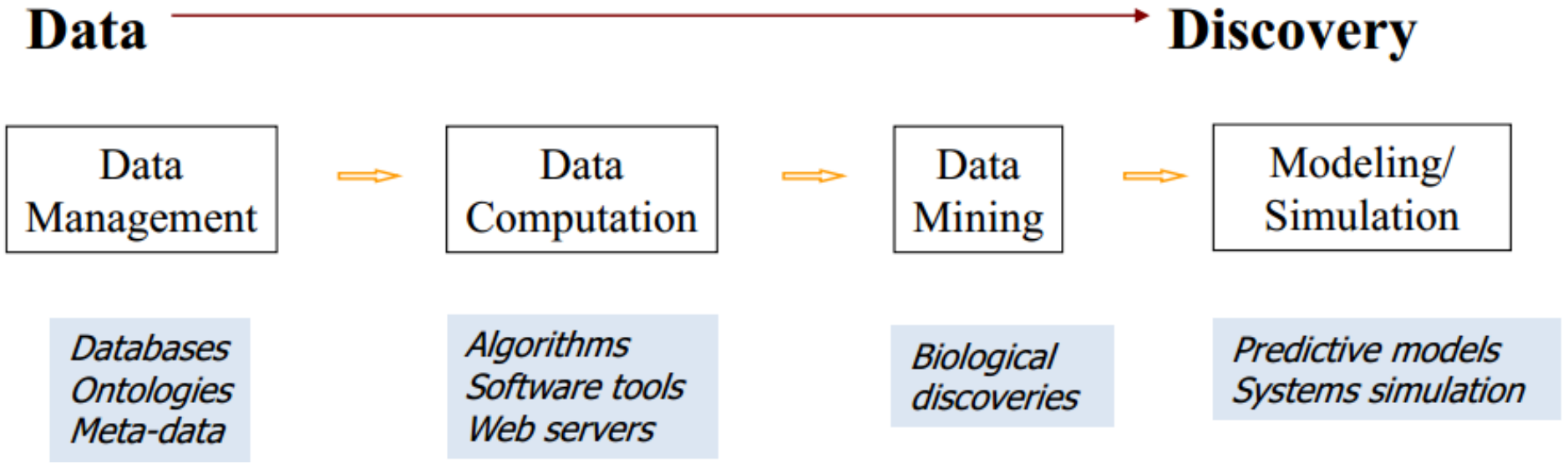
Bioinformatics is an interdisciplinary field



The Bio- in Bioinformatics



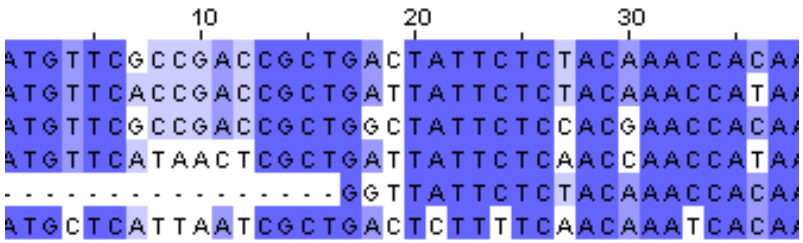
The –informatics in Bioinformatics



Genetic Research Deals with Inherited Traits

Genetic researchers study inherited traits by analyzing DNA sequences.

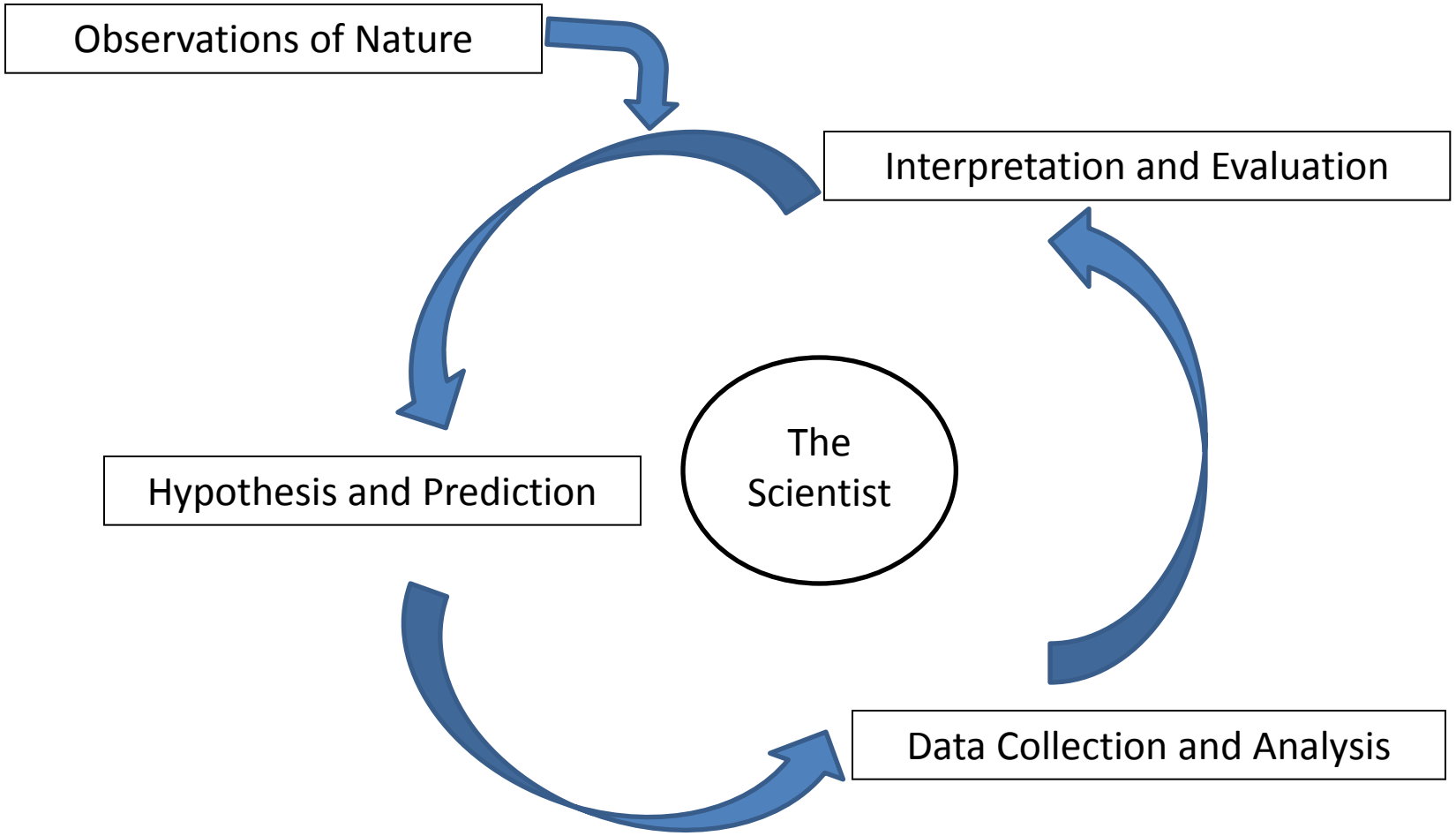
How are we similar?
How are we different?



Use bioinformatics to research differences in DNA sequences.

DNA isolation.

The Practices of Scientific Research



What is Bioinformatics?

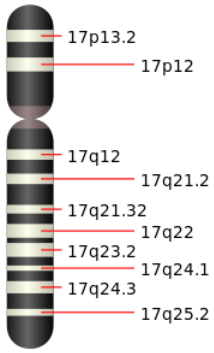
Bioinformatics is the application of computer science and information technology to biology and medicine.

Bioinformatics makes it possible to analyze large quantities of complex biological data and can be used to ***search biological databases, compare sequences,*** *and* ***draw molecular structures****.*

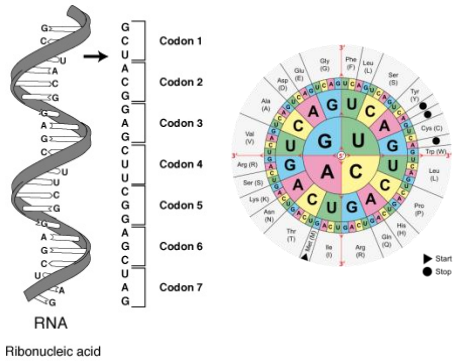
Bioinformatics Tools Help Scientists:

Organize, Process, and Make Sense of Complex Biological Data Sets

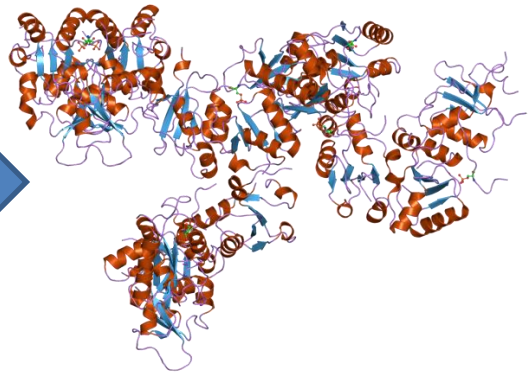
DNA



RNA



Protein



Bioinformatics Tools:

DNA Sequencing
Identify Mutations in DNA.

Bioinformatics Tools:

RNA Sequencing
Identify tissue specific gene expression.

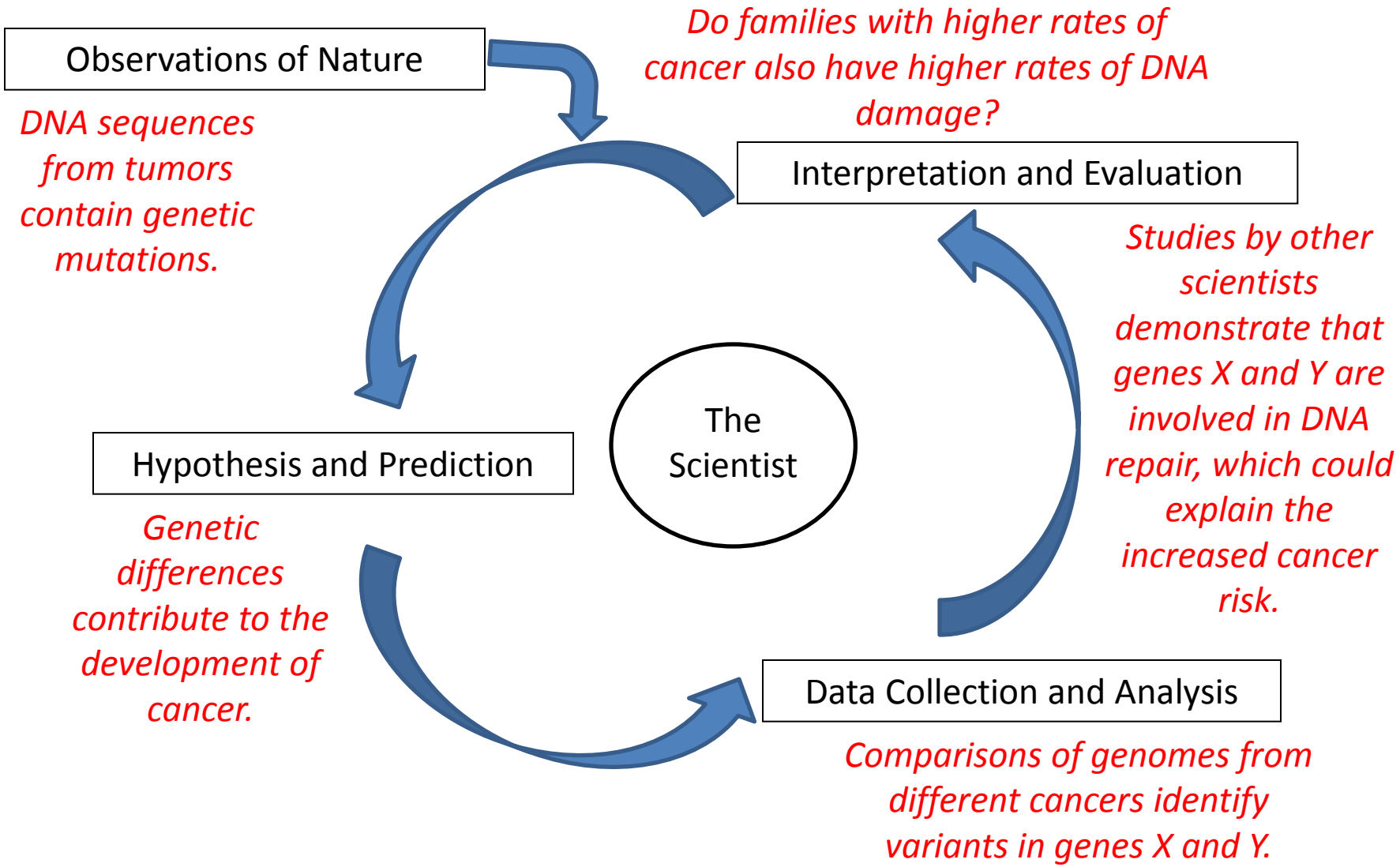
Bioinformatics Tools:

Protein 3D Structure visualization.

Question: What kinds of scientific questions can we answer with bioinformatics tools?

Image Source: Wikipedia Commons

The Practices of Scientific Research



Inside the Gene Machine:

How Information from DNA is Acquired and Used for Genetic Testing

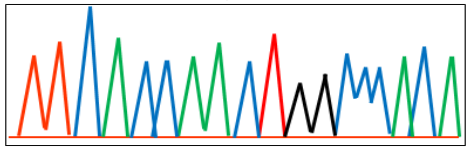
Patient Sample: Blood or Saliva.

Genetic Counselors work with patients to help them decide whether or not to have a genetic test, and help them understand the results of the test.

Extract DNA from Cells.

Lab Technicians work with patient samples in the lab, purifying and sequencing the DNA.

Sequence DNA.



…TTCACCAACAGGCCCACA…

Bioinformatics programmers create computer programs to help biologists analyze data.

Compare Patient DNA Sequence to Reference Sequence.

```
Reference  …TTCACCAACATGCCCACA…
               F  T  N  M  P  T

Patient    …TTCACCAACAGGCCCACA…
               F  T  N  R  P  T
```

Biomedical Researchers perform experiments with patient samples to find different variations of genes that might cause disease. Bioinformatics tools like **BLAST** and **ClustalW** are used to compare sequences.

Search Database to Determine if Patient Mutation is Associated with Disease.

Medical Doctors and *Veterinarians* use the knowledge gained from genetic testing to care for their patients.

Image Source: Microsoft Clip Art

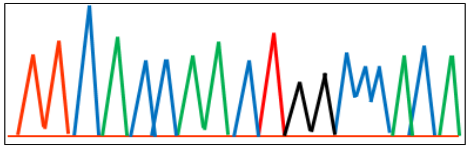
How DNA Sequence Data is Obtained for Genetic Research

Genetic Data

Obtain Samples:
Blood , Saliva, Hair
Follicles, Feathers, Scales.

Extract DNA from Cells.

Sequence DNA.



...TTCACCAACAGGCCCACA...

Compare DNA Sequences to One Another.

TTCAACAACAGGCCCAC
TTCACCAACAGGCCCAC
TTCATCAACAGGCCCAC

Image Source: Wikipedia Commons

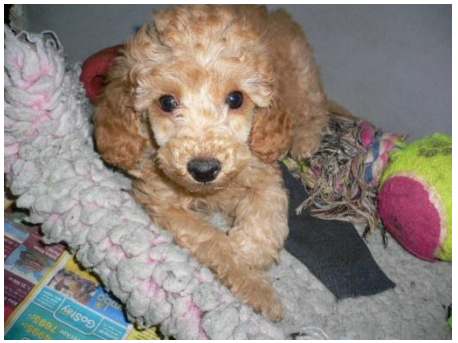
Which Animals are Most Closely Related to One Another?

Coyote

Jack Russell

Labradoodle

English Shepherd



Toy Poodle

Gray wolf

Red fox

Cocker Spaniel

Image Source: Wikipedia Commons

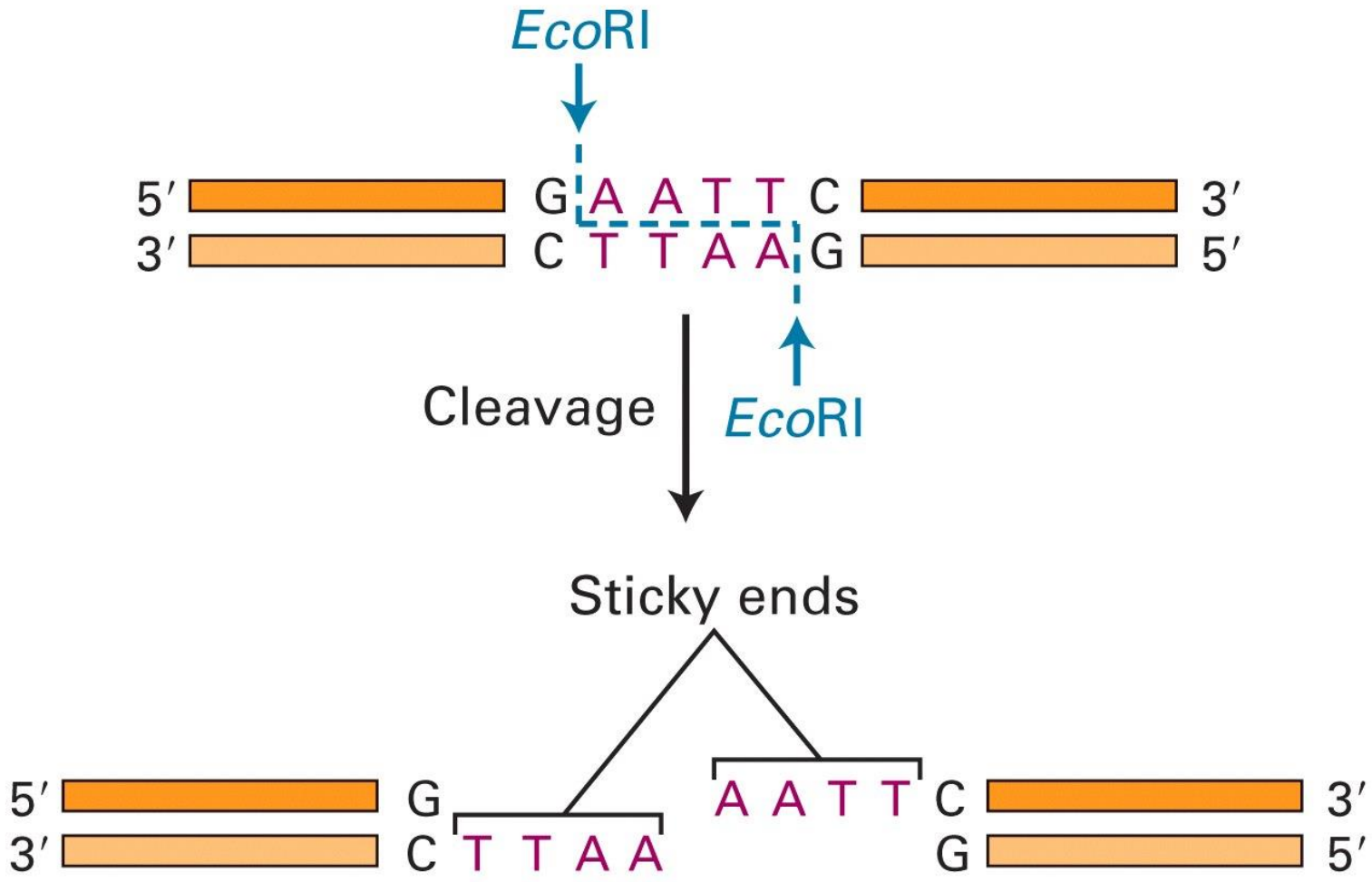
Multiple Sequence Alignment of Canine DNA Sequences

```
GreyWolf/1-883                305 TTCCTCCATCCTTTCTTCTACTATTAGCATCTTCTATGGTAGAAGCAGGTGCAGGAACGGGATGAACCGTATACCC 380
RedFox/1-880                  302 TTCCCCCATCCTTTCTTCTACTATTAGCATCTTCCATAGTAGAAGCGGGTGCGGGAACTGGGTGAACCGTATATCC 377
Coyote/1-883                  305 TTCCTCCATCTTTTCTTCTACTATTAGCATCTTCTATGGTAGAAGCAGGTGCAGGAACGGGATGGACTGTATATCC 380
Labradoodle/1-883             305 TTCCTCCATCCTTTCTTCTACTATTAGCATCTTCTATGGTAGAAGCAGGTGCAGGAACGGGATGAACCGTATACCC 380
JackRussell/1-883             305 TTCCTCCATCCTTTCTTCTACTATTAGCATCTTCTATGGTAGAAGCAGGTGCAGGAACGGGATGAACCGTATACCC 380
CockerSpaniel/1-883           305 TTCCTCCATCCTTTCTTCTACTATTAGCATCTTCTATGGTAGAAGCAGGTGCAGGAACGGGATGAACCGTATACCC 380
English Shepherd/1-883        305 TTCCTCCATCCTTTCTTCTACTATTAGCATCTTCTATGGTAGAAGCAGGTGCAGGAACGGGATGAACCGTATACCC 380
ToyPoodle/1-883               305 TTCCTCCATCCTTTCTTCTACTATTAGCATCTTCTATGGTAGAAGCAGGTGCAGGAACGGGATGAACCGTATACCC 380

GreyWolf/1-883                381 CCCACTGGCTGGCAATCTGGCCCATGCAGGAGCATCCGTTGACCTTACAATTTTCTCCTTACACTTAGCCGGAGTC 456
RedFox/1-880                  378 CCCATTAGCTGGTAACCTGGCTCATGCTGGAGCATCAGTGGACCTTACAATTTTCTCCCTGCACCTGGCCGGAGTC 453
Coyote/1-883                  381 TCCACTGGCTGGCAATCTGGCCCATGCAGGAGCATCCGTTGATCTTACAATTTTCTCCTTACATCTAGCTGGAGTC 456
Labradoodle/1-883             381 CCCACTGGCTGGCAATCTGGCCCATGCAGGAGCATCCGTTGACCTTACAATTTTCTCCTTACACTTAGCCGGAGTC 456
JackRussell/1-883             381 CCCACTGGCTGGCAATCTGGCCCATGCAGGAGCATCCGTTGACCTTACAATTTTCTCCTTACACTTAGCCGGAGTC 456
CockerSpaniel/1-883           381 CCCACTGGCTGGCAATCTGGCCCATGCAGGAGCATCCGTTGACCTTACAATTTTCTCCTTACACTTAGCCGGAGTC 456
English Shepherd/1-883        381 CCCACTGGCTGGCAATCTGGCCCATGCAGGAGCATCCGTTGACCTTACAATTTTCTCCTTACACTTAGCCGGAGTC 456
ToyPoodle/1-883               381 CCCACTGGCTGGCAATCTGGCCCATGCAGGAGCATCCGTTGACCTTACAATTTTCTCCTTACACTTAGCCGGAGTC 456

GreyWolf/1-883                457 TCTTCTATTTTAGGGGCAATTAATTTCATCACTACTATTATCAACATAAAACCCCCTGCAATATCCCAGTATCAAA 532
RedFox/1-880                  454 TCTTCAATTTTAGGAGCTATTAATTTCATCACTACTATTATCAATATAAAACCTCCCGCCATATCCCAATACCAAA 529
Coyote/1-883                  457 TCTTCTATTTTAGGGGCAATCAATTTCATCACTACTATTATCAACATAAAACCCCCTGCAATATCCCAGTATCAAA 532
Labradoodle/1-883             457 TCTTCTATTTTAGGGGCAATTAATTTCATCACTACTATTATCAACATAAAACCCCCTGCAATATCCCAGTATCAAA 532
JackRussell/1-883             457 TCTTCTATTTTAGGGGCAATTAATTTCATCACTACTATTATCAACATAAAACCCCCTGCAATATCCCAGTATCAAA 532
CockerSpaniel/1-883           457 TCTTCTATTTTAGGGGCAATTAATTTCATCACTACTATTATCAACATAAAACCCCCTGCAATATCCCAGTATCAAA 532
English Shepherd/1-883        457 TCTTCTATTTTAGGGGCAATTAATTTCATCACTACTATTATCAACATAAAACCCCCTGCAATATCCCAGTATCAAA 532
ToyPoodle/1-883               457 TCTTCTATTTTAGGGGCAATTAATTTCATCACTACTATTATCAACATAAAACCCCCTGCAATATCCCAGTATCAAA 532
```

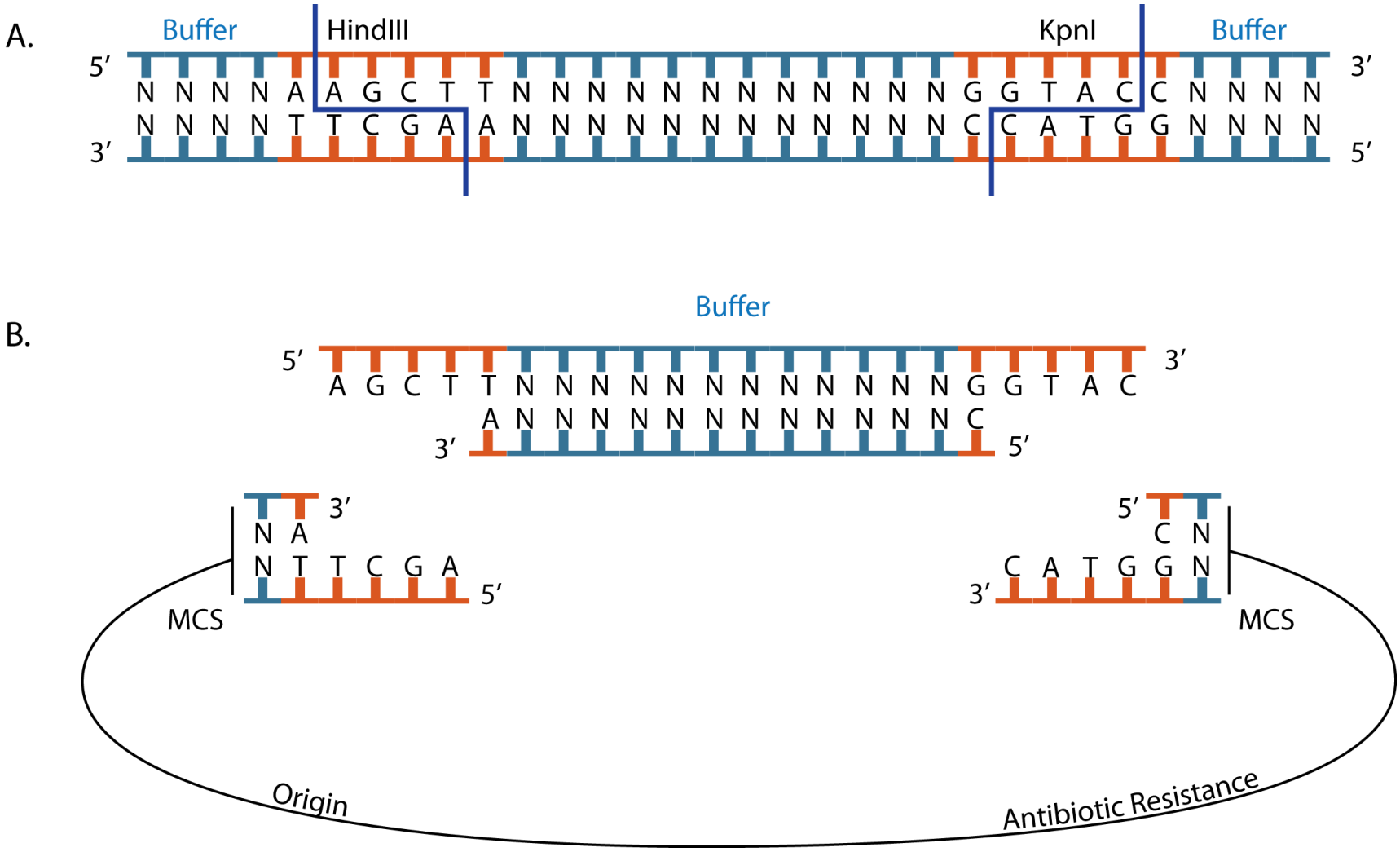
Color Coding Reveals Differences

Application of Bioinformatics

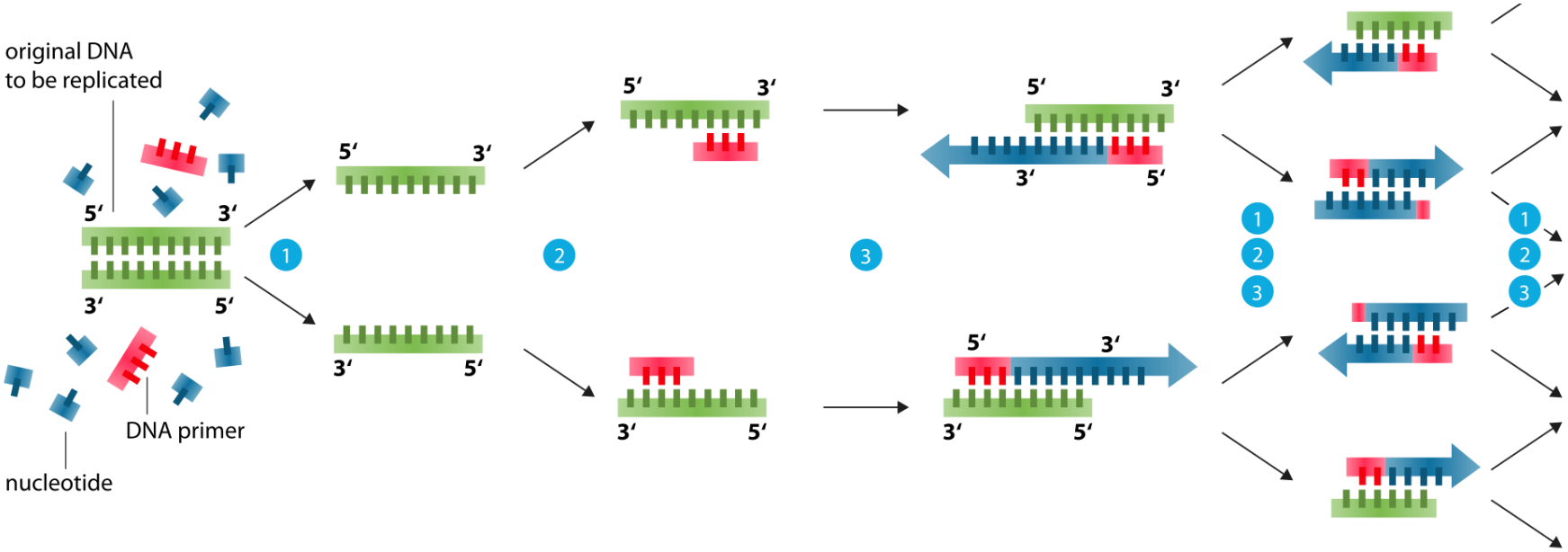
- Retrieving DNA sequences from databases .
- Identifying restriction sites.
- Designing polymerase chain-reaction (PCR) primers.
- Identifying open reading frames (ORFs).
- Predicting elements of DNA/RNA secondary structure .
- Finding repeats.
- Computing the optimal alignment between two or more DNA sequences
- Finding polymorphic sites in genes (single nucleotide polymorphisms, SNPs)
- Assembling sequence fragments
- Finding orthologous and paralogous genes.


EcoRI
5′
3′
G A A T T C
C T T A A G
3′
5′
Cleavage
EcoRI
Sticky ends
5′
3′
G
C T T A A
A A T T C
G
3′
5′

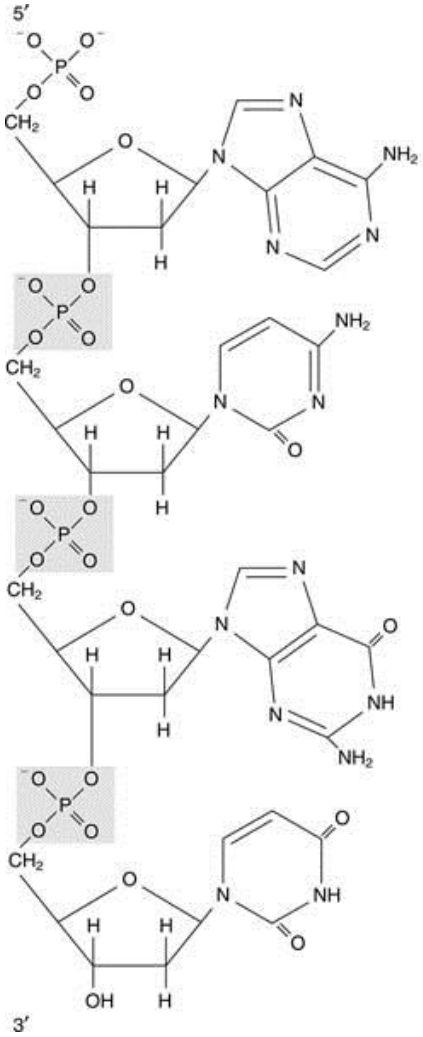
Restriction sites

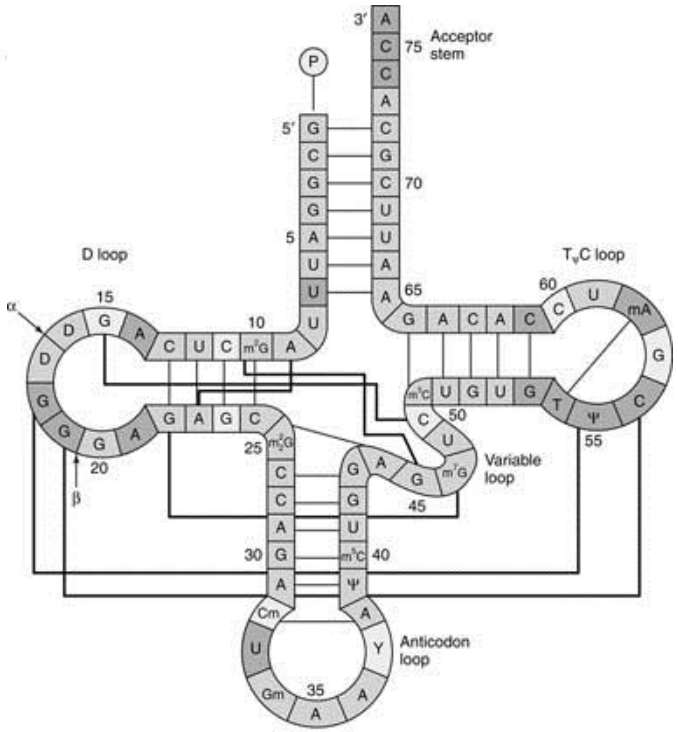


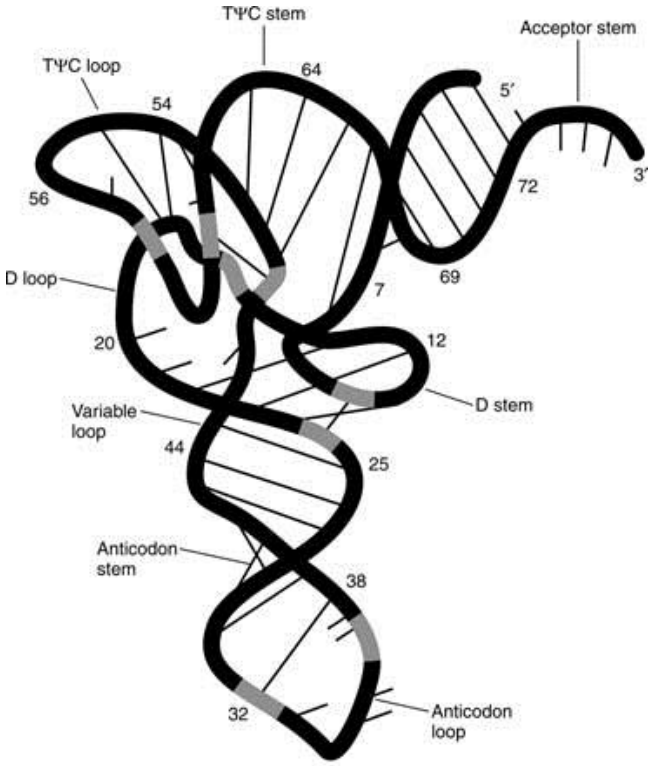
Polymerase chain reaction (PCR)



1. **Denaturation** at 94-96°C
2. **Annealing** at ~68°C
3. **Elongation** at ca. 72 °C


5′
3′


Acceptor stem
D loop
Variable loop
Anticodon loop
Secondary = Base pairs
(a)

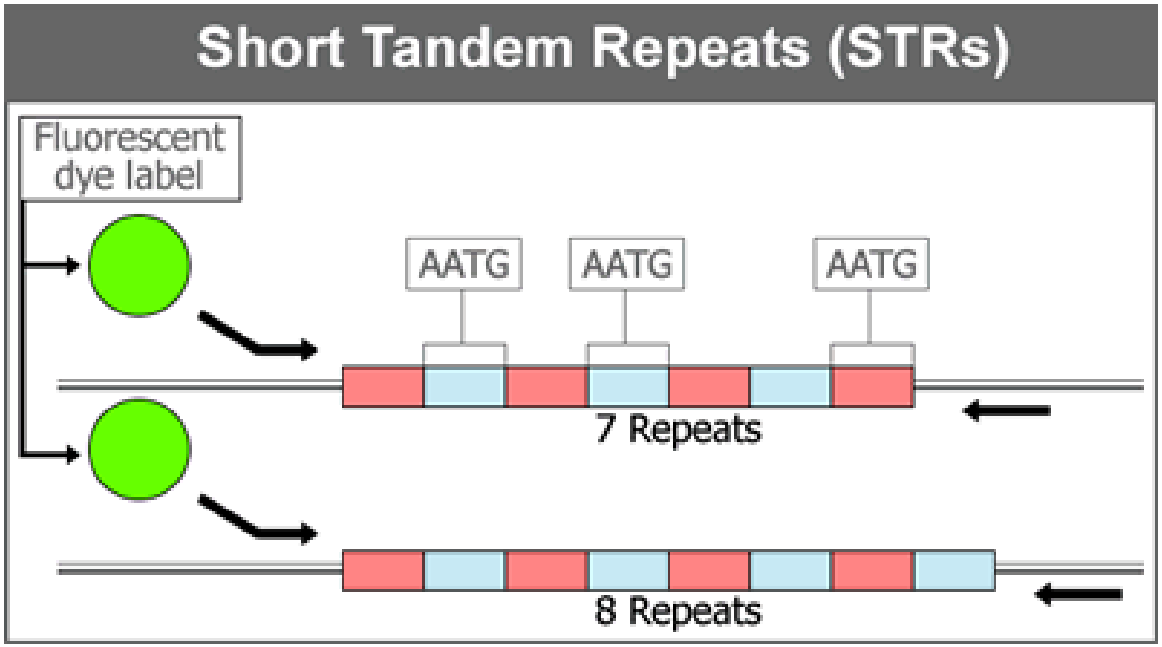

TΨC stem
Acceptor stem
TΨC loop
D loop
Variable loop
D stem
Anticodon stem
Anticodon loop
Tertiary = Folded in 3 dimensions
(b)

Tandem repetitive DNA

Tandem repeats appear in genomic DNA with a wide variety.

A tandem repeat in DNA is two or more similar copies of a DNA sequence.

Tandem repeats may not have any functional role.
Tandemly repeated DNA sequences are widespread throughout the human genome.



The STR loci carry lots of desirable features which makes them an ideal candidate for diverse applications including:

1- Forensic applications

2- Individual identification

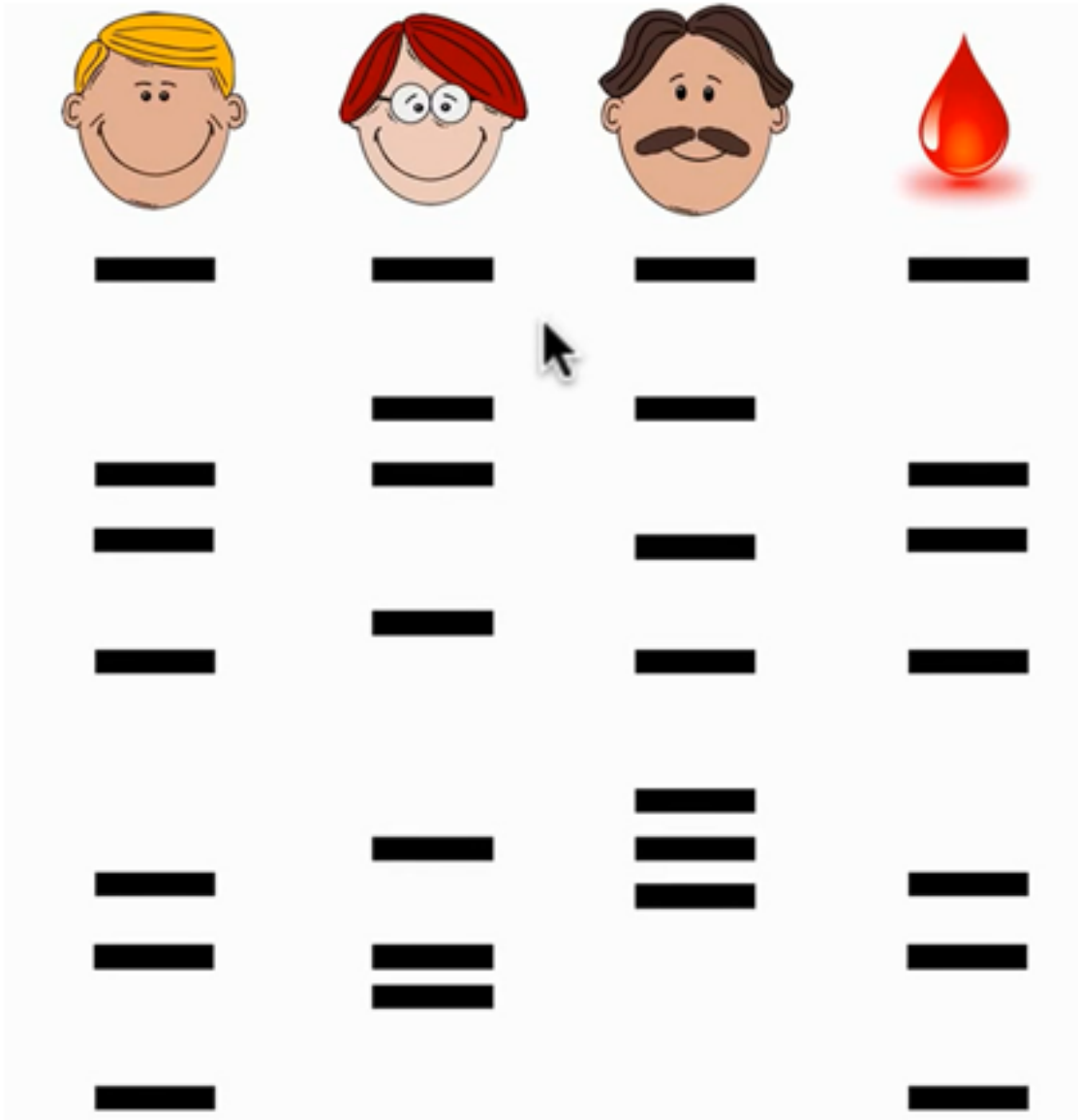
3- True paternity/maternity detection

4- Fine scale genetic mapping

5- Inter and intra group phylogenetic reconstruction.

Who is the Murder?

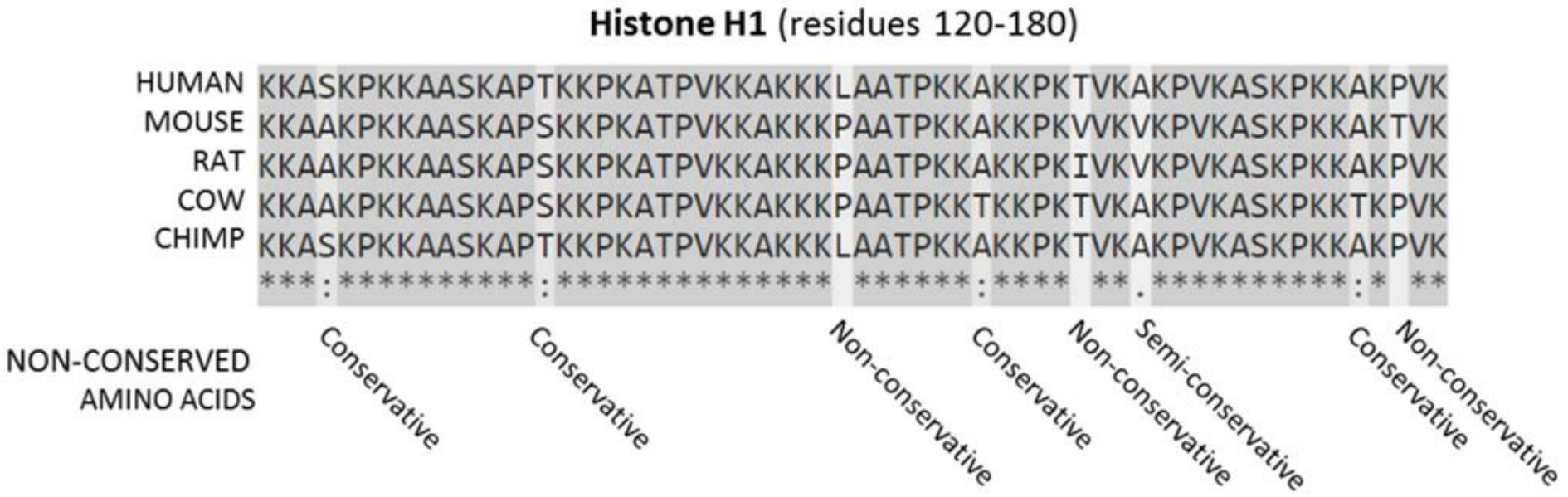
Hambozo	Cokrat	Dongule



Sequence Alignment

It is a way of arranging the sequences of DNA, RNA, or protein to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences.

Aligned sequences of nucleotide or amino acid residues are typically represented as rows within a matrix.

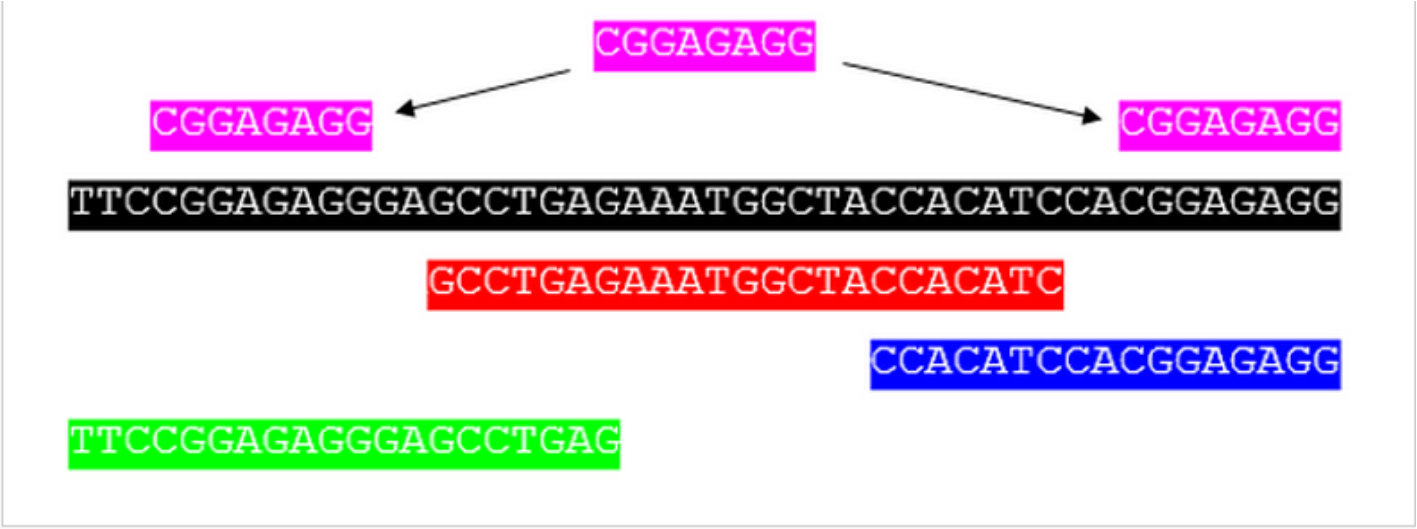


A sequence alignment, produced by ClustalO, of mammalian histone proteins.
Sequences are the amino acids for residues 120-180 of the proteins. Residues that are conserved across all sequences are highlighted in grey. Below the protein sequences is a key denoting conserved sequence (*), conservative mutations (:), semi-conservative mutations (.), and non-conservative mutations ( ).[2]

Sequence Assembly

Aligning and merging fragments of a much longer DNA sequence in order to reconstruct the original sequence.

In DNA sequencing technology cannot read whole genomes in one go, but rather reads small pieces of between 20 and 30000 bases, depending on the technology used.



Sample sequence showing how a sequence assembler would take fragments and match by overlaps. Image also shows the potential problem of repeats in the sequence.

Orthologs vs Paralogs

Orthologs

They are genes of common ancestry between species.

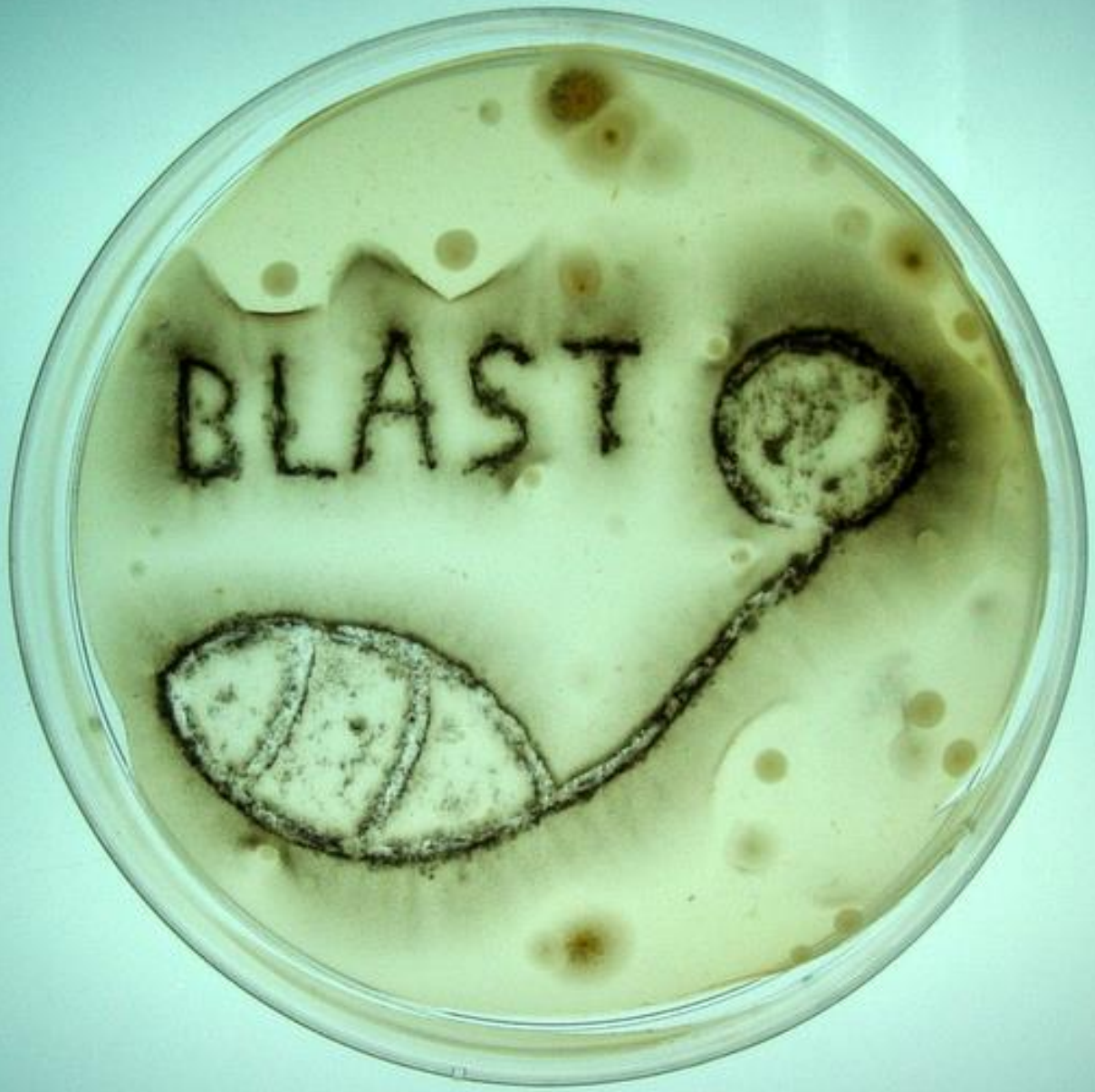
Eg: Dog hemoglobin and human hemoglobin, for example are orthologous.

Paralogs

They are when genes are duplicated then one of the copies evolves a new function.

Eg: Humans would be myoglobin and hemoglobin; both are oxygen storage proteins derived from the same ancestor gene, but have different functions (one is a transport protein in red blood cells, the other is a storage protein in muscle cells).

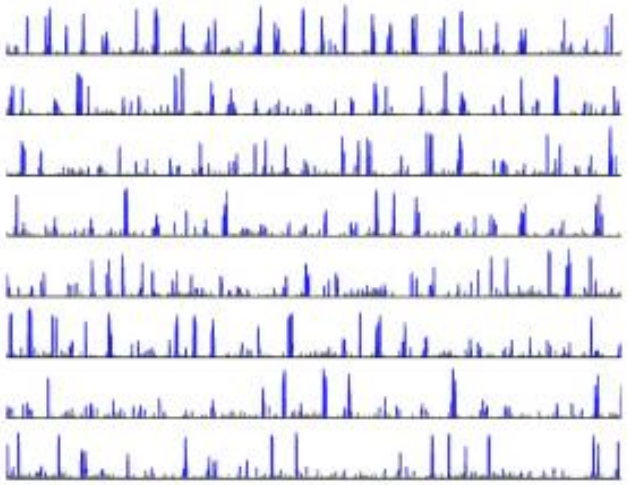
Impact of Genomics in Medicine


BLAST

Fungal bioinformatics

Programme Leader(s): Murray Cox
Team Member(s): Pierre-Yves Dupont

We are identifying lateral gene transfer events in fungi by developing evolutionary algorithms and computational tools.



Lateral gene transfer

Organisms typically evolve by vertical transmission, where genes are passed from parents to offspring. However, genes are sometimes passed between organisms in a process known as lateral, or horizontal, gene transfer.

Fungal adaptability and gene transfer

Assoc Prof Murray Cox and colleagues at Massey University are investigating whether lateral gene transfer between fungi is a key driving force in the adaptability of fungal species. Laterally transferred genes have been show to confer new functions or traits, such as pathogenicity. Recent research has shown that entire metabolic pathways, as well as individual genes, have been transferred between fungi.

Bioinformatics and fungal genomes

The researchers are comparing genomes in two case study systems with major relevance to the New Zealand economy:

1. *Dothistroma septosporum*, a fungus which causes pine needle blight, costing New Zealand's pine plantation industry $200 million each year.
2. *Epichloë festucae*, a fungal endophyte which protects New Zealand's pasture grasses from insect damage.

Fungal & Microbial Genomics

Welcome to the MIPS Fungal & Microbial Genomics Group

The group focuses on the analysis of fungal and microbial genomes. With former efforts on the model systems yeast and Neurospora we now work on plant pathogenic fungi and advanced model systems, mainly Fusarium spec., Ustilaginaceae and selected human pathogenic bacteria. We also provide comprehensive analysis of further public genome data for comparative analysis.

The basis to explore the species specific properties are correct structural annotations. Thus we improve the gene sets of our core projects and related genome data by applying comparative gene calling procedures. The resulting structurally annotated genomes are used for integration of further omics datasets and comparative studies.

Contact

Dr. Ulrich Güldener
Fungal & Microbial Genomics
Institute of Bioinformatics and Systems Biology
Helmholtz Zentrum München (GmbH)
Ingolstädter Landstrasse 1
D-85764 Neuherberg
Germany

Email

Phone: +49 (0) 89 3187-3582
Fax: +49 (0) 89 3187-3585

Featured Resources

List all analyzed Fungal Genomes - PEDANT ›

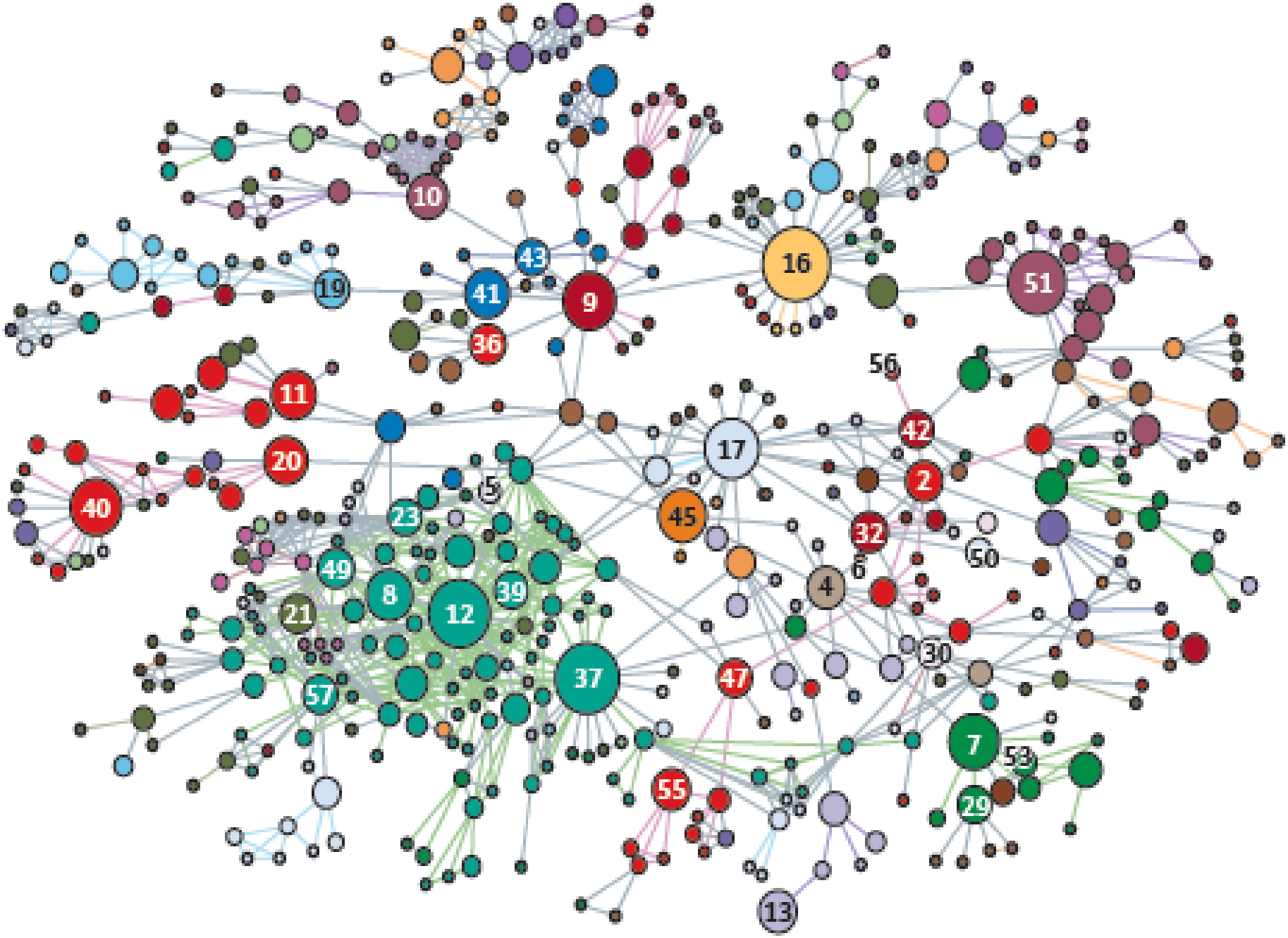
ftp://ftpmips.gsf.de/fungi/ ›

Fusarium species groups ›

Network medicine: a network-based approach to human disease

Albert-László Barabási[*‡§], *Natali Gulbahce*[*‡||] *and Joseph Loscalzo*[§]

Abstract | Given the functional interdependencies between the molecular components in a human cell, a disease is rarely a consequence of an abnormality in a single gene, but reflects the perturbations of the complex intracellular and intercellular network that links tissue and organ systems. The emerging tools of network medicine offer a platform to explore systematically not only the molecular complexity of a particular disease, leading to the identification of disease modules and pathways, but also the molecular relationships among apparently distinct (patho)phenotypes. Advances in this direction are essential for identifying new disease genes, for uncovering the biological significance of disease-associated mutations identified by genome-wide association studies and full-genome sequencing, and for identifying drug targets and biomarkers for complex diseases.


10
43
16
51
19
41
9
36
56
11
42
17
20
2
5
40
45
23
32
50
6
49
8
12
39
4
21
30
37
47
57
7
53
55
29
13

1. Aldosteronism
2. Alzheimer's disease
3. Anaemia, congenital deserythropoietic
4. Asthma
5. Ataxia-telangiectasia
6. Atherosclerosis
7. Blood group
8. Breast cancer
9. Cardiomyopathy
10. Cataract
11. Charcot–Marie–Tooth disease
12. Colon cancer
13. Complement component deficiency
14. Coronary artery disease
15. Coronary spasm
16. Deafness
17. Diabetes mellitus
18. Enolase-β deficiency
19. Epidermolysis bullosa
20. Epilepsy
21. Fanconi's anaemia
22. Fatty liver
23. Gastric cancer
24. Gilbert's syndrome
25. Glaucoma 1A
26. Goitre congenital
27. HARP syndrome
28. HELLP syndrome
29. Haemolytic anaemia
30. Hirschprung disease
31. Hyperbilirubinaemia
32. Hypertension
33. Hypertension diastolic
34. Hyperthyroidism
35. Hypoaldosteronism
36. Leigh syndrome
37. Leukaemia
38. Low renin hypertension
39. Lymphoma
40. Mental retardation
41. Muscular dystrophy
42. Myocardial infarction
43. Myopathy
44. Nucleoside phosphorylase deficiency
45. Obesity
46. Paraganglioma
47. Parkinson's disease
48. Pheochromocytoma
49. Prostate cancer
50. Pseudohypoaldosteronism
51. Retinitis pigmentosa
52. Schizoaffective disorder
53. Spherocytosis
54. Spina bifida
55. Spinocerebellar ataxia
56. Stroke
57. Thyroid carcinoma
58. Total iodide organification defect
59. Trifunctional protein deficiency
60. Unipolar depression

Top Reasons To study Bioinformatics/Computational Biology

Computing is the key skill set for 21st century biology

Biology is becoming a more quantitative science.

In the future, new discoveries will require leveraging big datasets and using advanced analytical methods.

Big data and complex models require computational skills.

There is no way to escape this reality.

Nobel-prize winning pioneer

- Walter Gilbert:

"To use this flood of [sequence] knowledge, which will pour across the computer networks of the world, biologists not only must become computer literate, but also change their approach to the problem of understanding life"

- Sydney Brenner:

"I spent many hours persuading people that computing was not only going to be the essential tool for biological research but would also provide models for analyzing"

Computational skills are highly transferable

Computational skills transfer across this sector, plus a much wider market outside of the (bio)science.

Increasing your computational chops will give you a better chance at landing a job.

You will have a deeper appreciation for how computers work and more mastery of when you interact with computers in your daily life

Computing will help improve your core scientific skills

Computing forces you to confront and tame the very human tendency to do science in *ad hoc* ways and therefore it naturally develops core scientific skills such as:

- Logically planning experiments
- Collecting data consistently
- Developing reproducible methodology
- Analyzing your data with proper statistical methodology.

You should use you Ph.D./Post-Doc to develop new skills

Majority of Biology Ph.D. students have no training in scientific computing skills beyond using Excel or statistics package.

So you have to start training in something new, not just further developing a skill set that you already have.

You will develop a more unique skill set in Biology

Majority of Biologists have experimental training, but very few have advanced Computational training.

And because you will be able to get results that many others cannot, plus the fact that you will have skills that set you apart from the herd, you will be more competitive on the job market.

You will publish more papers

Computational Biology will provide you more than your fair share of failed experiments.

As a result, you are very likely to publish more papers per unit time in Computational Biology.

Funding agencies want to see the junior researchers who have good ideas and can take them to completion.

Publication is the proof that you can finish projects.

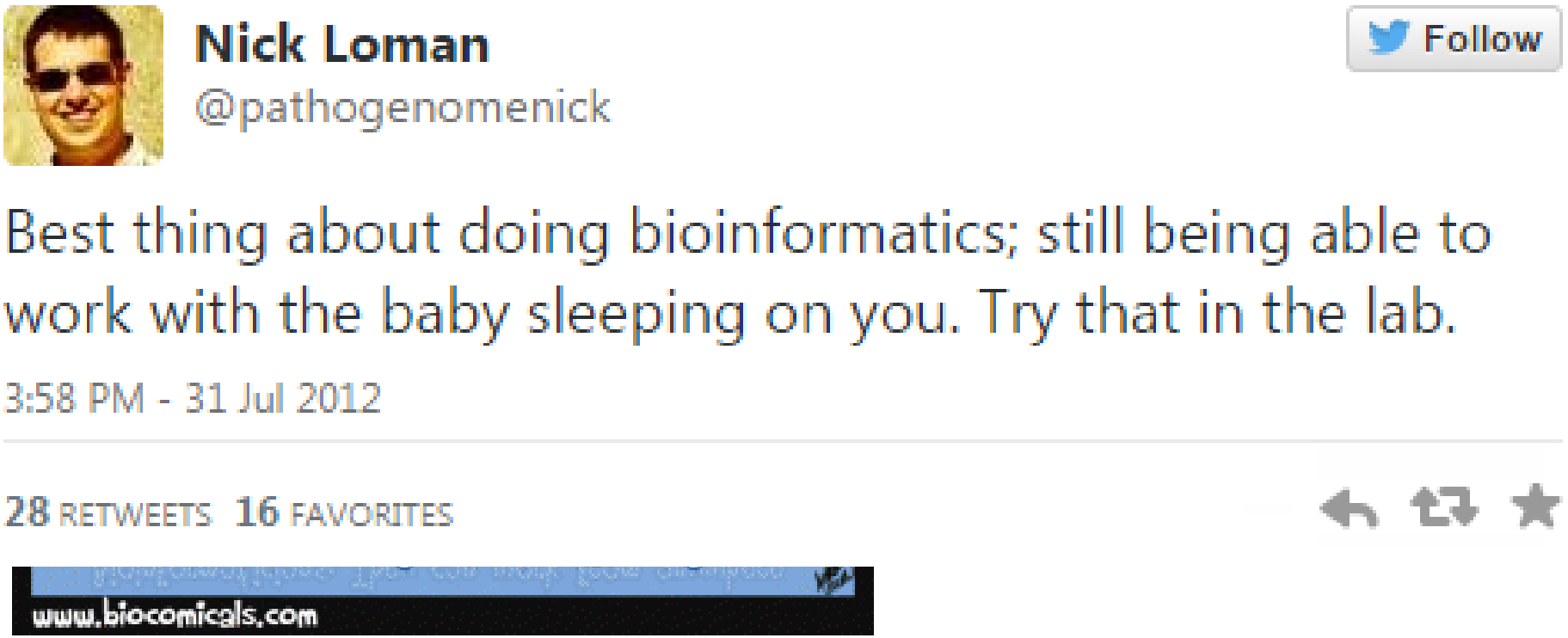
You will have more flexibility in your research

You can move from flexibly from topic to topic more easily than you can if your skill set is linked to specific experimental techniques.

This flexibility in scope allows you to satisfy your intellectual curiosity or chase the latest trend as you wish.

You will have more flexibility in working practices

- Unlike being chained to the bench, you can do working from home.



Computational research is cost-effective

- Computational Biology research is cheaper than most experimental work that requires a large consumables budget.

- Cost-efficiency is also very important when you are starting your group and for maintaining continuity of productivity when riding out troughs in funding or group size.

A successful scientist ends up in an office

- The truth is that the native habitat for an academic researchers is sitting in their office in front of their computer.

Webb Miller (most highly-cited bioinformaticians)

- If you've really wanted to do research since you were young, then ask yourself: why train in skills you will never ultimately use for the majority of your career, while somebody else in your lab gets to have fun making all the discoveries?

According to Fast Future report

- Computational biology is going to be among the most popular jobs of the future.

- Besides, most NGS data will sit untouched and unloved unless there are better methods to collect information about phenotype.

Safer and Environment friend

Mason Vail :

“You’re far less likely to spill a pathogen on yourself if you avoid the wet lab altogether “.

Another reason that came up in recent discussions is that computational work is more environmentally friendly.

Less plastic waste, less energy spent on autoclaving, etc.


The End