

# ***APPLICATIONS OF COMPUTER AND MATHEMATICS IN BIOLOGICAL RESEARCH***



**PROF. (DR.) AMAR P. GARG**  
Vice Chancellor

Shobhit University, Meerut - 250110

Email: [vicechancellor@shobhituniversity.ac.in](mailto:vicechancellor@shobhituniversity.ac.in)

# ***BASIC PRINCIPLE OF RESEARCH***

- ***It is not enough to identify a problem***
- ***Not enough to solve a problem***
- ***Need to identify and fix the root cause-otherwise problem is likely recur***

***IDENTIFICATION OF  
ROOT CAUSE IS NECESSARY***

***“Computers are like Old Testament gods; lots of rules and no mercy”***

***Joseph Campbell***

***“Math is the language in which GOD has written the universe”***

***- Albert Einstein***

# ***The Facts***

***Few mathematical scientists are  
biologically educated***

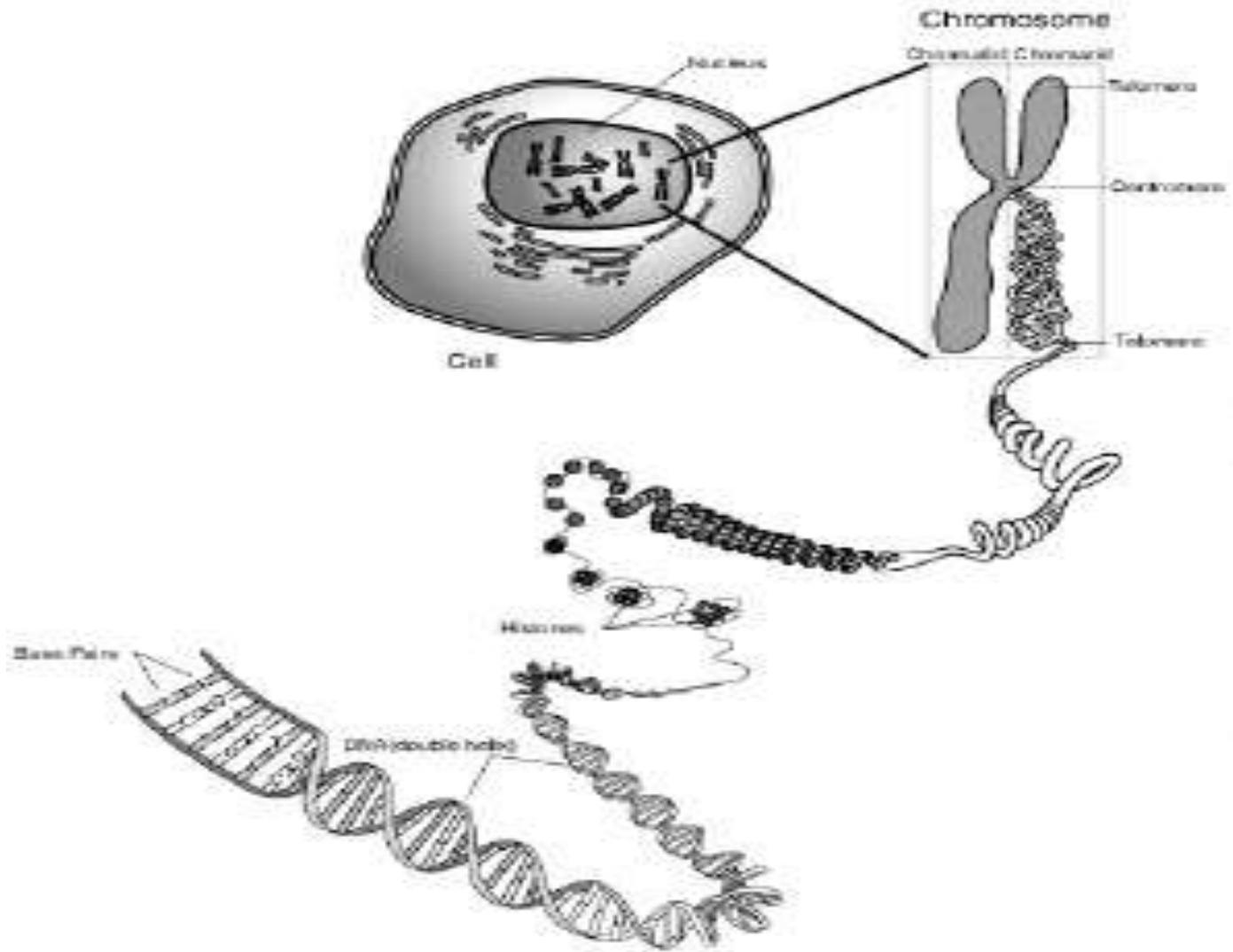
***Few biological scientists are  
mathematically educated***

***But Science needs mathematically  
educated biologists***

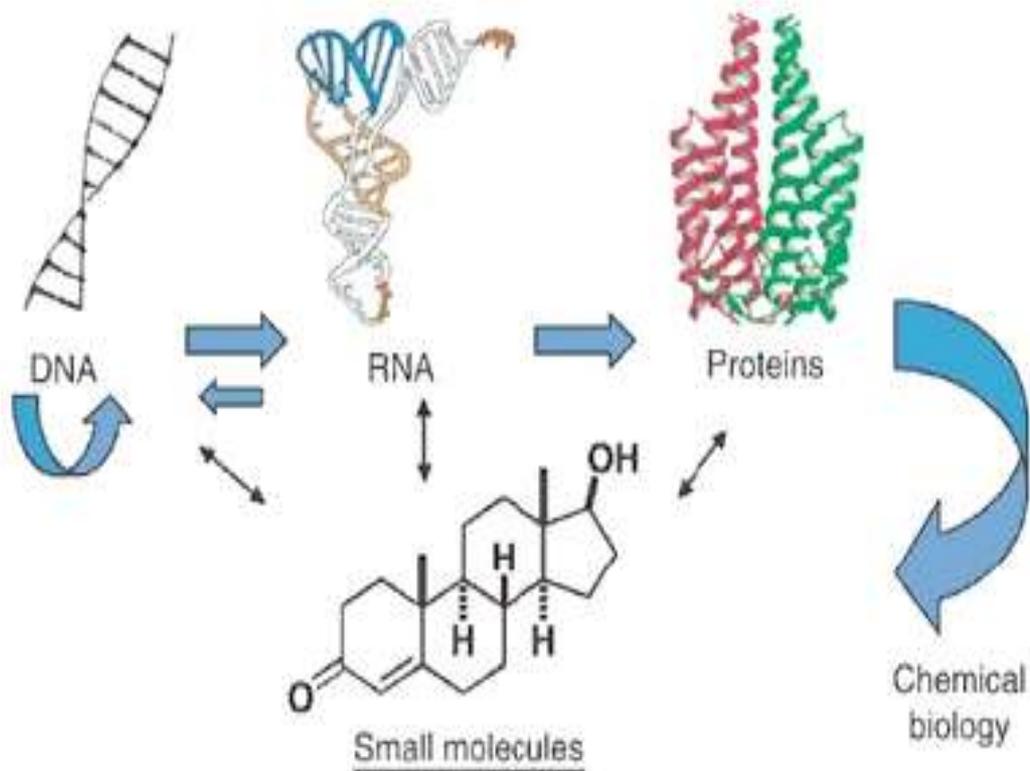
***We are all now connected by the  
Internet, like neurons in a giant brain***

***Stephen Hawking***

# ***DATA STORAGE IN CELL***



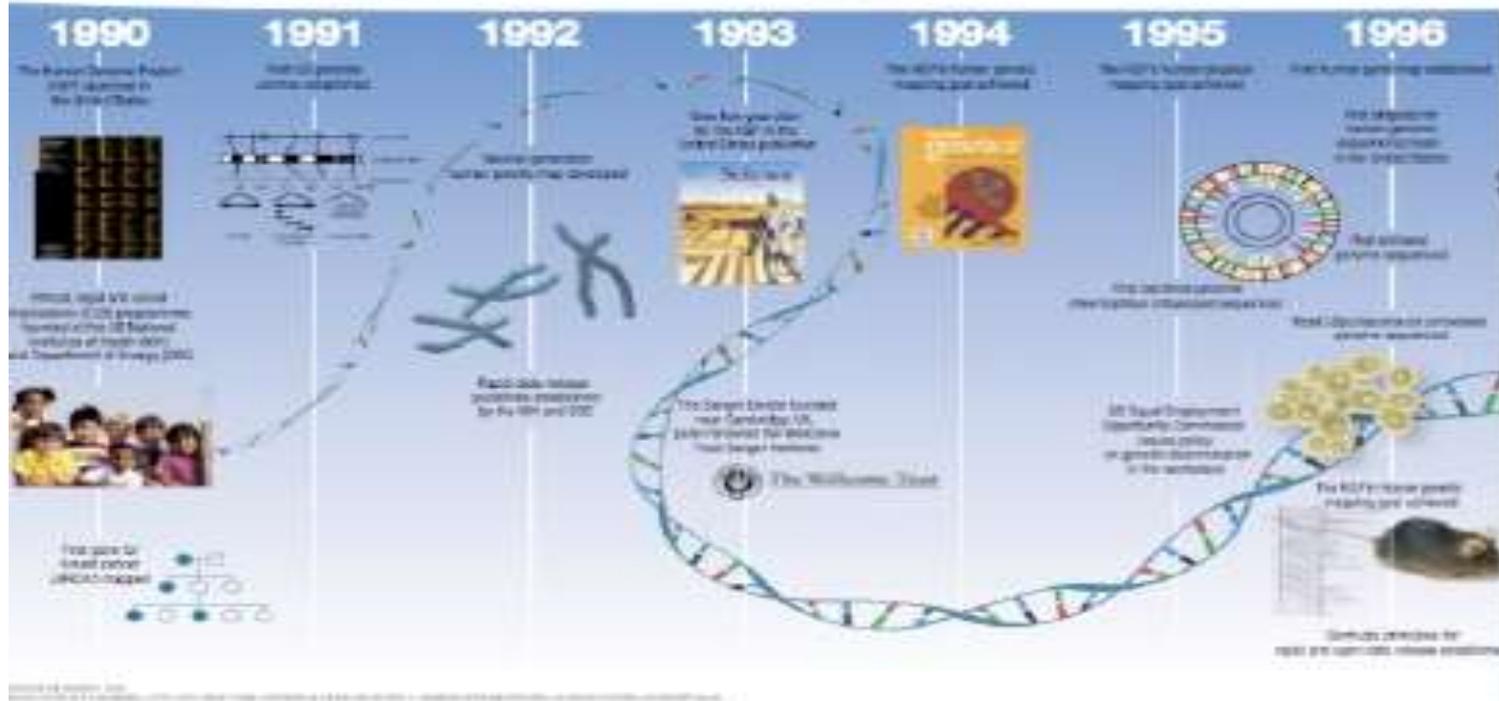
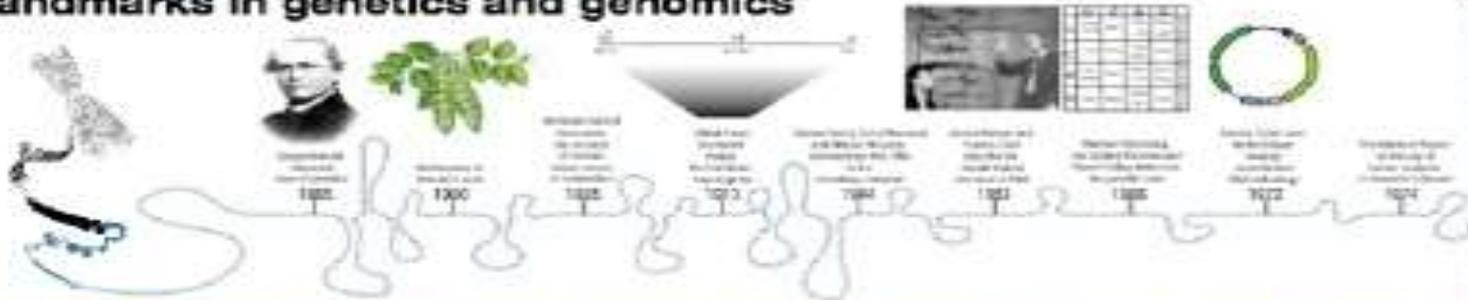
# ***TRANSFER OF DATA IN A CELL***



Cognition, signaling, life's origins, probes, drugs

# GENOMICS

## Landmarks in genetics and genomics





# *How Computers Help in Research*

## ***Save time***

***We can automate, make code run faster***

## ***Save efforts***

***May be someone has worked on a similar problem before?***

## ***Do better science***

***A more appropriate program, an overlooked parameter, better designing, better understanding, planning and management***

***This is the most important slide of the day***

# ***WHERE A BIOLOGIST NEEDS COMPUTER AND MATHEMATICS***

- ***Population Genetics***
- ***Structural Biology***
- ***Developmental Biology***
- ***Design of Equipments like Fermentor***
- ***Experimental Design***
- ***Bioinformatics***
- ***3-d Structural Analysis in Medical Biotechnology | Diagnostics | Bio-molecules***
- ***Various Pathways***

***Some More Applications of  
Maths/Stats/Computer in the  
Biosciences***

***Genomics and proteomics***

***Description of intra- and inter-  
cellular processes***

***Growth and morphology***

***Epidemiology and population  
dynamics, Neuroscience***

***and many more.....***

# **COMPUTATIONAL BIOLOGY**

- ***A new combined branch of computer technology and biology is Computational biology.***
- ***It's a recently developed new area of research that uses exciting tools from first principles of quantum mechanics to describe the electronic structure of atoms and their chemical properties.***
- ***Computational biology allows researchers to construct quantum simulations to see within biochemical processes to learn how reactions are taking place on a molecular and amazingly at atomic level too.***

- ***Computational biology uses mathematical and informational techniques including statistics to solve biological problems.***
- ***It does so, by using computer programs or mathematical models or creating both.***
- ***One of the major areas of computational biology is data mining which includes the analysis of the data collected by several genome projects.***
- ***Genome projects are scientific projects that are utilized to map the genome of a living being in other words the complete set of genes carried by this being or virus.***
- ***It has great application in medical and industrial research with vast potential***

# ***Impact of Computers***

- ***History of biology dates to the rise of various civilization***
  - ***while computers are relatively new, these have a monumental impact on biological research***

## ***3 examples of impact:***

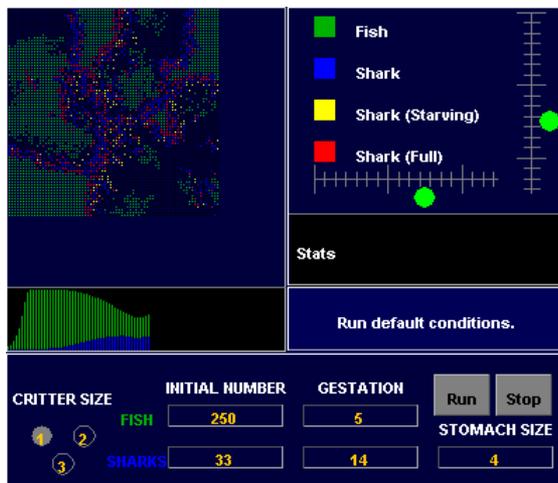
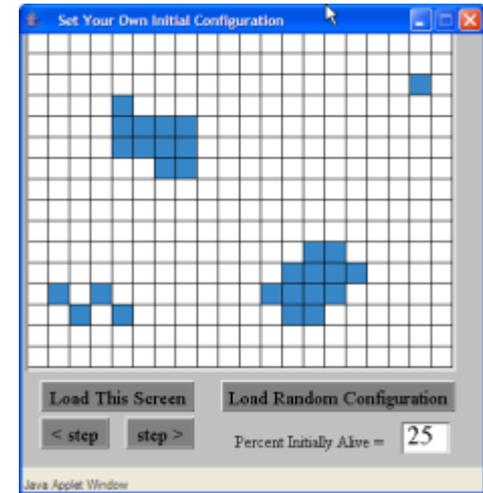
- 1. Computer technology is rapidly advancing the tools of scientific research***
- 2. Computer models are being used to study complex systems***
- 3. Computers are being used to store, process, and analyze large collections of biological data***
  - ***Many aspects of biology and computer science are converging***
  - ***Biological researchers must be savvy computer users and even programmers***
  - ***Computer scientists must know basic principles of various branches of Science including biology and must be able to solve interdisciplinary problems***

# Ecosystem Modeling

In late 1960s, John Conway showed that a simple model of an environment could produce complex and interesting behavior

- the environment is modeled as a 2-D grid of cells
- a cell can be alive (contain an organism) or dead
- simple rules model evolution

1. A dead cell becomes alive in the next generation if it has exactly 3 neighbors
2. A living cell survives in the next generation if it has 2 or 3 neighbors



Conway's ideas have been extended to a variety of ecosystems

- Different colored cells denote different organisms (sharks & fish)
- Other systems have modeled:
  - ✓ the growth of viruses
  - ✓ the spread of infectious diseases in a population
  - ✓ the behavior of an ant colony

# **Bioinformatics**

- **Bioinformatics bridges biology and computer science (or informatics, as it is known in Europe)**
    - **Broad definition of bioinformatics: the use of computer science techniques to solve biological problems**
    - **narrower but common definition: the application of computer science techniques to the representation and processing of biological data**
  - **With the advancement of research tools, biologists are generating enormous amount of data**
    - **a single experiment with genetic material can produce thousands or millions of data points**
    - **computational and statistical tools are needed to analyze and understand such large volumes of data**
- The biggest impact of computers in biology is in storing, accessing, and processing of large amounts of biological data**

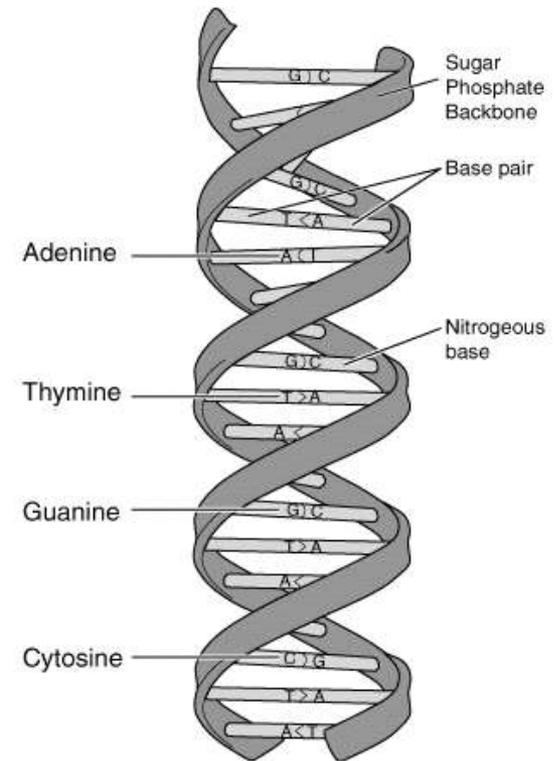
# Bioinformatics Tools

- **Many tools are available for searching and manipulating genetic sequences e.g. the GeneBoy program ([www.dnai.org/geneboy](http://www.dnai.org/geneboy))**
  - **demonstrates DNA → RNA transformations**
  - **analyzes the composition of a sequence**
  - **searches for specific patterns in the sequence**



# DNA Overview

- **DNA is the genetic blue-print of life**
  - **made of nucleotides with four bases (A, T, G, C), organized in a double-helix**
  - **the two strands match A+T and C+G base pairs**
  - **can think of DNA as encoding information in 4 bases**
- **A gene is a region of DNA that encodes the chemical structure of a protein**
- **It is currently believed that there are 20,000-30,000 different genes in human DNA**
  - **Means roughly 3 billion base pairs**



"If our strands of DNA were stretched out in a line, the 46 chromosomes making up the human genome would extend more than six feet. If the ... length of the 100 trillion cells could be stretched out, it would be ... over 113 billion miles. That is enough material to reach to the sun and back 610 times." [Source: Centre for Integrated Genomics]

# DNA → RNA → Proteins

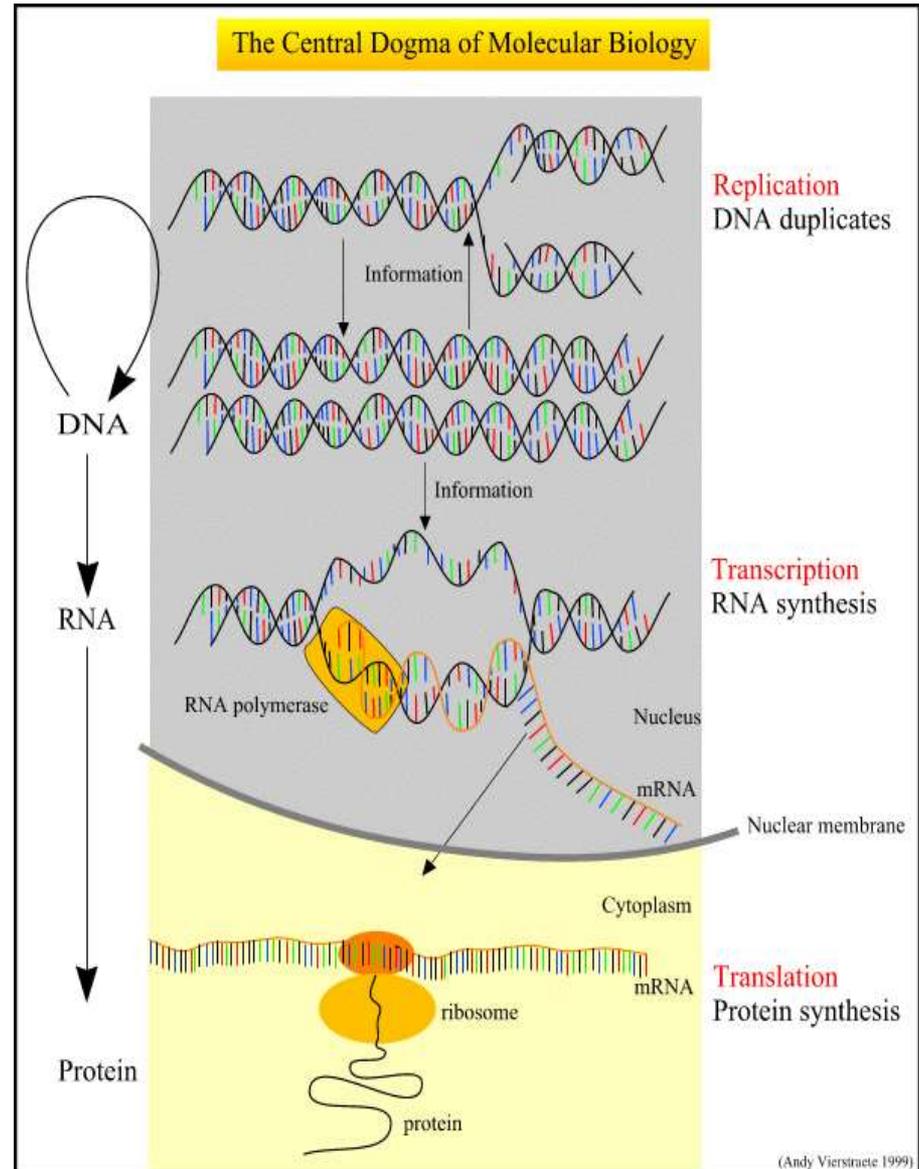
## • During cell division,

- the two strands of DNA split
- each strand is paired with free nucleotides in the nucleus to complete copies of the original DNA
- each cell gets a complete set of the DNA

## • in mapping DNA to proteins,

- the DNA strands are split and copied into mRNA (using the same bases except U replaces T)
- this mRNA is then "read" by a ribosome to build the specified protein
- proteins are commonly represented using a 20 letter alphabet (for the different types of amino acids)

[www.dnai.org/alindex.html](http://www.dnai.org/alindex.html)



# DNA Databases and Tools

- **often, the source or purpose of a DNA sequence can be determined by comparing it with documented genetic material**
  - **several large databases are available online**
  - **tools for visualizing and/or searching the databases are also available**
- **e.g. the Ensemble site ([www.ensembl.org](http://www.ensembl.org)) contains visualizations of the human genome and other DNA sequences**

The screenshot shows the Ensembl Genome Browser website in Internet Explorer. The browser window title is "Ensembl Genome Browser - Internet Explorer Provided by Cox High Speed Internet". The address bar shows "http://www.ensembl.org/index.html". The website features the Ensembl logo and a search bar. The main content area is titled "browse a genome" and is organized into three columns: "Mammals", "Other chordates", and "Other eukaryotes". Each column lists various species with their Ensembl IDs and links to "browse" or "what's new".

| Mammals   | Other chordates   | Other eukaryotes  |
|---|---|---|
| <b>Homo sapiens</b><br>[NCBI 35]<br>browse   what's new   Vega  | <b>Gallus gallus</b><br>[WASHUC1]<br>browse   what's new          | <b>Drosophila melanogaster</b><br>[BGDP 4]<br>browse   what's new |
| <b>Pan troglodytes</b><br>[CHIMP1]<br>browse   what's new       | <b>Xenopus tropicalis</b> [JGI 3]<br>browse   what's new          | <b>Anopheles gambiae</b><br>[MOZ 2]<br>browse   what's new        |
| <b>Macaca mulatta</b> [Mmul 0.1]<br>browse <i>pre!</i> site     | <b>Danio rerio</b><br>[WTSI Zv5]<br>browse   what's new   Vega    | <b>Apis mellifera</b><br>[Amel 2.0]<br>browse   what's new        |
| <b>Mus musculus</b><br>[NCBI m34]<br>browse   what's new   Vega | <b>Takifugu rubripes</b><br>[Fugu 2.0]<br>browse   what's new     | <b>Caenorhabditis elegans</b> [WS140]<br>browse   what's new      |
| <b>Rattus norvegicus</b><br>[RGSC 3.4]<br>browse   what's       | <b>Tetraodon nigroviridis</b><br>[TETRAODON 7]<br>browse   what's |   |

# GenBank

- **the GenBank public repository of DNA and RNA sequence data contains**
  - **partial or complete genomes for more than 165,000 organisms**
  - **more than 1 trillion bases of sequence data**
  - **roughly 3 million new DNA sequences are added per month**
- **the database can be accessed and searched using various tools at [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)**

NCBI BLAST - Internet Explorer Provided by Cox High Speed Internet

Address: <http://www.ncbi.nlm.nih.gov/BLAST/>

NCBI → BLAST Latest news: 28 August 2005: BLAST 2.2.12 released

**About**

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

|  |  |
|--|--|
| <b>Nucleotide</b> <ul style="list-style-type: none"><li>Quickly search for highly similar sequences (megablast)</li><li>Quickly search for divergent sequences (discontiguous megablast)</li><li>Nucleotide-nucleotide BLAST (blastn)</li><li>Search for short, nearly exact matches</li><li>Search trace archives with megablast or discontiguous megablast</li></ul> | <b>Protein</b> <ul style="list-style-type: none"><li>Protein-protein BLAST (blastp)</li><li>Position-specific iterated and pattern-hit initiated BLAST (PSI- and PHI-BLAST)</li><li>Search for short, nearly exact matches</li><li>Search the conserved domain database (rpsblast)</li><li>Protein homology by domain architecture (cdart)</li></ul> |
| <b>Translated</b> <ul style="list-style-type: none"><li>Translated query vs. protein database (blastx)</li><li>Protein query vs. translated database (tblastn)</li><li>Translated query vs. translated database (tblastx)</li></ul>  | <b>Genomes</b> <ul style="list-style-type: none"><li>Human, mouse, rat, chimp <b>NEW</b>, cow, pig, dog, sheep, cat</li><li>Chicken, puffer fish, zebrafish</li><li>Environmental samples</li><li>Malaria</li><li>Insects, nematodes, plants, fungi, microbial genomes, other eukaryotic genomes</li></ul>   |
| <b>Special</b> <ul style="list-style-type: none"><li>Search for gene expression data (GEO BLAST)</li><li>Align two sequences (blast2seq)</li></ul>   | <b>Meta</b> <ul style="list-style-type: none"><li>Retrieve results</li></ul>   |

**Software**

- Downloads
- Developer info

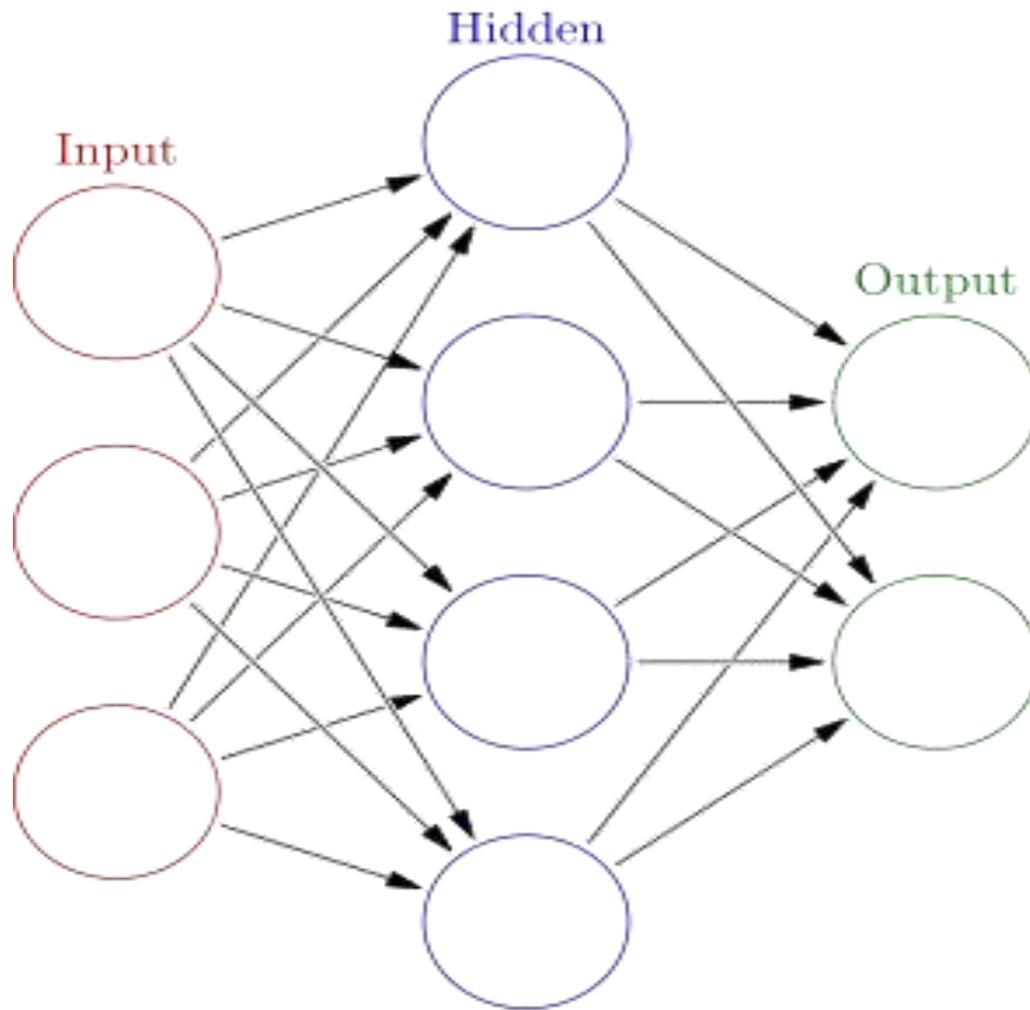
**Other resources**

- References
- NCBI Contributors
- Mailing list
- Contact us

# ***Need of Mathematics and Computer Science in Solving Biological Puzzles***

## ***Target 1***

***For building a cancer predictive model using ANN (Artificial Neural Network) based tool.***



An artificial neural network (ANN) is an interconnected group of nodes, akin to the vast network of [neurons](#) in a [brain](#). Circular node represents an [artificial neuron](#) and arrows represent a connection with each other.

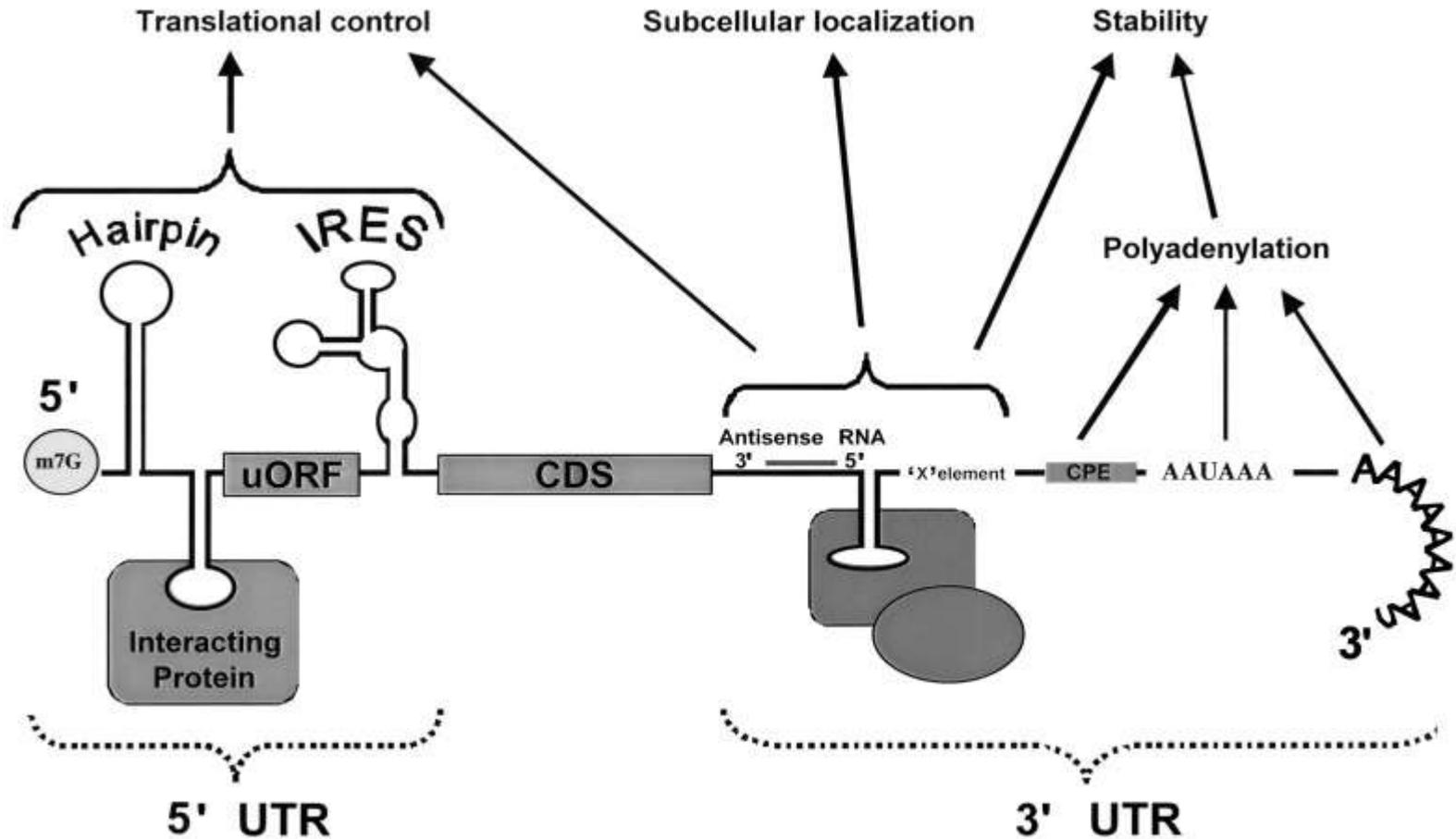
## **Target 2.**

**To find out those genomic features in UTR (Un-translated Regions) as well as the concerned genes which are actually controlling their expression.**

## **Target 3.**

**For target 2 we have to have those UTRs as well as the concerned genes.**

**“Highly Interdisciplinary Work”**

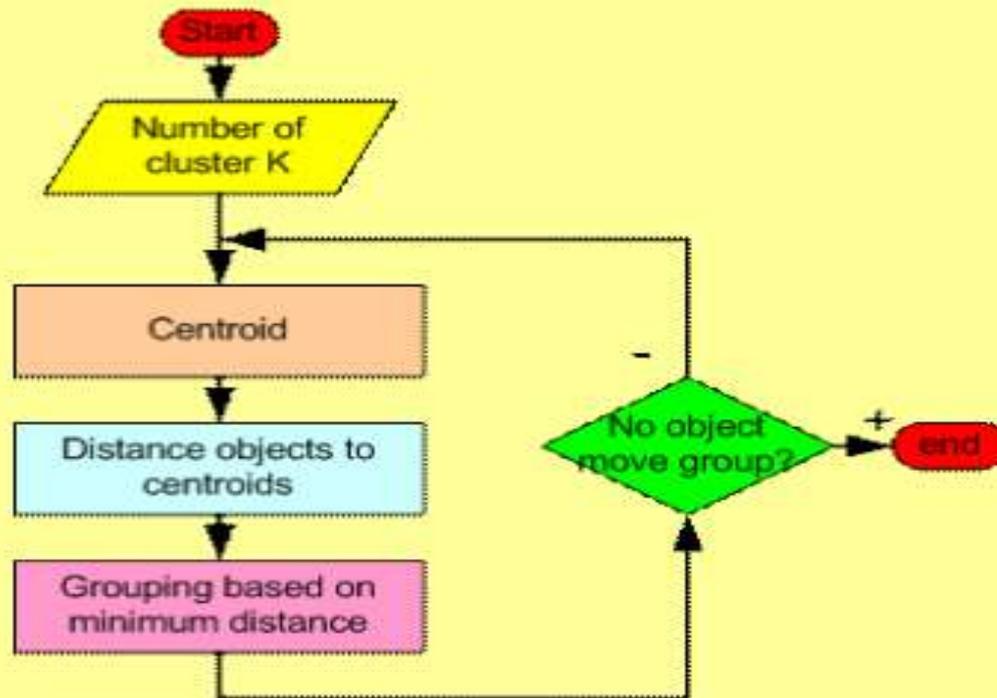


General structure of an eukaryotic [mRNA](#) illustrating some post-transcriptional [regulatory elements](#) for gene expression and their activity.  
**[5'UTR](#) and [3'UTR](#) mediated regulation**

# Use of K-mean Clustering on Biological Data

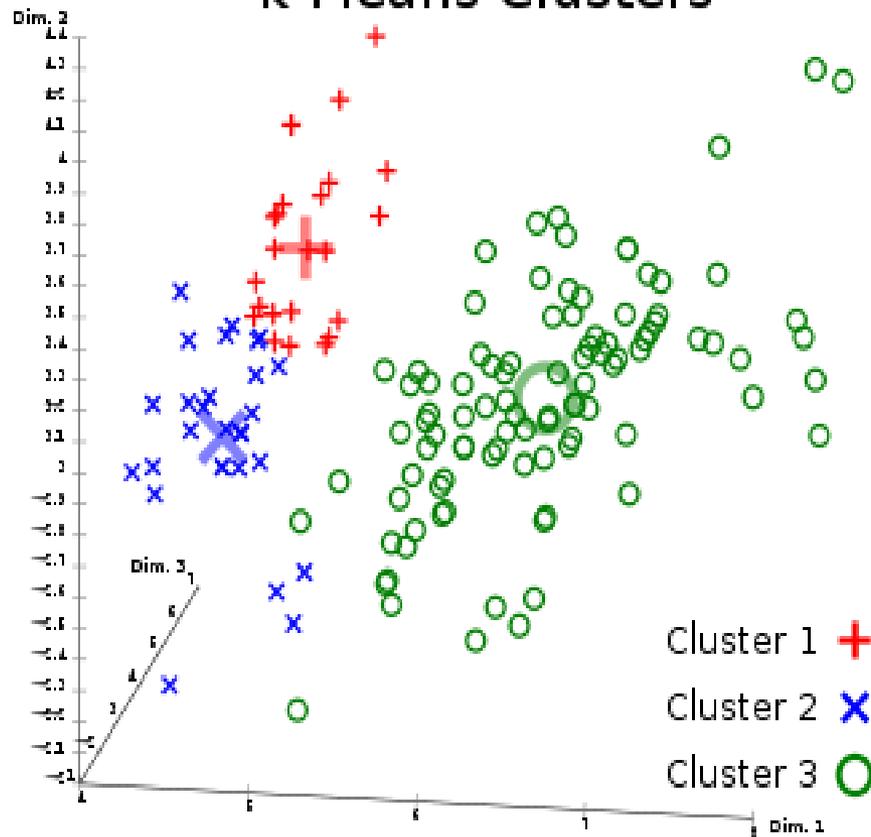
Iterate until *stable* (= no object move group):

1. Determine the centroid coordinate
2. Determine the distance of each object to the centroids
3. Group the object based on minimum distance

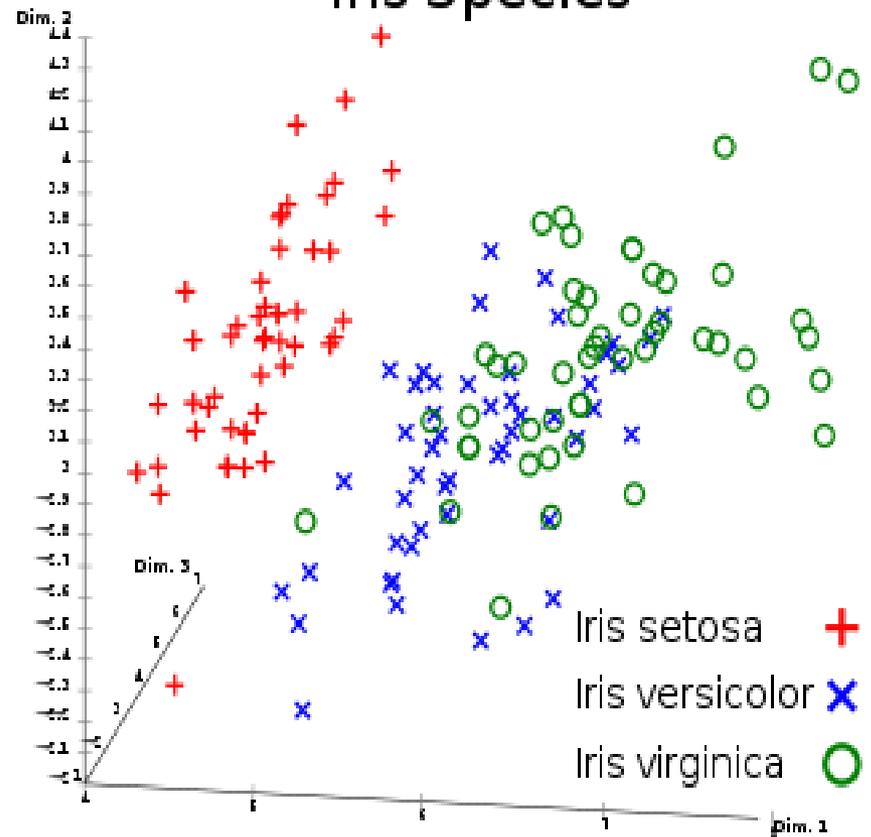


***Our first effort to classify genes in 2014***

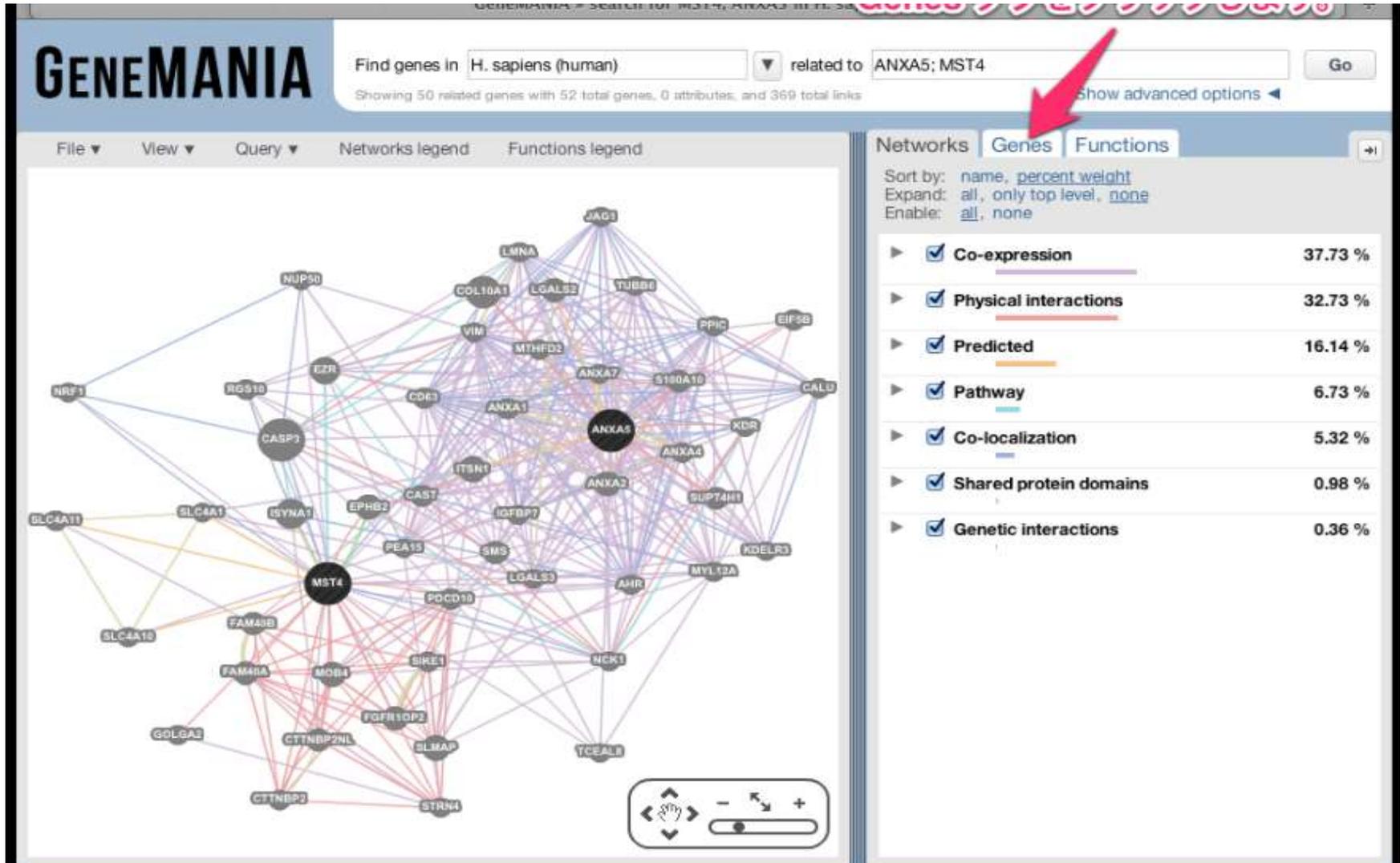
## k-Means Clusters



## Iris Species



# Use of Rank Correlation Coefficient in Gene Interaction



# ***Use of Mathematical Modeling in Biological Research***

Data Extraction appropriate

Data validation

Screening of sequences

Identification of genomic features and calculation.

Normalization and correlation between various features

Identification of genes (normal, stressed, and diseased condition)

Gene features calculation

Microarray data extraction for genes

Statistical analysis and correlation.

Multivariate analysis and model generation

Testing of generated mathematical models

Correlation of gene expression data to the UTR genomic features (Classified).

Simulation of gene expression using these genomic features

Finding of best model

Classification of data

# ***CT Scans and MRI's***

***X-rays have been used for about a century in medical diagnostics. These are high frequency, high energy electromagnetic waves like visible light, except that the electromagnetic waves of visible light are of lower frequency and lower energy.***

***Since we see the differences of frequency in light as 'color' and differences of energy as 'brightness', so one may say that X-rays are very bright and of a different color from what we see.***

***When X-rays strike film, the chemicals in the film are changed – and the amount of the change is related to the strength of the rays that strike the film. When the film is developed, an image appears that records the intensity of the X-rays reaching the film at that spot.***

***The process is exactly analogous to the creation of a black-and-white negative in ordinary cameras—the bright areas of the picture are black, the dark areas are white.***

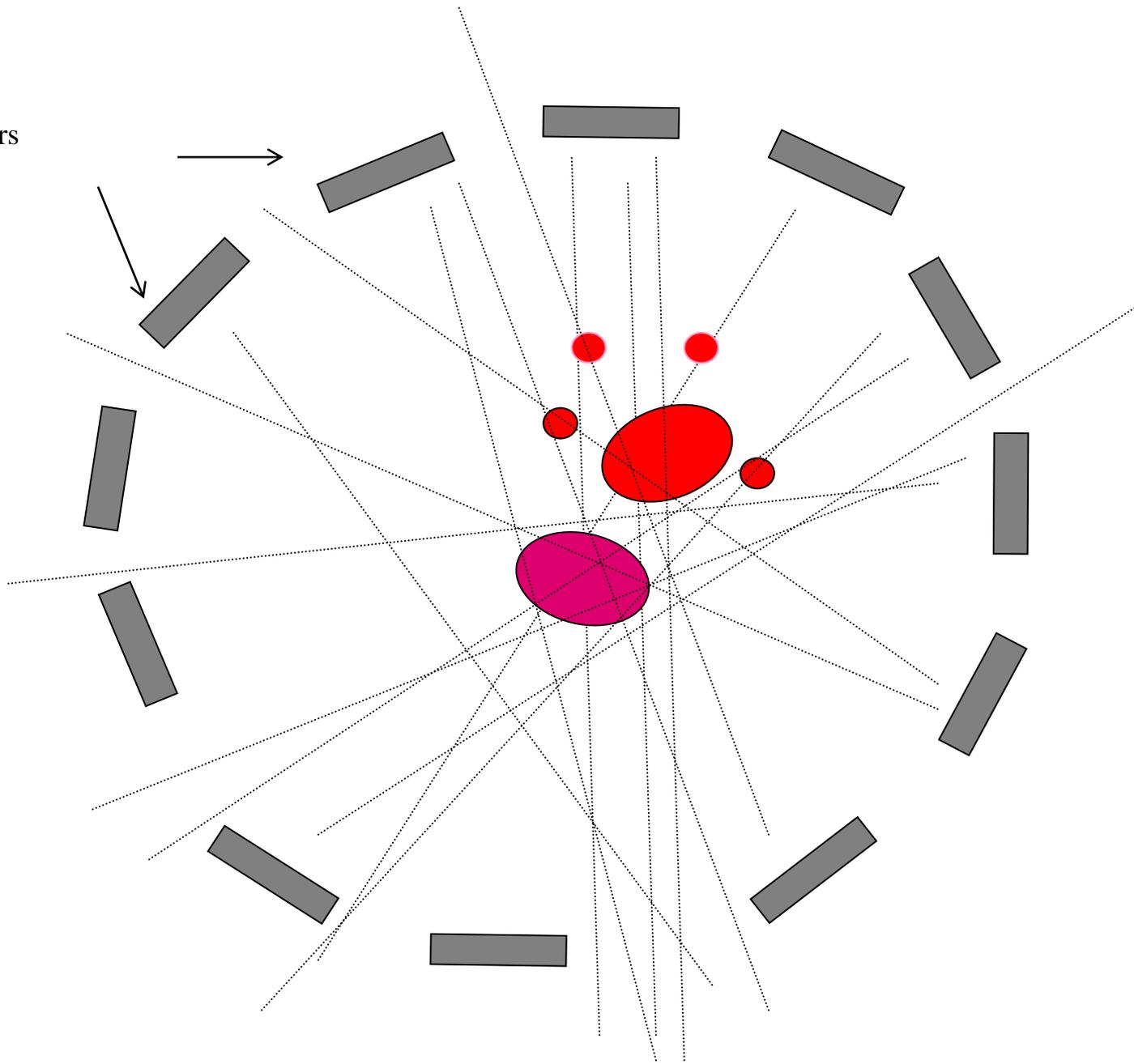
***Just as we have replaced film in our cameras with a light detector that creates a digital image, in some situations X-ray film has been replaced by X-ray detectors that create a digital image.***

***The trouble with X-rays photographs is that two-dimensional images are created, so that it is often hard to tell which piece of a broken bone is on top of the other. It is for this reason that doctors often take several X-rays of the same site to get different perspectives on the situation.***

***In CAT Scans (Computerized Axial Tomography) and MRI's (Magnetic Resonance Imaging), and similar diagnostic technologies, images are made from thousands, even millions of different perspectives – and mathematics is used to integrate the information from all these perspectives into a single **THREE-dimensional digital image.*****

***(Literally the integration what we teach our calculus students, is just a sophisticated use of it)***

detectors



***The pictures your doctor shows you, are from slices of the three-dimensional digital image.***

***Sometimes, they have computers to show you the whole three dimensional image on the screen and then turn it or slice it to get different views of particular features of interest.***

***This new use of mathematics has revolutionized diagnostic medicine, yet few in our society recognize it as new mathematics helping all of us.***

# ***SARS, HIV, Cancer, Parkinson's and Epidemics***

***Increasingly, scientists are calling on mathematicians to help them understand the spread of disease or the mechanisms of disease by the creation of a mathematical model of the situation.***

***Just as an architect might create a cardboard model of a structure being designed in order to understand it better, scientists use mathematical models to understand problems.***

***The architect's cardboard model is a smaller, simplified version of the real thing, and because it is smaller and simpler, it is easier to understand.***

***Scientists and mathematicians create mathematical models that are smaller and simpler than the real thing so that they will be easier to understand.***

***We often use geometric models of our houses to help us compute the amount of carpet we need to buy to cover our floors.***

***The mathematical models used in epidemiology or in creating treatments of cancer or Parkinson's are often differential equations models, matrix models, or models from discrete (sometimes called finite) math. Moreover, they often must incorporate statistical tools into the models, because biological processes are rarely deterministic.***

***These models are used to predict the way the real situations develop, and then used to understand how to modify those outcomes.***

## ***Hypothesis Suggested by Neuro-physiologists***

- ***The brain processes signals together that arrive from the two eyes at the same time.***
- ***The signal from a darker image is sent later than the signal from a brighter image, that is, signals from darker images are delayed.***

# **CONCLUSIONS**

- ***Using Computer and mathematical models useful descriptions of biological phenomena are possible***
- ***Computer based models can be used as evidence to support or refute biological hypotheses***
- ***Computerized models can suggest new experiments, simulate experiments or treatments that have not yet been carried out, or estimate parameters that are experimentally inaccessible***
- ***Working together, biologists, computer scientists, statisticians, and mathematicians can contribute more to science than any group can contribute separately.***
- ***Can save, time, money and energy***
- ***Can produce more reliable results of multiple applications***

Thank You!

