

GENOMICS & PROTEOMICS

WHAT DO YOU MEAN BY GENOMICS?

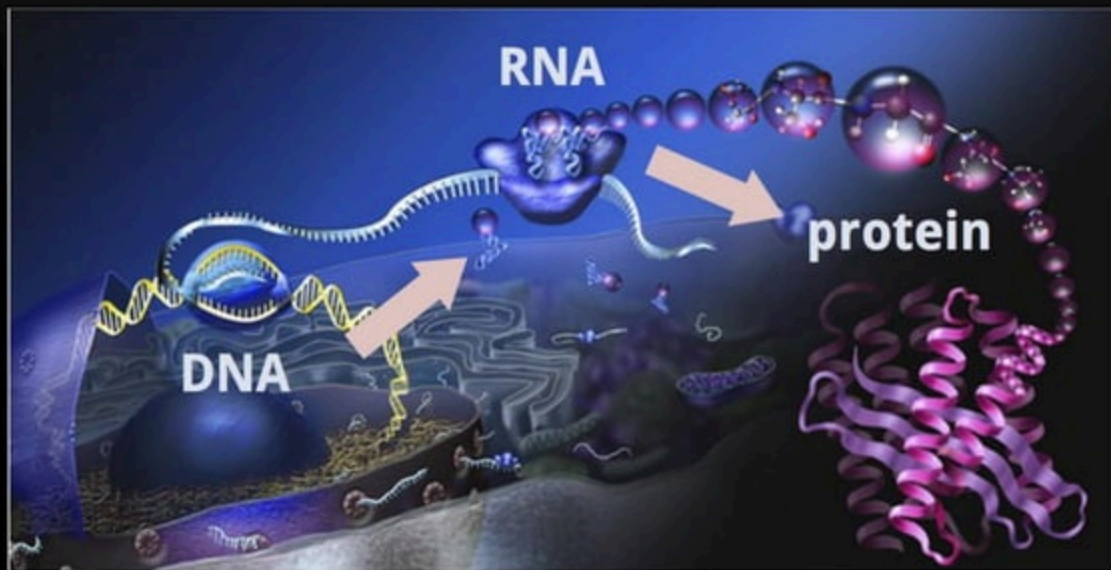
The term genome introduced by **H. Winkler** in 1920

The term genomics coined by **T.H. Roderick** in 1987

Genome + Omics → Genomics

Genomics is a an area of life science that deals with
the study of the genomes of organisms

CENTRAL DOGMA OF MOLECULAR BIOLOGY



- **Today genomics includes:**

- ✓ sequencing of genomes
- ✓ determination of the complete set of proteins encoded by an organism
- ✓ the functioning of genes and metabolic pathways in an organism

Where do we get these sequences from?

- ✓ Through genome sequencing projects



The **Genome** Is **All The DNA** In A Cell

- All the DNA on all the chromosomes
- Includes genes, intergenic sequences, repeats
- Specifically, it is all the DNA in an organelle
- **Eukaryotes can have 2-3 genomes**
 - Nuclear genome
 - Mitochondrial genome
 - Plastid genome
- If not specified, "genome" usually refers to the nuclear genome

How Many Types Of Genome???

- **Prokaryotic genomes**
- **Eukaryotic Genomes**
 - *Nuclear Genomes*
 - *Mitochondrial genomes*
 - *Choloroplast genomes*

GENOME SEQUENCING- HISTORY

- The first genome to be sequenced was that of *Haemophilus influenzae* in 1995.
- The *E. coli* genome was completely sequenced in 1997.
- Yeast (*Saccharomyces cerevisiae*) (12.8 x 10⁶ bp) and worm (*Caenorhabditis elegans*) genomes were the first eukaryotic genomes to be sequenced in 1999.
- Genomes of *Drosophila melanogaster* and *Arabidopsis thaliana* were sequenced in 2000.

GENOME SEQUENCING PROJECT

Human Genome project

- The Human Genome Project officially began on Oct. 1, 1990.
- Completed in 13 years
- **Mission of HGP:**
- To understand the human genome and the role it plays in both health and disease.
- The U.S. govt. project coordinated by the Department of Energy and the National Institutes of Health
- **Francis Collins**, Director of the HGP and the National Human Genome Research Institute (**NHGRI**)



THE GENOME IS OUR GENETIC BLUEPRINT

- Nearly every human cell contains 23 pairs of chromosomes
 - 1 - 22 and XY or XX
 - XY = Male**
 - XX = Female**
- Length of chr 1-22, X, Y together is ~3.2 billion bases



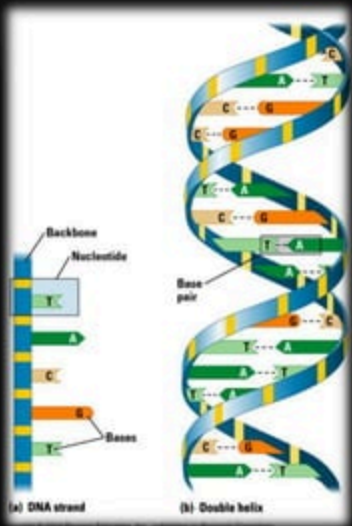
The genome is who we are on the inside!

- **Chromosomes consist of DNA**

- molecular strings of A, C, G, & T
- base pairs, A-T, C-G

- **Genes**

- DNA sequences that encode proteins
- less than 3% of human genome



AIMS OF THE PROJECT:

- To identify the approximate 100,000 genes in the human DNA.
 - Determine the sequences of the 3 billion bases that make up human DNA.
 - Store this information in databases.
 - Develop tools for data analysis.
 - Address the ethical, legal, and social issues that arise from genome research.
-

Beginning of project

- The first 10 years of the project were spent improving the technology to sequence and analyze DNA.
- Scientists all around the world worked to make detailed maps of our chromosomes and sequence model organisms, like worm, fruit fly, and mouse.



How was it done...

First there was the Assembly

- The DNA sequence is so long that no technology can read it all at once, so it was broken into pieces.
 - There were millions of **clones** (small sequence fragments).
 - The assembly process included finding where the pieces overlapped in order to put the draft together.
-

UCSC put the human genome sequence on the web July 7, 2000

- UCSC put the human genome sequence on CD in October 2000, with varying results



The screenshot shows the UCSC Genome Browser website. At the top, there are logos for Santa Cruz and UCSC, followed by the text "Genome Browser". Below this is a navigation bar with links: Home, Genesets, Genome Browser, Tools, Mirrors, Downloads, My Data, Help, and About Us. The main content area is divided into two columns. The left column features a large, glowing blue DNA double helix graphic. The right column is titled "Our tools" and lists several tools with brief descriptions:

- **Genome Browser**
Interactively visualize genomic data
- **BLAT**
rapidly align sequences to the genome
- **Table Browser**
download data from the Genome Browser databases
- **Variant Annotation Integrator**
get functional effect predictions for variant calls
- **Data Integrator**
combine data sources from the Genome Browser databases
- **Gene Sorter**
find genes that are similar by expression and other metrics
- **Genome Browser in a Box (GBIB)**
run the Genome Browser on your laptop or server
- **In-Silico PCR**
rapidly align PCR primer pairs to the genome
- **LiftOver**
convert genome coordinates between assemblies
- **VisiGene**
Interactively view in situ images of mouse and frog

More tools...

Next ...the Annotation

- Where are the **genes**?
- **How do genes work?**
- How do scientists use this information for scientific understanding and to benefit us?
- **What do genes do anyway?**
- We only have ~27,000 genes, so that means that each gene has to do a lot.
- **Genes make proteins that make up nearly all we are (muscles, hair, eyes).**
- Almost everything that happens in our body happens because of proteins
- **(walking, digestion, fighting disease).**



or



Eye Color is determined by genes

From our genome so far...

- Relatively small number of human genes, less than 30,000
- Have a complex architecture (which is yet to be analyzed completely)
- We know where 85% of genes are in the sequence.
- We don't know where the other 15% are because we haven't seen them "on" (they may only be expressed during fetal development).
- We only know what about 20% of our genes do so far.

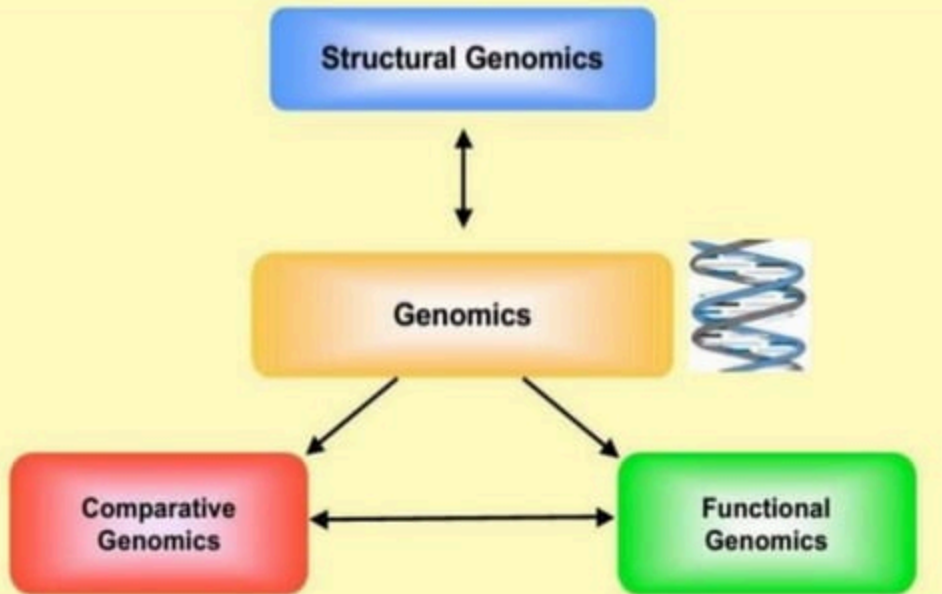


What Does The Draft Human Genome Sequence Tell Us?

- The human genome contains 3.2 billion chemical nucleotide bases (A, C, T, & G)
- Takes **95** years to read



Types of genomics :



Model
systems



Agriculture
Medicine

(synteny/allelic diversity)

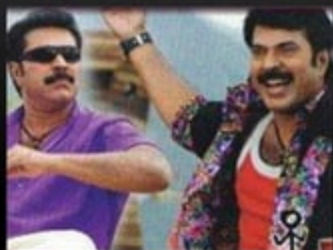
(αλληλεγγύη/αλληλική ποικιλία)

Genome



Phenome

Sequence Similarity/ Dissimilarity??



0.001%



7%



95-98%



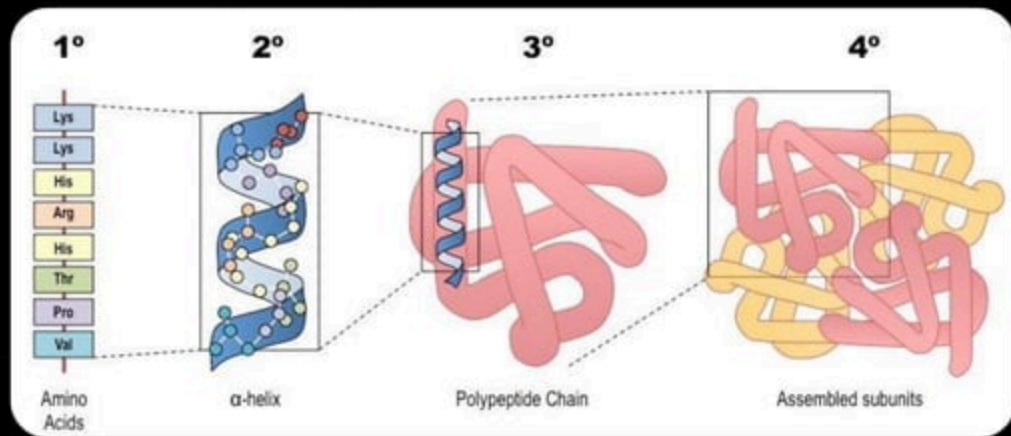
36%



90%

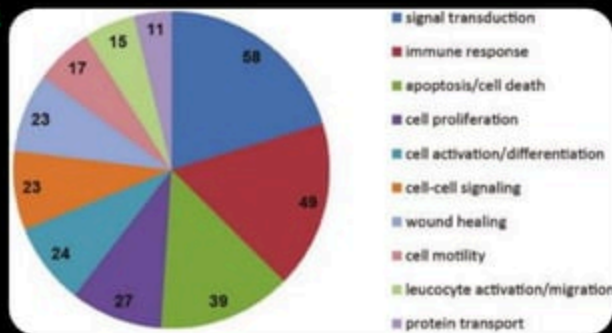
STRUCTURAL GENOMICS

- Effort aimed at determining the three-dimensional **structures** of gene products
- Using efficient and high-throughput mode
- **For Proteins- Structural proteomics!**
- Understanding novel proteins and 3D structures



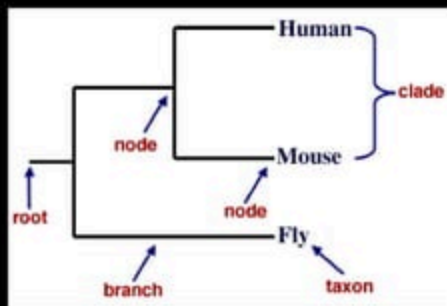
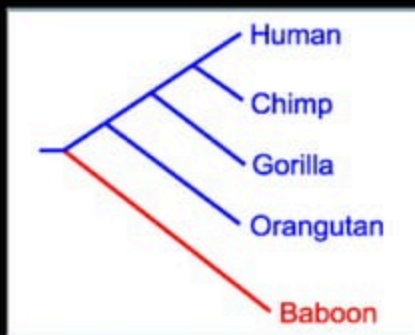
FUNCTIONAL GENOMICS

- Identify functions of gene and non-gene sequences
- Describe gene & protein functions
- Gene & Protein interaction
- Genotype- Phenotype



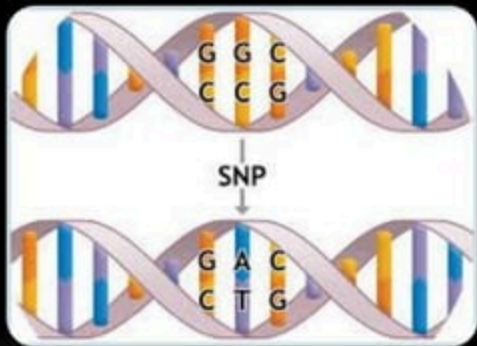
COMPARATIVE GENOMICS

- Compare genome sequence between different species
- To better understand the evolutionary relationships
- Determine the function of each genome



MUTATIONAL GENOMICS

- Study of genome in terms of mutations that occur in the DNA or Genome of an individual
- Also termed as **gene function determination**
- Understand the mutations in
 - ✓ Coding sequences
 - ✓ Non coding sequences
- Due to Repeat sequences:
 - ✓ Minisatellites
 - ✓ Microsatellites
- **SNP**



TRANSCRIPTOMICS

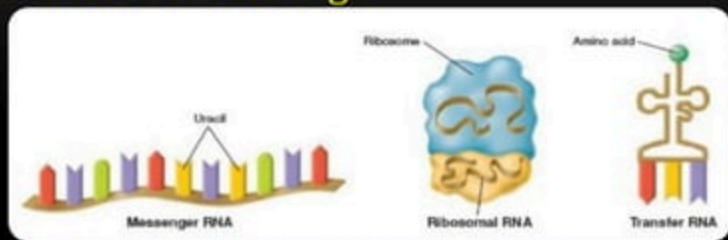
- The set of all RNA molecules including:

- ✓ mRNA

- ✓ rRNA

- ✓ tRNA

- ✓ non-coding RNA produced in one or a population of cells



Transcriptomics, is a global way of looking at gene expression patterns

TRANSCRIPTOME PROFILING

- Deep investigation of the transcriptome
- Study the transcriptional activity
- Proteins coded by the RNA transcript
- Study gene fusions etc...

Annotate the RNA transcript

PROTEOMICS

We all are made
up of proteins

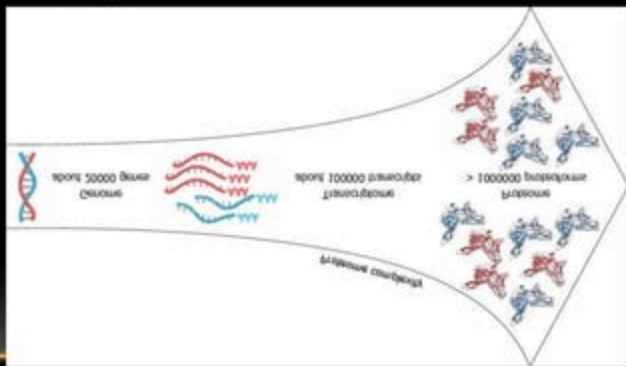


WHY PROTEOMICS?

Fact:

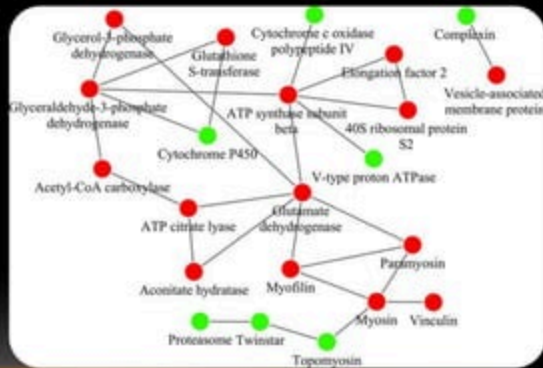
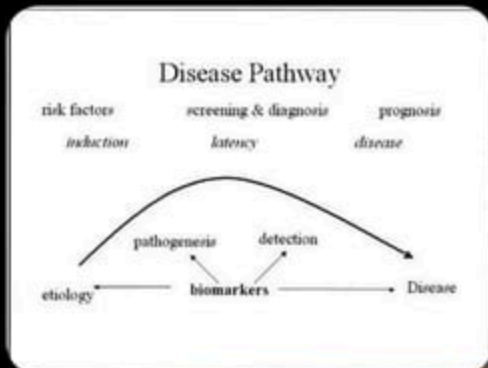
- Genome ~ 26,000-31,000 protein encoding genes
- Human proteins ≥ 1 million
- Proteomics –
- Study of the full protein complement of organisms

e.g. plasma, cells and tissue



UNDERSTANDING THE PROTEOME ALLOWS...

- Characterisation of proteins
- Understanding protein interactions
- Identification of disease biomarkers



MAJOR APPLICATIONS...

GENOMICS, TRANSCRIPTOMICS, PROTEOMICS

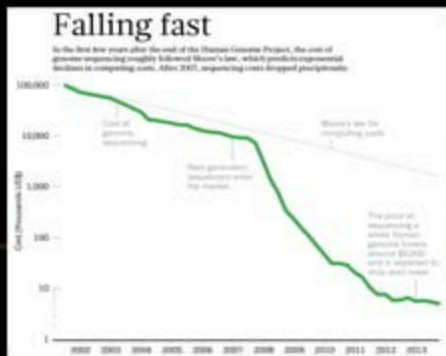
- Gene prediction
- ORF Finding
- Metagenomics
- Next Generation Sequencing
- Computer Aided Drug Design

NEXT GENERATION SEQUENCING

- DNA sequencing technology which has revolutionised genomic research
- Determining the number and order of nucleotides that make up a given molecule of DNA.
- Using NGS an entire human genome can be sequenced within a single day.
- In contrast to the previous Sanger sequencing technology
- A number of different modern sequencing technologies including:

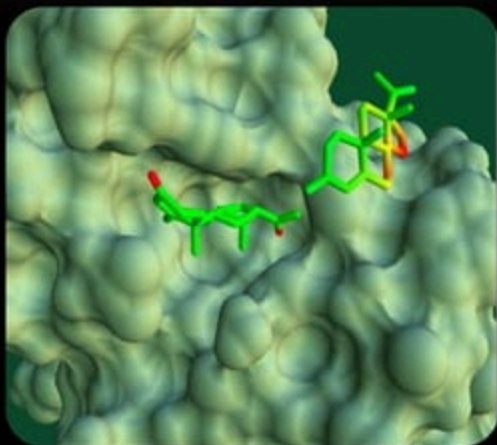
Illunmina, Roche 454 sequencing, Ion Torrent , PacBio etc.

- Cost



Method	Read Length
Sanger	600-1000 bp
454	300-500 bp
Illumina	~100 bp
Ion Torrent	~200 bp

COMPUTER AIDED DRUG DESIGN



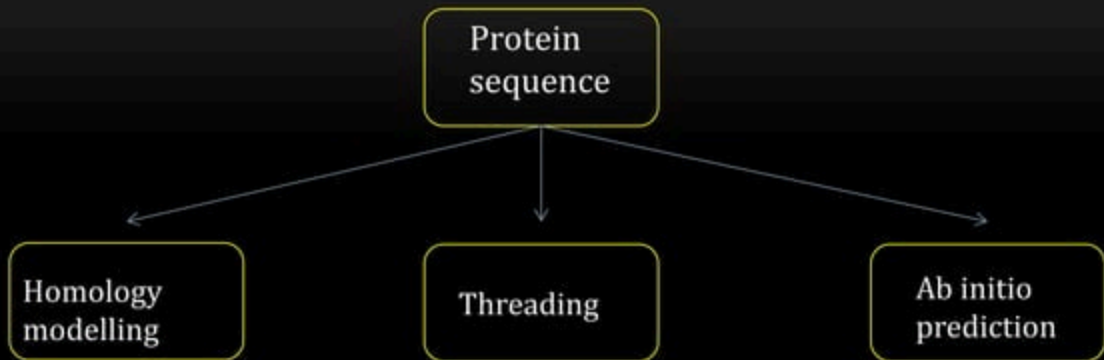
COMPUTER AIDED DRUG DISCOVERY



**Choosing the right
“one” among
millions: *virtual
screening***

**Designing the right
“one” from chemical
building blocks:
*de novo drug design***

PREDICT TERTIARY STRUCTURE



- Find homologous sequence
- Homology > 30%
- Keeping in view of the template structure

**Swiss PDB Viewer ,
MODELLER**

If homologous sequence is < 30 % similar we use this method

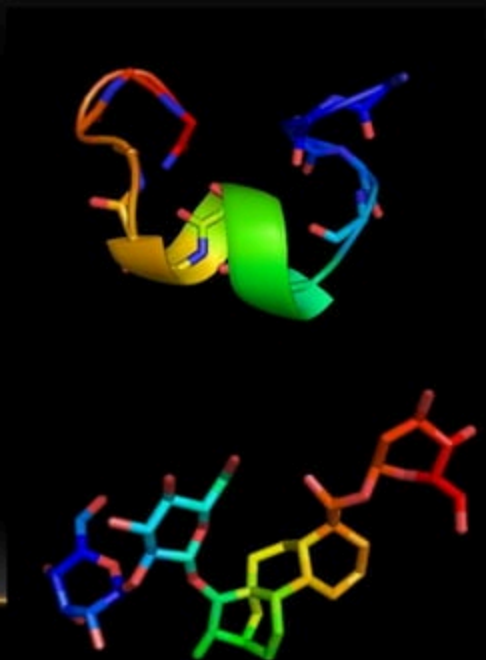
**iTASSER,
PHYRE**

Prediction of structure from scratch using the knowledge of amino acid properties

ROSETTA

STRUCTURE VISUALIZATION

RASMOL
MOLMOL
PYMOL
SPDBV



TIME & MONEY ...

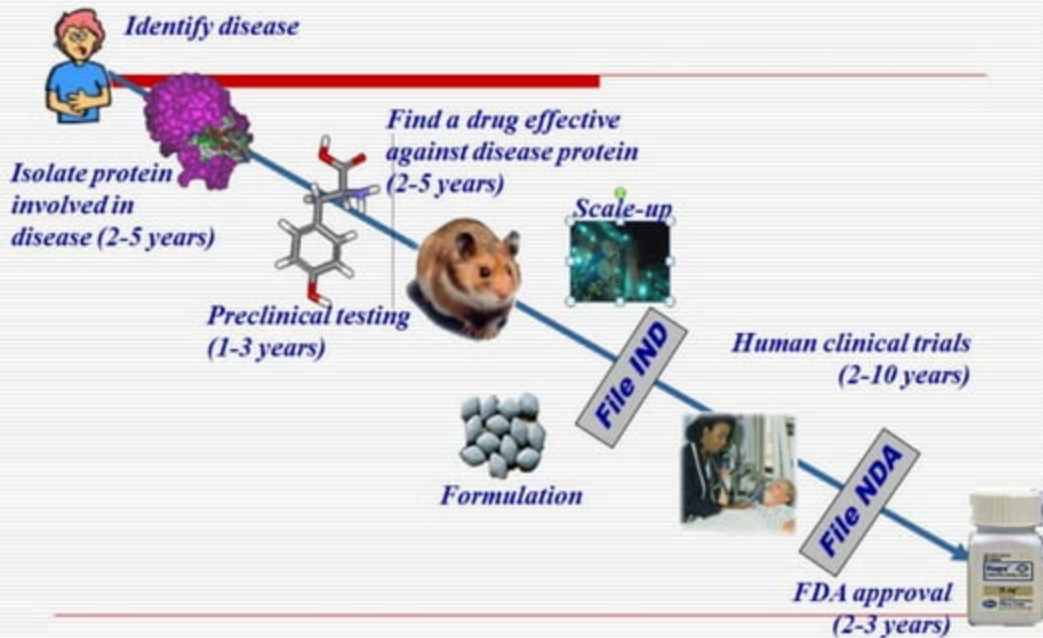
- 10-12 Years
- 1 Drug/Year
- Rs 400 Crores
- 5000 to even 50000 screenings

Returns too are striking...

- *Lipitor*, cholesterol reducer from Pfizer sold for 8.6 Billion US\$ in 2001



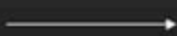
DRUG PIPELINE



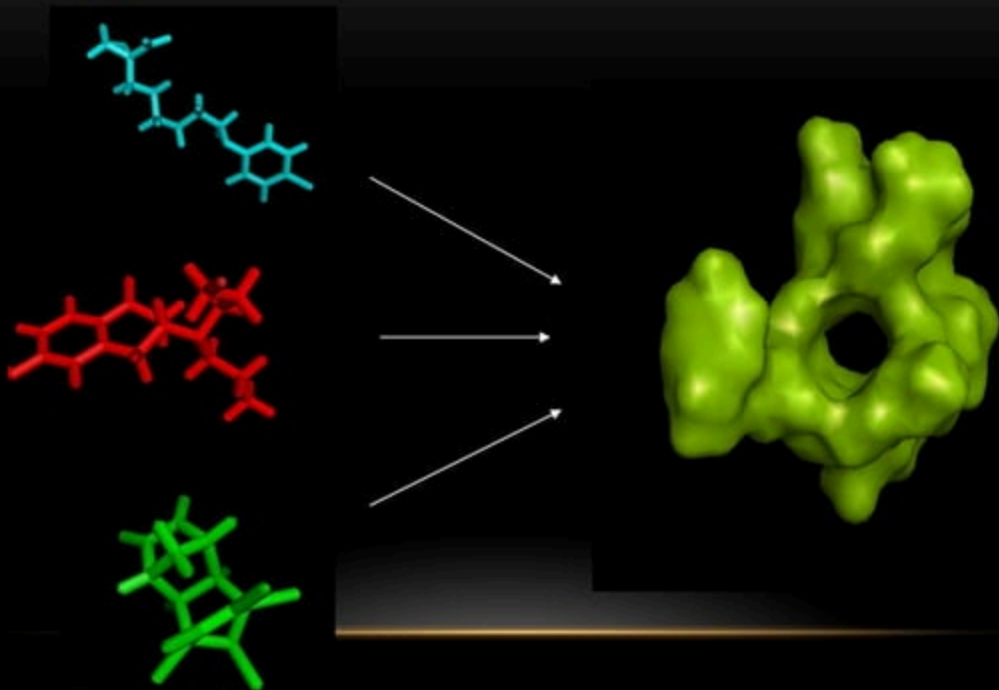
IMPORTANT TERMS

- **Target**- a molecule important in a disease-usually a protein
- **Ligand**- a small molecule binds to a larger one
- **Active site**- ligand binding site
- **Hit**- a ligand which can geometrically fit to the binding site
- **Lead**- hit with biological activity
- **DRUG**- Ligand that can modulate the function of target in desired way

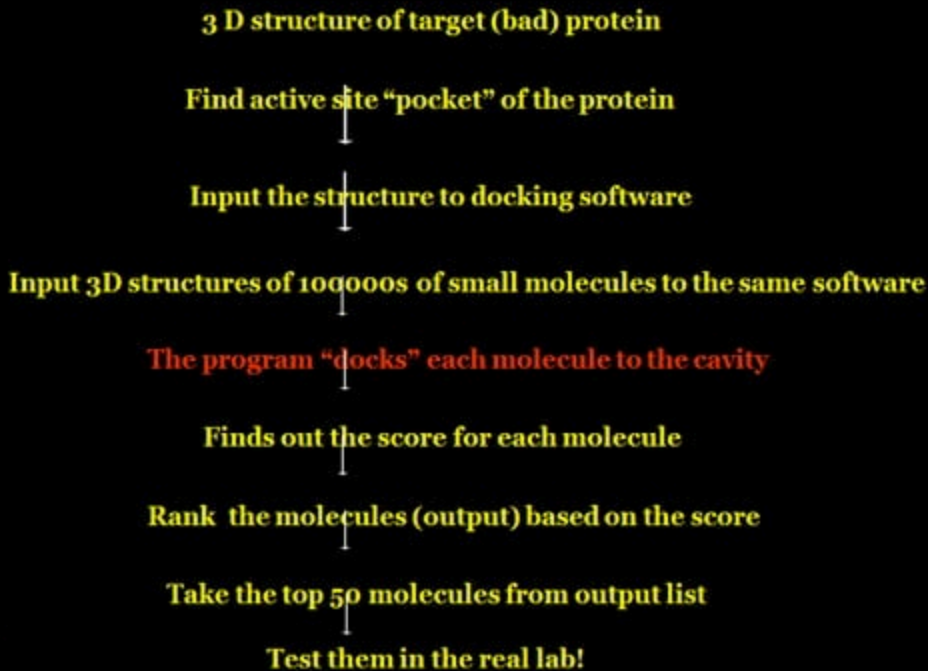
20000 molecules



30 - 50

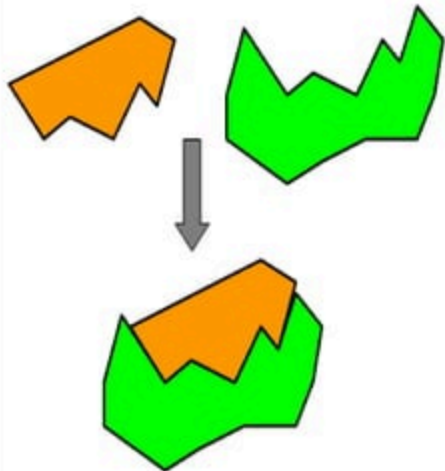


STEPS



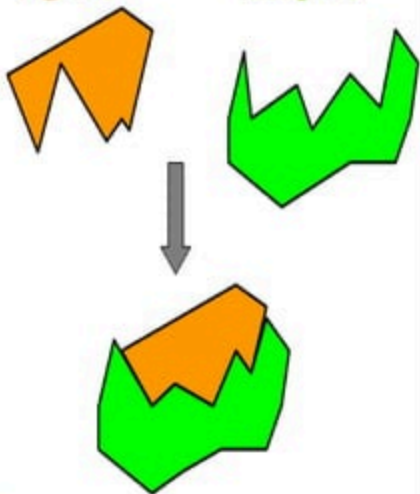
ENZYME – SUBSTRATE BINDING: 2 MODELS

Substrate (ligand) + *Enzyme (receptor)*



Lock and Key

Substrate (ligand) + *Enzyme (receptor)*



Induced Fit

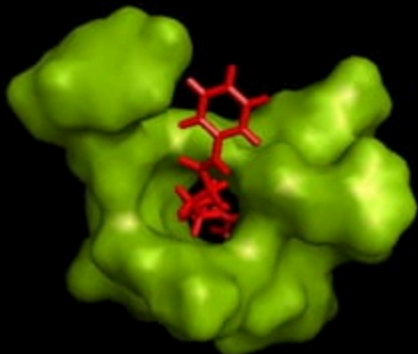
- **Docking Software-**

- ✓ Discovery Studio, Schrodinger
- ✓ Auto Dock, PhyreDock, Patch dock

- **Mostly drug activity is obtained through binding of one molecule to the pocket of another.**

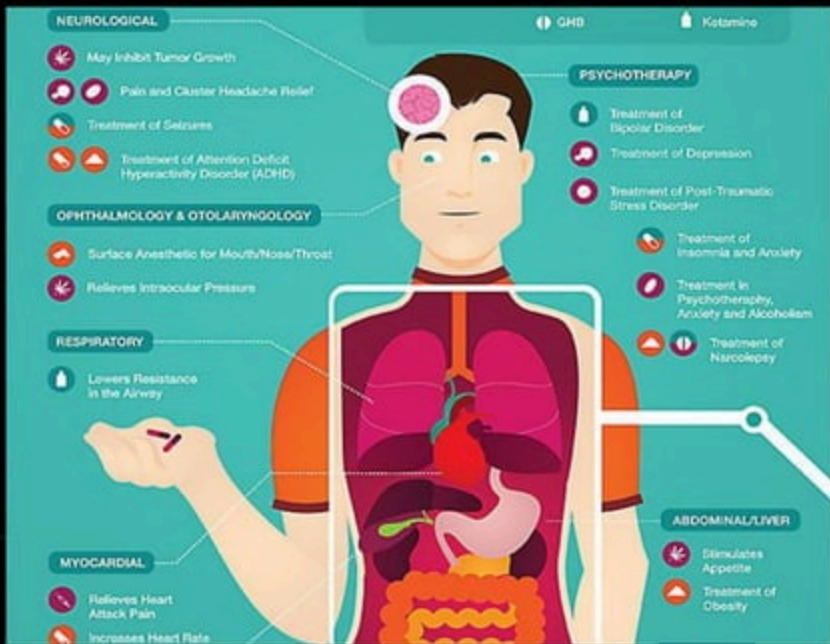
- **ADME Test**

- ✓ Absorption, Distribution, Metabolization, & Excretion



MORE IS NOT ALWAYS BETTER

- Be careful about dosage amounts





Thank You ...

DNA MICROARRA



Course Teacher

Dr. Shamamma S.

Professor

Dept, of Plant Biotechnology

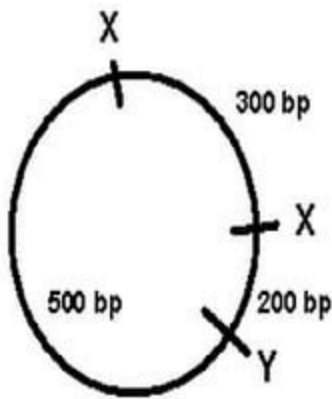
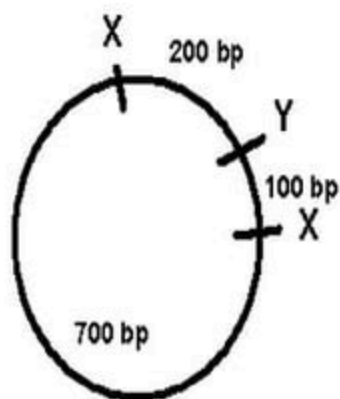
Presented By

Mahendra Reddy

Jn.M.Sc. Plant Biotechnology

PALB5270

Restriction Mapping





Welcome

PROTEIN MICROARRAY

K.KRUPA SAGAR

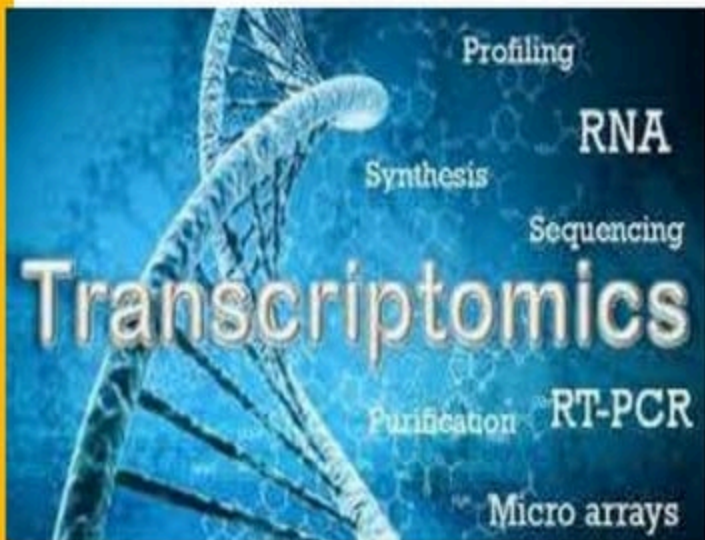
M.Pharm 1st year

STRUCTURAL GENOMICS



By:

Vaibhav Kumar Maurya




TRANSCRIPTOME ANALYSIS

Er. Divya Srivastava

DNA SEQUENCING



SURENDER RAWAT
Msc. MICROBIAL BIOTECH
ROLL NO. 1784



Chromosome Walking

Presented by: Aleena Ahmad Khan

