

# Mathematical Modeling in Molecular Biology

Junior Barrera

DCC-IME-USP/BIOINFO-USP

# Projects

Cell cycle,

Tripanosoma-cruzi,

Dictyostelium

....

CAGE

Bioquímica-USP

Cancer

NHGRI

Virus

Hemocentro

Agriculture

Ribeirão-USP

BIOINFO-USP

VGDN

ICB-USP

Agriculture

ESALQ

OMS

ICB-USP

Malária

ICB-USP

Cancer

Ludwig

# Layout

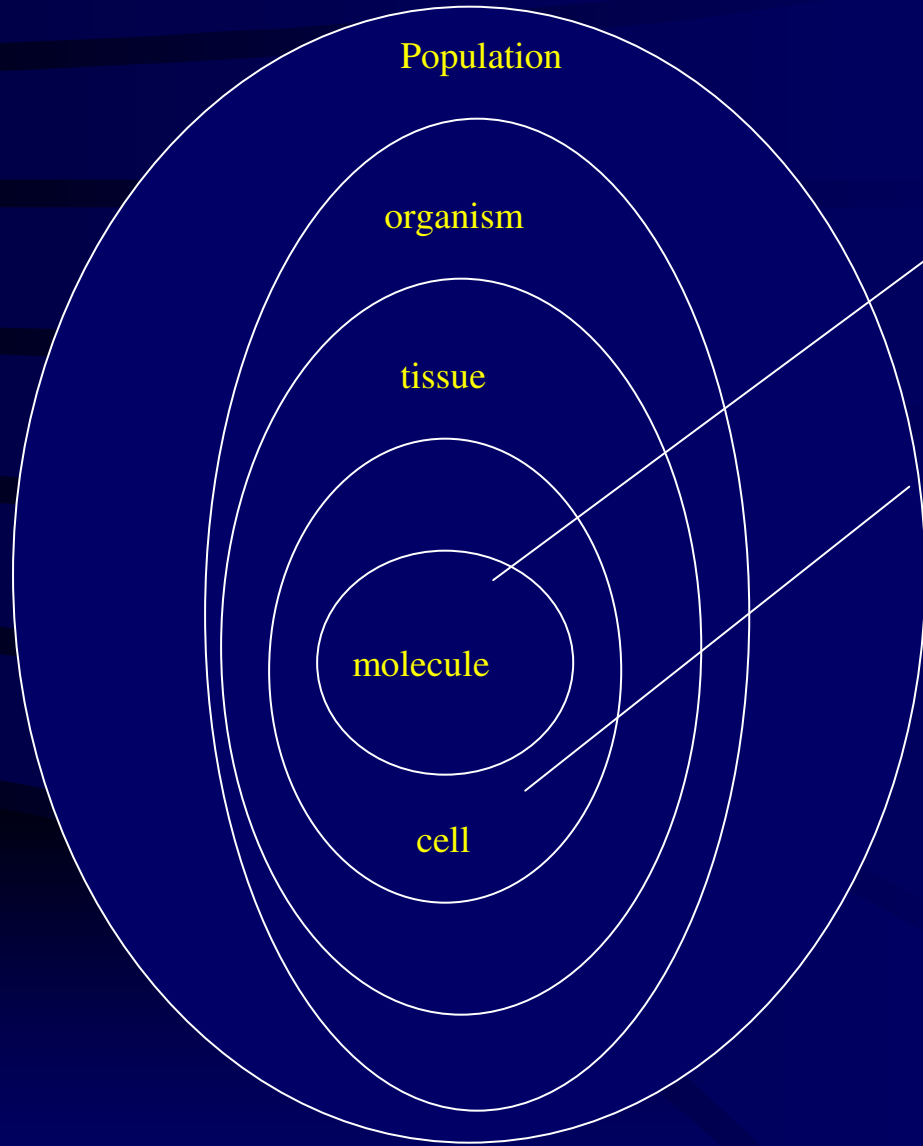
- Economics perspective
- Fundamentals of Molecular Biology
- Experimental techniques
- Data mining environments
- Some Biological problems
- Formalizing Biological problems
- A family of Mathematical problems
- Other areas with similar problems

# Economics Perspective

- Atomic program (1930-1945)
- Space program (1960-1975)
- Genetics (1990- ...): international investment, public and private

Brazilian investment is proportionally greater than the investment of some developed countries

# Fundamentals of Molecular Biology

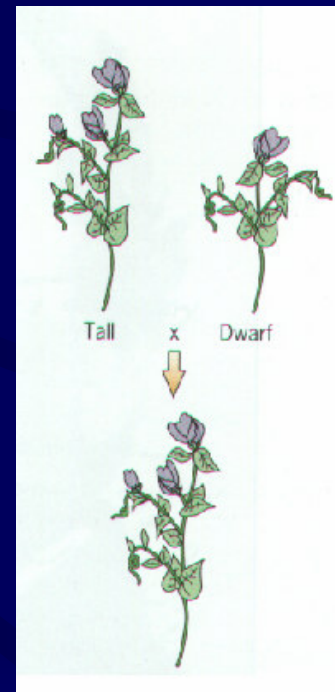


Protein structure  
and dynamics

DNA, Protein,  
Gene Expression,  
Gene Networks

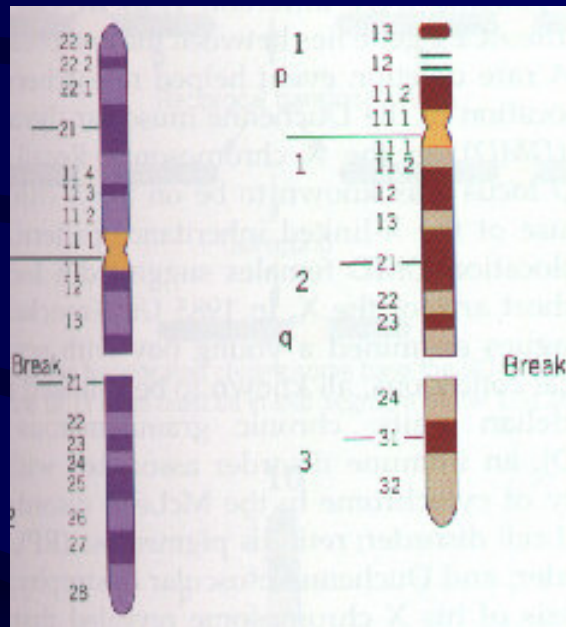
## Heredity - Mendel (1866)

The phenotypes of an individual depends on genes of his parents.



# Chromosome Theory - Morgan (1910)

Genes were situated in chromosomes

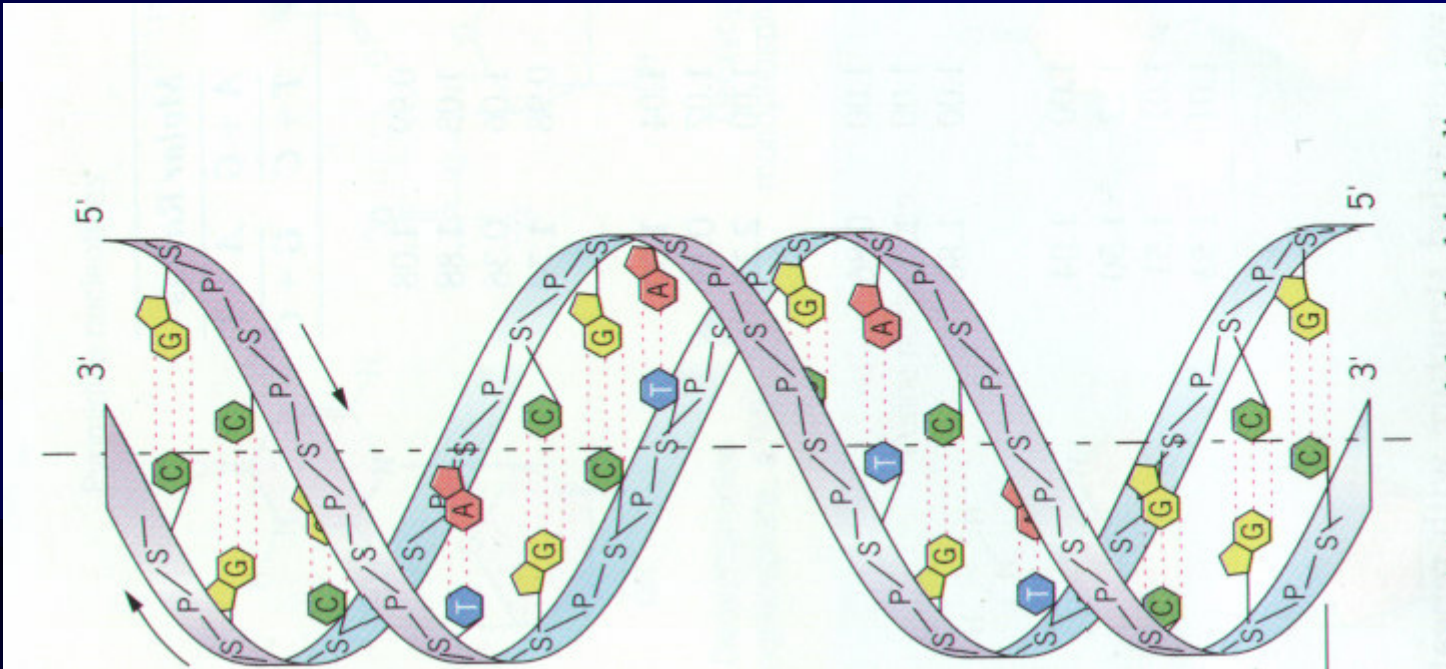




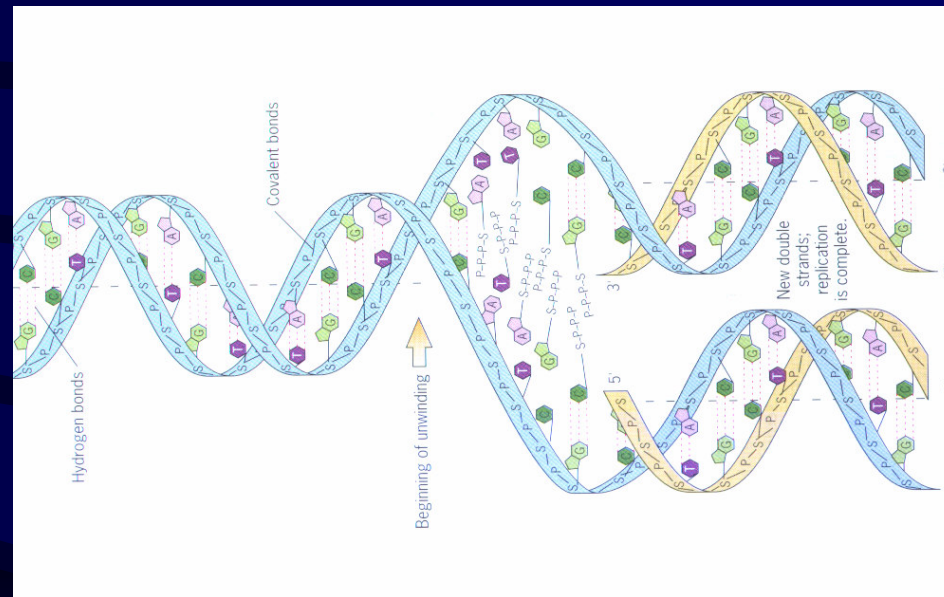
The molecular structure of chromosomes (Watson and Crick - 1953)

DNA structure: the double helix

Four basis: adenine(A), guanine(G), thymine(T), cytosine(C)  
genes are sequences of nucleotides



# cut, replication and decoding

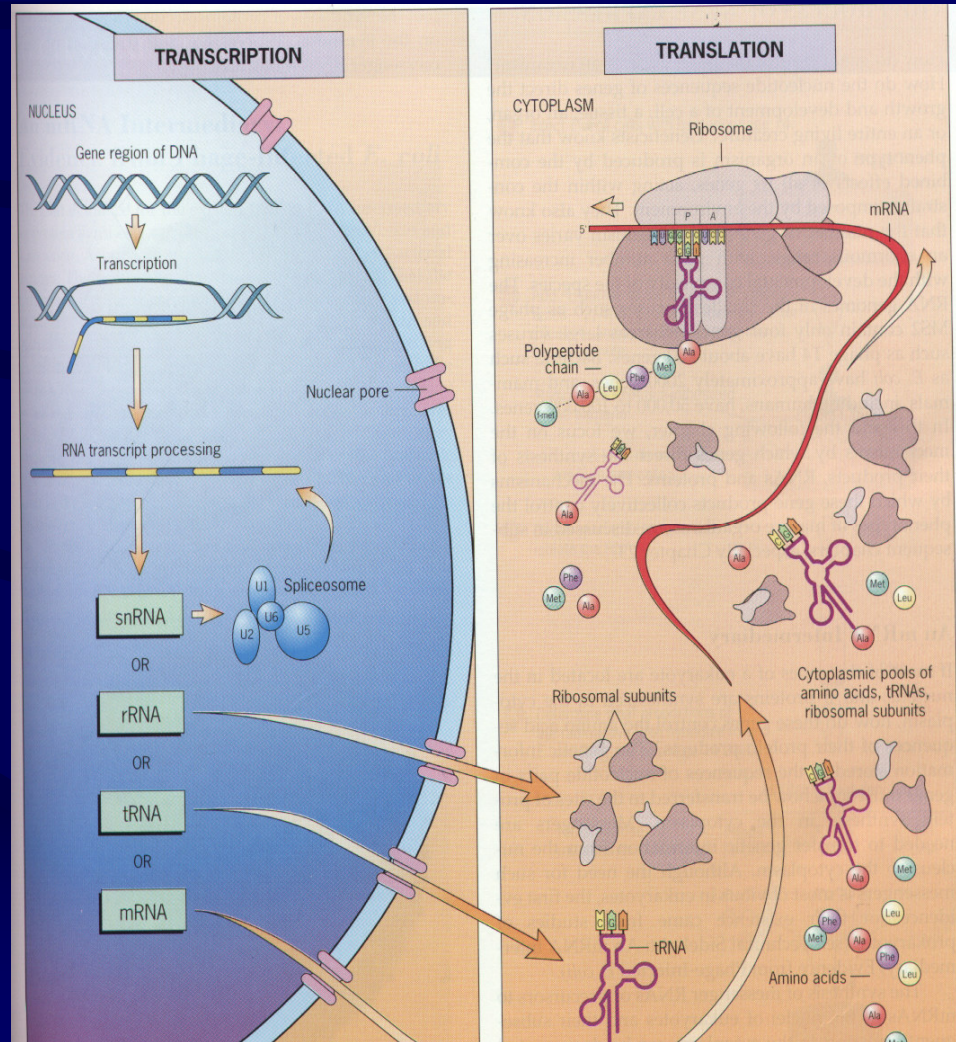


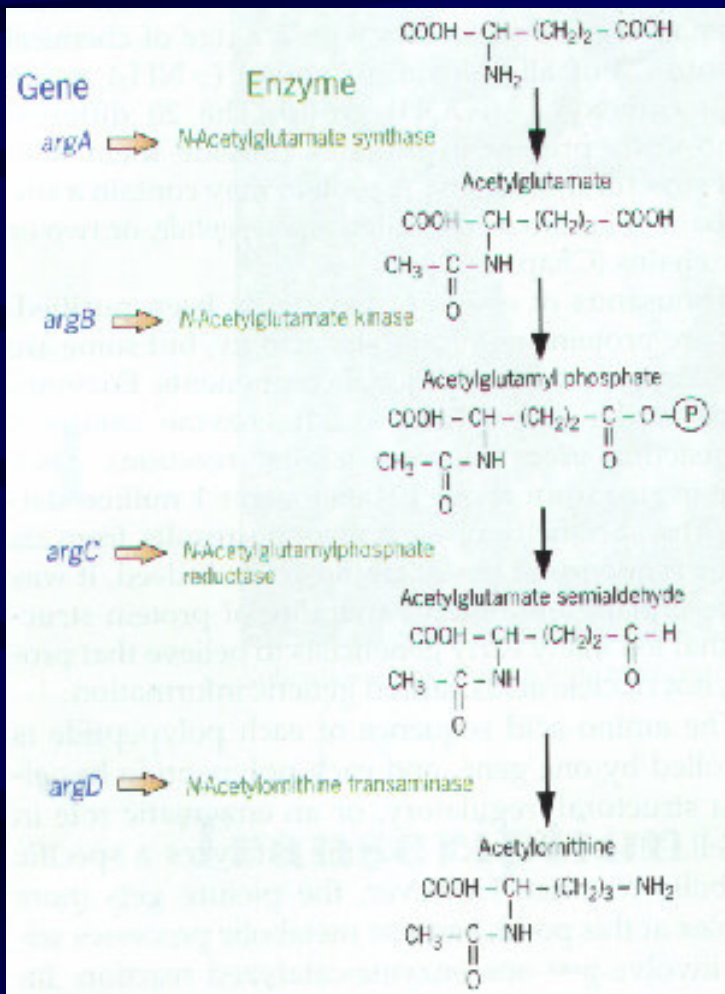
Genes control the metabolism

Metabolism occurs by sequences of enzyme-catalyzed reactions.

Enzymes are specified by one or more genes

# Gene expression



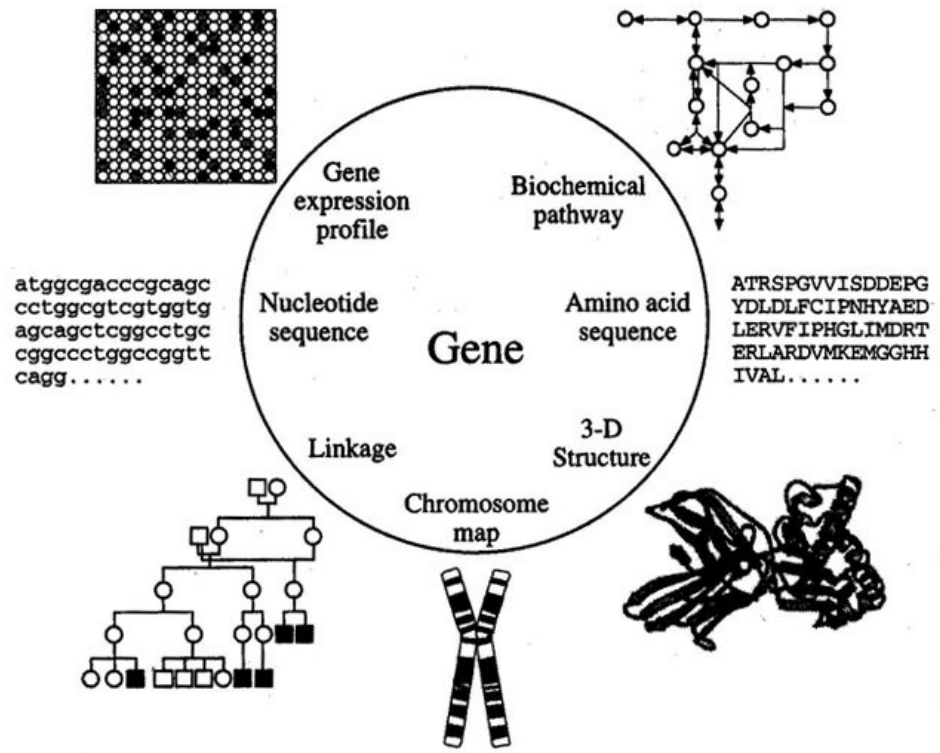


species modification, diagnostics, drug production

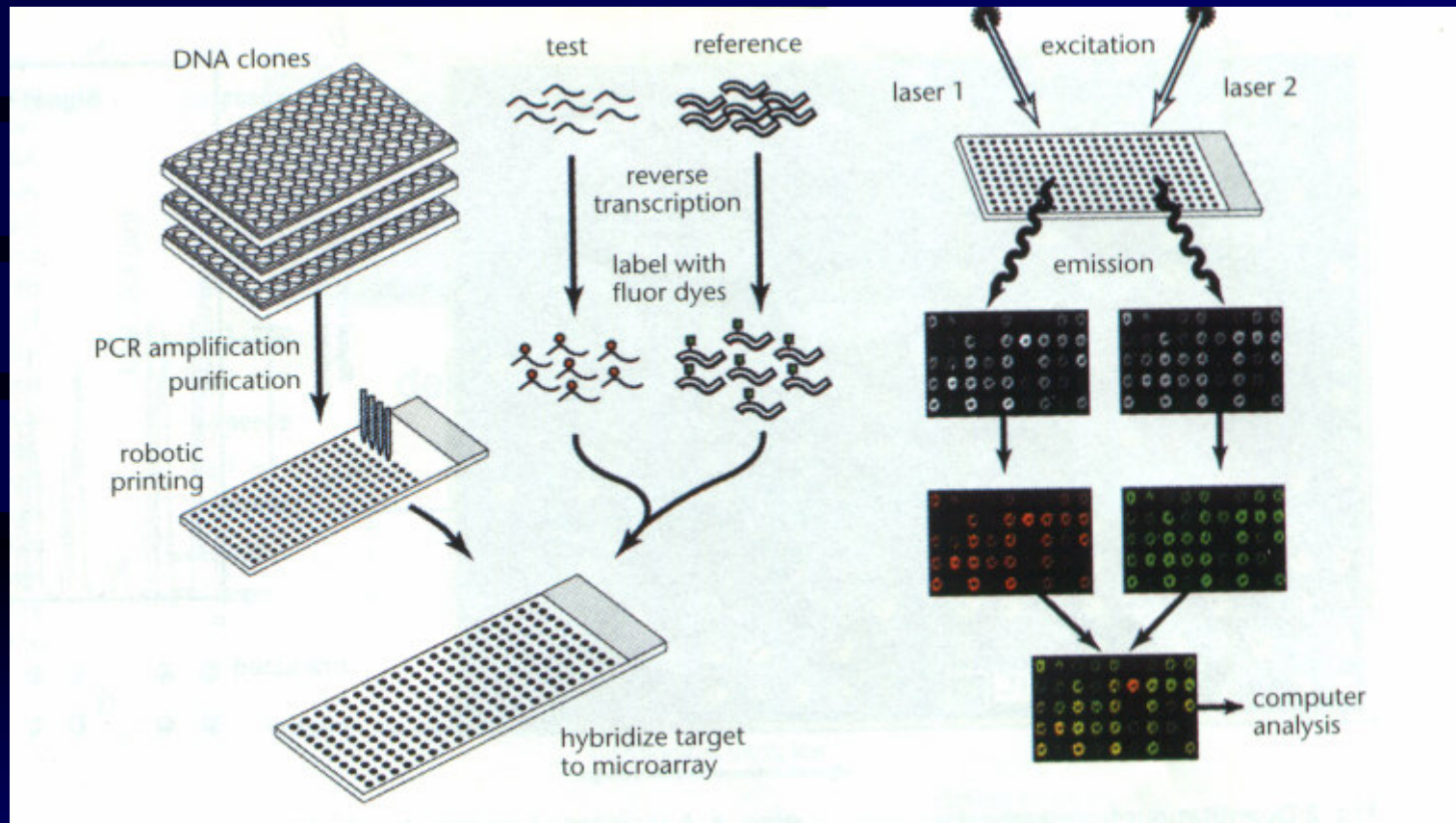


# Experimental techniques

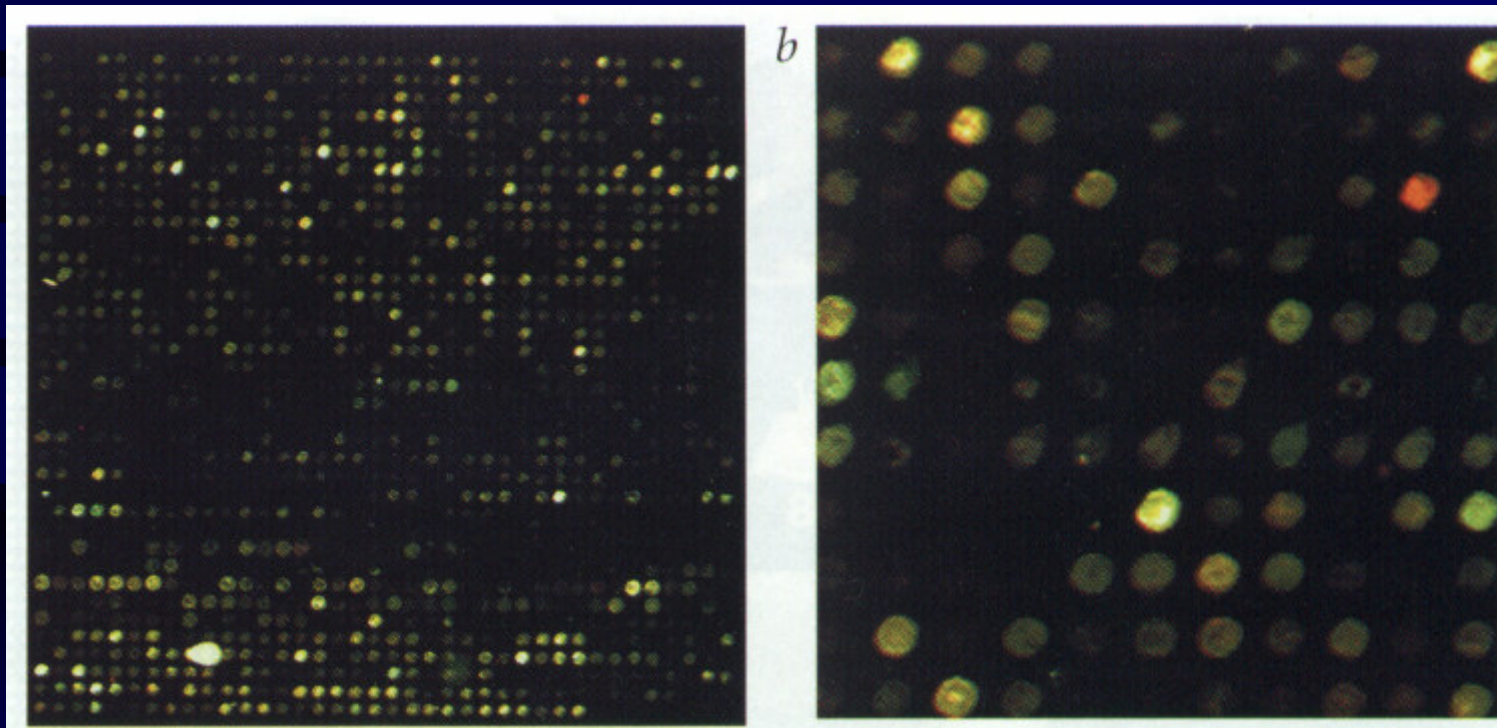




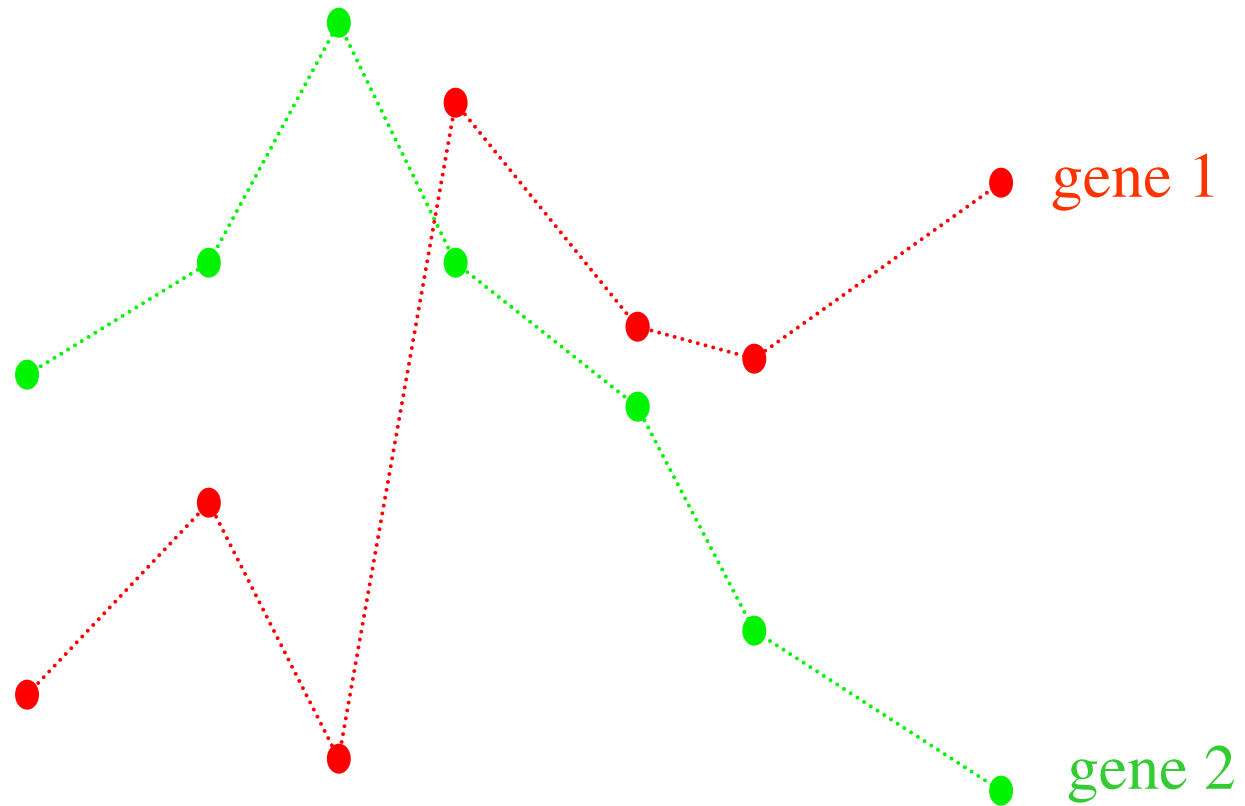
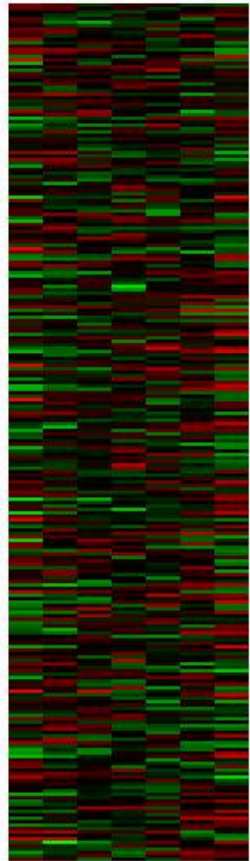
# Microarrays



# Image Analysis

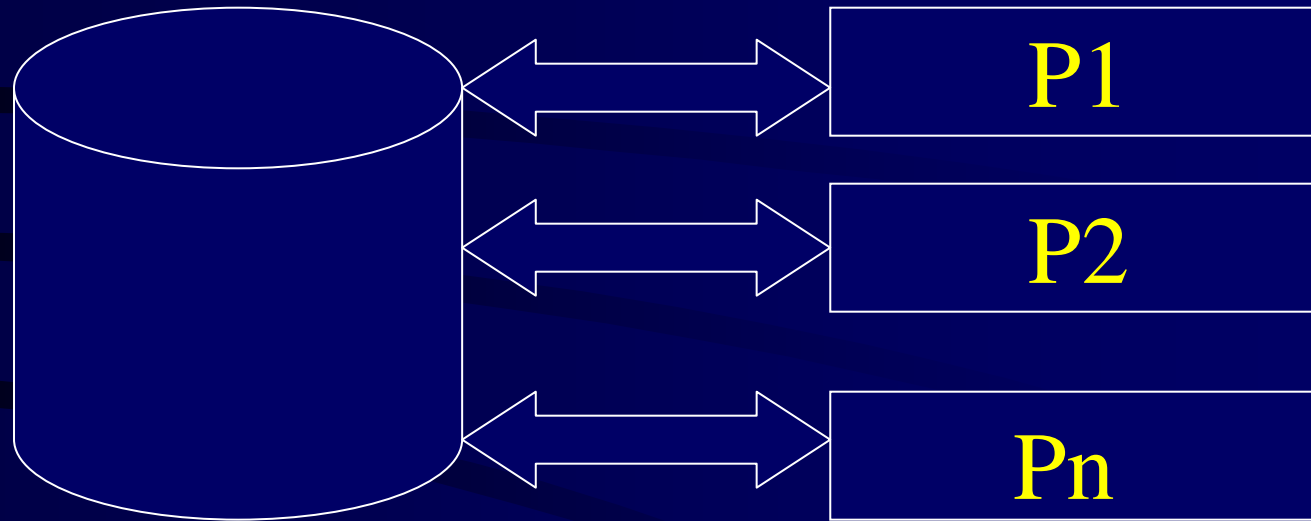


# Signal



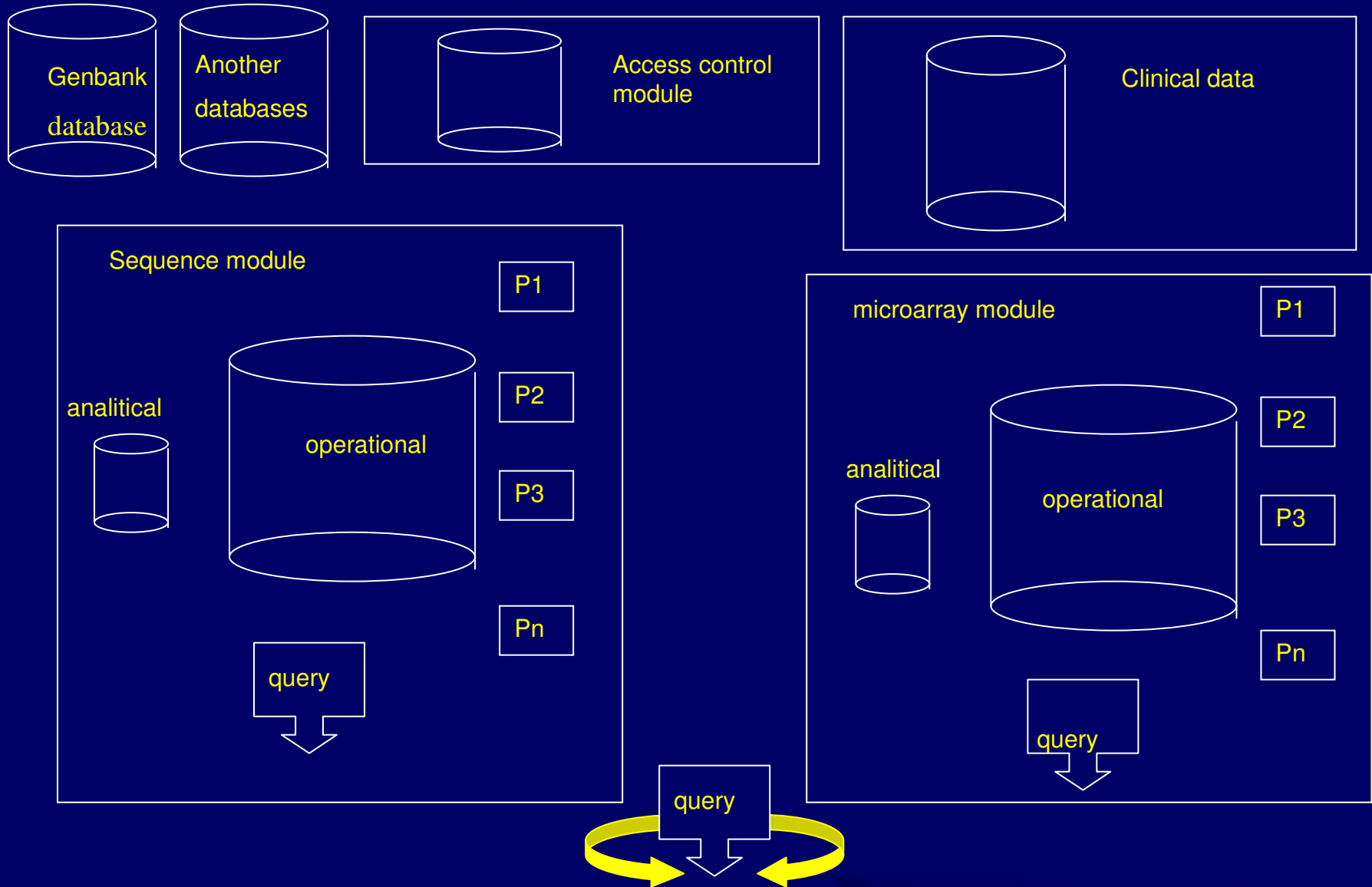
# Data Mining Environments

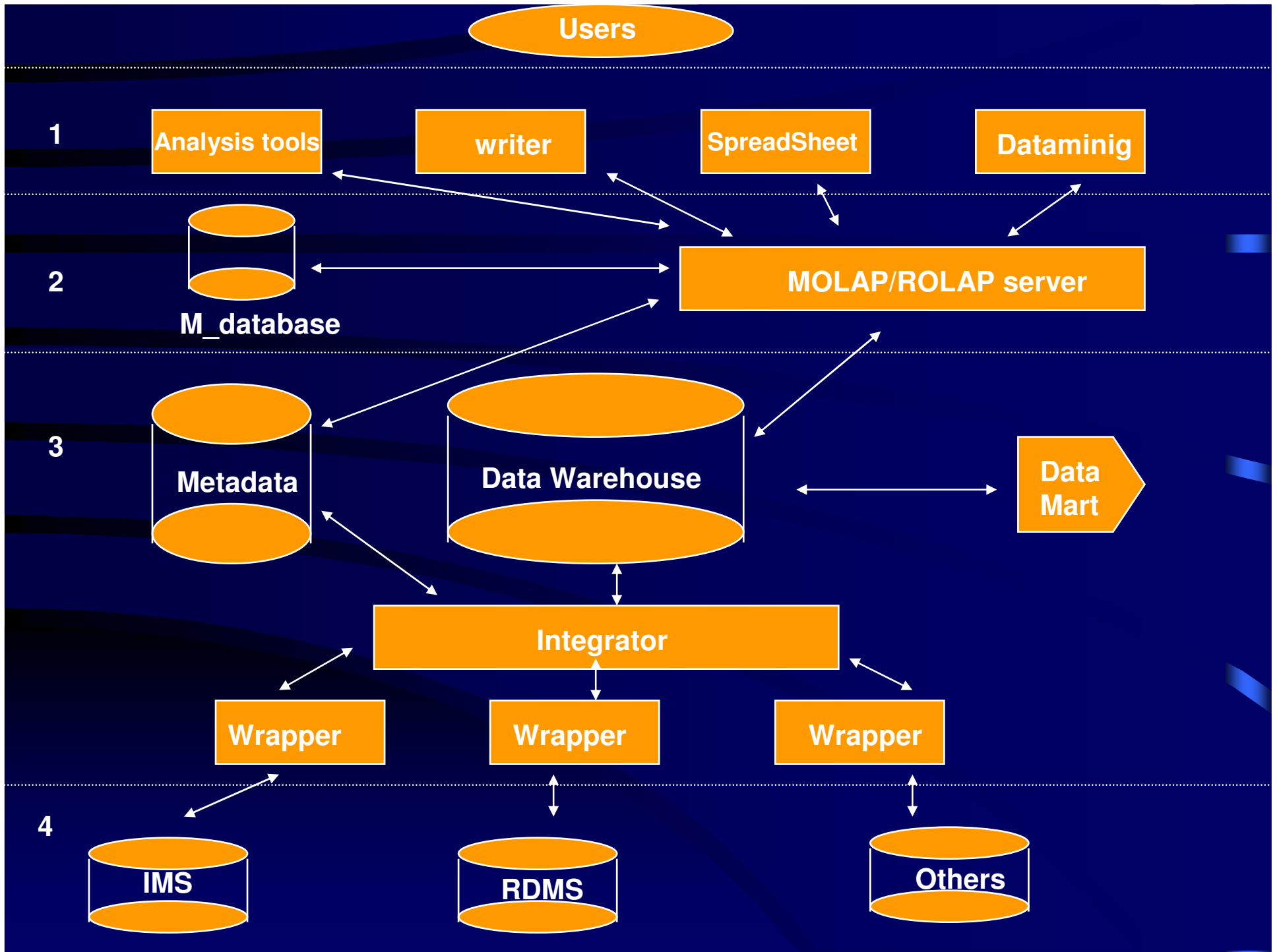
## Objected oriented database



$P_i$  : analytical and mining procedures (kernel parallel)

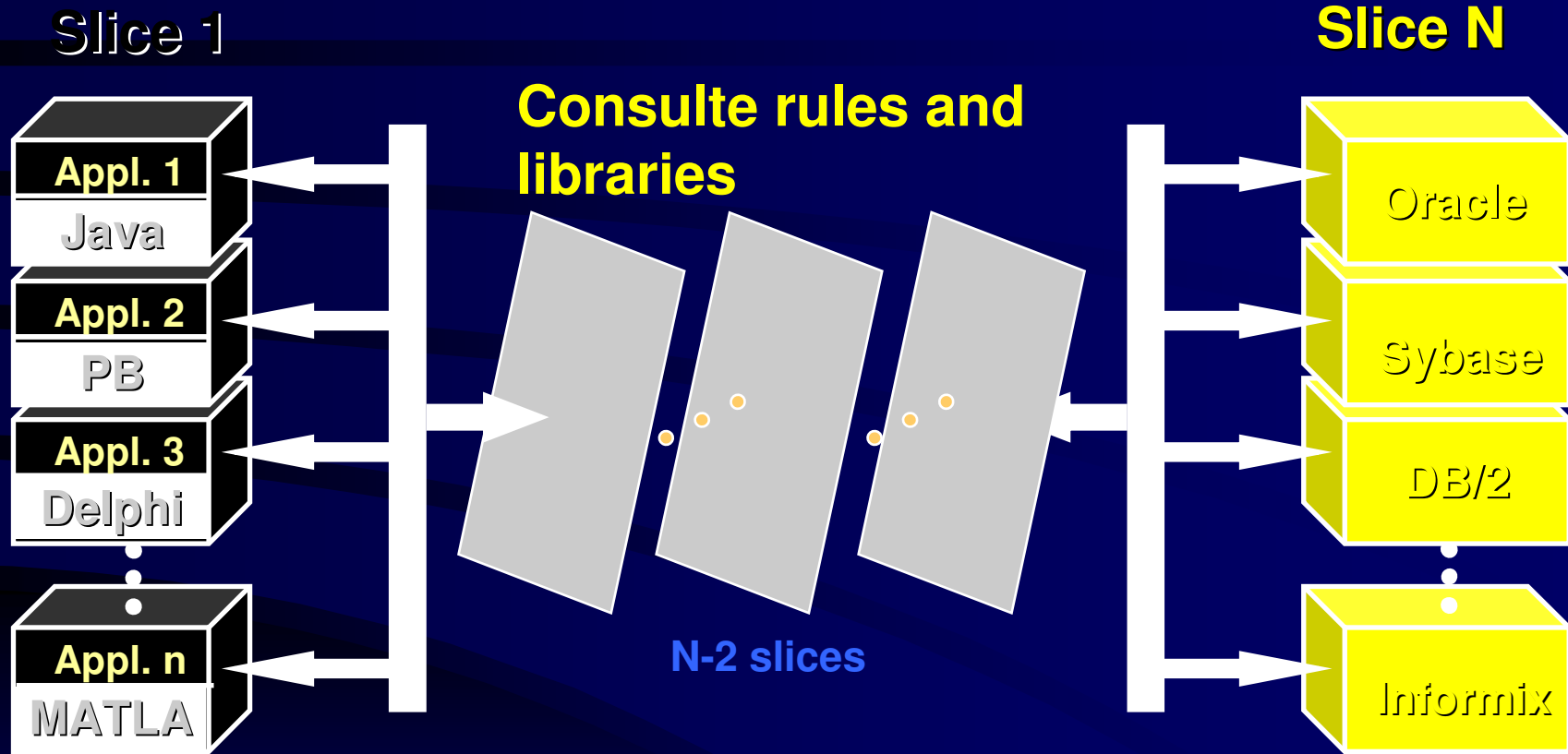
# Integrated Environment







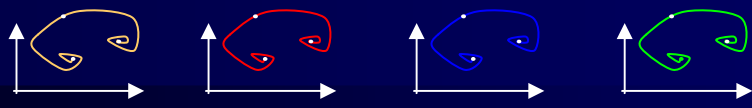
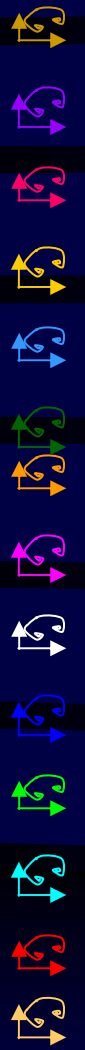
# System Architecture



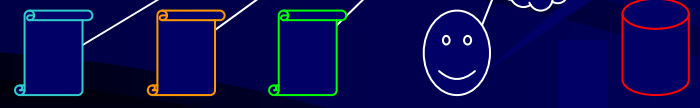
# Some Biological problems

- Modification of Sugar Cane, Eucalyptus and chickens
- Cancer diagnostics
- Drug performance against HIV and Malaria
- Understanding of the cell division cycle
- Reconstitution of nervous tissue
- Design of drugs
- Prediction of new virus

# Formalizing Biological Problems

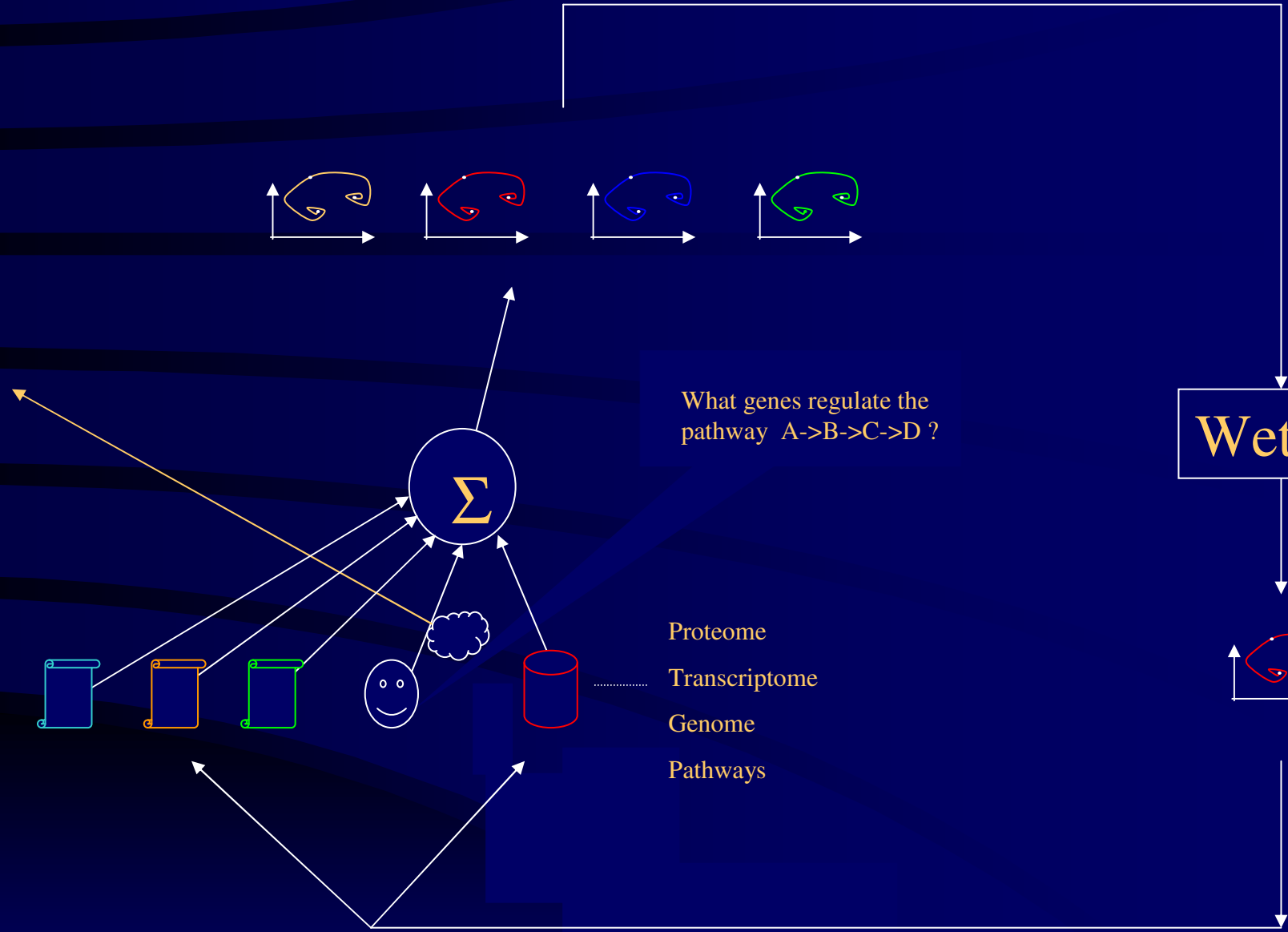
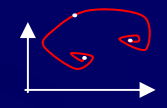


What genes regulate the pathway A->B->C->D ?



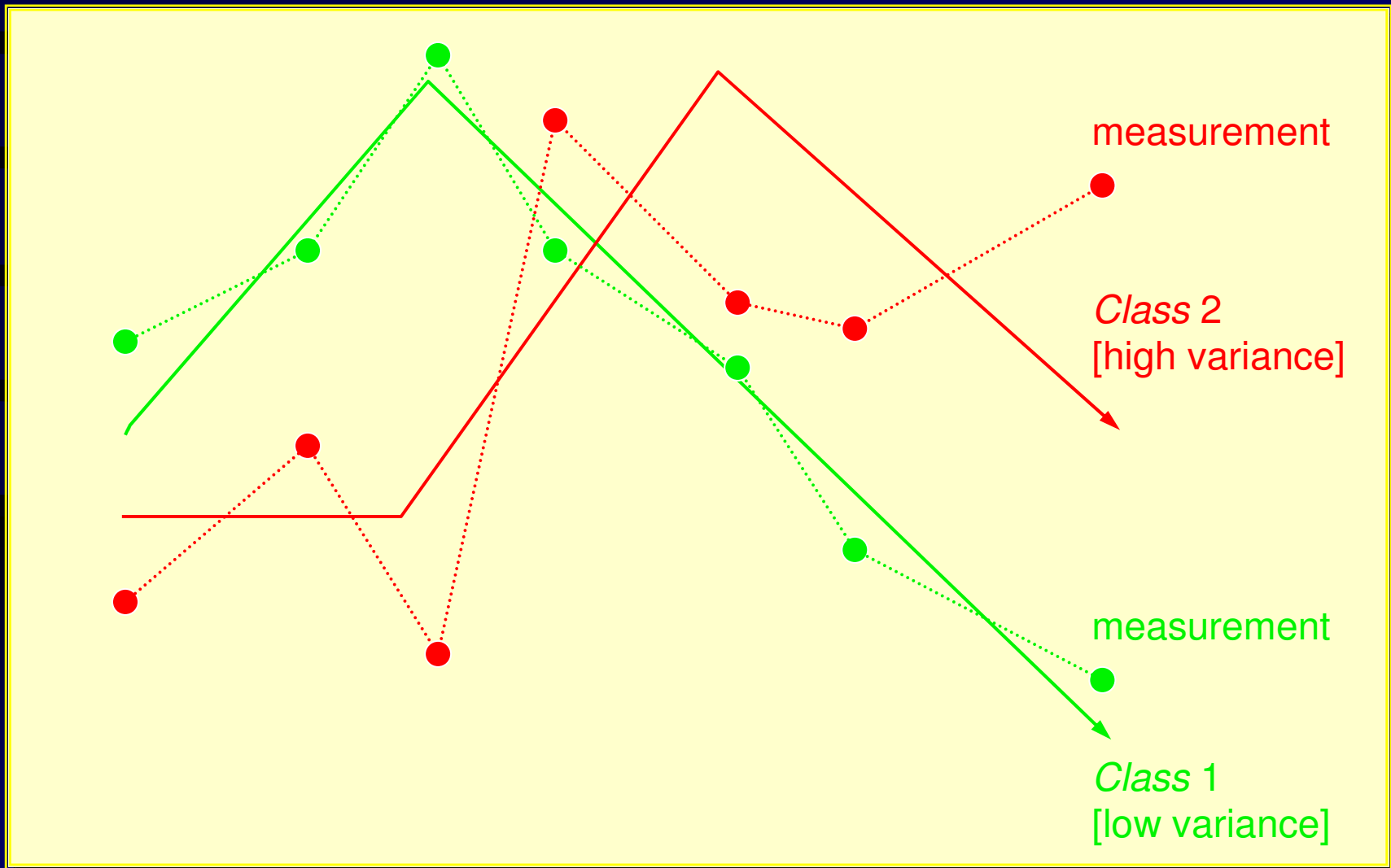
- Proteome
- Transcriptome
- Genome
- Pathways

Wet Lab



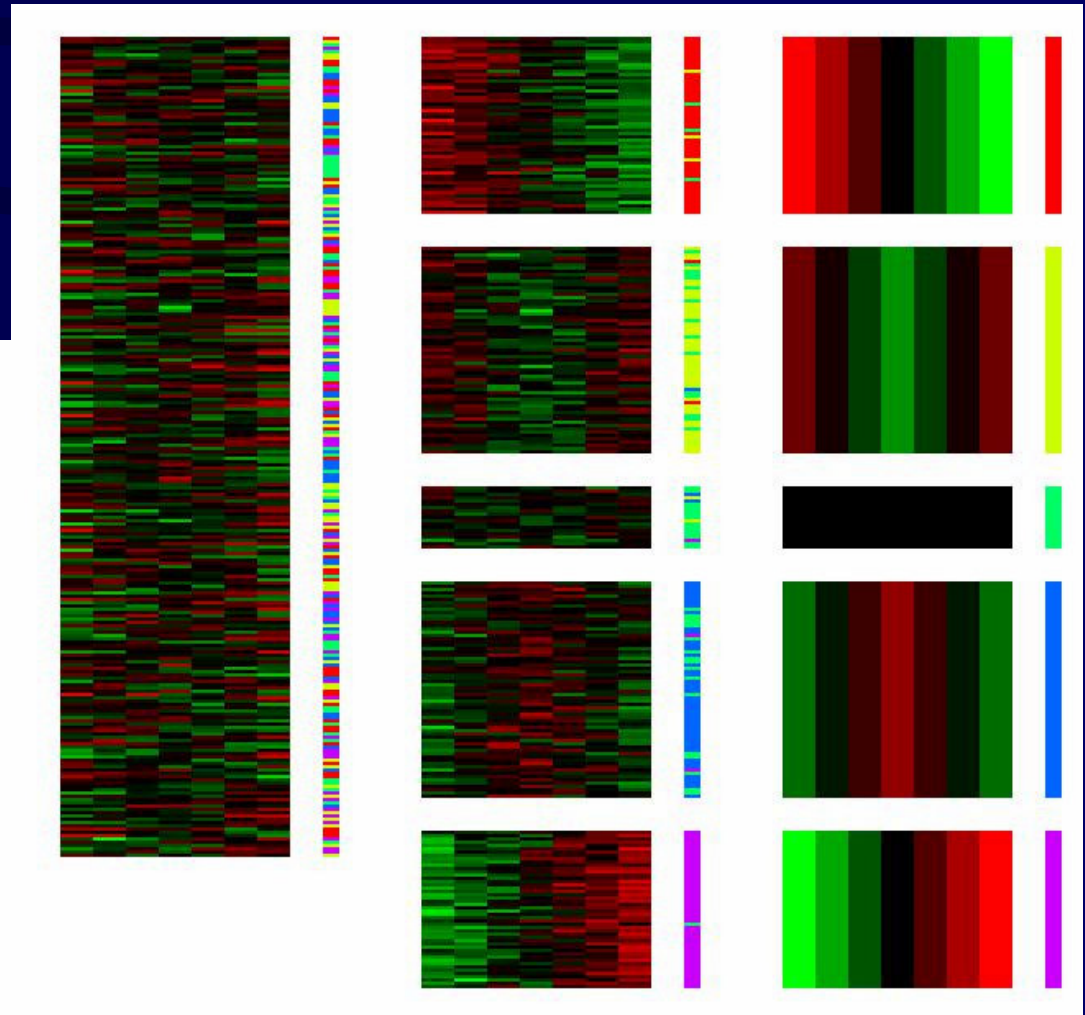
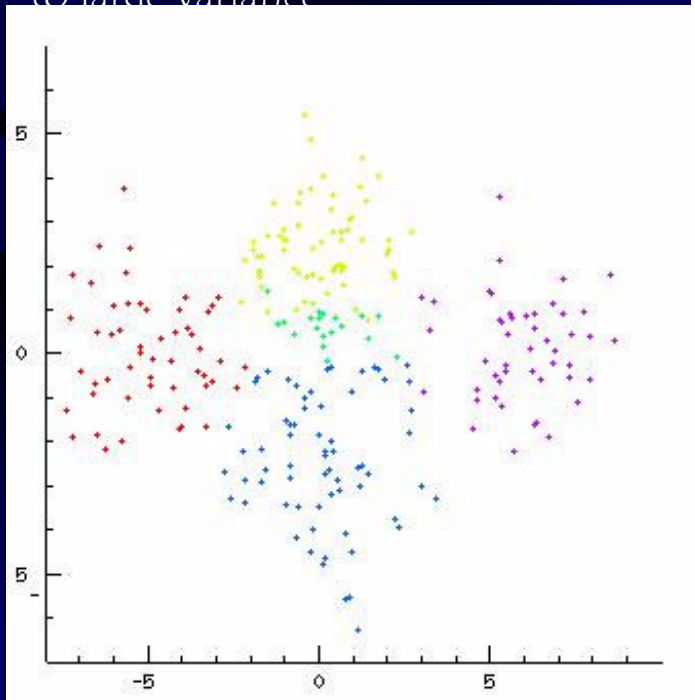
# Choice of adequate clustering technique

# Clustering

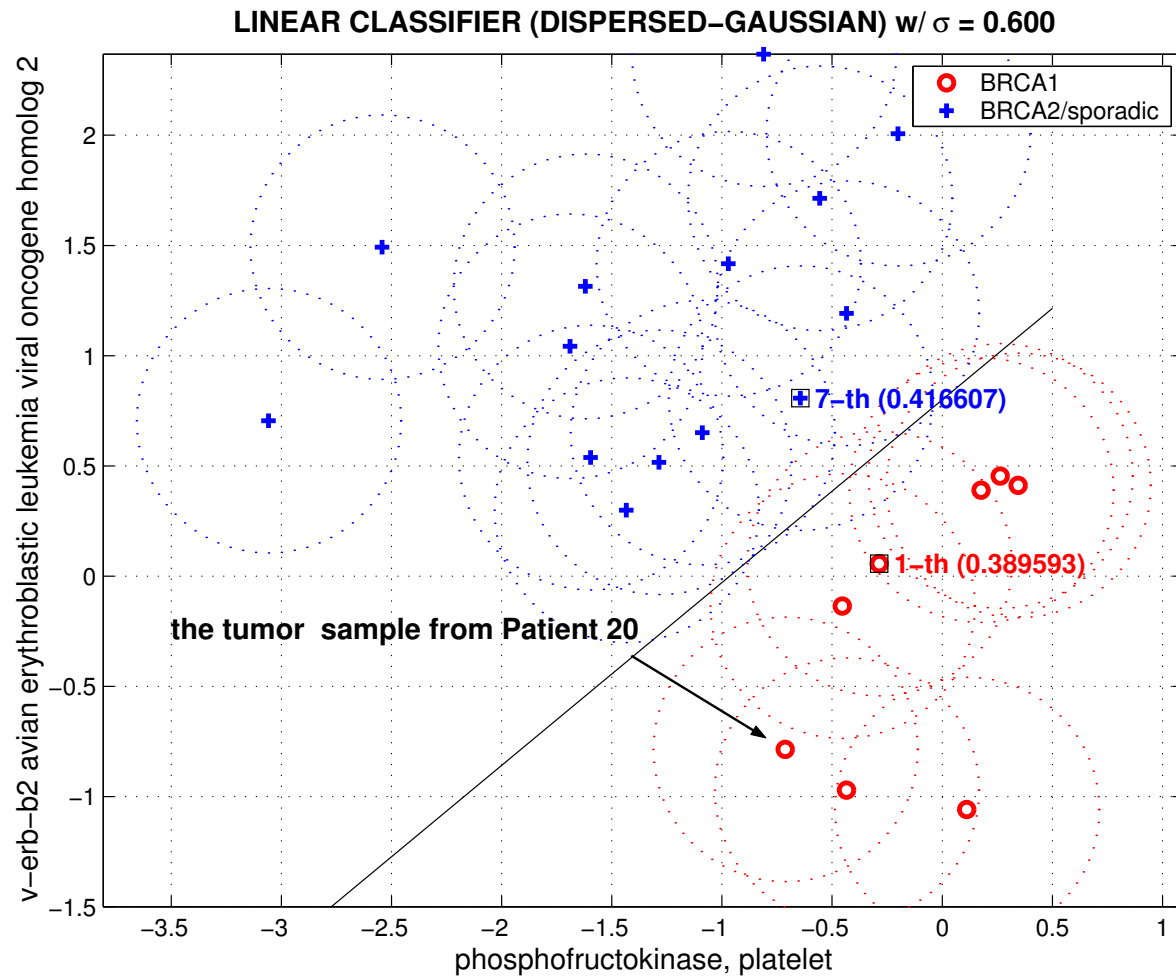


# Clustering

Looser clusters due  
to large variance

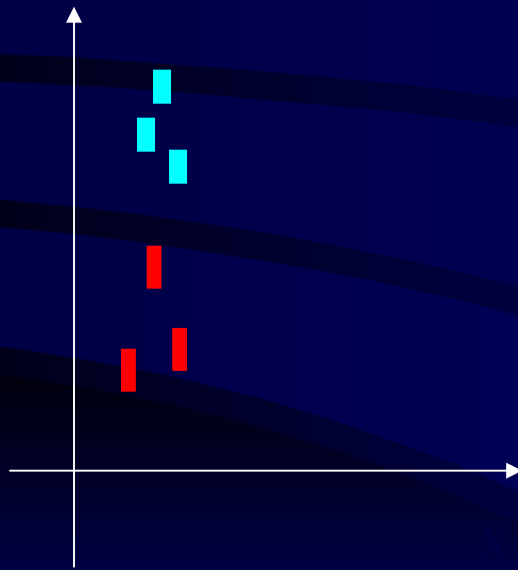


# Design classifier



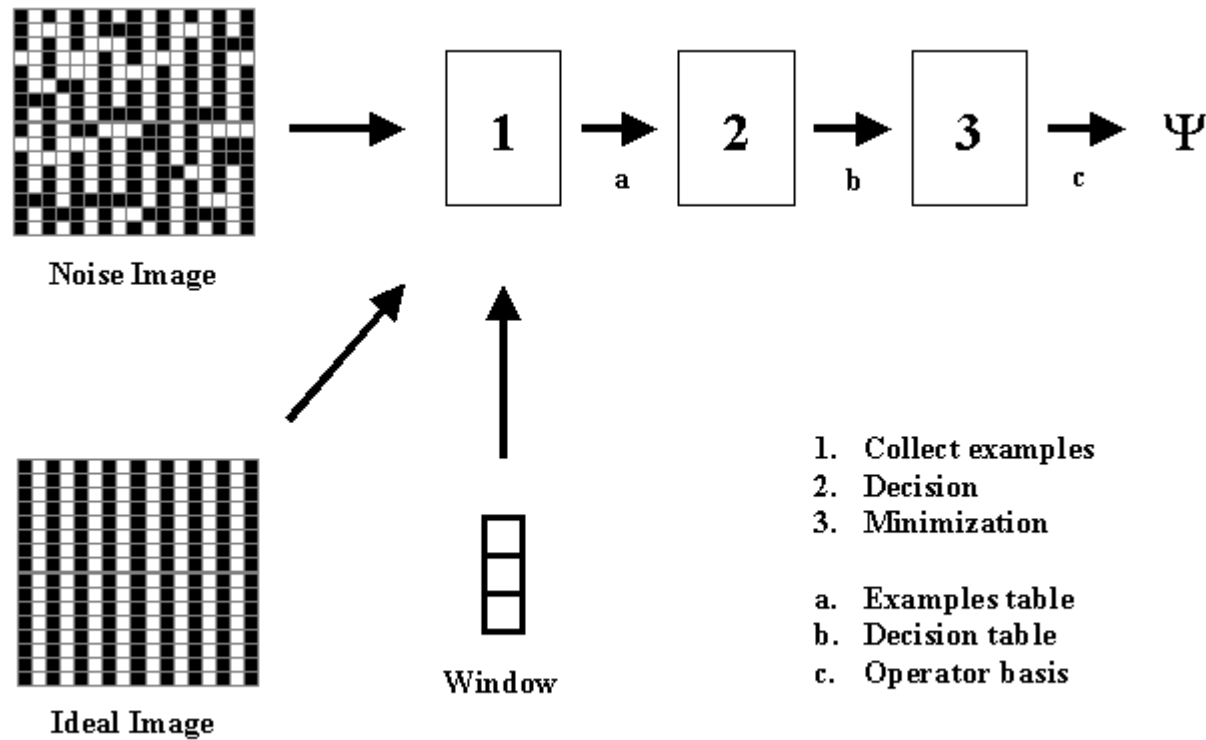


# Dimensionality Reduction

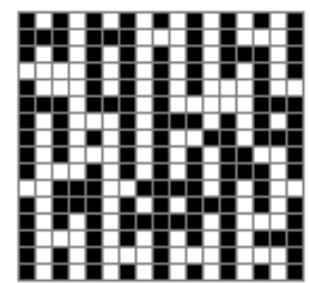


What is the minimum number of genes that is enough to distinguish two Biological states?

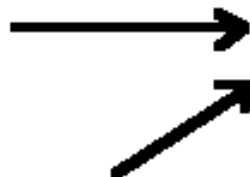
# Filter Design



# Application

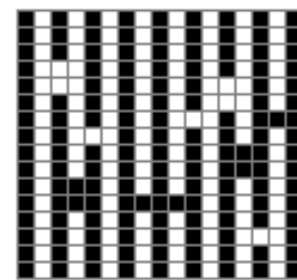


**Noise Image**



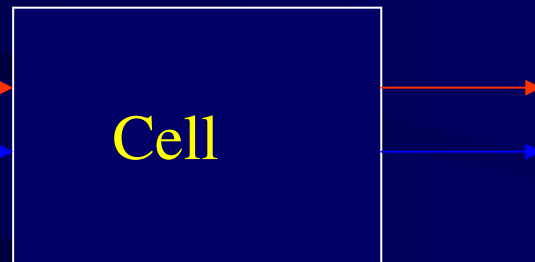
Operator basis  
 $\{[110,111], [101,111], [011,111]\}$

$\Psi$



**Restored Image**

# Modeling Dynamical Systems



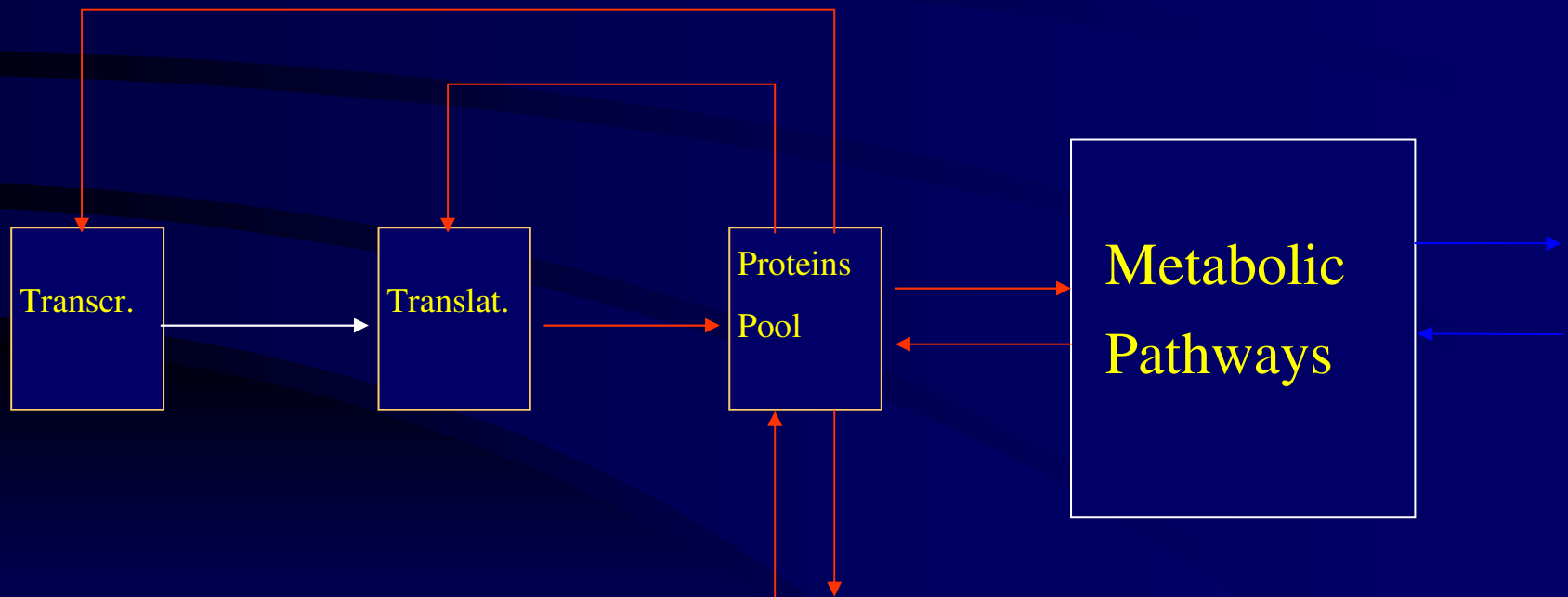
 peptide

 other signals

■ peptide

■ other signals

■ mRNA

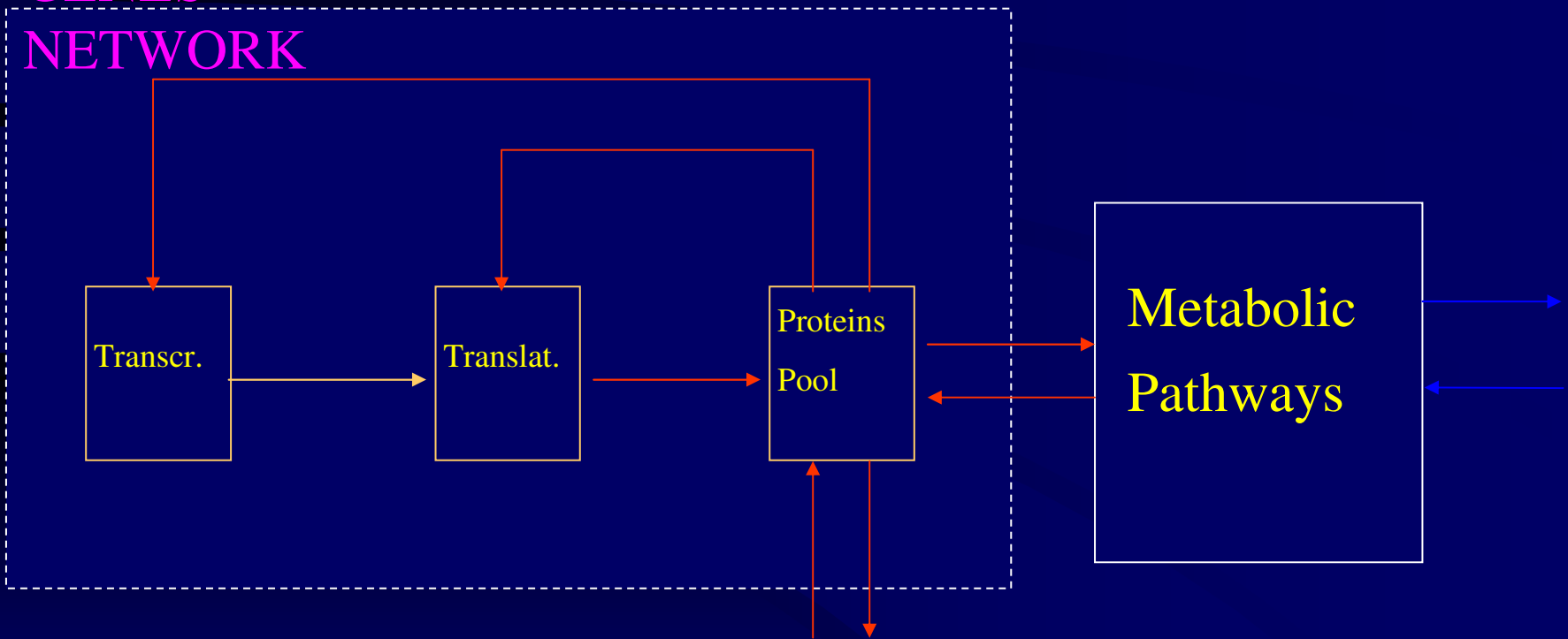


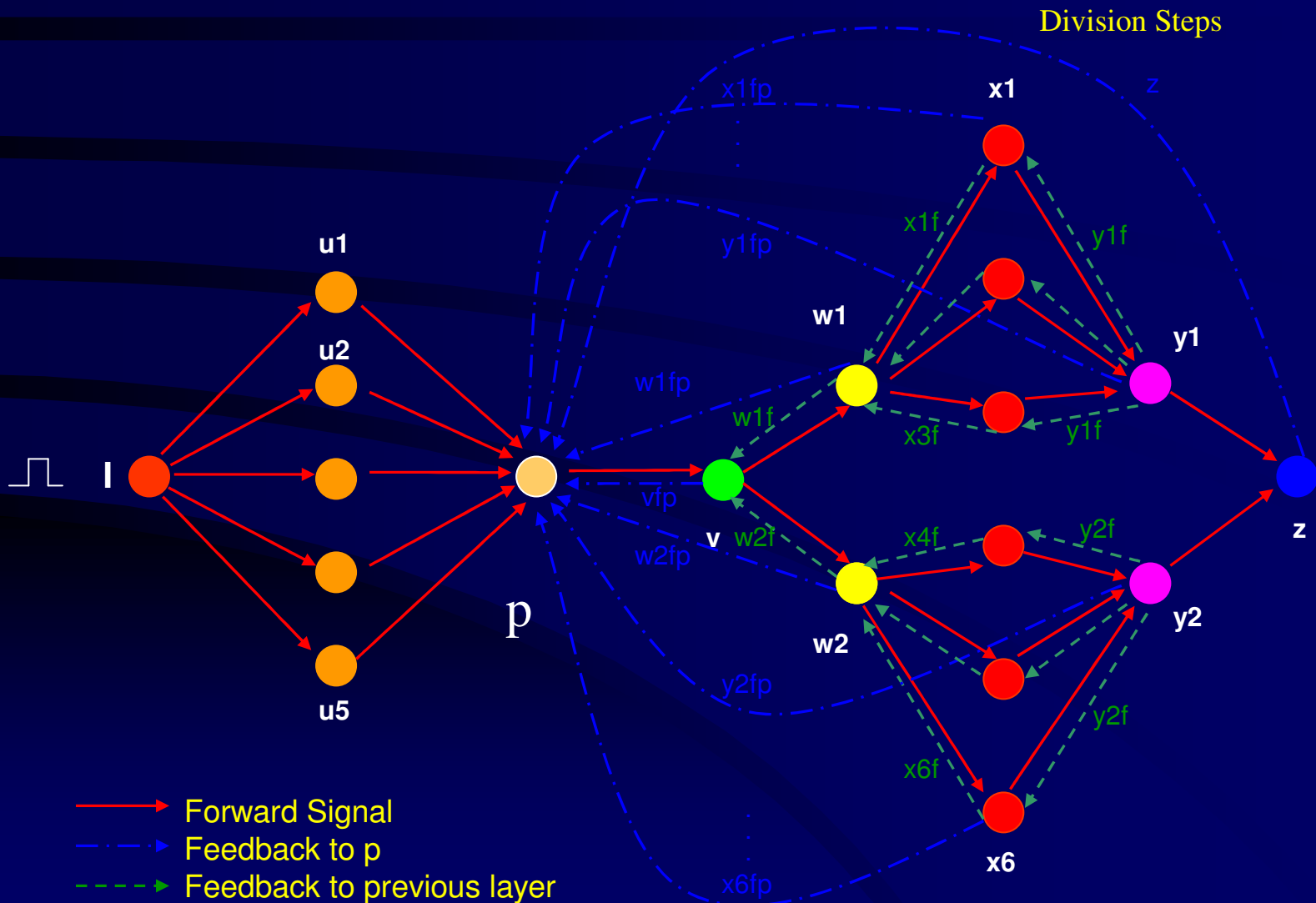
■ peptide

■ other signals

■ mRNA

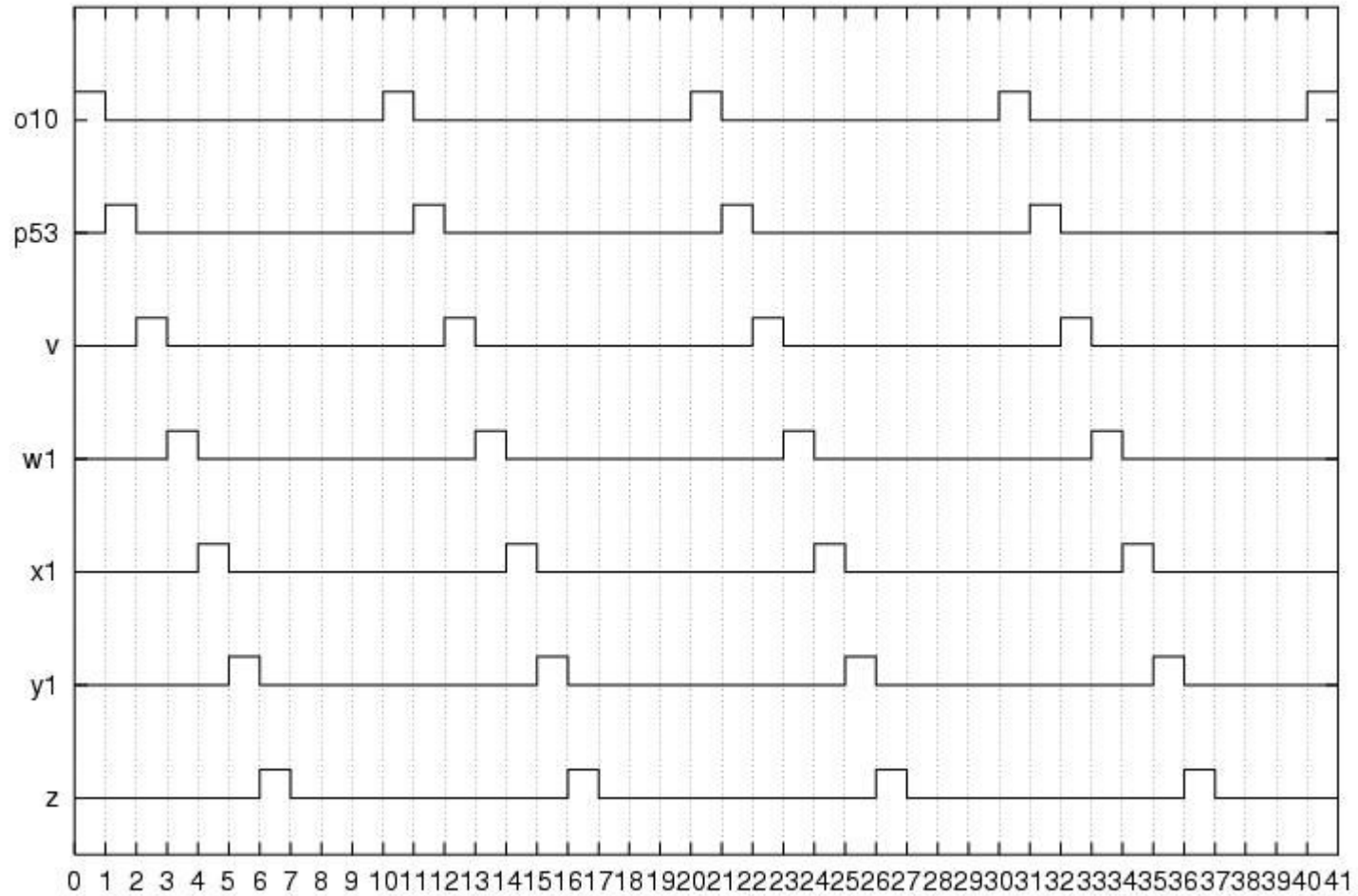
## GENES NETWORK



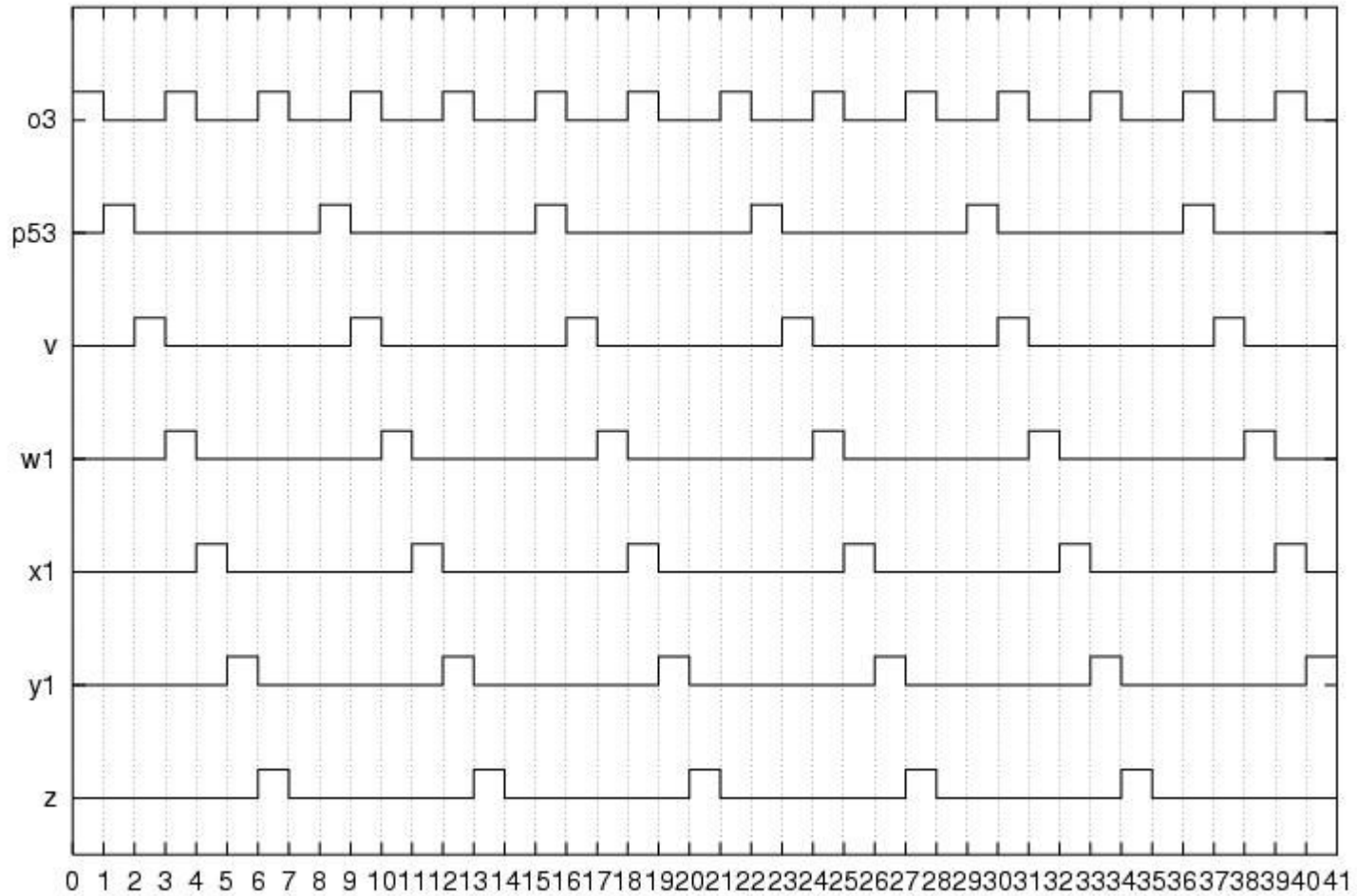




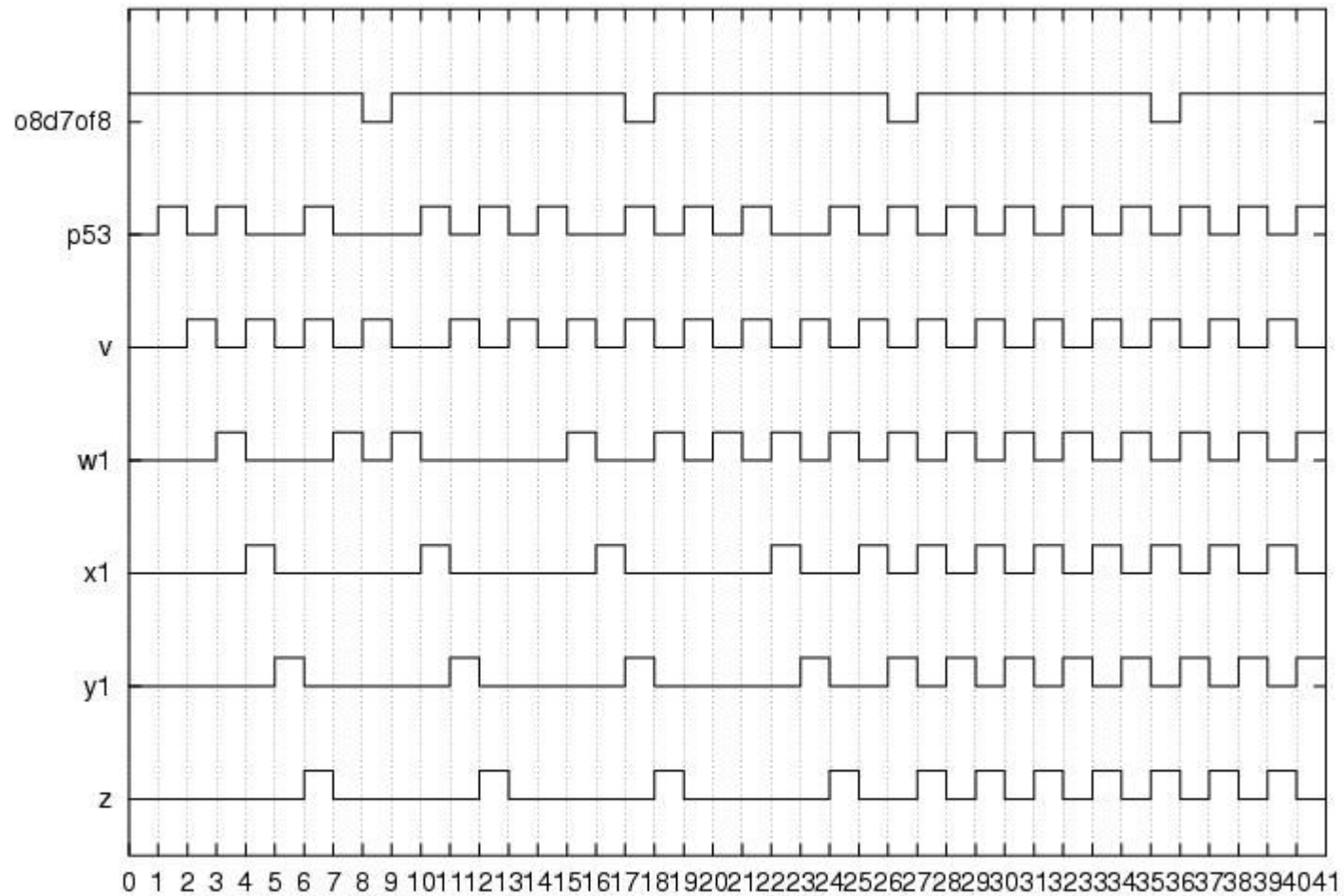
Oscilador de Período 10: FUNCIONAMENTO GERAL (parte\_B-t4A-o10.sim)



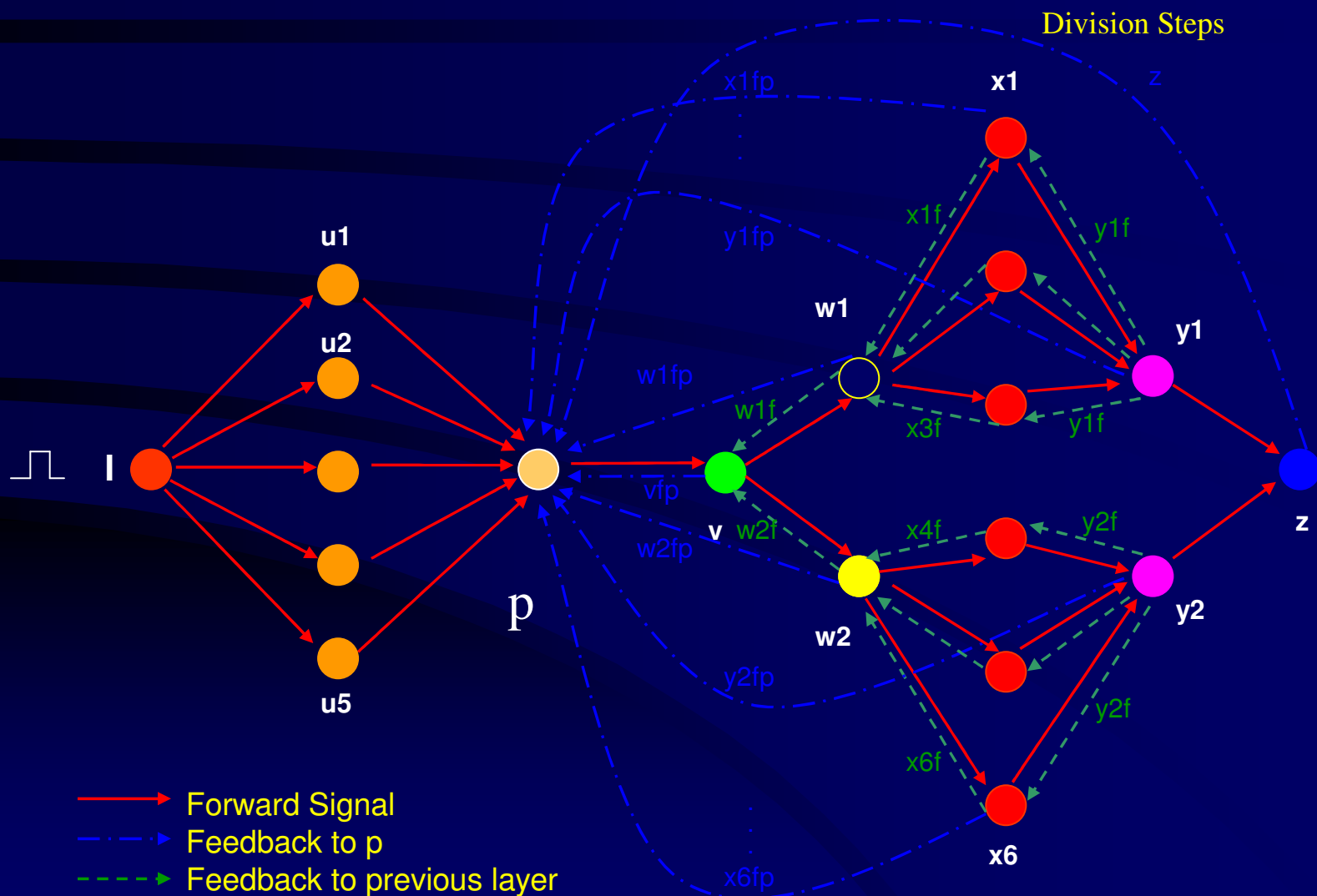
Oscilador de Período 3: FUNCIONAMENTO GERAL (parte\_B-t4A-o3.sim)



Sinal Periodico 7 ligados 1desligado: FUNCIONAMENTO GERAL (parte\_B-t4-o8-7of8.sim)



# Knockout



# System identification

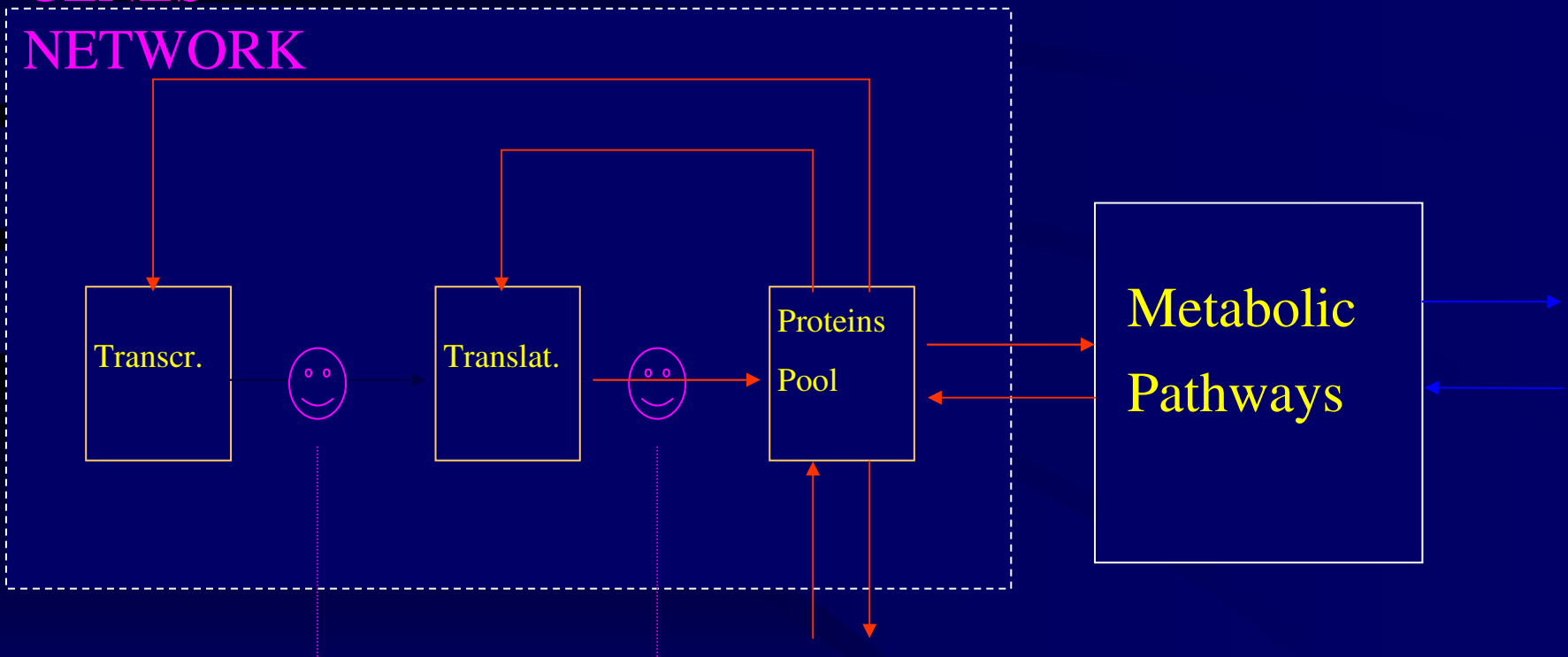
Cell

peptide

other signals

mRNA

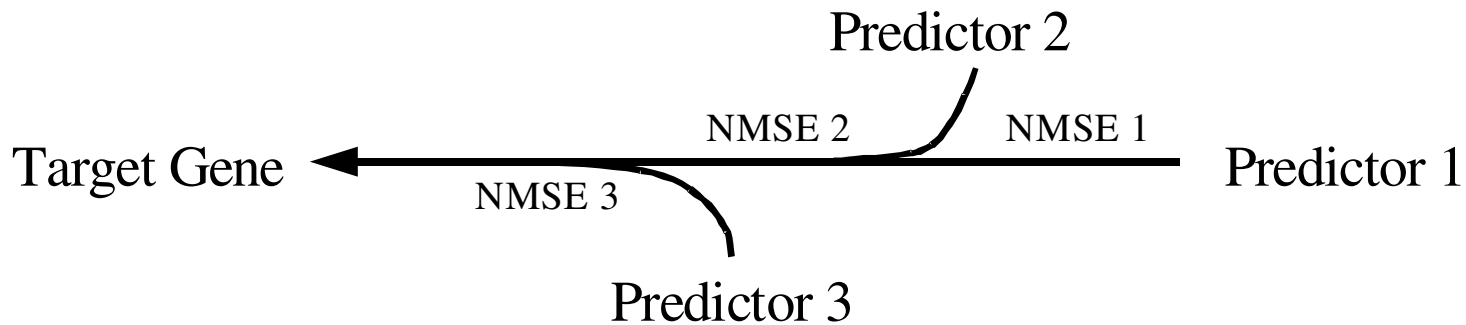
GENES  
NETWORK



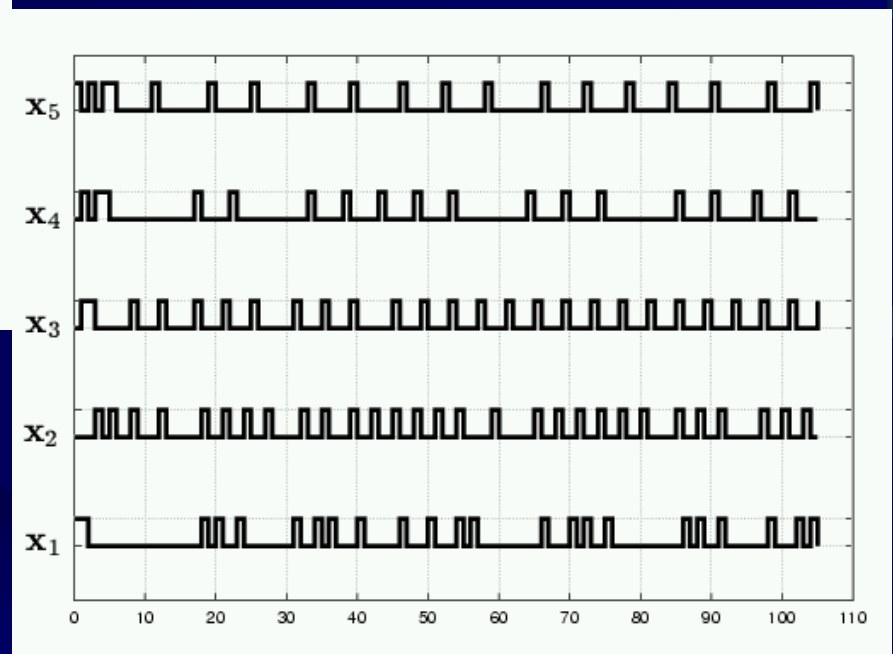
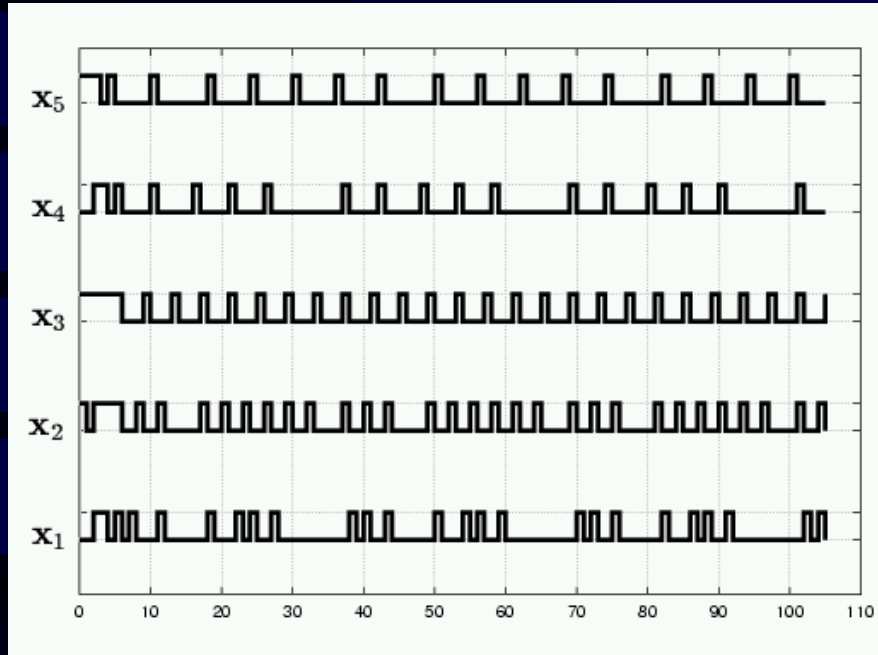
microarray

proteome

Find the architecture of a gene regulation network from microarray data.

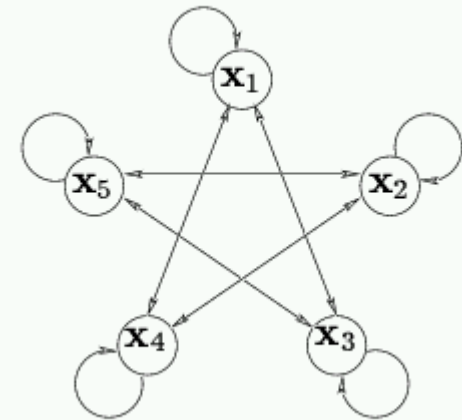


# System dynamics simple





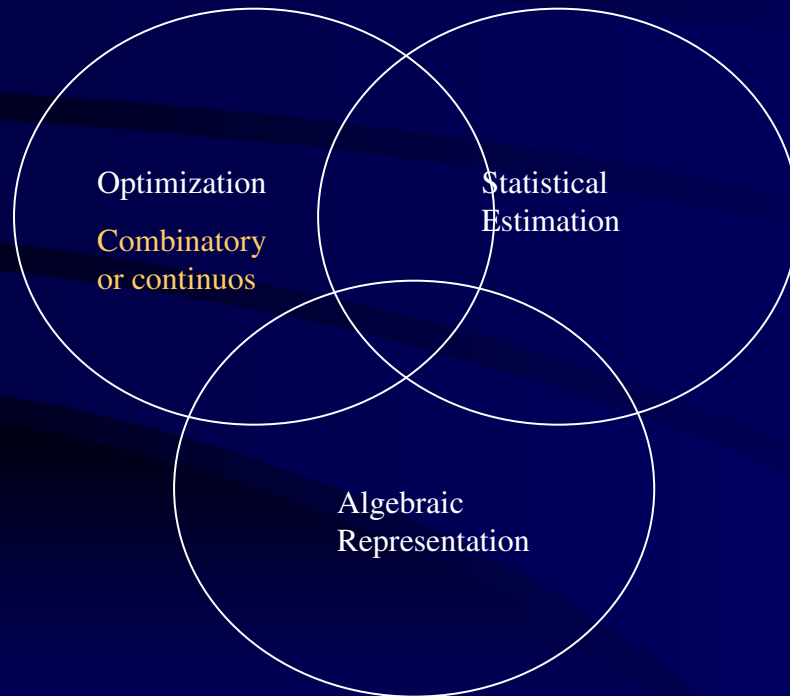
# System identified



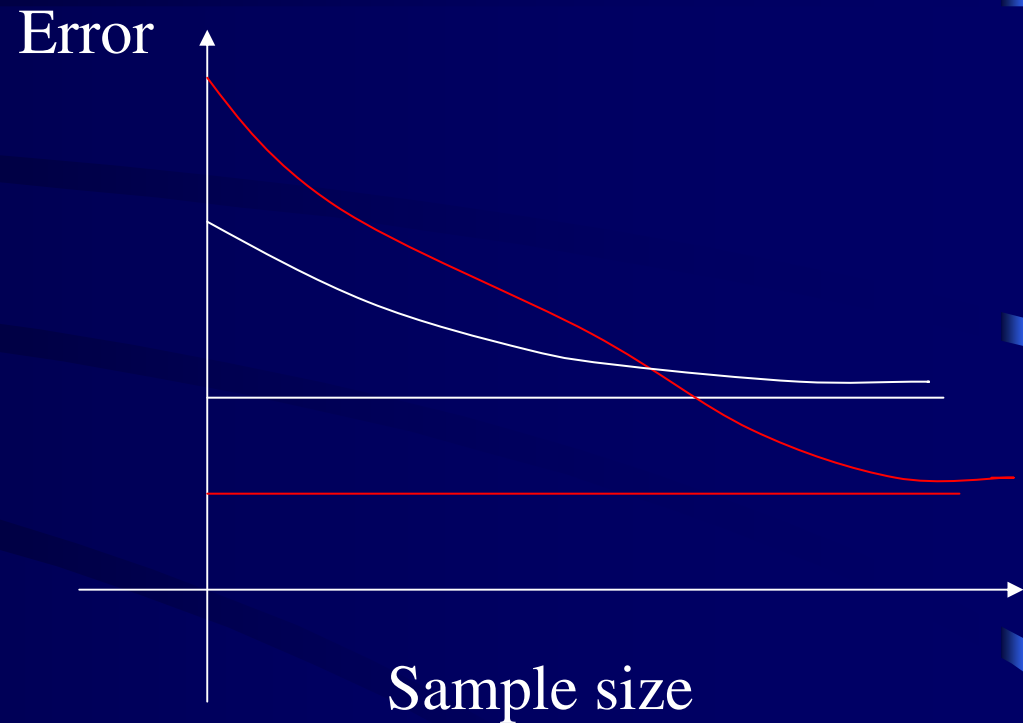
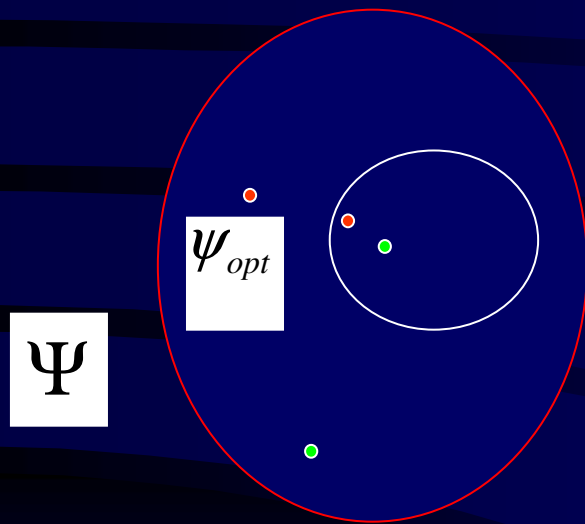
$$\mathbf{x}_1[t+1] = 1 \iff \left\{ \begin{array}{l} \mathbf{x}_1[t] = 0 \\ \text{and} \\ \left[ \left( (\mathbf{x}_3[t] = 1 \text{ or } \mathbf{x}_3[t-1] = 1 \text{ or } \mathbf{x}_3[t-2] = 1) \text{ and} \right. \right. \\ \quad \left. \left. (\mathbf{x}_4[t] = 1 \text{ or } \mathbf{x}_4[t-1] = 1 \text{ or } \mathbf{x}_4[t-2] = 1) \right) \right] \\ \text{or} \\ \left( \mathbf{x}_3[t] = \mathbf{x}_3[t-1] = \mathbf{x}_3[t-2] = \mathbf{x}_3[t-3] = \mathbf{x}_3[t-4] = 0 \text{ and} \right. \\ \quad \left. \mathbf{x}_4[t] = \mathbf{x}_4[t-1] = \mathbf{x}_4[t-2] = \mathbf{x}_4[t-3] = \mathbf{x}_4[t-4] = 0 \right) \end{array} \right.$$

# A family of Mathematical problems

# Design of classifier, filter or dynamical system



# The constrained estimation problem



# Other areas with similar problems

- Finances
- Marketing
- Digital TV
- Petrol Industry
- Neuro Sciences
- ...