Perspectives in Bioinformatics

Junior Barrera

Associate Professor: IME-USP

Layout

- Introduction
- Molecular Biology: historical overview
- Measure procedures
- Data storage
- Information extraction
- Educational impact

Introduction

Great Scientific Programs

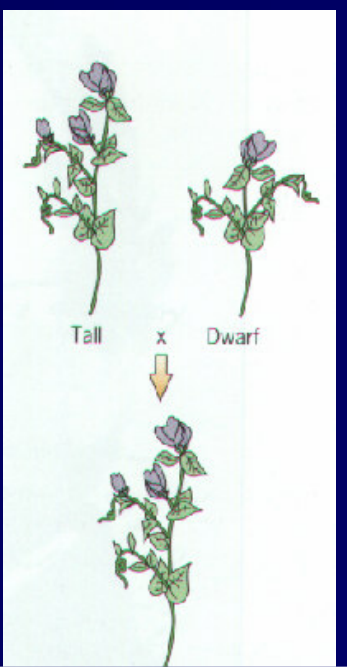
- Manhattan project - Nuclear bomb - 1945
- Space program - USA/URSS - 1965
- Molecular Biology Research - 1990 ....

Mocelular Biology Research

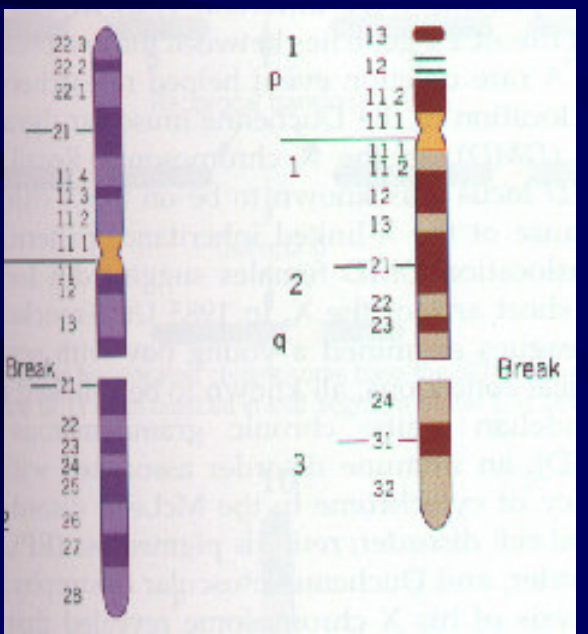
- Impact in many fields: medicine (drug discovery, diagnostics); agriculture (animal and vegetal phenotype modification)
- World Movement
- Governmental and private money
- A historical moment: the approximation between Biology and Mathematics (BIOINFORMATICS)

Molecular Biology: a historical overview

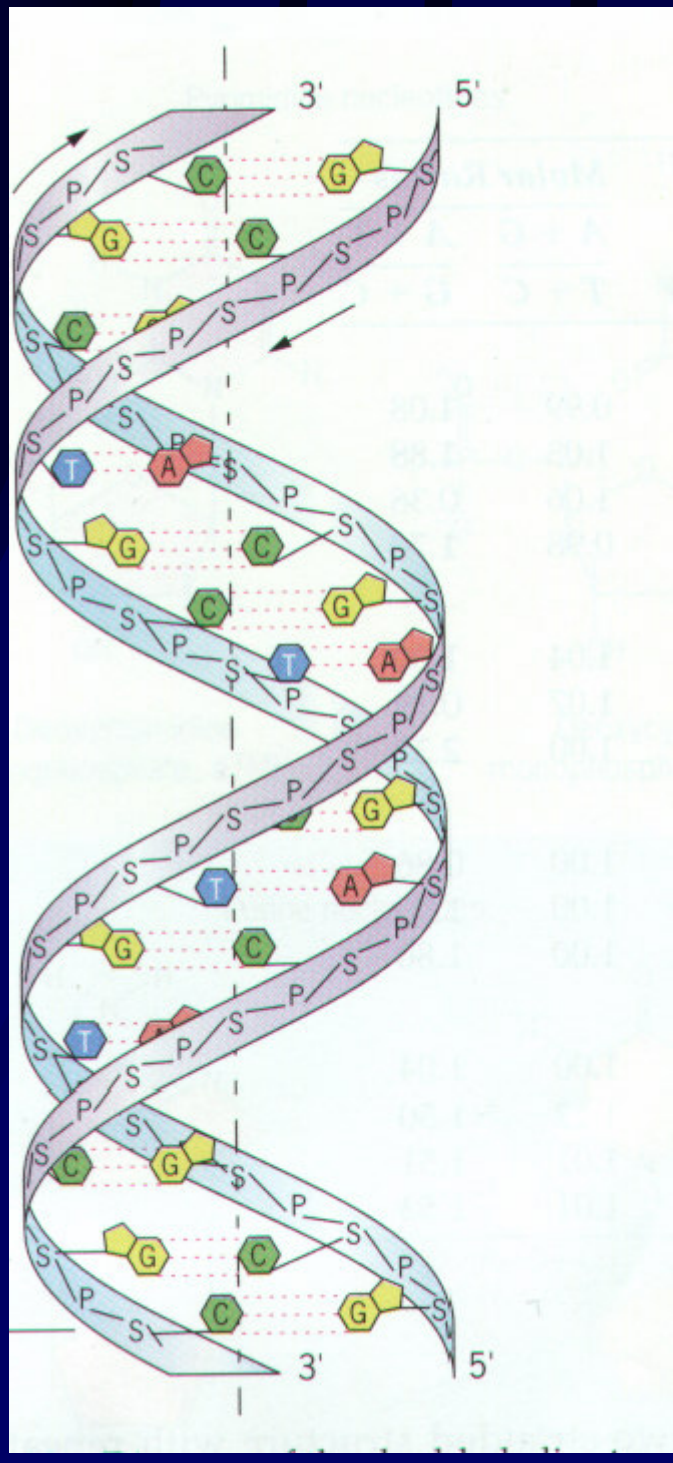
- Heredity Mendel (1866)
- The phenotypes of an individual depends on genes of his parents.



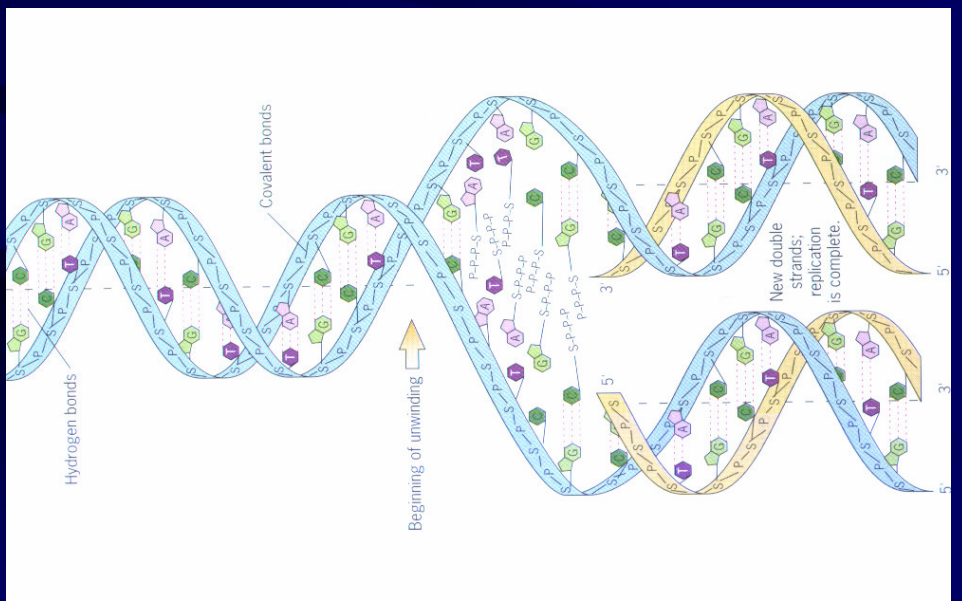
- Chromosome Theory - Morgan (1910)
- Genes were situated in chromosomes



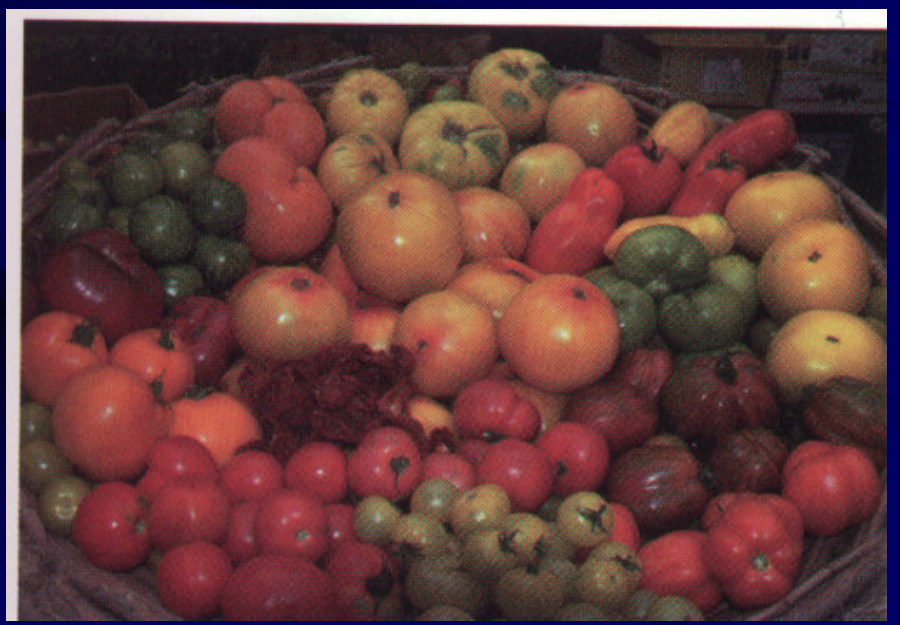
- The molecular structure of chromosomes (Watson and Crick - 1953)
- DNA structure: the double helix
- Four basis: adenine(A), guanine(G), thymine(T), cytosine(C)
- genes are sequences of nucleotides


3'
5'
C
G
T
A
S
P
3'
5'

- cut, replication and decoding

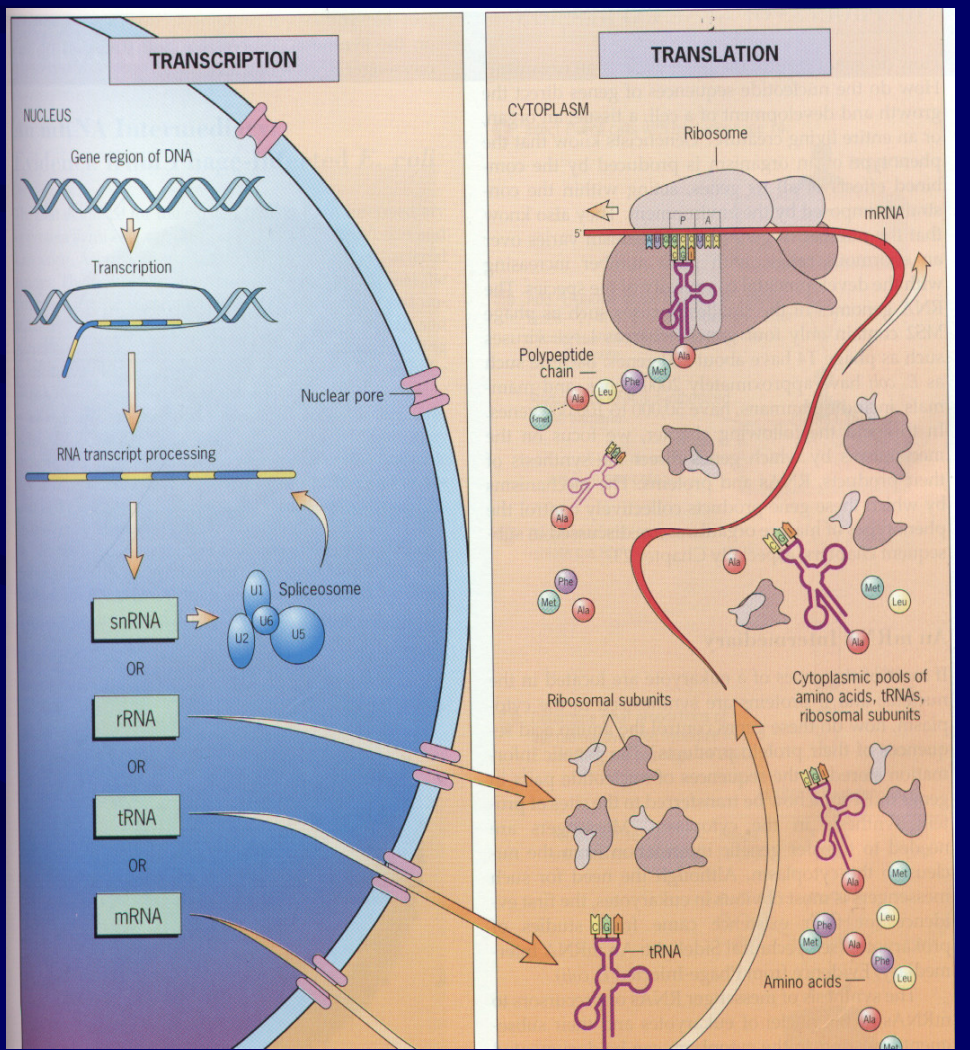


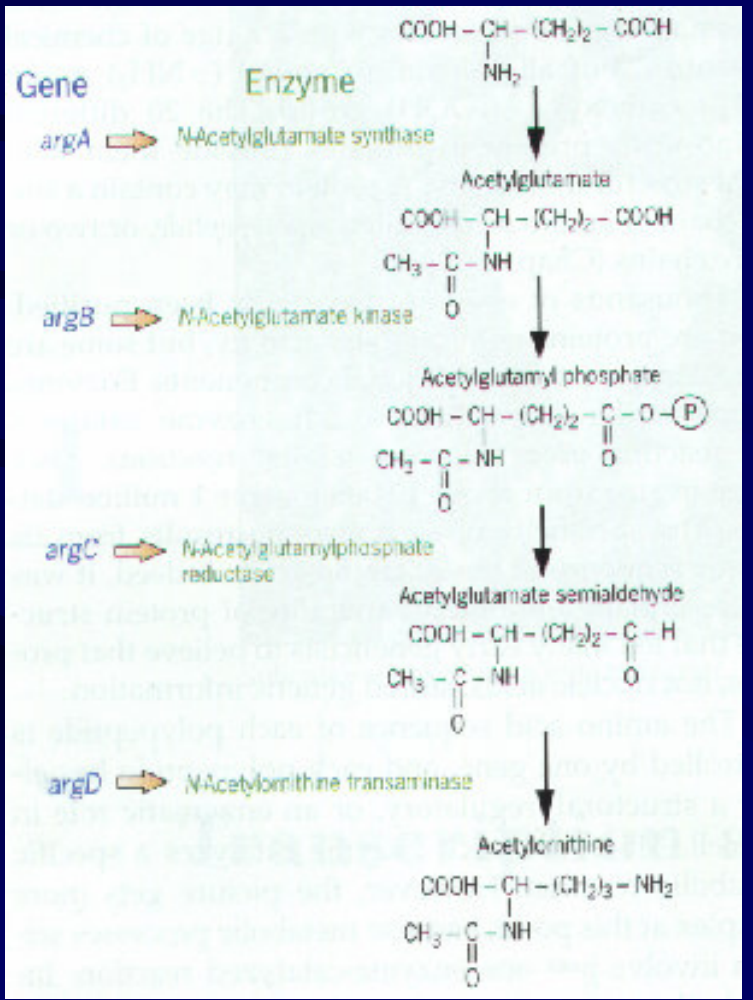
- species modification, drug production



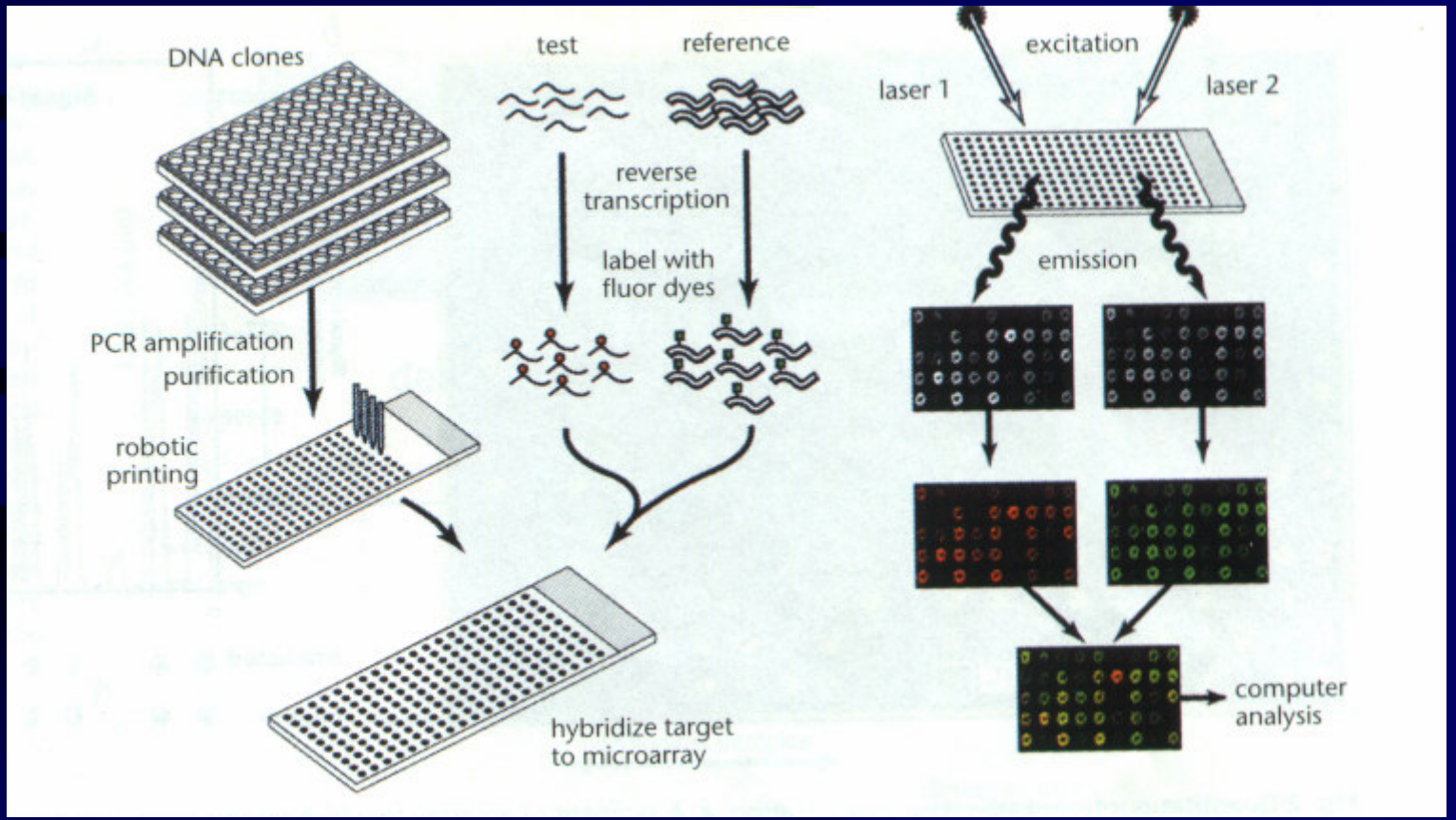
- Genes control the metabolism
- Metabolism occurs by sequences of enzyme-catalyzed reactions.
- Enzymes are specified by one or more genes

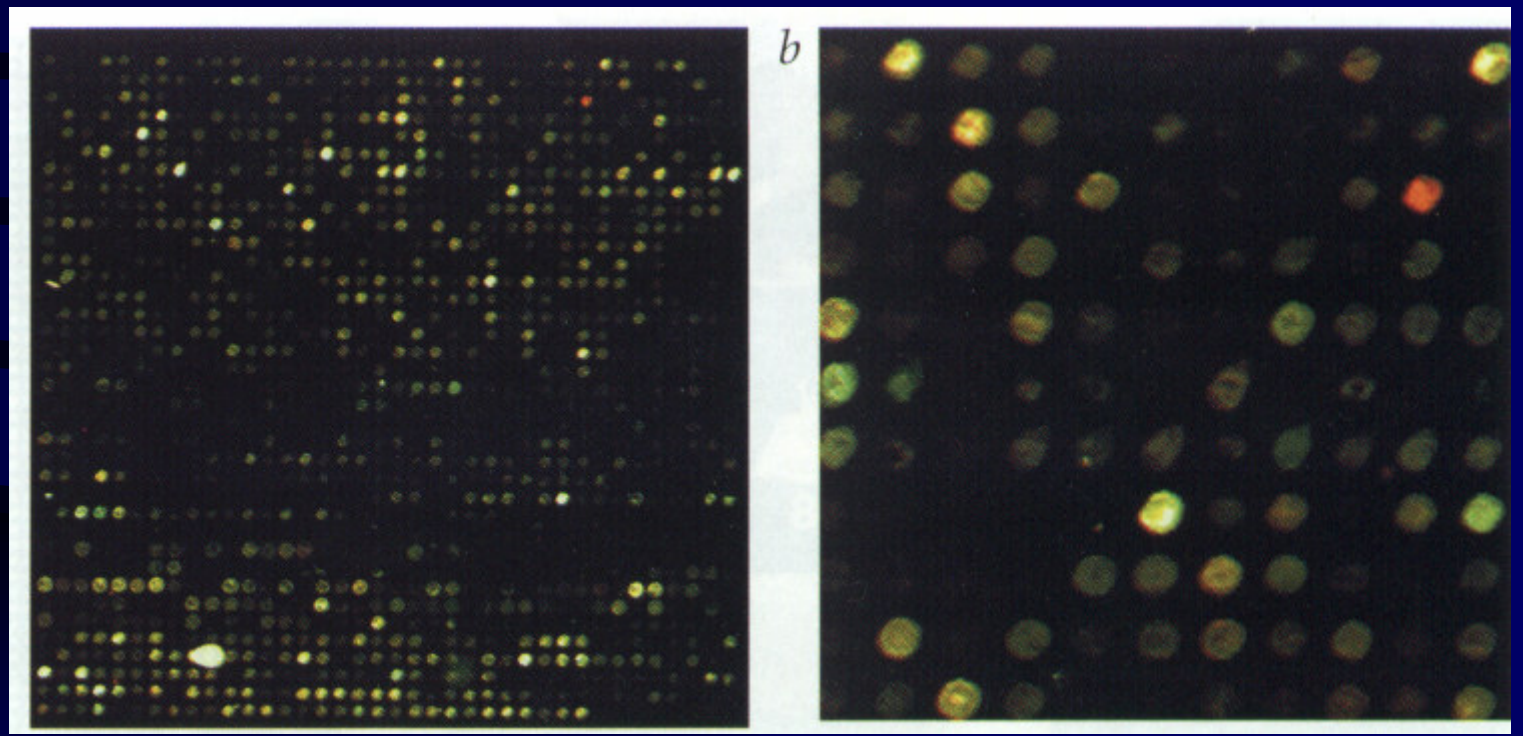
- Gene expression

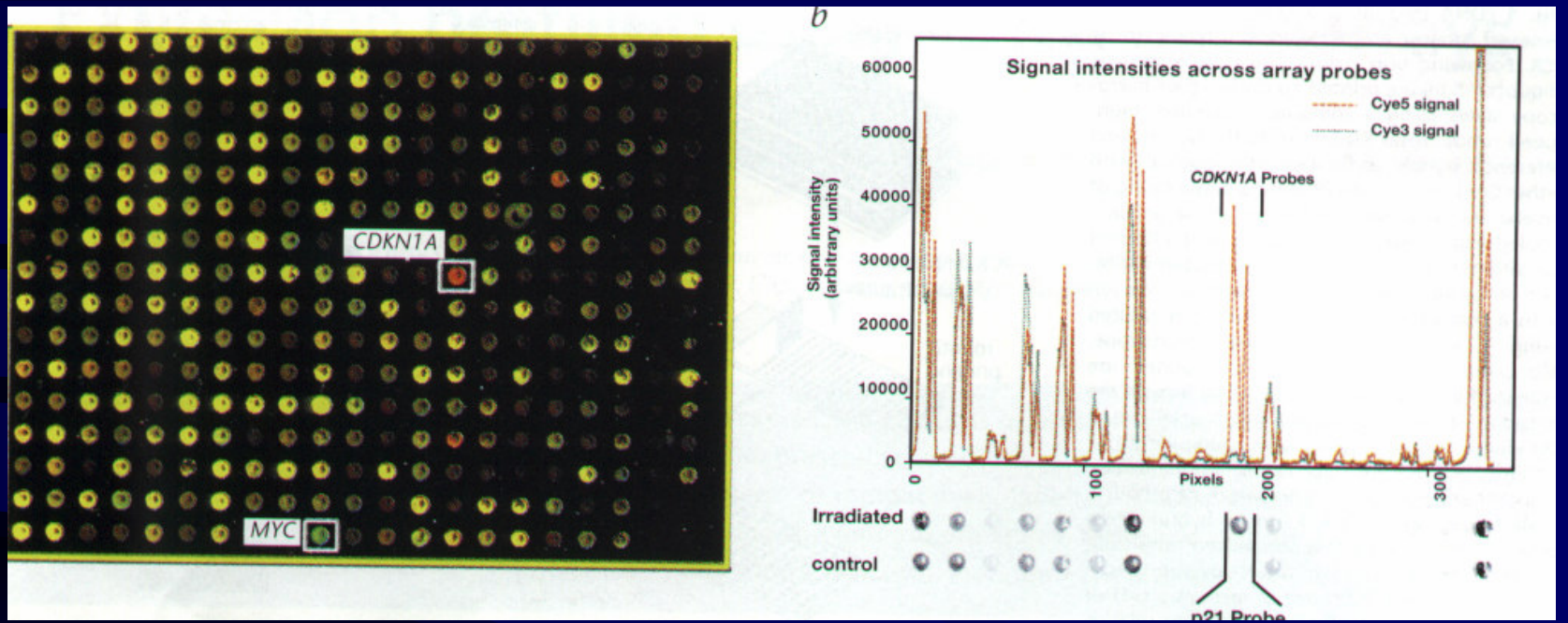



Gene
Enzyme
argA
N-Acetylglutamate synthase
Acetylglutamate
argB
N-Acetylglutamate kinase
Acetylglutamyl phosphate
argC
N-Acetylglutamylphosphate reductase
Acetylglutamate semialdehyde
argD
N-Acetylornithine transaminase
Acetylornithine

Measure Procedures


DNA clones
test
reference
excitation
laser 1
laser 2
reverse transcription
label with fluor dyes
emission
PCR amplification purification
robotic printing
hybridize target to microarray
computer analysis


b

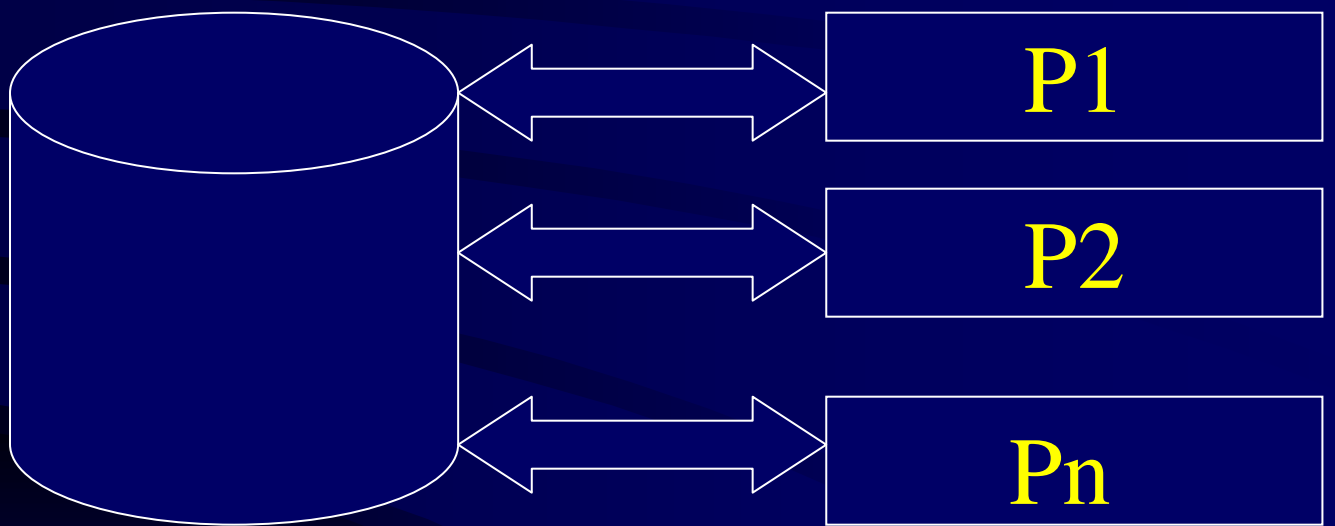


Quantization - {-1,0,1}

Data Storage

System

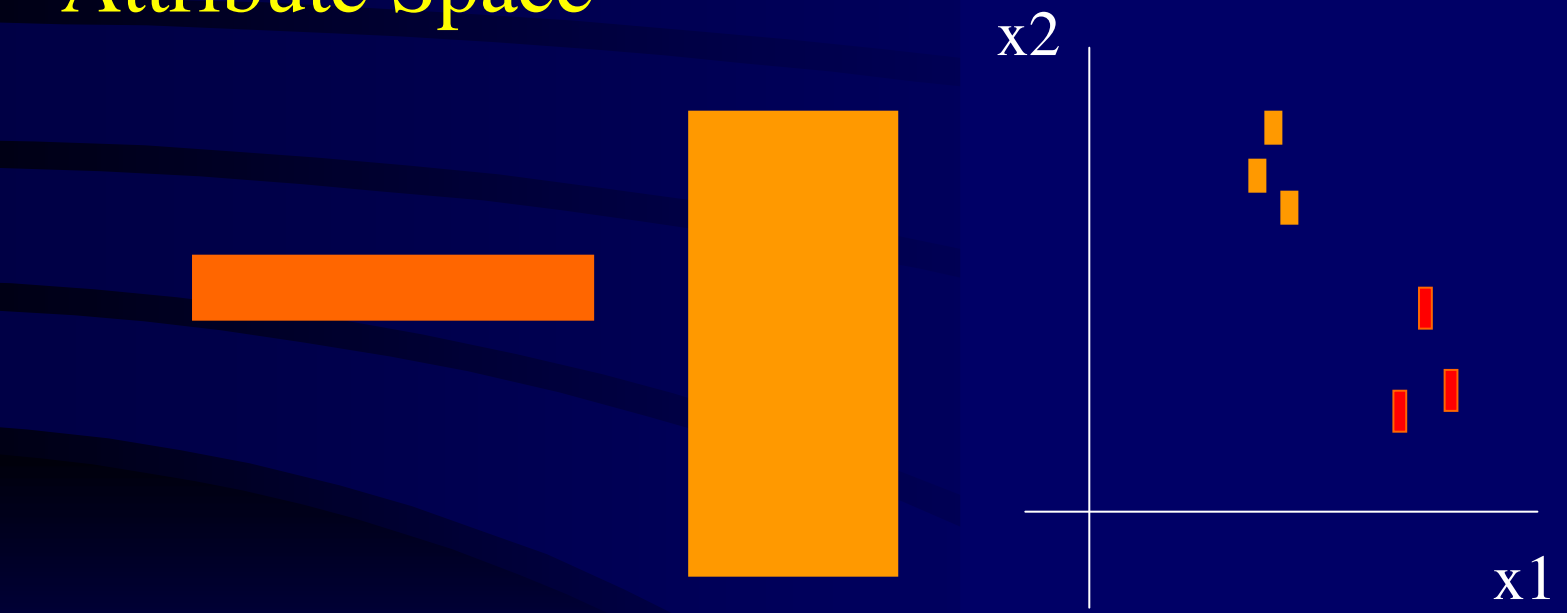
Objected oriented database



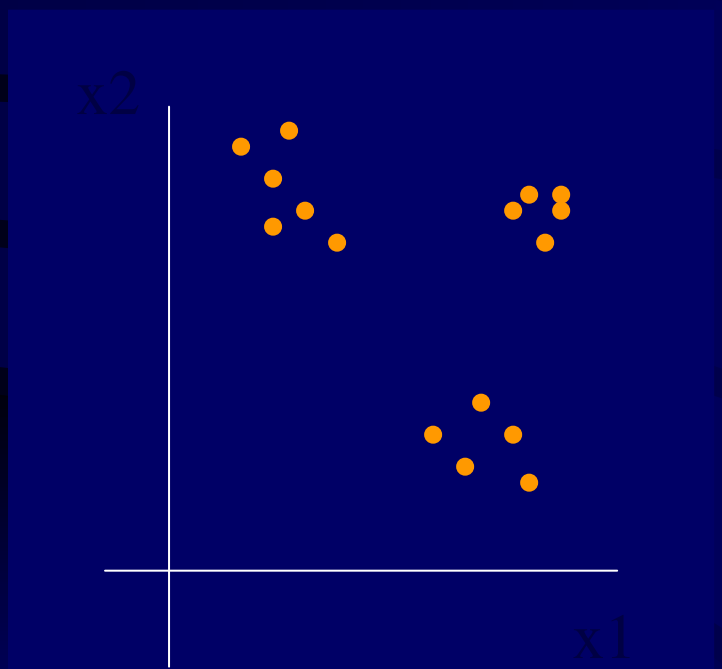
Pi : analytical and mining procedures (kernel parallel)

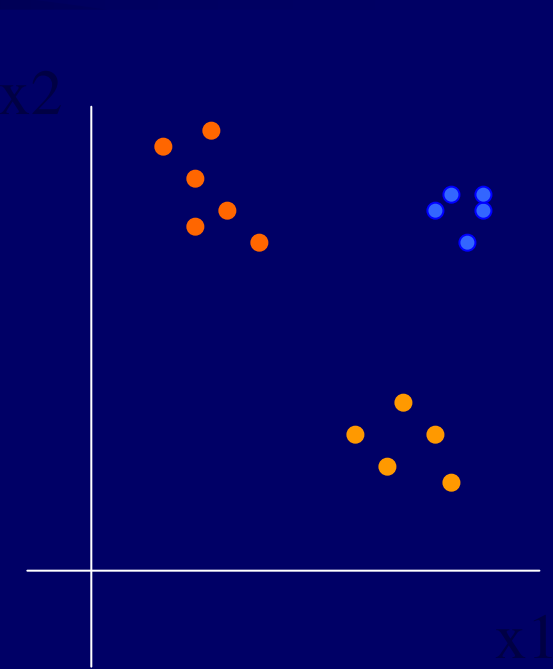
Information Extraction

- Attribute Space

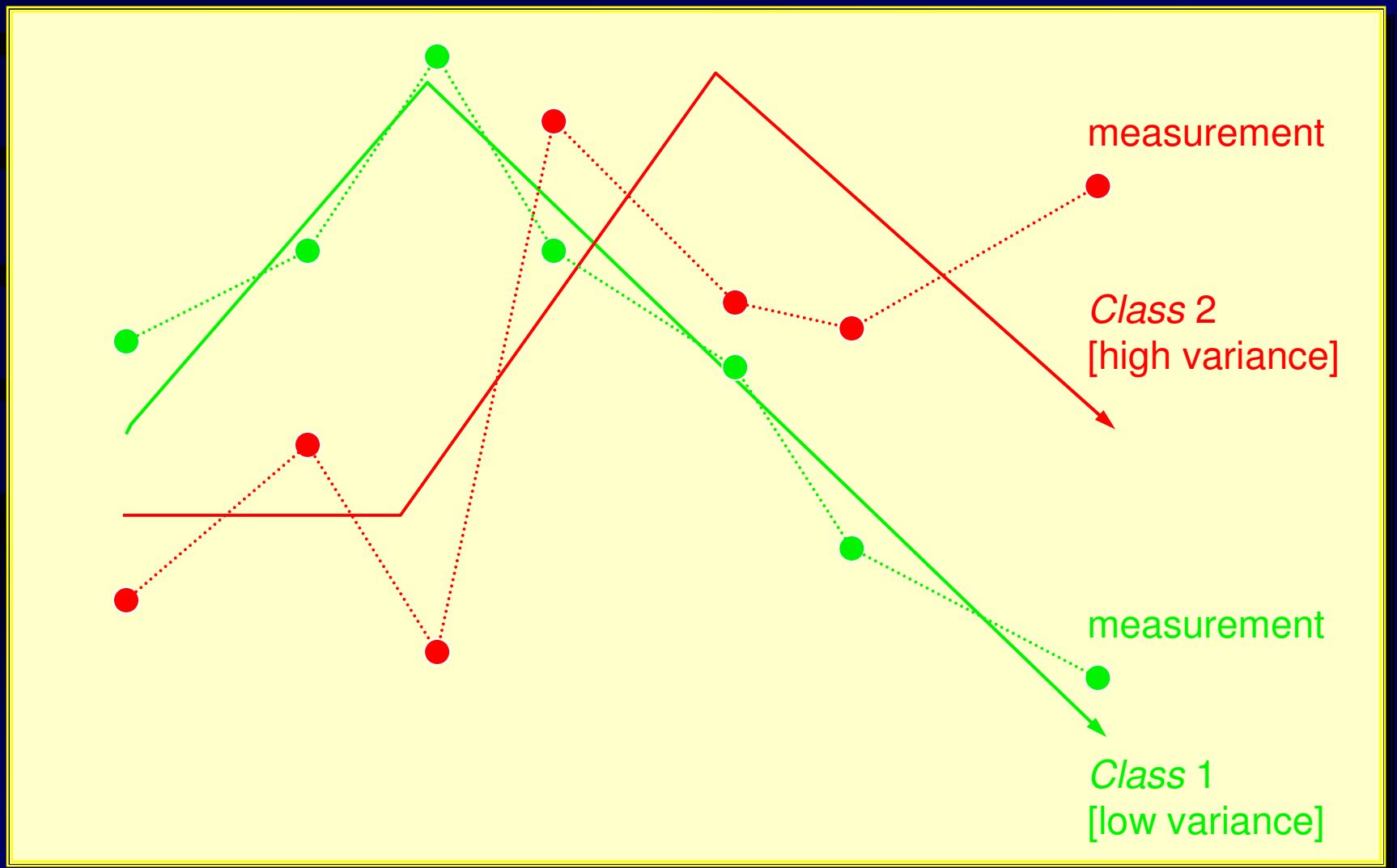


- Clustering

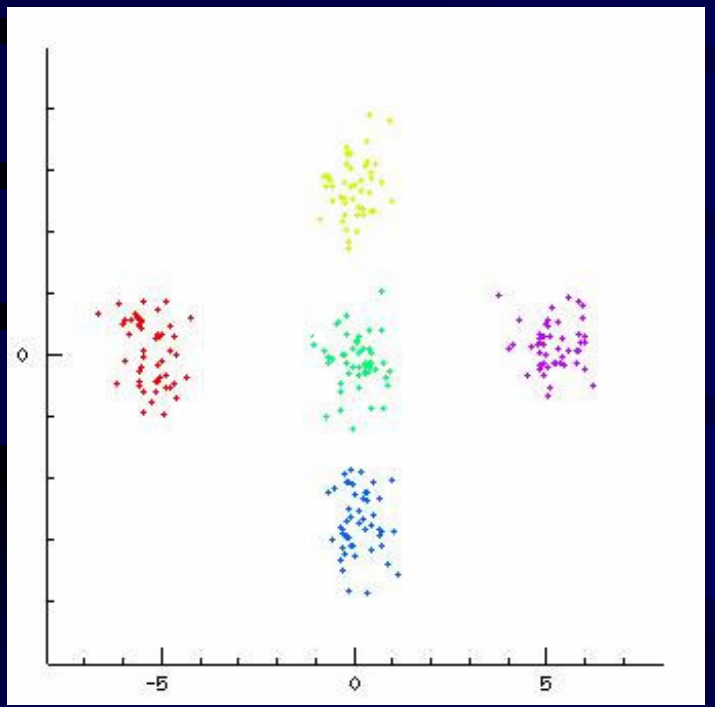


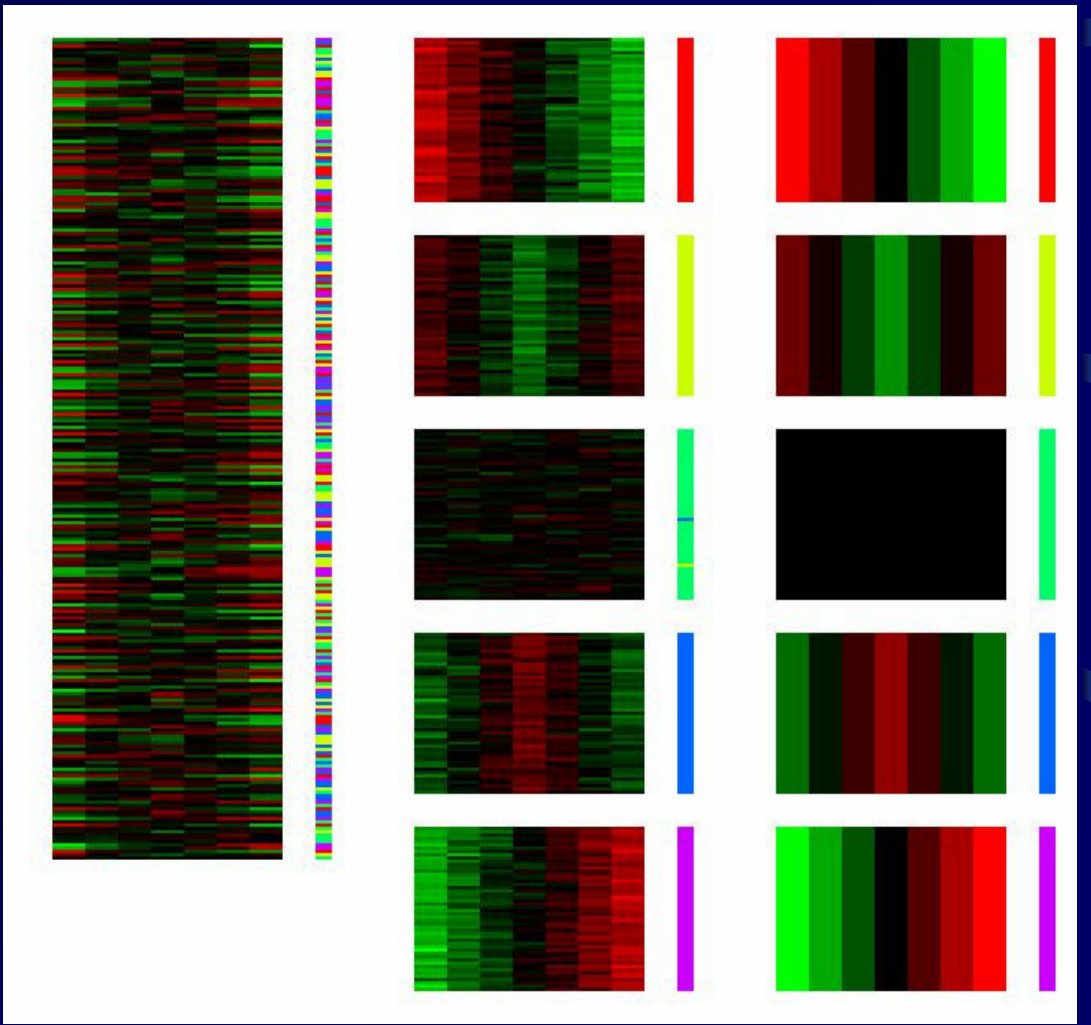


Time Course Model



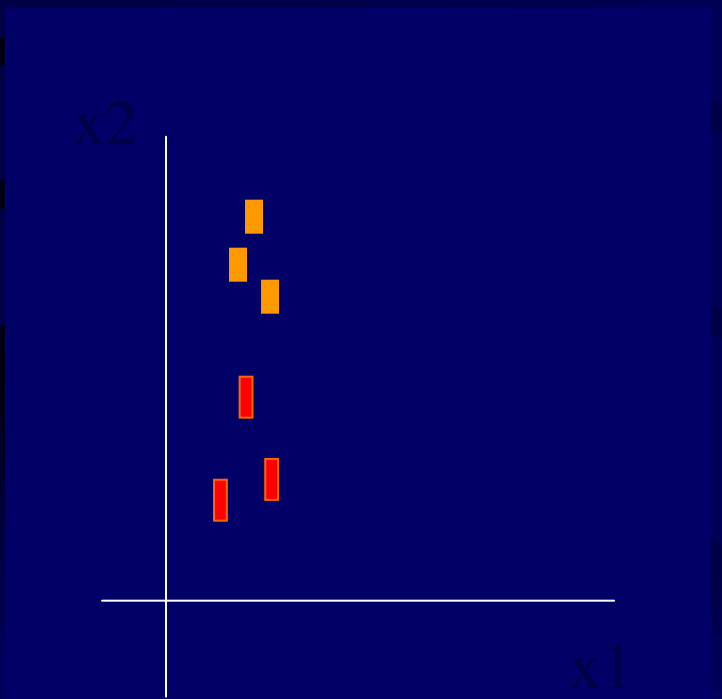
Time course clustering





Results from Fuzzy c-means

- Attribute Space Dimension



CANCER
DIAGNOSTICS

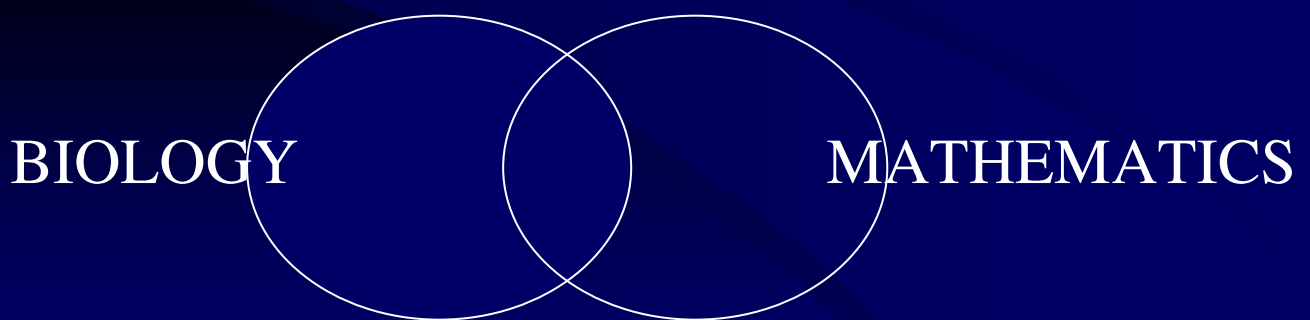
Educational Impact

Multidisciplinary teams

- Biologists
- Biochemists
- Doctors
- Agronomists
- Computer Scientists
- Engineers
- Statisticians

Develop a common language

- Mathematians need to learn Biology and Biochemestry
- Biologists need to learn Mathematics, Statistics and Computer Science



USP PhD program on Bioinformatics

- 8 Institutes of USP
- Students from Bio and Math Sciences
- 2 advisers (Biol/Math) by student
- Mathematicians learn Bio
- Biologists learn Math
- Several application projects