Introduction

- **Genome**[1]
  - A genome is the full set of instructions needed to make every cell, tissue, and organ in your body.
  - four-letter language of DNA (A, C, T, and G).
  - The human genome contains 3 billion of these "letters" or bases.
  - 23 pair of chromosomes
  - Each chromosome contains genes
  - 32,185 genes in human body
  - These genes are the deciding factors for the development of protein

Genome

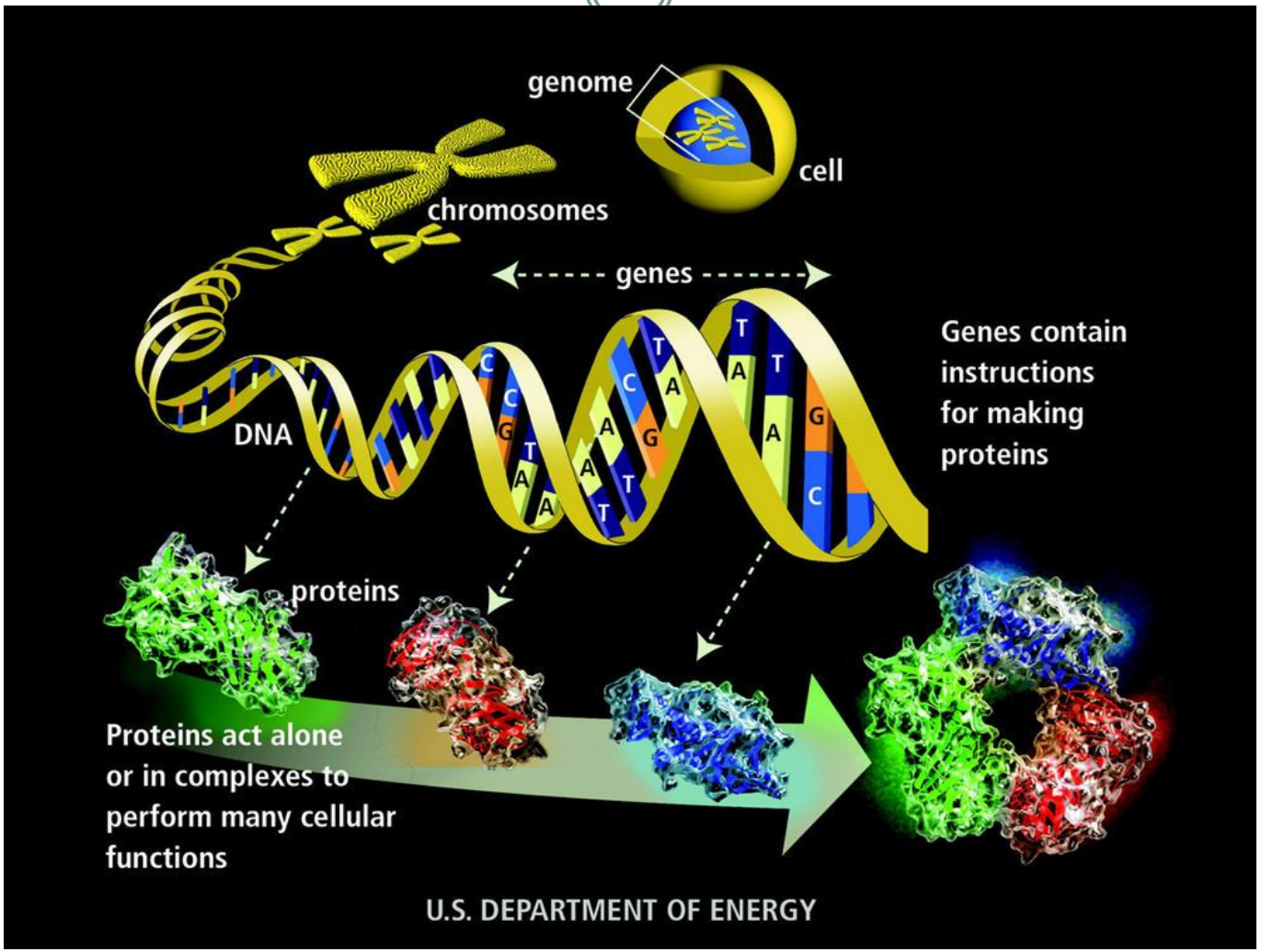

Image courtesy of U.S. Department of Energy Genome Programs

Introduction

- Bioinformatics
  - Field of science in which the computer science and mathematics are used to model and analyze the biological system.
  - Bioinformatics derives knowledge from computer analysis of biological data.
    - Information stored in the genetic code and experimental results from various sources such as patient statistics, scientific literature etc.
  - Areas of research
    - Methods for storage
    - Retrieval
    - Analysis of the data

Introduction

- Analysis of the Biological data
  - Sequence Analysis
    - DNA sequences
      - Adenine(A), Guanine(G), Thymine(T), Cytosine(C)
    - RNA sequences
      - Adenine(A), Guanine(G), Uracil(U), Cytosine(C)
    - Protein sequences
      - Sequence of 20 amino acids
        - Glycine, Alanine, Valine, Leucine, Isoleucine, Serine, Cysteine, Selenocysteine, Threonine, Methionine, Proline, Phenylalanine, Tyrosine, Tryptophan, Histidine, Lysine, Arginine, Aspartate, Glutamate, Asparagine, Glutamine

Introduction

- Applications[3]
  - Comparison of sequences in order to find similarity(Homologous)
  - Identification of intrinsic features of the sequence
    - Active sites-small port of an enzyme where substrate molecules bind
    - Post translational modification sites-polypeptide chains undergo PTM (such as folding, cutting and other processes) before becoming the mature protein product
    - Gene structures-Structure of gene (ATCG)
    - Reading frames-Dividing the sequence of nucleotides in a nucleic acid molecule into non-overlapping triplets.
    - Distributions of introns and exons-Non coding and coding area
    - Regulatory elements-capable of increasing or decreasing the expression of specific genes
  - Identification of sequence differences and variations such as point mutations and single nucleotide polymorphism (SNP) in order to get the genetic marker.
  - Revealing the evolution and genetic diversity of sequences and organisms

Analysis of the Biological data

- Structural analysis[4]
  - Analysis and prediction of the three-dimensional structure of biological macromolecules such as proteins, RNA, and DNA.
- Applications
  - Comparisons of overall folds and local motifs
    - Structural motif:- a pattern in a protein structure formed by the spatial arrangement of amino acids
  - Principles of molecular folding
    - Physical process by which a polypeptide folds into its characteristic and functional three-dimensional structure
  - Evolution
    - Change in the inherited characteristics of biological populations over successive generations.
  - Binding interactions
    - Interaction between two molecules that results in a stable association in which the molecules are close to each other.
  - Structure/function relationships

Analysis of the Biological data

- Expression analysis [5]
  - Gene expression
    - Gene expression is the process by which information from a gene is used in the synthesis of a functional gene product.
      - DNA transcribed to messenger RNA (mRNA)
      - mRNA is then translated into polypeptide chains
      - which are ultimately folded into proteins
  - Protein expression
    - subcomponent of gene expression

Gene expression analysis

- Gene expression is the most fundamental level at which the genotype of an organism (internal blueprint of genetic information) gives rise to the phenotype (the outward physical appearance).

- How each gene is expressed under a particular condition.

- The expression of each gene in a particular biological condition can be expressed as Gene expression profile

Gene expression analysis

- Gene expression profiling
  - Measurement of the expression of thousands of genes at once, to create a global picture of cellular function.
  - If a gene is used to produce mRNA, it is considered "on", otherwise "off".
  - The sequence tells us what the cell could possibly do, while the expression profile tells us what it is actually doing at a point in time.
  - Many factors determine whether a gene is on or off, such as the time of day, whether or not the cell is actively dividing, its local environment, and chemical signals from other cells.
  - The technology that is used to develop Gene profiling is microarray technology.

Gene expression analysis

- Microarray Technology
  - DNA microarray is a modern technology, which is used to analyze the interactions between thousands of genes in parallel.
  - A DNA microarray (also commonly known as gene chip, DNA chip, or biochip) is a collection of microscopic DNA spots attached to a solid surface.
  - Steps in microarray technology

1. Prepare a gene chip with probe sequences
2. Label green for the uninfected sample 1
3. Label red for the infected sample 2
4. Mix both the samples
5. Hybridize with the probe mixture
6. Measure the amount or intensity of green and red in the filtered mixture and compare the amounts to determine the relative gene expression ratio

MICROARRAY Technology

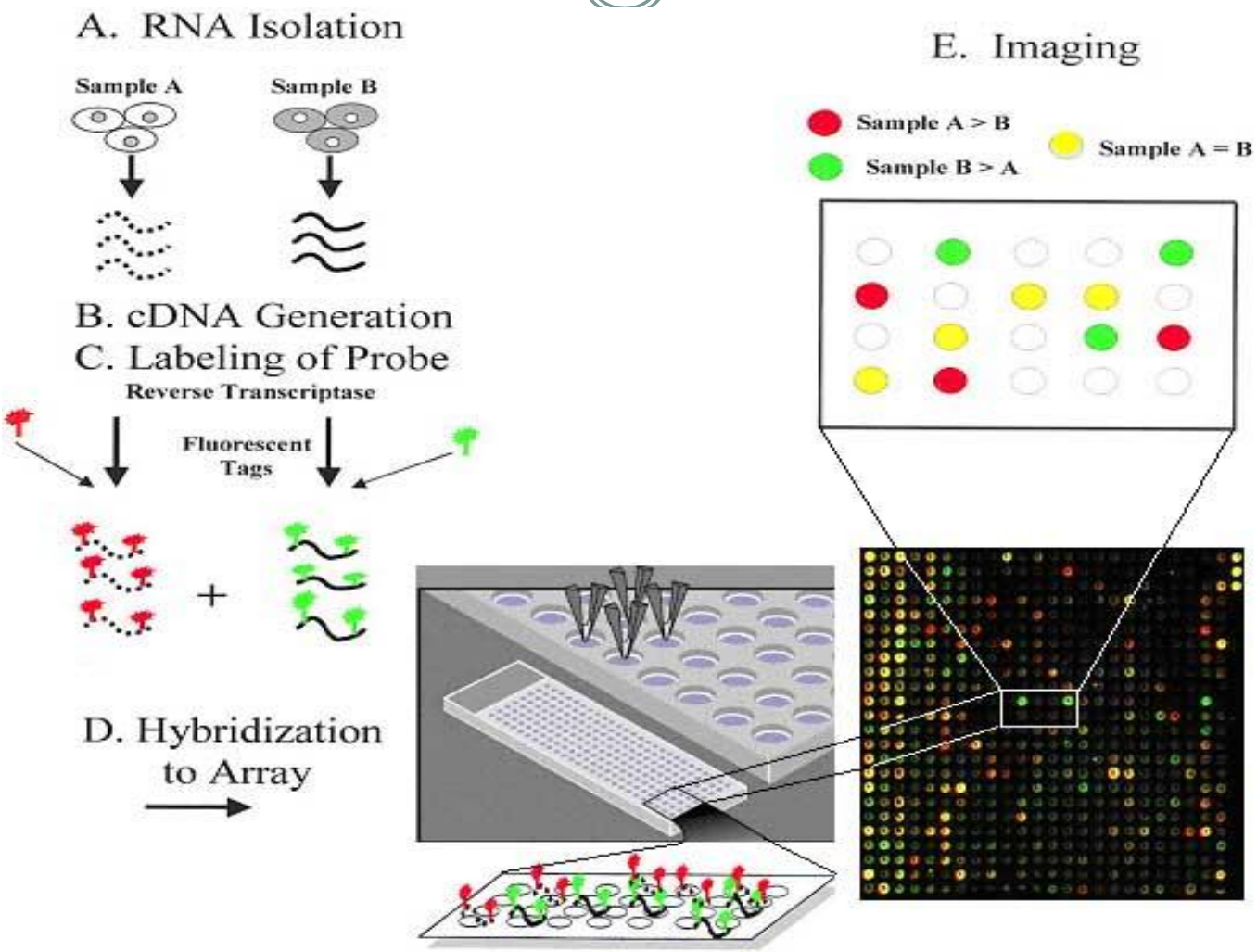

Image courtesy of www.eufic.org

MICROARRAY Technology

Time / Genes	T1	T2	T3	T4	T5	T6	.....	Tm
Gene 1	..	..	..	..	..	..	..	..
Gene 2	..	..						
Gene 3								
Gene 4								
. . .								
Gene n						...	...	...