What is Bioinformatics?

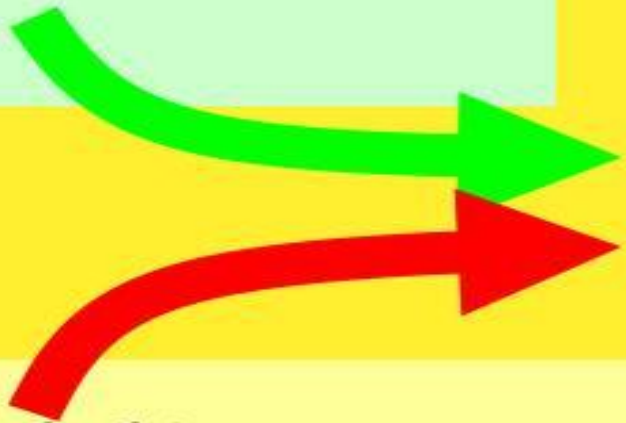
Bioinformatics = Biology + **Information**

Biology is becoming an information science.

Computational methods are necessary to analyze the Massive amount of information that are coming out of the genome projects.

The field of science in which **biology**, **computer science** and **information technology** merge into a single discipline

Biologists
collect molecular data:
DNA & Protein sequences,
gene expression, etc.



Bioinformaticians
Study biological questions by analyzing molecular data

Computer scientists
(+Mathematicians, Statisticians, etc.)
Develop tools, softwares, algorithms
to store and analyze the data.

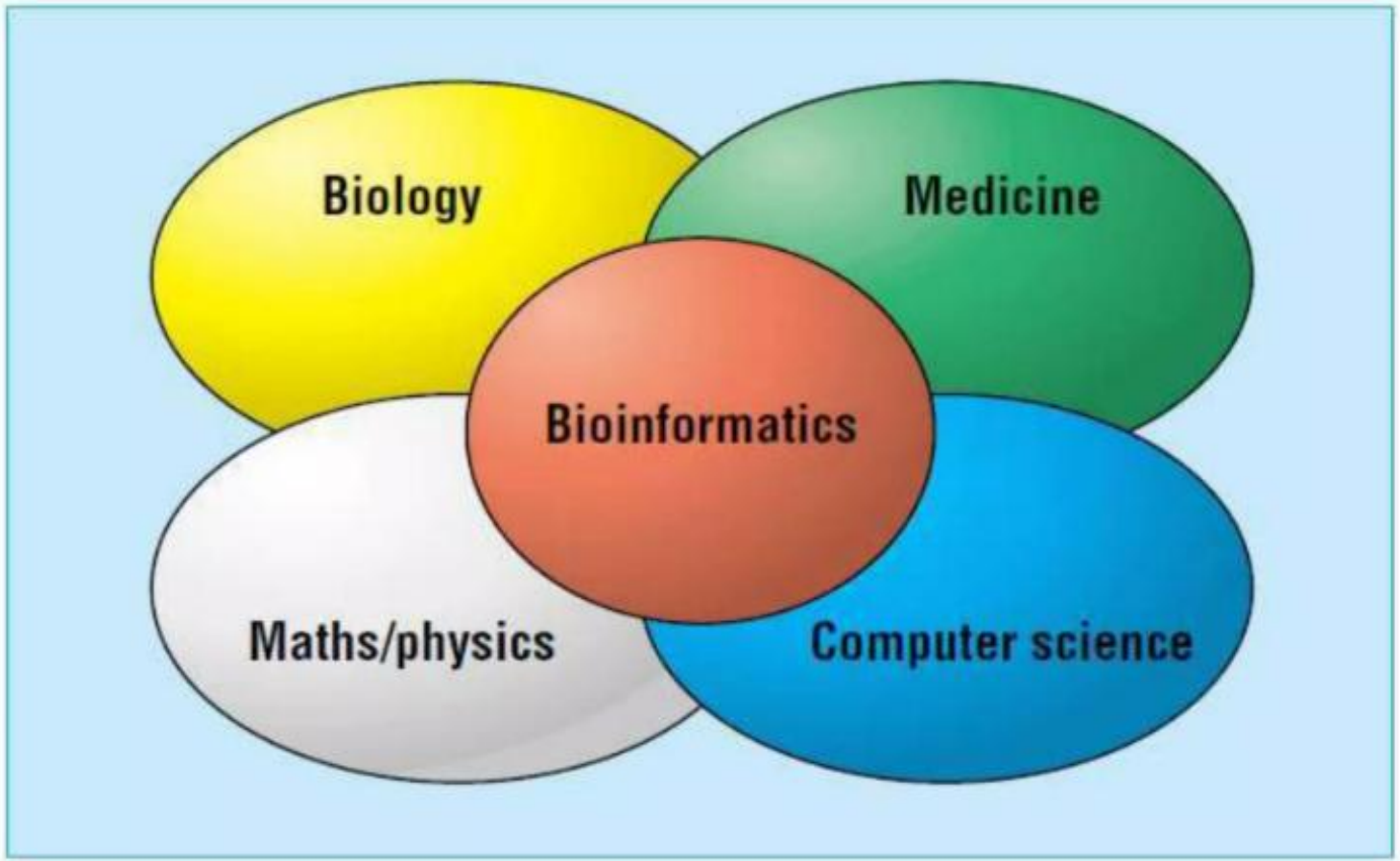
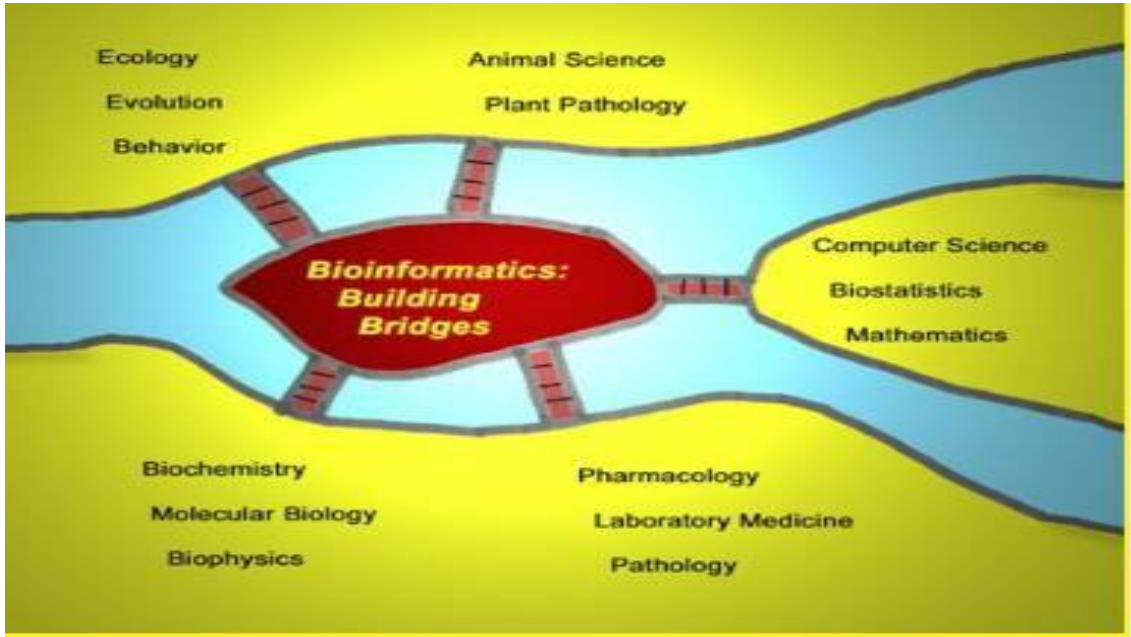
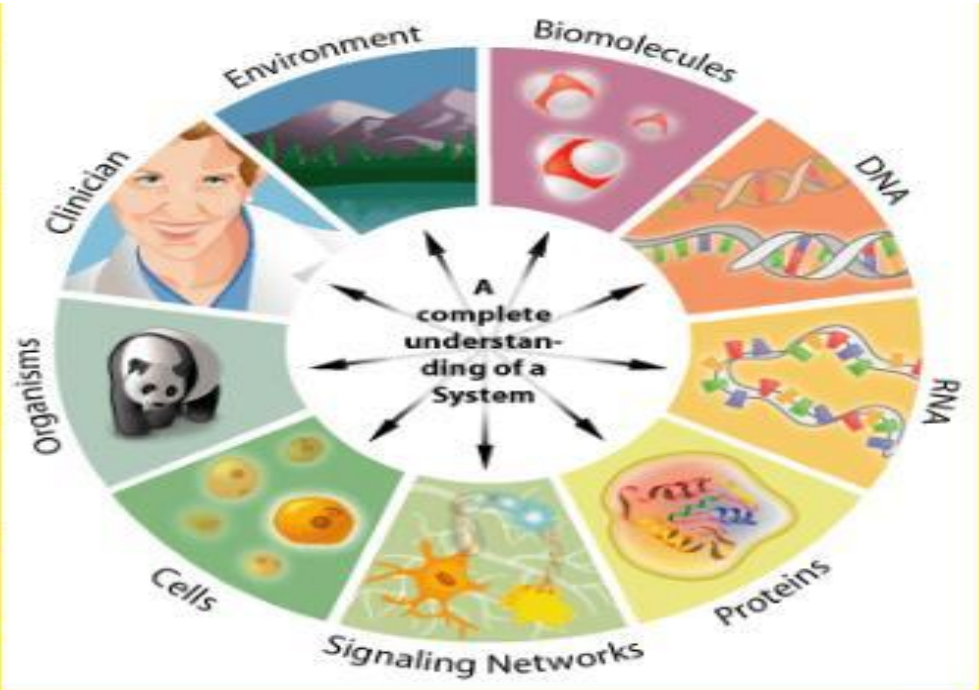


Fig 1 Interaction of disciplines that have contributed to the formation of bioinformatics





What is Bioinformatics?

Bioinformatics

- ***Bioinformatics***: is the development and use of computer applications for the *Analysis*, *Interpretation*, and *Prediction* of biological Systems and corresponding experimental methods in nature sciences.

Bioinformatics:
Studying Biology on Computer

data management; data mining; modeling;
prediction; theory formulation.

Computer system

bioinformatics

genes, proteins, protein complexes, pathways,
cells, organisms, ecosystem

scientific aspect

What is bioinformatics?

- Main goal is to convert complex data into information and knowledge
- Data includes gene and protein sequences, nucleotide sequences .
- Information used to build synthetic and predictive models.
- Future applications in biology, chemistry, pharmaceuticals, medicine, and agriculture

What is Bioinformatics?

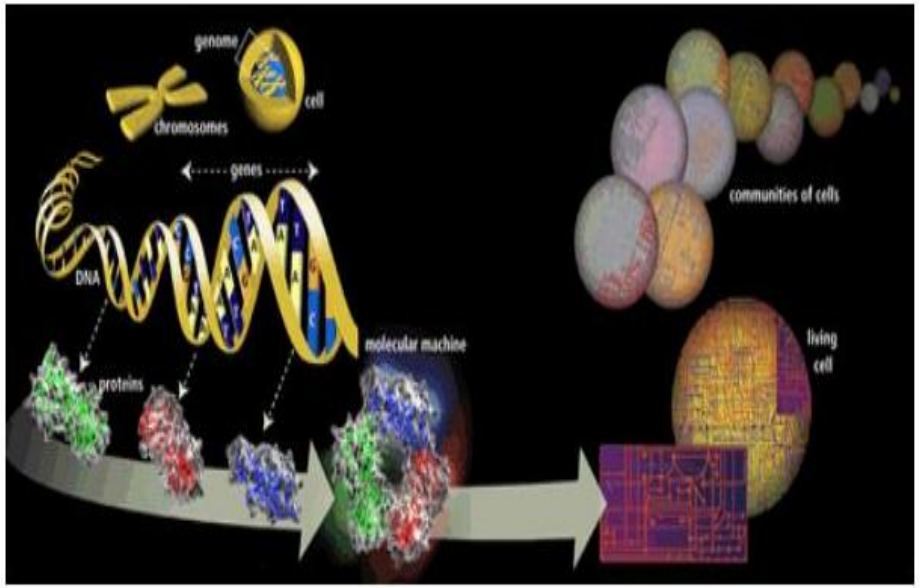
- The field of science in which biology, computer science, and information technology merge into a single discipline.
- The science of managing and analyzing biological data using advanced computing techniques.
- Its goal is to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology can be discerned.

What is done in bioinformatics?

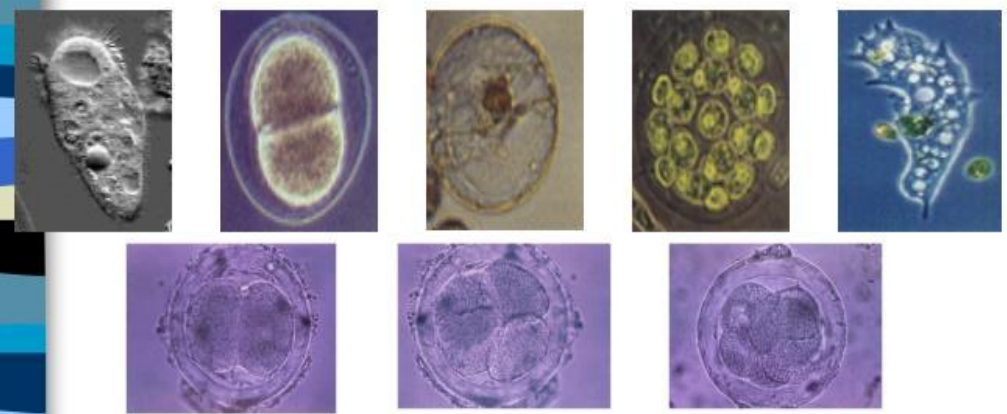
- Development and implementation of tools that enable efficient access and management of different types of information, such as various databases, integrated mapping information.

What "units of information" do we deal with in bioinformatics?

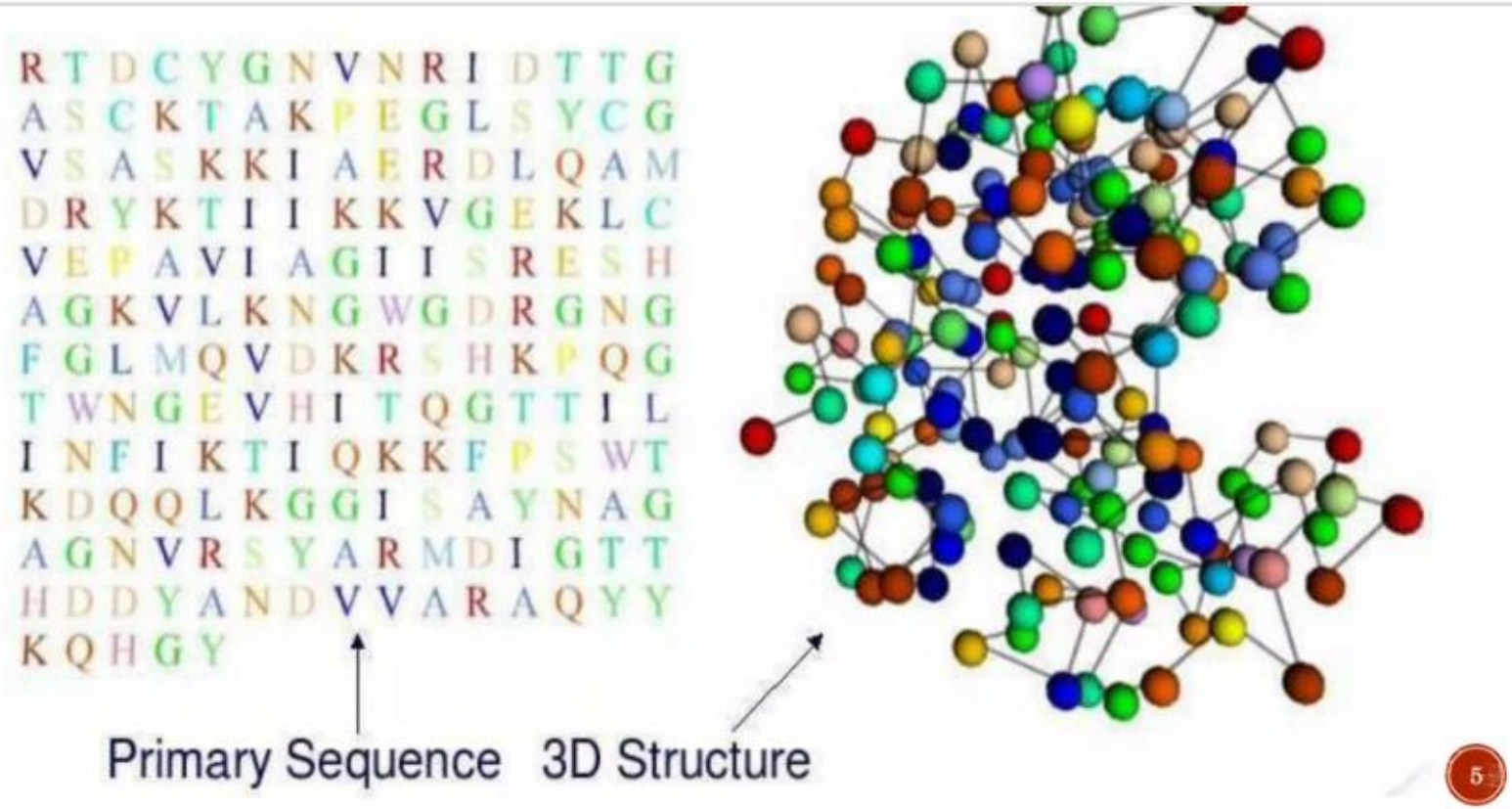
- DNA
- RNA
- Protein
- Sequence
- Structure
- Evolution
- Pathways
- Interactions
- Mutations



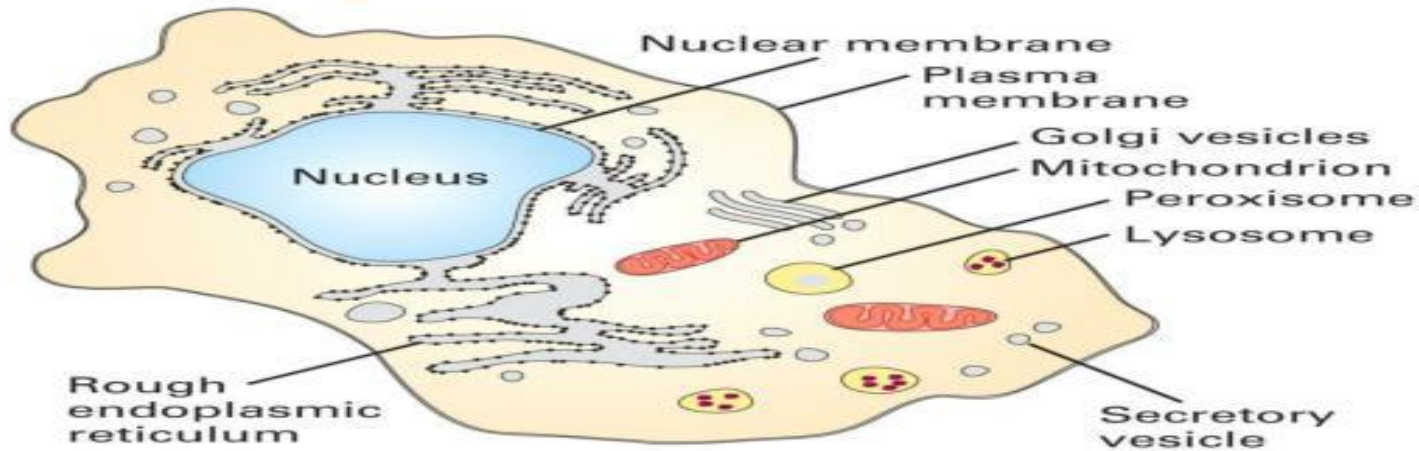
What Is Life Made Of?



All living things are made of cells.

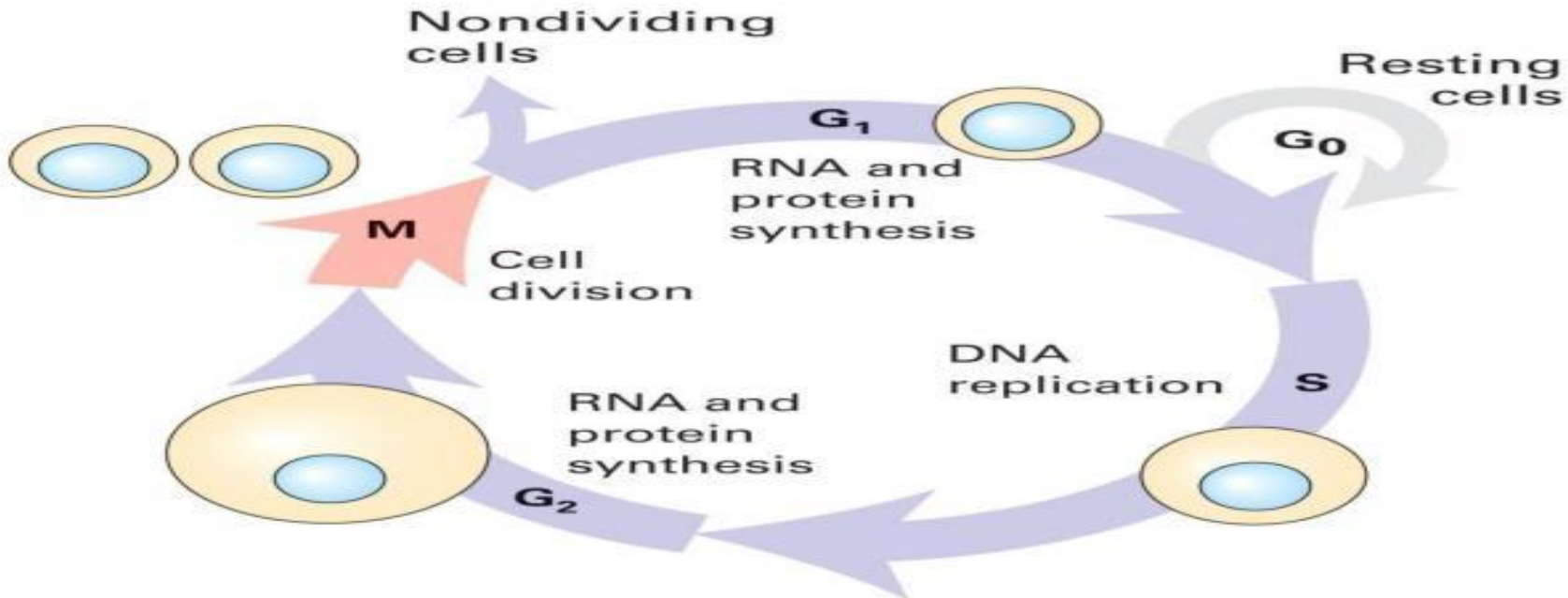

R T D C Y G N V N R I D T T G
A S C K T A K P E G L S Y C G
V S A S K K I A E R D L Q A M
D R Y K T I I K K V G E K L C
V E P A V I A G I I S R E S H
A G K V L K N G W G D R G N G
F G L M Q V D K R S H K P Q G
T W N G E V H I T Q G T T I L
I N F I K T I Q K K F P S W T
K D Q Q L K G G I S A Y N A G
A G N V R S Y A R M D I G T T
H D D Y A N D V V A R A Q Y Y
K Q H G Y
Primary Sequence
3D Structure

Life begins with the cell



- A cell is a smallest structural unit of an organism that is capable of independent functioning
- All cells have some common features

All cells have common cycles



- Born, eat, replicate, and die

Signaling Pathways: Control Gene Activity

- Instead of having brains, cells make decision through complex networks of chemical reactions, called pathways
  - Synthesize new materials
  - Break other materials down for spare parts
  - Signal to eat or die

Cell Information and Machinery

- A cell stores all information to replicate itself
  - Human genome is around 3 billion base pairs long
  - Almost every cell in human body contains same set of genes
  - But not all genes are used or expressed by those cells
- Machinery:
  - Collect and manufacture components
  - Carry out replication
  - Kick-start its new offspring

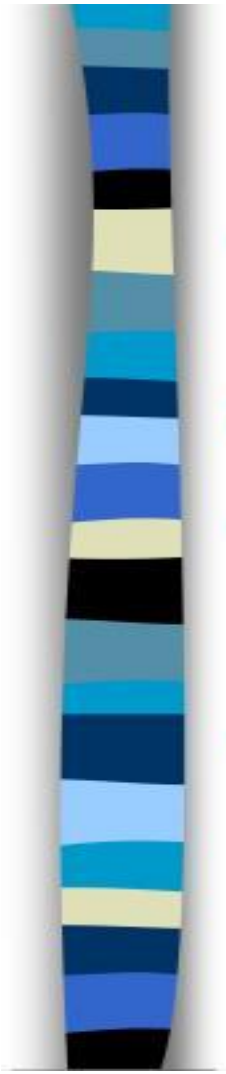
  (A cell is like a car factory)

Overview of organization of life

- **Nucleus = library**
- **Chromosomes = bookshelves**
- **Genes = books**
- Almost every cell in an organism contains the same libraries and the same sets of books.
- Books represent all the information (DNA) that every cell in the body needs so it can grow and carry out its various functions.

Some Terminology

- Genome: an organism's genetic material
- Gene: a discrete units of hereditary information located on the chromosomes and consisting of DNA.
- Genotype: the genetic makeup of an organism
- Phenotype: the physical expressed traits of an organism
- Nucleic acid: biological molecules (RNA and DNA) that allow organisms to reproduce



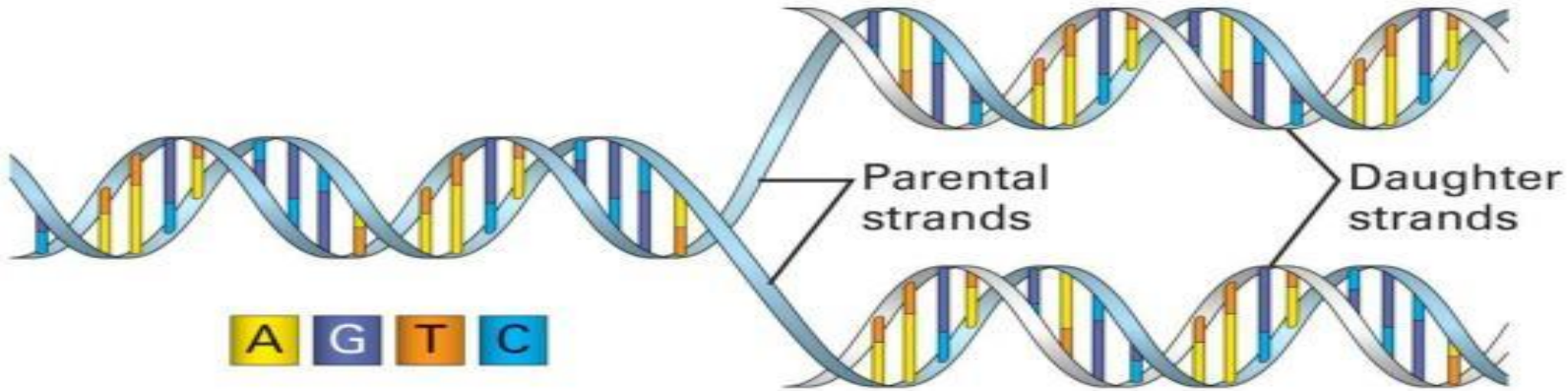
More Terminology

- The **genome** is an organism's complete set of DNA.
  - a bacterium contains about 600,000 DNA base pairs
  - human and mouse genomes have some 3 billion.
- The human genome has 24 distinct chromosomes.
  - Each chromosome contains many **genes**.
- **Gene**
  - basic physical and functional units of heredity.
  - specific sequences of DNA bases that encode instructions on how to make **proteins**.
- **Proteins**
  - Make up the cellular structure
  - large, complex molecules made up of smaller subunits called **amino acids.**

All life depends on 3 critical molecules

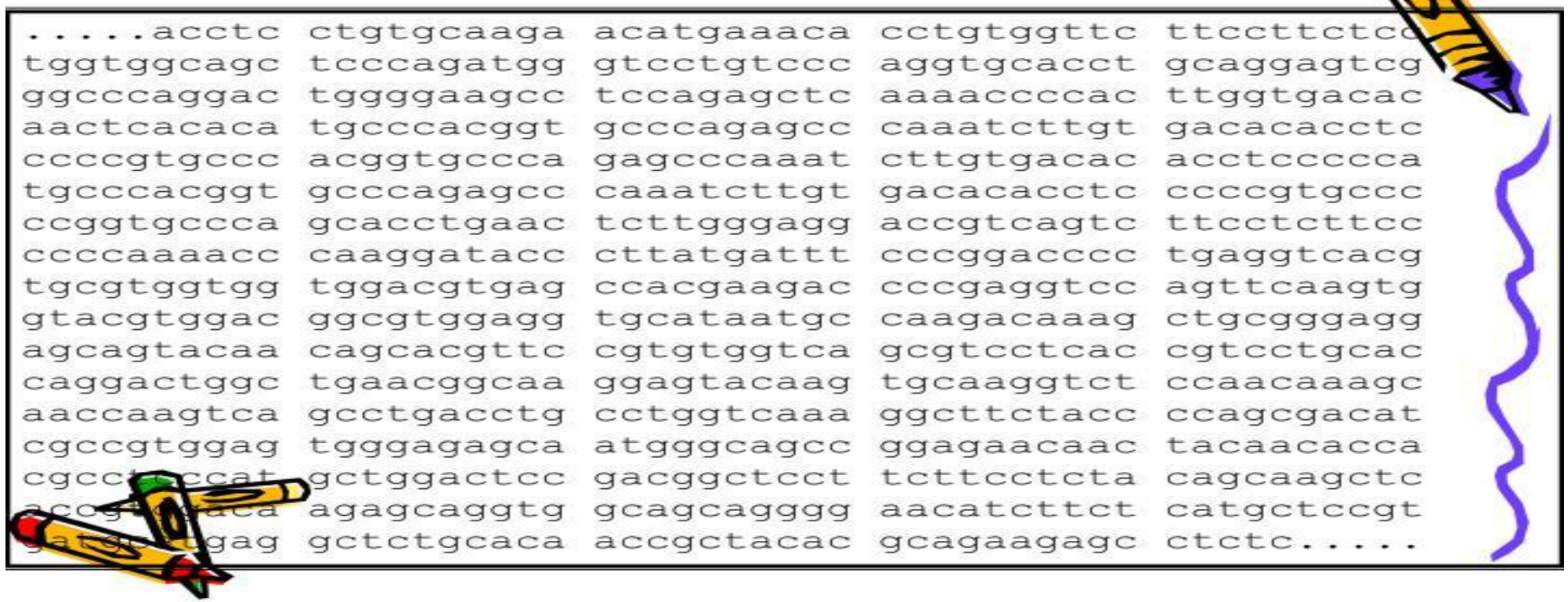
- DNAs
  - Hold information on how cell works
- RNAs
  - Act to transfer short pieces of information to different parts of cell
  - Provide templates to synthesize into protein
- Proteins
  - Form enzymes that send signals to other cells and regulate gene activity
  - Form body's major components (e.g. hair, skin, etc.)

DNA: The Code of Life

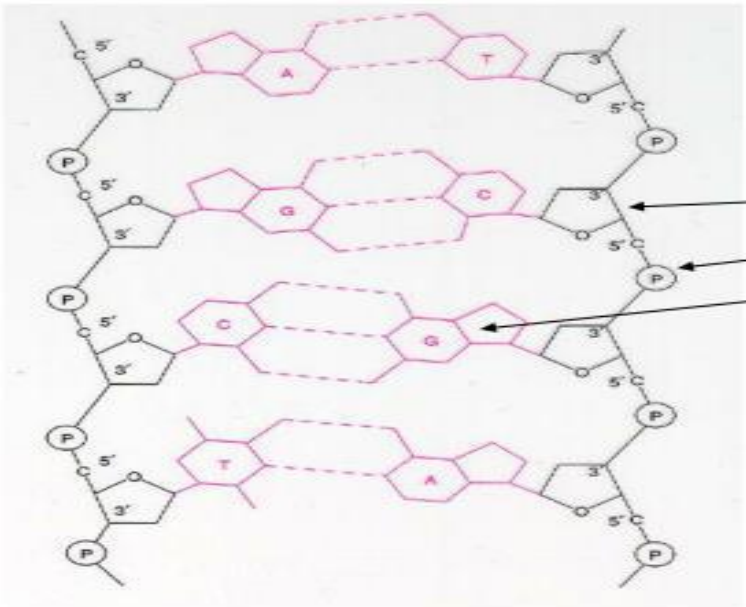


- The structure and the four genomic letters code for all living organisms
- Adenine, Guanine, Thymine, and Cytosine which pair A-T and C-G on complimentary strands.

DNA - Sequence

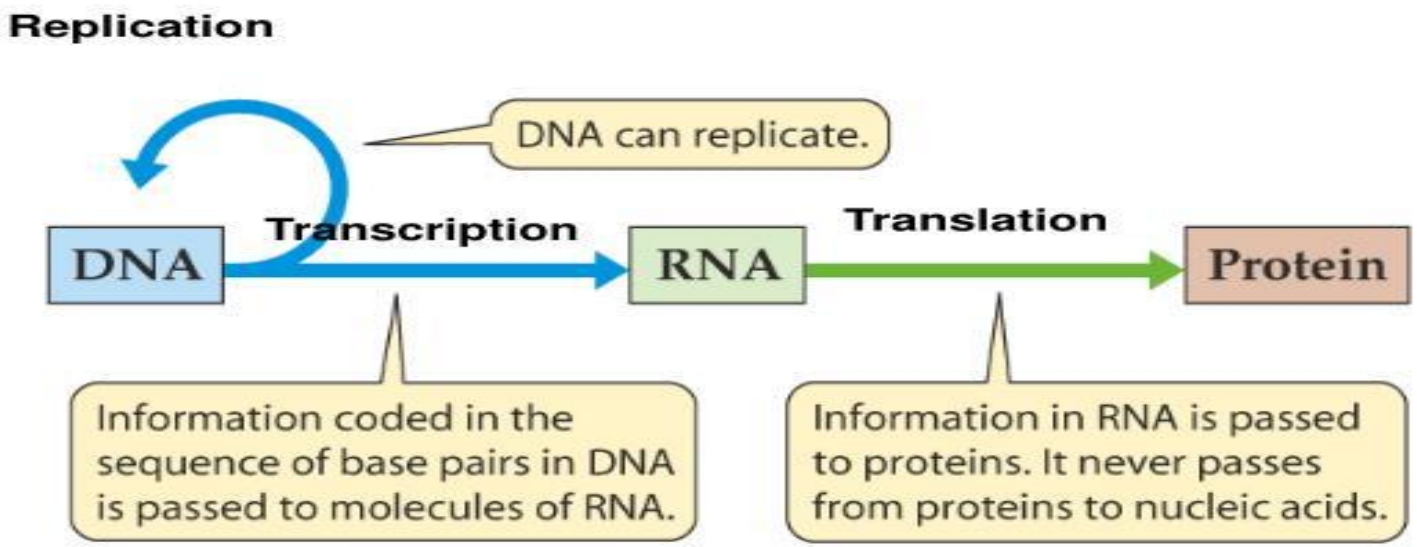


DNA, continued

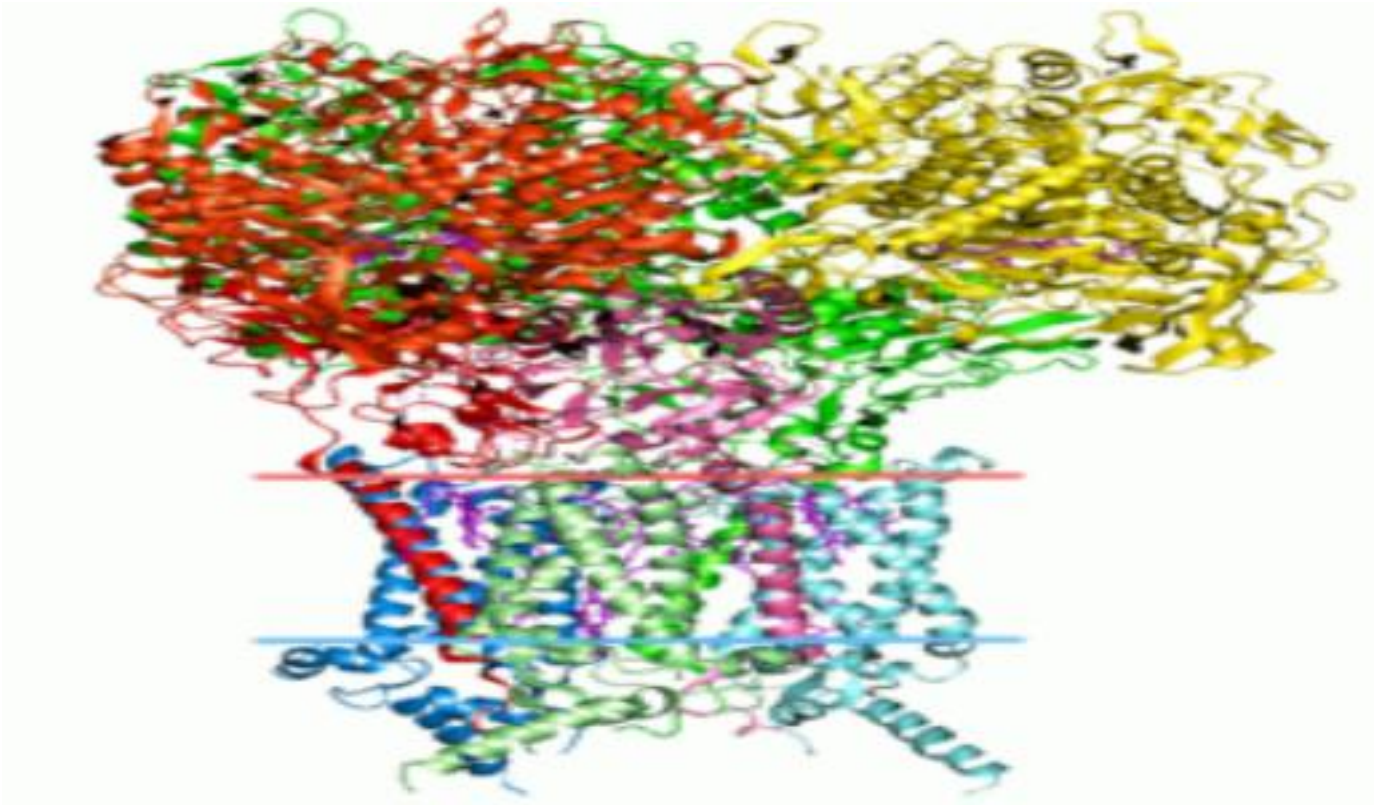


- DNA has a double helix structure which is composed of
  - sugar molecule
  - phosphate group
  - and a base (A,C,G,T)
- DNA always reads from 5' end to 3' end for transcription replication
  5' ATTTAGGCC 3'
  3' TAAATCCGG 5'

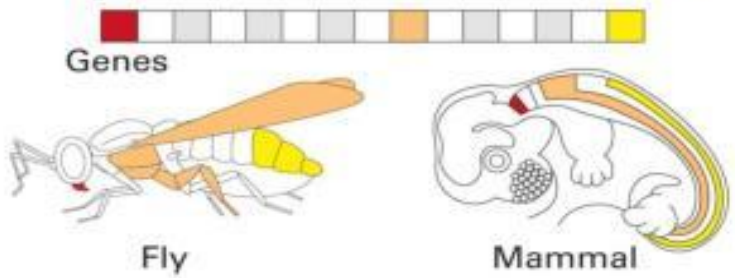
DNA, RNA, and the Flow of Information



3D protein structure



DNA the Genetics Makeup

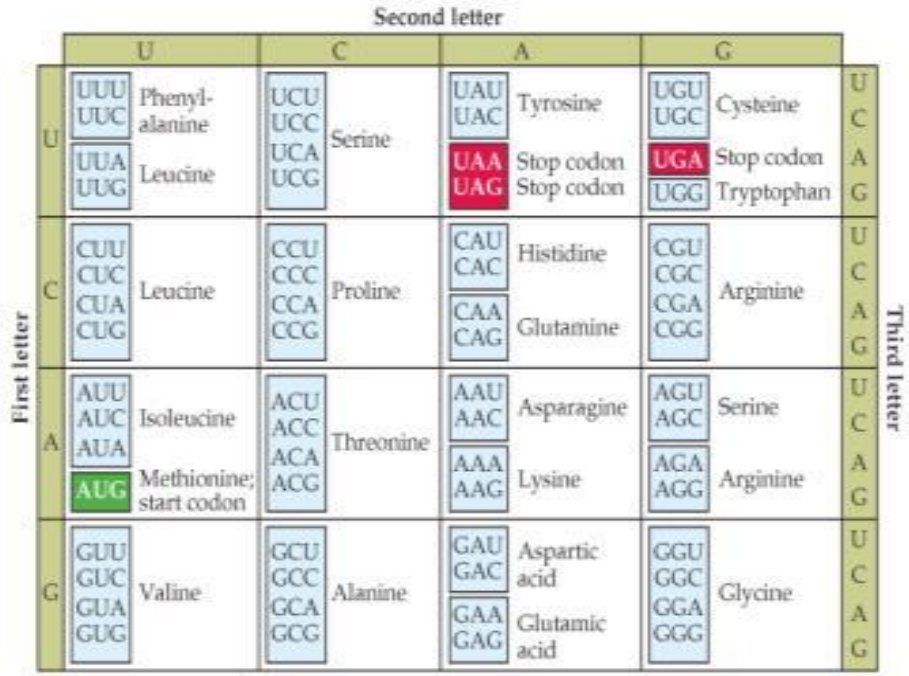


- Genes are inherited and are expressed
  - **genotype** (genetic makeup)
  - **phenotype** (physical expression)

- On the left, is the eye's phenotypes of green and black eye genes.

Cell Information: Instruction Book of Life

- DNA, RNA, and Proteins are examples of strings written in either the four-letter nucleotide alphabet of DNA and RNA (A C G T/U)
- or the twenty-letter amino acid alphabet of proteins. Each amino acid is coded by 3 nucleotides called codon. (Leu, Arg, Met, etc.)



First letter	Second letter: U	Second letter: C	Second letter: A	Second letter: G	Third letter
U	UUU, UUC: Phenylalanine	UCU, UCC, UCA, UCG: Serine	UAU, UAC: Tyrosine	UGU, UGC: Cysteine	U C A G
	UUA, UUG: Leucine		UAA: Stop codon; UAG: Stop codon	UGA: Stop codon; UGG: Tryptophan	
C	CUU, CUC, CUA, CUG: Leucine	CCU, CCC, CCA, CCG: Proline	CAU, CAC: Histidine	CGU, CGC, CGA, CGG: Arginine	U C A G
			CAA, CAG: Glutamine		
A	AUU, AUC, AUA: Isoleucine	ACU, ACC, ACA, ACG: Threonine	AAU, AAC: Asparagine	AGU, AGC: Serine	U C A G
	AUG: Methionine; start codon		AAA, AAG: Lysine	AGA, AGG: Arginine	
G	GUU, GUC, GUA, GUG: Valine	GCU, GCC, GCA, GCG: Alanine	GAU, GAC: Aspartic acid	GGU, GGC, GGA, GGG: Glycine	U C A G
			GAA, GAG: Glutamic acid		

GOALS OF BIOINFORMATICS

- To uncover the wealth of Biological information hidden in the mass of sequence, structure, literature and biological data.
- It is being used now and in the foreseeable future in the areas of molecular medicine.
- It has environmental benefits in identifying waste and clean up bacteria.
- In agriculture, it can be used to produce high yield, low maintenance crops.

FIELD OF BIOINFORMATICS

- Molecular Medicine
- Gene Therapy
- Drug Development
- Microbial genome applications
- Crop Improvement
- Forensic Analysis of Microbes
- Biotechnology
- Evolutionary Studies
- Bio-Weapon Creation

WHERE BIOINFORMATICS HELP?

- In Experimental Molecular Biology
- In Genetics and Genomics
- In generating Biological Data
- Analysis of gene and protein expression
- Comparison of genomic data
- In Simulation & Modeling of DNA, RNA & Protein

OTHER APPLICATIONS

- Waste cleanup
- Microbial genome applications
- Antibiotic resistance
- Alternative energy sources
- Crop improvement and development of resistant varieties
- Forensic analysis
- Bio-weapon creation
- Insect resistance
- Sequence analysis
- Literature analysis

SOFTWARE AND TOOLS OF BIOINFORMATICS

Blast

Basic **L**ocal **A**lignment **S**earch **T**ool.

- It is an algorithm for comparing biological sequences information, such as amino acid sequence of different proteins or the nucleotides of DNA sequences.
- BLAST is used to identify library sequences that resembles the query sequences.

Aims of Bioinformatics

- The various important ways in which bioinformatics can be used.
- The aim of bioinformatics is fourfold and includes data acquisition, tool and database development, data analysis, and data integration.

Data Acquisition:

- Data Acquisition is primarily concerned with accessing and storing data generated directly from the biological experiments.
- The data generated by various sequencing projects have to be retrieved in the appropriate format, and capable of being linked to all the information related to the DNA samples.
- The data are organized in different databases so that the researchers can access existing information.

Tool and Database Development

- Many laboratories generate large volumes of data such as DNA sequences, gene expression information, 3D molecular structure , and highly-throughput screening.
- Consequently, they must develop effective databases for storing and quickly accessing data. The other aim is to develop tools and resources that aid in the analysis of data.

Data Analysis:

- The third aim is to use these tool to analyse the data and interpret the results in a biologically meaningful manner. Efficient analysis require an efficiently deigned database.
- It must allow researchers to place their query effectively and provide them with all the information they need to begin their data analysis.

Conti...

- If queries cannot be performed , or if the performance is too slow, the whole system breaks down since scientists will not be inclined to use the database.

Data Integration :

- Once information has been analysed , a researcher must often associate or integrate it with the related data from the other databases.
- For e.g a scientist may run a series of gene expression analysis experiments and observe that a particular et of 100 genes is more highly expressed in a cancerous lung tissue than in a normal lung tissue.
- The scientist may wonder which of the genes is most likely to be truly related to the disease.

Bioinformatics Applications

Molecular medicine :

- The human genome will have profound effects on the fields of biomedical research and clinical medicine. Every disease has a genetic component.
- This may be inherited or a result of the body's response to an environmental stress which causes alterations in the genome (eg. cancers, heart disease, diabetes.)
- The completion of the human genome means that we can search for the genes directly associated with different diseases and begin to understand the molecular basis of these diseases more clearly.
- This new knowledge of the molecular mechanisms of disease will enable better treatments, cures and even preventative tests to be developed.

Conti...

- **Personalised medicine:**
- Clinical medicine will become more personalised with the development of the field of pharmacogenomics.
  This is the study of how an individual's genetic inheritence affects the body's response to drugs.
- At present, some drugs fail to make it to the market because a small percentage of the clinical patient population show adverse affects to a drug due to sequence variants in their DNA.
- As a result, potentially life saving drugs never make it to the marketplace. Today, doctors have to use trial and error to find the best drug to treat a particular patient as those with the same clinical symptoms can show a wide range of responses to the same treatment.

Conti...

Drug development :

- At present all drugs on the market target only about 500 proteins.
- With an improved understanding of disease mechanisms and using computational tools to identify and validate new drug targets, more specific medicines that act on the cause, not merely the symptoms, of the disease can be developed.
- These highly specific drugs promise to have fewer side effects than many of today's medicines.

Conti...

Gene therapy :

- In the not too distant future, the potential for using genes themselves to treat disease may become a reality.
- Gene therapy is the approach used to treat, cure or even prevent disease by changing the expression of a persons genes.
- Currently, this field is in its infantile stage with clinical trials for many different types of cancer and other diseases ongoing.

NEETHU ASOKAN

Conti...

- **The reality of bioweapon creation :**
- Scientists have recently built the virus poliomyelitis using entirely artificial means.
- They did this using genomic data available on the Internet and materials from a mail-order chemical supply.
- The research was financed by the US Department of Defence as part of a biowarfare response program to prove to the world the reality of bioweapons. The researchers also hope their work will discourage officials from ever relaxing programs of immunisation. This project has been met with very mixed feelings.

NEETHU ASOKAN

Conti.....

Antibiotic resistance :

- Scientists have been examining the genome of Enterococcus faecalisa leading cause of bacterial infection among hospital patients.
- They have discovered a virulence region made up of a number of antibiotic-resistant genes that may contribute to the bacterium's transformation from a harmless gut bacteria to a menacing invader.
- The discovery of the region, known as a pathogenicity island, could provide useful markers for detecting pathogenic strains and help to establish controls to prevent the spread of infection in wards.

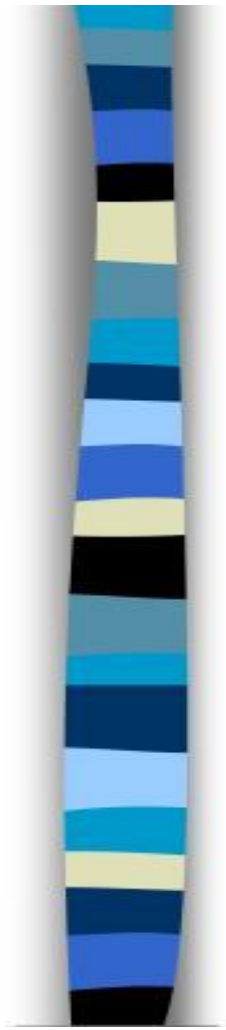
NEETHU ASOKAN

Why Bioinformatics?

- Bioinformatics is the combination of biology and computing.
- DNA sequencing technologies have created massive amounts of information that can only be efficiently analyzed with computers.
- So far 70 species sequenced
  - Human, rat, chimpanzee, chicken, and many others.
- As the information becomes ever larger and more complex, more computational tools are needed to sort through the data.
  - Bioinformatics to the rescue!!!
- Bioinformatics is the electronic infrastructure of molecular biology.

Bio-Information

- Since discovering how DNA acts as the instructional blueprints behind life, biology has become an information science
- Now that many different organisms have been sequenced, we are able to find meaning in DNA through *comparative genomics,* not unlike comparative linguistics.
- Slowly, we are learning the syntax of DNA



What is done in bioinformatics?

- Analysis and interpretation of various types of biological data including: nucleotide and amino acid sequences, protein domains, and protein structures.
- Development of new algorithms and statistics with which to assess biological information, such as relationships among members of large data sets.
- Development and implementation of tools that enable efficient access and management of different types of information, such as various databases, integrated mapping information.

The future…

- Bioinformatics is still in its infancy
- Much is still to be learned about how proteins can manipulate a sequence of base pairs in such a peculiar way that results in a fully functional organism.
- How can we then use this information to benefit humanity without abusing it?