

# Design of PGNs for malaria parasite and cell cycle control systems

Junior Barrera

DCC-IME-USP / BIOINFO-USP

# Layout

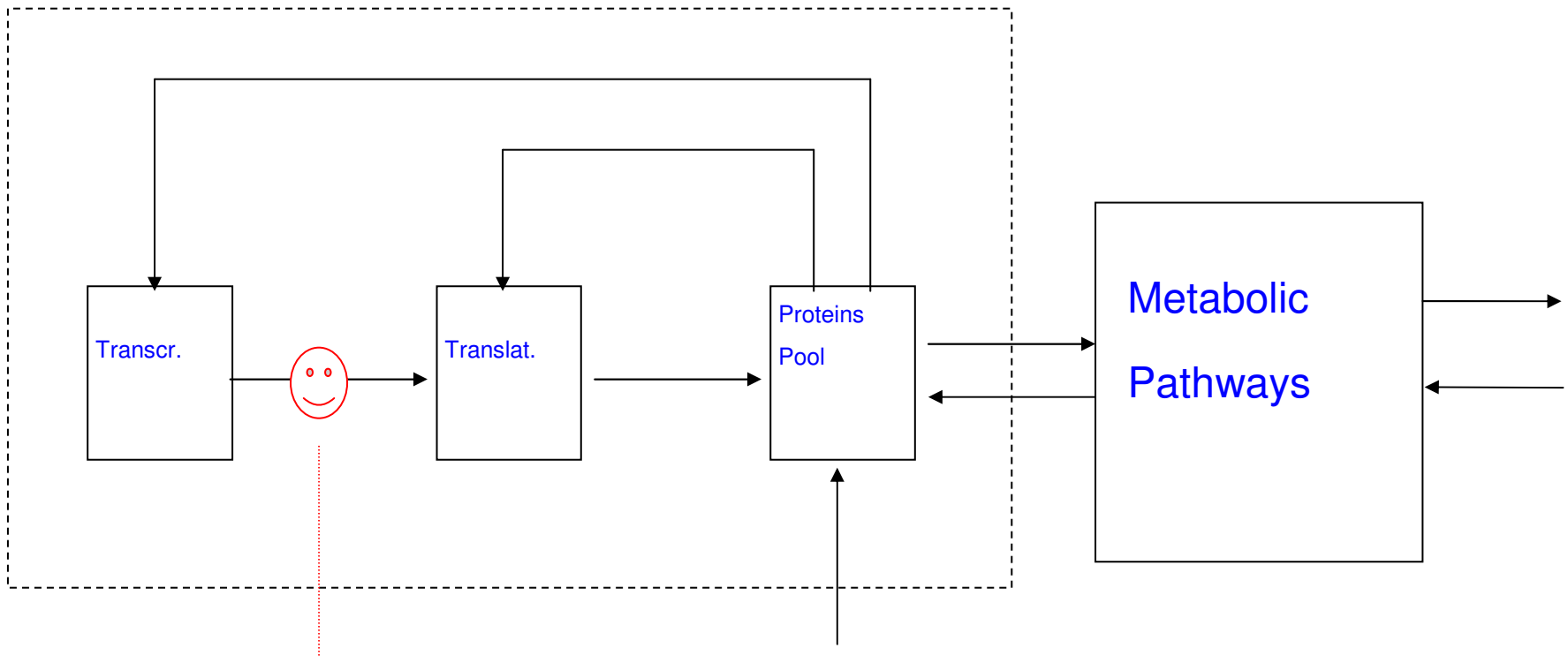
- Introduction
- Probabilistic genetic network (PGN)
- Cell cycle regulation
- Design of PGN from data
- Malaria parasite

# Layout

- Introduction
- Probabilistic genetic network (PGN)
- Cell cycle regulation
- Design of PGN from data
- Malaria parasite

# Regulatory System

## GENES NETWORK



microarray

# Layout

- Introduction
- Probabilistic genetic network (PGN)
- Cell cycle regulation
- Design of PGN from data
- Malaria parasite

**GENE is a**  
**NON LINEAR STOCHASTIC GATE**

**Expression of a Gene depends on**  
**Activator and Inhibitory Signals**

**SYSTEM: built by compiling these gates**

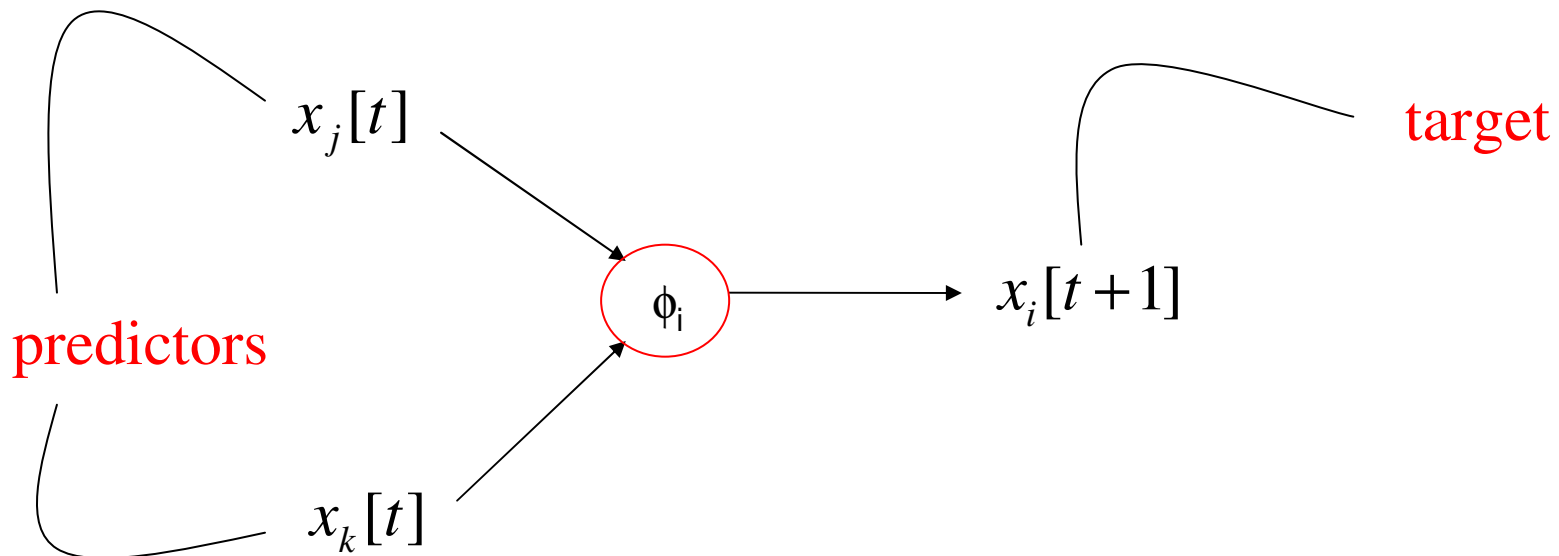
Expression of gene  $i$  at time  $t$ :  $x_i[t] \in \{-1, 0, +1\}$

State of the regulatory network at time  $t$ :  $x[t] = \begin{bmatrix} x_1[t] \\ x_2[t] \\ \cdot \\ \cdot \\ x_n[t] \end{bmatrix}$

Network dynamics:  $x[t + 1] = \phi(x[t])$

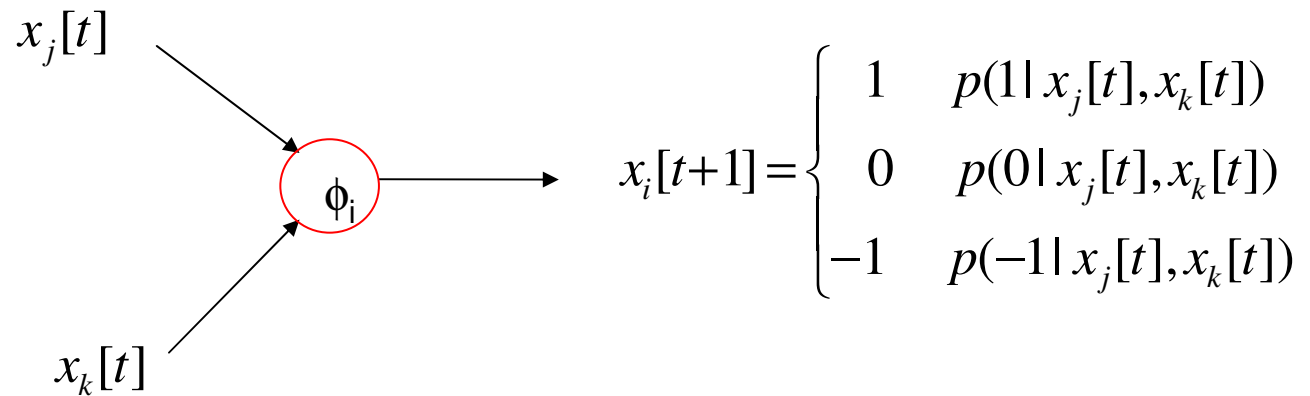
$$\phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \cdot \\ \cdot \\ \cdot \\ \phi_n \end{bmatrix}$$

$$x_i[t+1] = \phi_i(x[t])$$





# Probabilistic Genetic Network (PGN)



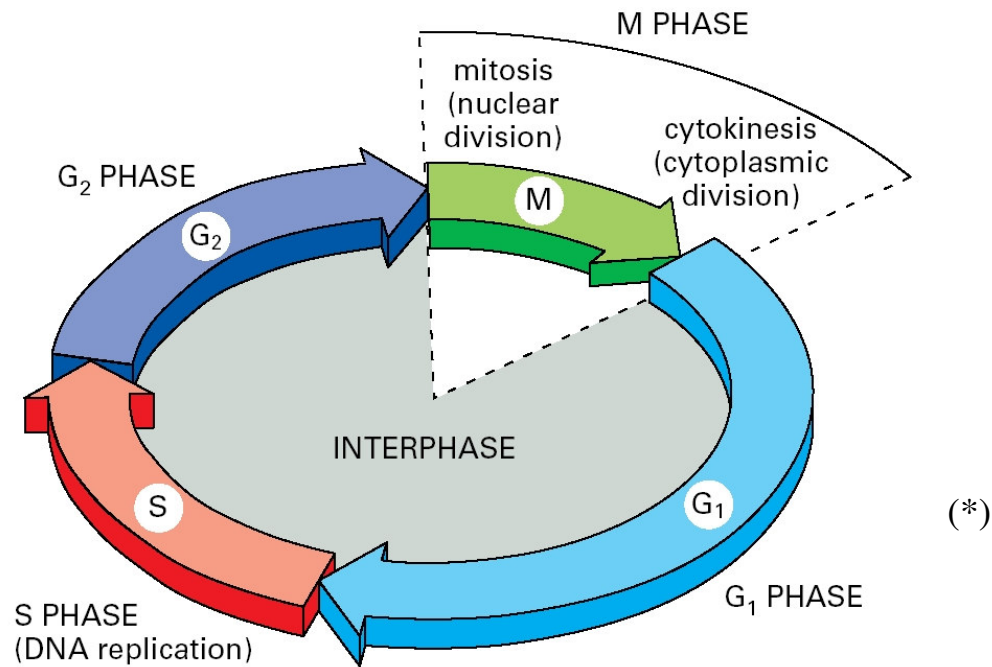
$\exists y, z, w \in \{-1, 0, 1\}, y \neq z \neq w:$

$$p(y | x_j[t], x_k[t]) \gg p(z | x_j[t], x_k[t]) + p(w | x_j[t], x_k[t])$$

# Layout

- Introduction
- Probabilistic genetic network (PGN)
- Cell cycle regulation
- Design of PGN from data
- Malaria parasite

# Phases of the Cell Cycle



# Model: architecture and dynamics

Li, 2004.

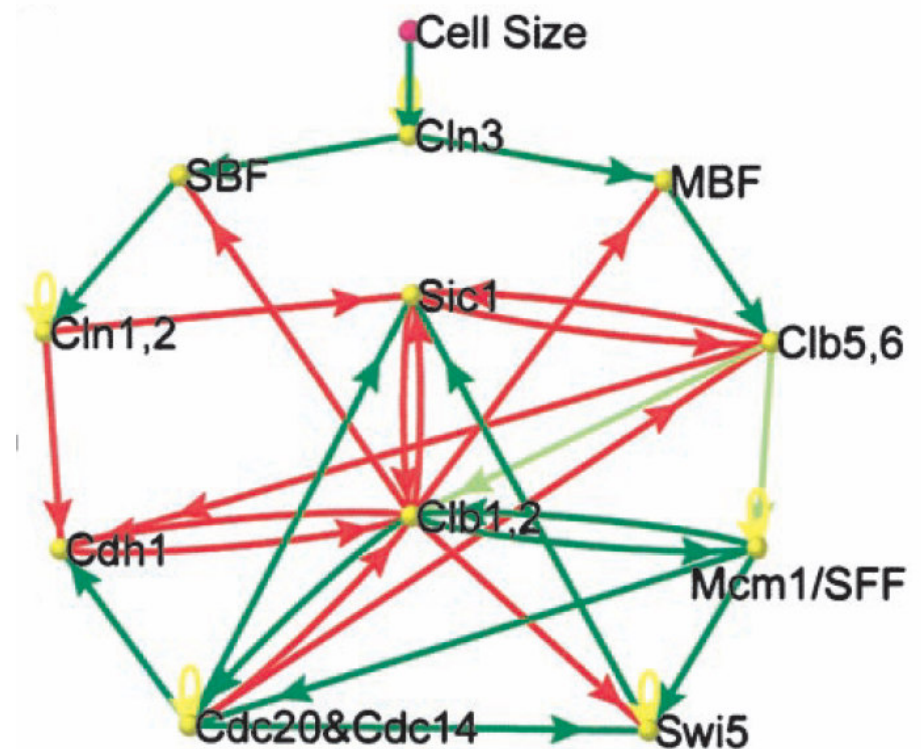
## Transition Function

$$S_i(t+1) = \begin{cases} 1, & \sum_j a_{ij} S_j(t) > 0 \\ 0, & \sum_j a_{ij} S_j(t) < 0 \\ S_i(t), & \sum_j a_{ij} S_j(t) = 0 \end{cases}$$

$a_{ij} = a_g$  green arrow from  $i$  to  $j$

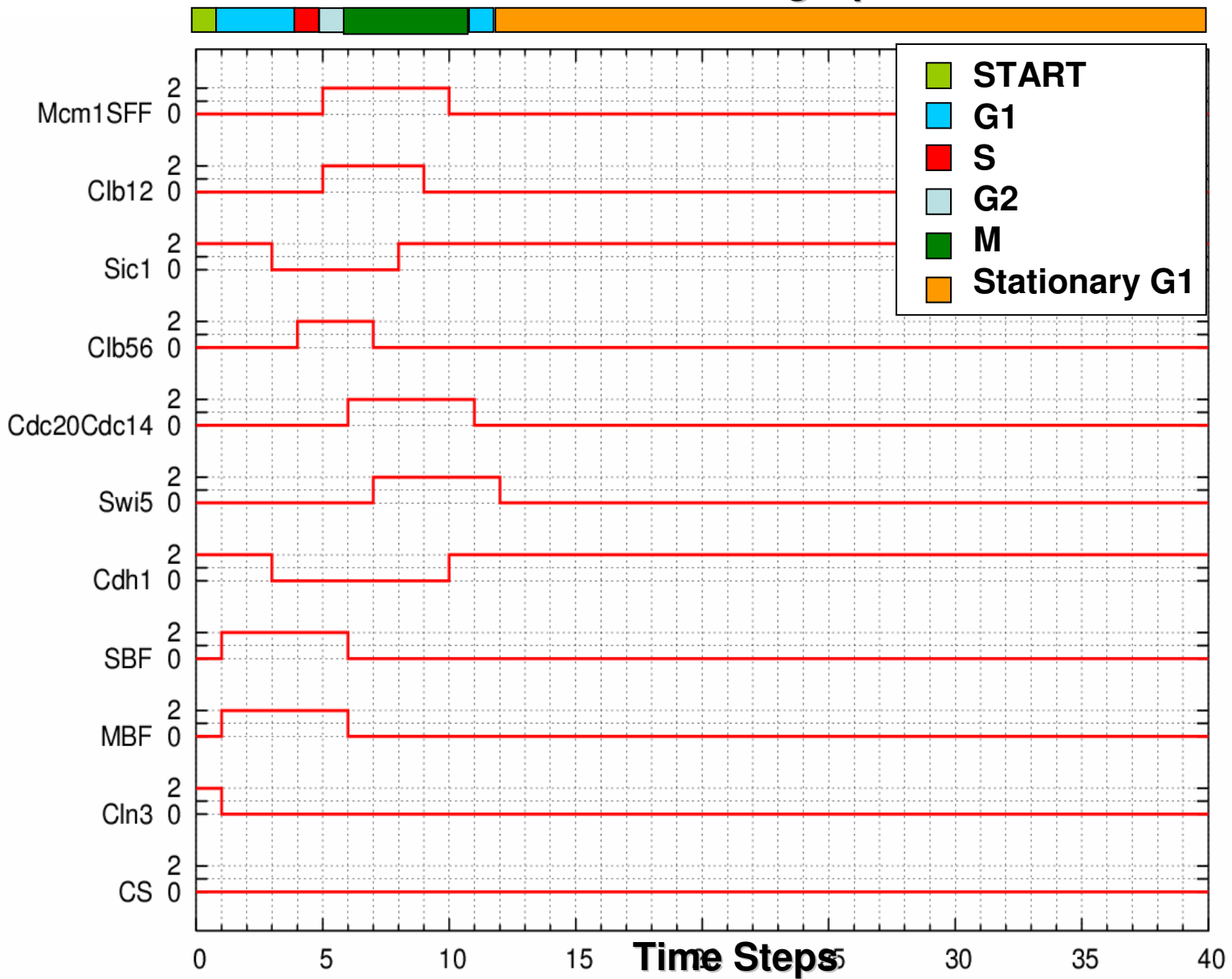
$a_{ij} = a_r$  red arrow from  $i$  to  $j$

Simulation parameters:  $a_g = -a_r = 1$



# Deterministic

One single pulse of  $CS = 2$  at  $t = -1$

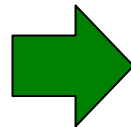


**Binary**

**1** → **2**  
**0** → **0**

**3 Levels (0, 1, 2)**

$\sum_j a_{ij} S_j(0)$	$S_j(t+1)$	
	$S_j(t)=0$	$S_j(t)=1$
⋮	⋮	⋮
3	1	1
2	1	1
1	1	1
0	0	1
-1	0	0
-2	0	0
-3	0	0
⋮	⋮	⋮



$\sum_j a_{ij} S_j(0)$	$y_j(t+1)$		
	$S_j(t)=0$	$S_j(t)=1$	$S_j(t)=2$
⋮	⋮	⋮	⋮
3	2	2	2
2	2	2	2
1	1	2	2
0	0	1	2
-1	0	0	1
-2	0	0	0
-3	0	0	0
⋮	⋮	⋮	⋮

# Stochastic Transition Function

$\sum_j a_{ij} S_j(0)$	$y_j(t+1)$		
	$S_j(t) = 0$	$S_j(t) = 1$	$S_j(t) = 2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
<b>3</b>	<b>2</b>	<b>2</b>	<b>2</b>
<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>
<b>1</b>	<b>1</b>	<b>2</b>	<b>2</b>
<b>0</b>	<b>0</b>	<b>1</b>	<b>2</b>
<b>-1</b>	<b>0</b>	<b>0</b>	<b>1</b>
<b>-2</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>-3</b>	<b>0</b>	<b>0</b>	<b>0</b>
$\vdots$	$\vdots$	$\vdots$	$\vdots$

$$x_i(t+1) = \begin{cases} y_i(t+1) & \text{with } P = 0.99 \\ a & \text{with } P = 0.005 \\ b & \text{with } P = 0.005 \end{cases}$$

where  $a, b \in \{0, 1, 2\} - \{y_i(t+1)\}$





# Our gene model

Total input signal driving a generic variable

$$x_i(t) \in \{0, 1, 2\} \quad (1 \leq i \leq N)$$

$$d_i(t) = \sum_{j=1}^N \sum_{k=1}^m a_{ji}^k x_j(t-k)$$

**Driving function** for  $x_i$

$m$ : memory of the system

$a_{ji}^k$ : weight for variable  $x_j$  at time  $t - k$

# Our gene model

$$y_i(t+1) = \begin{cases} 2 & \text{if } d_i(t) \geq th_{i2} \\ 1 & \text{if } th_{i1} \leq d_i(t) \leq th_{i2} \\ 0 & \text{if } d_i(t) < th_{i1} \end{cases}$$

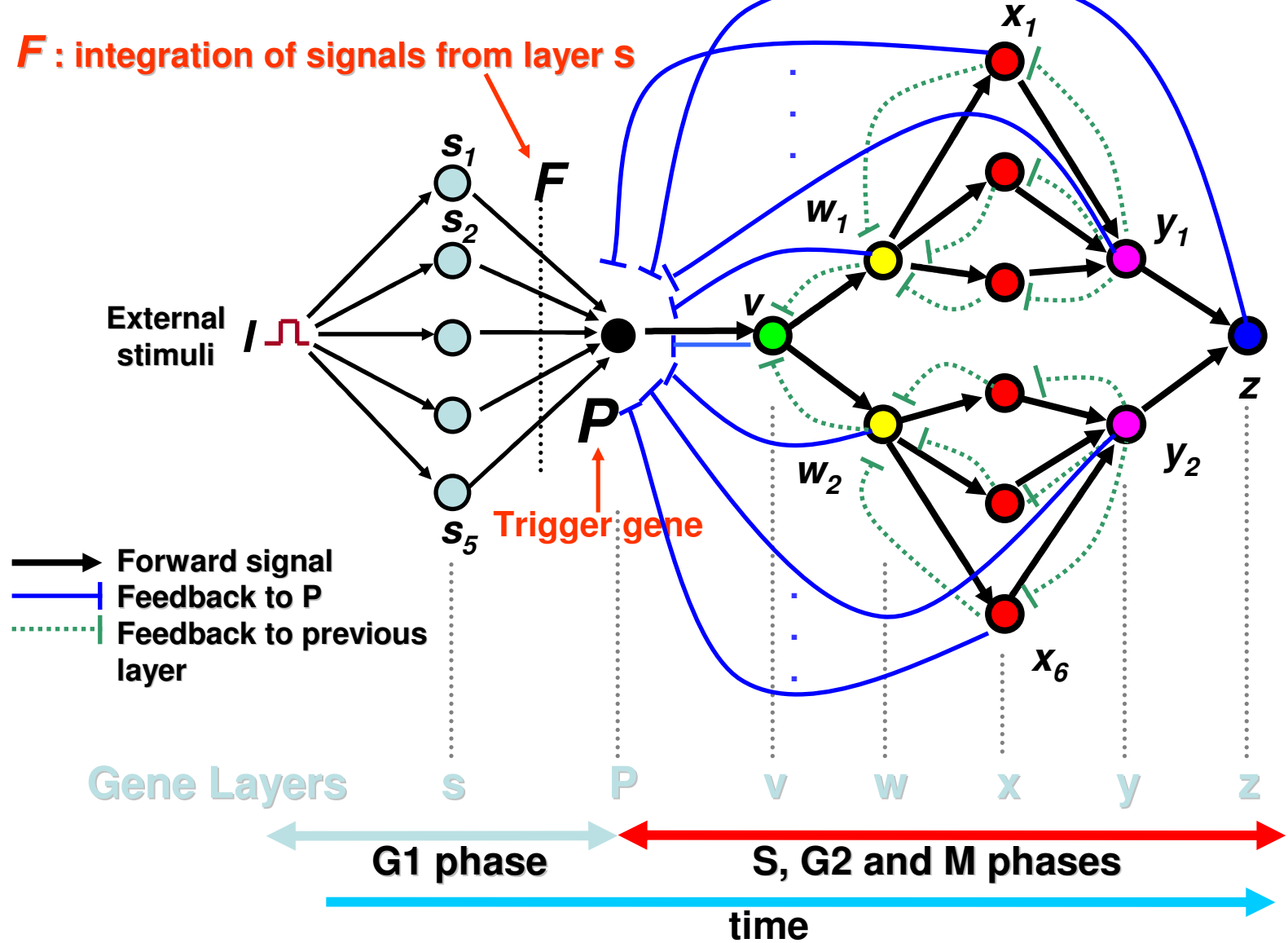
**Stochastic  
Transition  
Function**

$$x_i(t) = \begin{cases} y_i(t) & \text{with } P \approx 1 \\ a & \text{with } (1 - P)/2 \\ b & \text{with } (1 - P)/2. \end{cases}$$

$$a, b \in \{0, 1, 2\} - \{y_i(t)\}$$

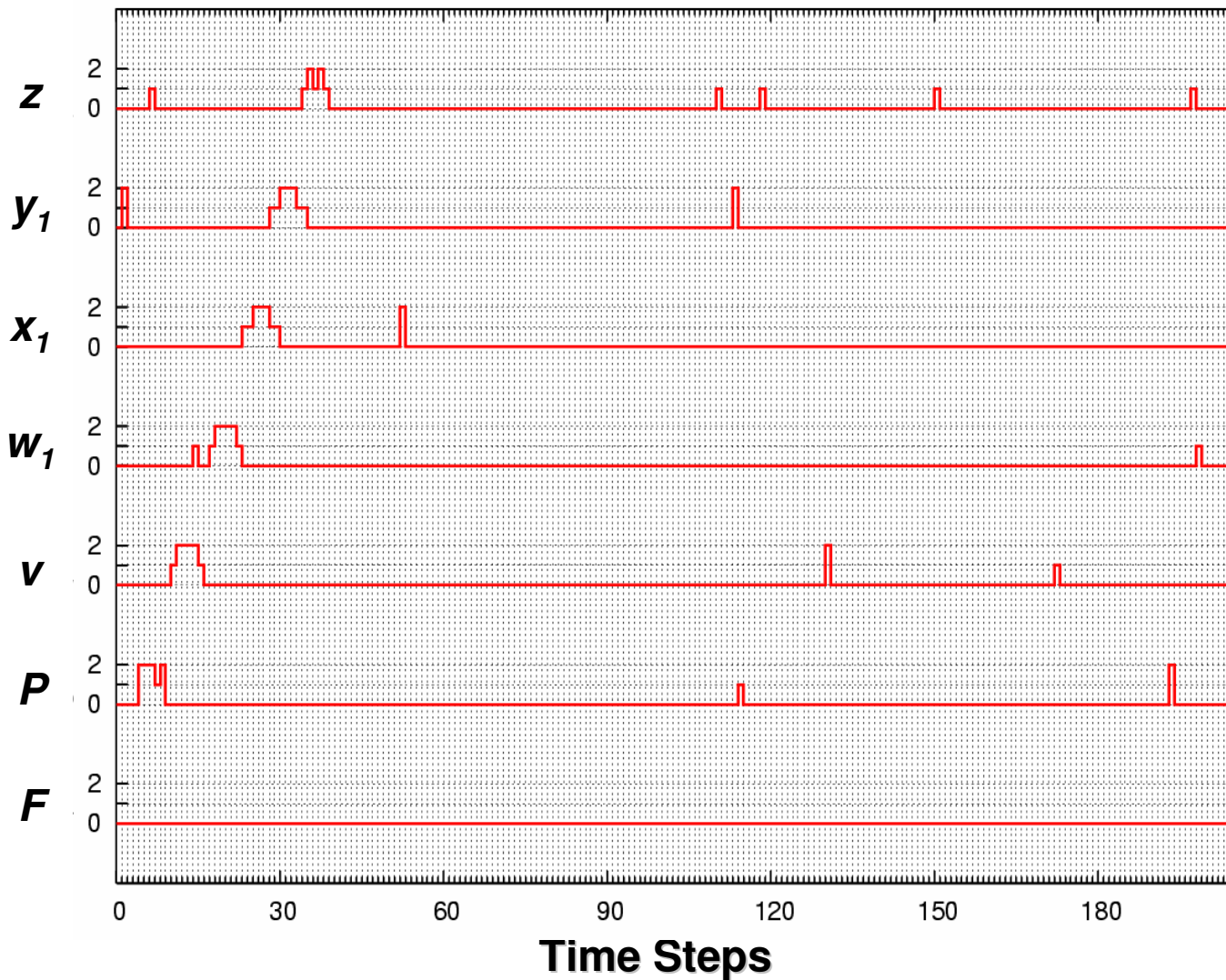
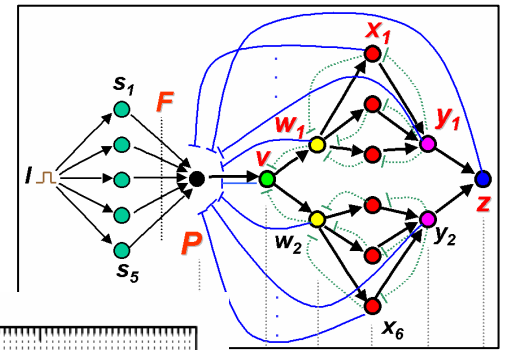
# Our network architecture

$F$  : integration of signals from layer  $s$



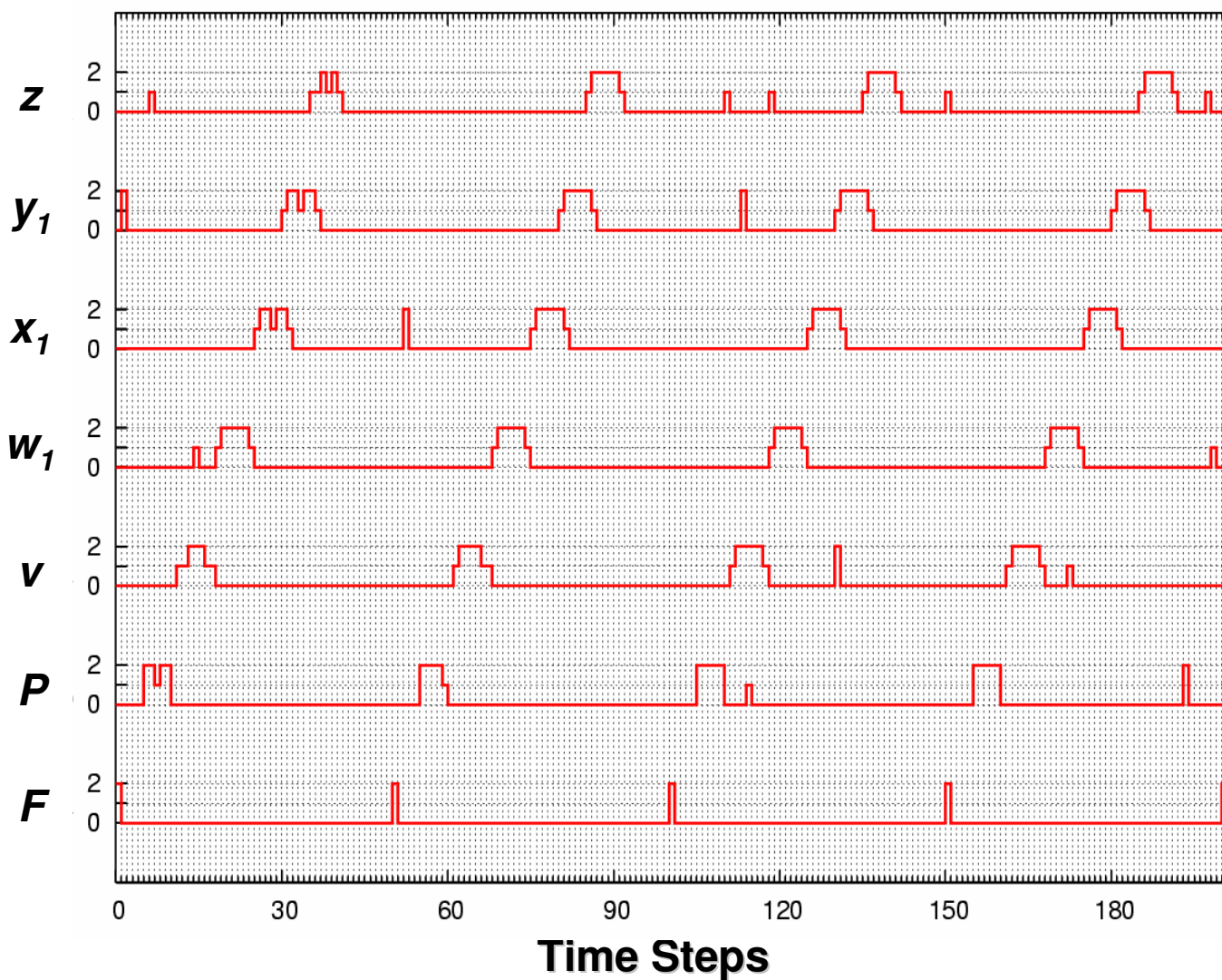
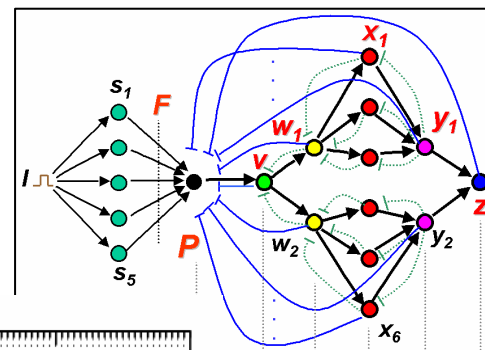
# PGN with $P = 0.99$

One single pulse of  $F = 2$  at  $t = -1$



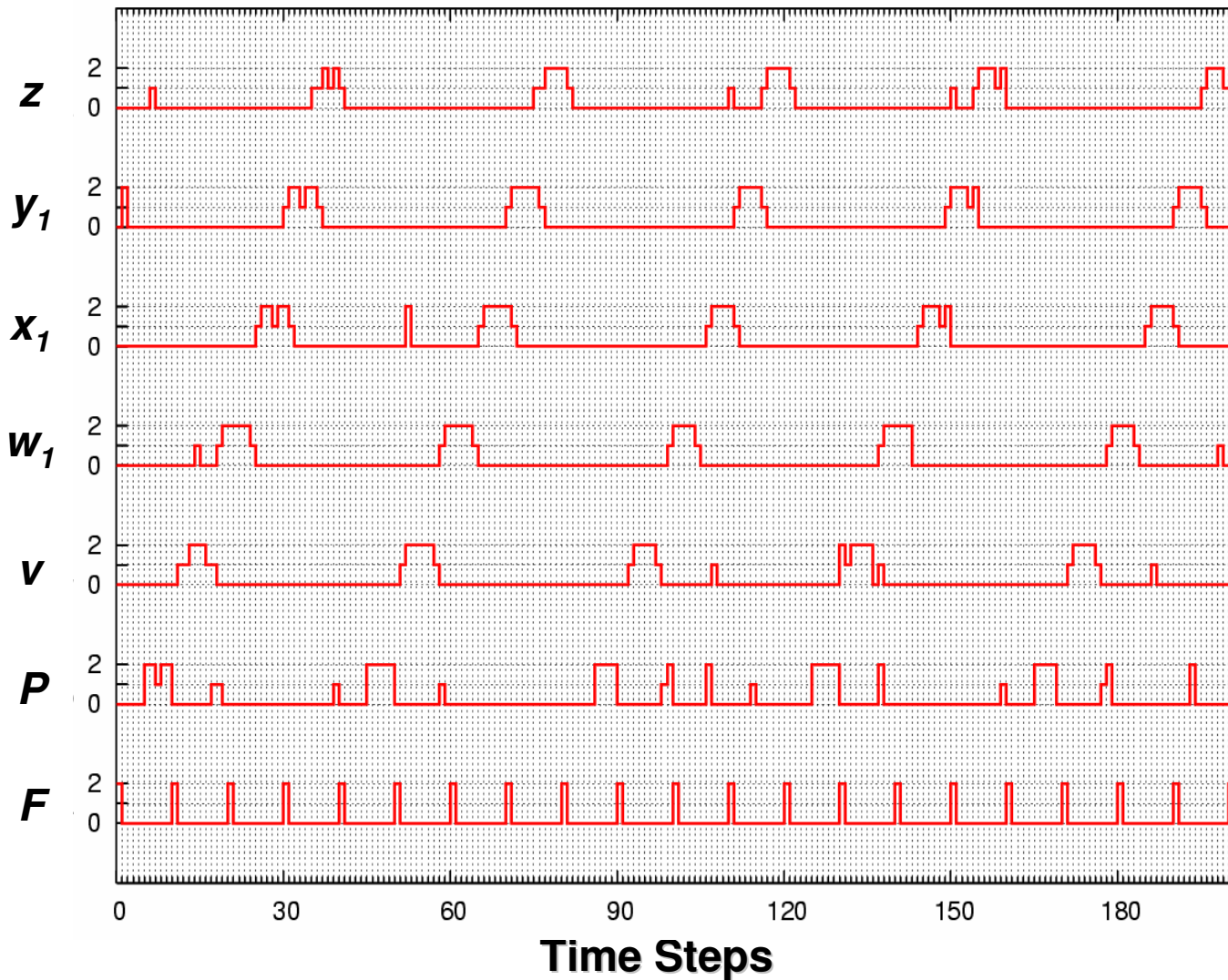
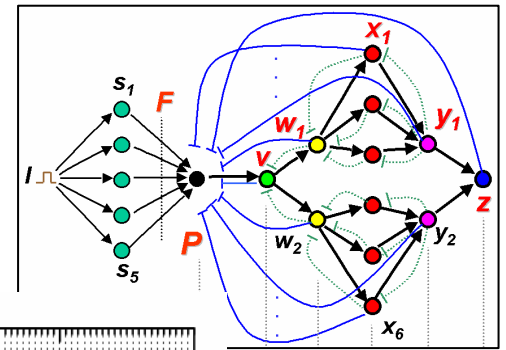
# PGN with $P = 0.99$

Signal  $F$  = period 50 oscillator



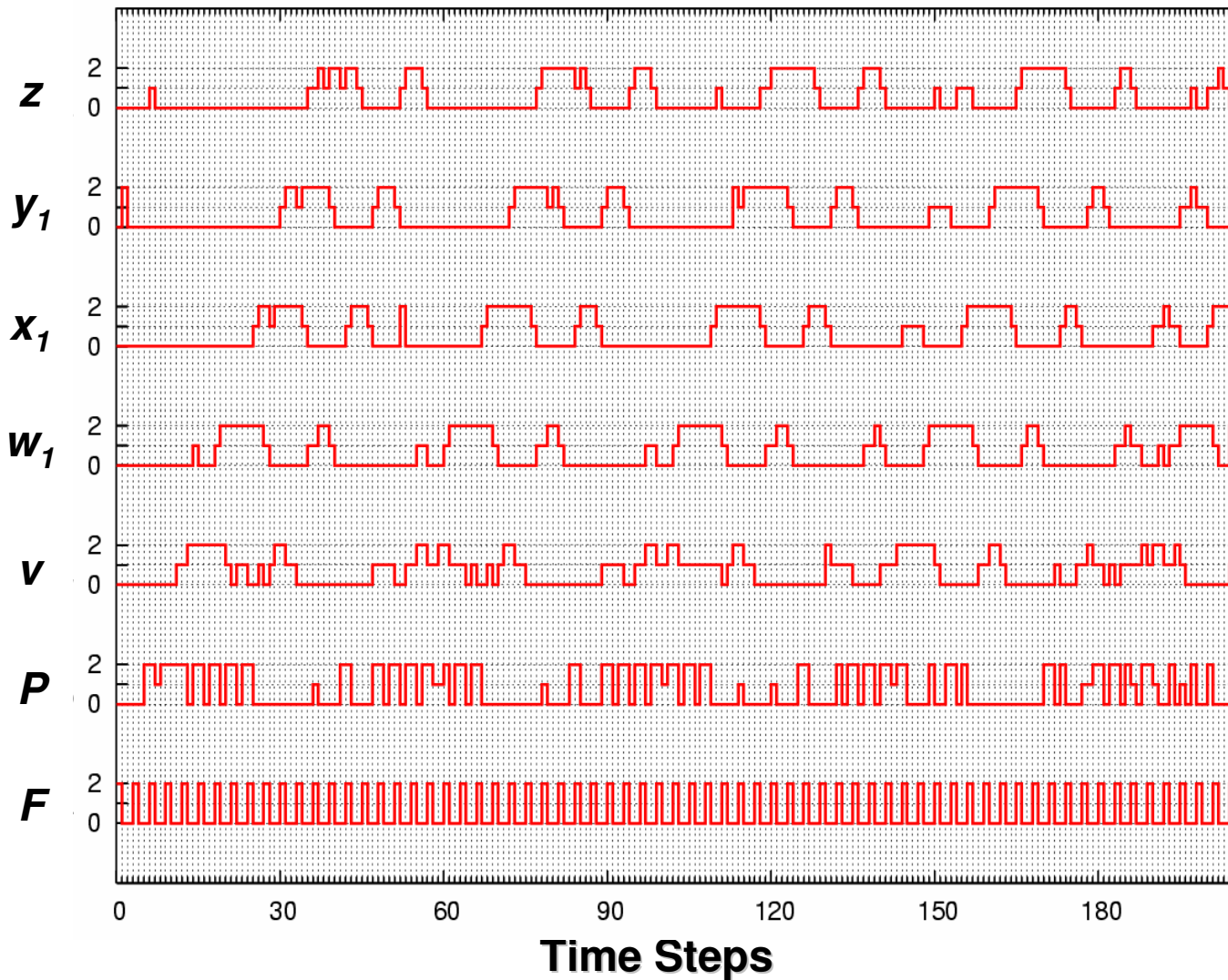
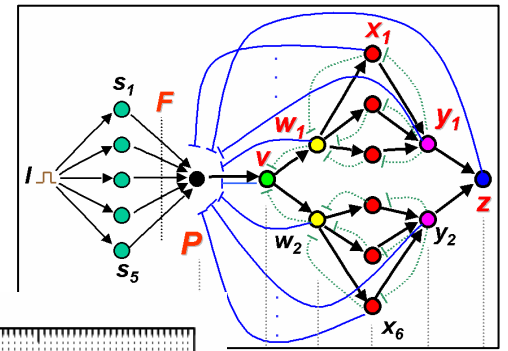
# PGN with $P = 0.99$

Signal  $F =$  period 10 oscillator



# PGN with $P = 0.99$

Signal  $F = \text{period 3 oscillator}$

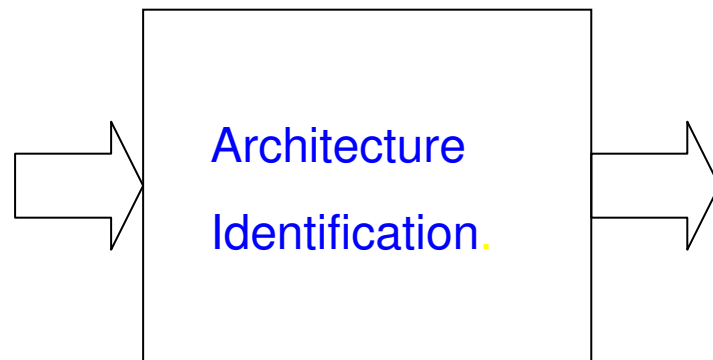


# Layout

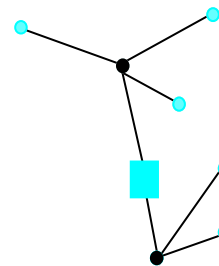
- Introduction
- Probabilistic genetic network (PGN)
- Cell cycle regulation
- Design of PGN from data
- Malaria parasite



$x[1], x[2], \dots, x[n]$



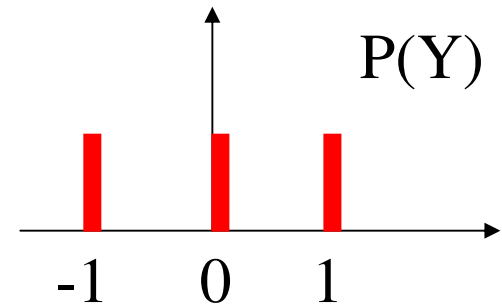
target genes



## Distribution of Y

$$P : \{-1, 0, 1\} \rightarrow [0, 1]$$

$$\sum_{y \in \{-1, 0, 1\}} P(y) = 1$$



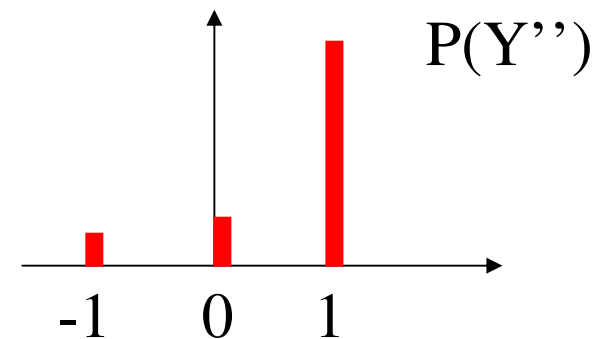
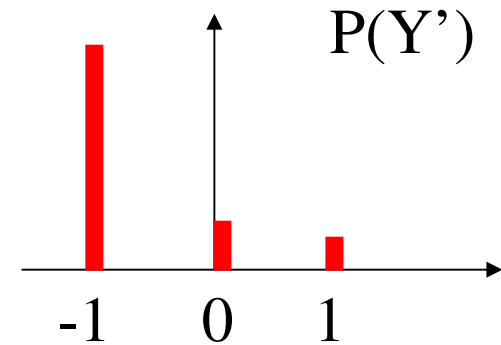
## Entropy

$$H(Y) = - \sum_{y \in \{-1, 0, 1\}} P(y) \log P(y)$$

$$H(Y) > H(Y') \quad H(Y') = H(Y'')$$

## Mutual information

$$I(X, Y) = H(Y) - H(Y | X) \geq 0$$



## Mean conditional entropy

$$E[H(Y | X)] = -\sum P(X) \sum P(Y | X) \log(P(Y | X))$$

## Mean mutual information

$$E[I(X, Y)] = H(Y) - E[H(Y | X)]$$

## Mean mutual information estimation

$$\hat{E}[H(Y | X)] = -\sum \hat{P}(X) \sum \hat{P}(Y | X) \log(\hat{P}(Y | X)).$$

$$\hat{E}[I(X, Y)] = H(\hat{Y}) - \hat{E}[H(Y | X)]$$

## Estimation of $P(Y|X)$

Y: the target gene at  $t+1$ , that is,  $Y = x_i[t+1]$

X: the predictors at  $t$ , that is,  $X = (x_j[t], x_k[t])$

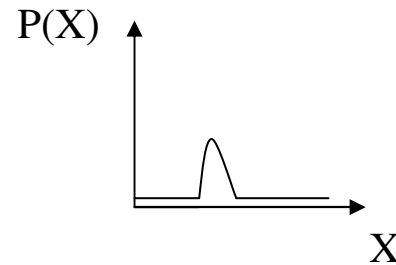
For a fixed parameter  $n$

If  $\#(X=(a,b)) \geq n$ , then  $\hat{P}(Y=c | X=(a,b)) = \frac{\#((Y=c) \wedge X=(a,b))}{\#(X=(a,b))}$

If  $\#(X=(a,b)) < n$ , then  $\hat{P}(Y | X=(a,b))$  is uniform

## Estimation of $P(X)$ for a fixed parameter $n$

$$X = (x_j[t], x_k[t])$$



$$N^+ = \sum_{\#(X=(a,b)) \geq n, \forall (a,b)} \#(X=(a,b))$$

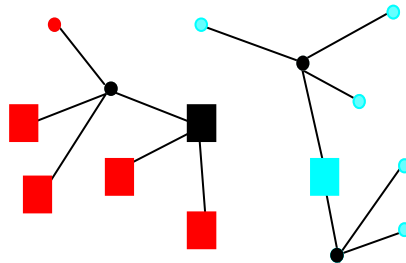
$$N^- = \sum_{\#(X=(a,b)) < n, \forall (a,b)} \#(X=(a,b))$$

$$\text{If } \#(X=(a,b)) \geq n, \text{ then } \hat{P}(X=(a,b)) = \frac{N^+}{N^- + N^+} \times \frac{\#(X=(a,b))}{N^+}$$

$$\text{If } \#(X=(a,b)) < n, \text{ then } \hat{P}(X=(a,b)) = \frac{N^-}{N^- + N^+} \times \frac{1}{3^2 - |\{(a,b) : \#(X=(a,b)) \geq n\}|}$$

## Buiding Interaction Graphes

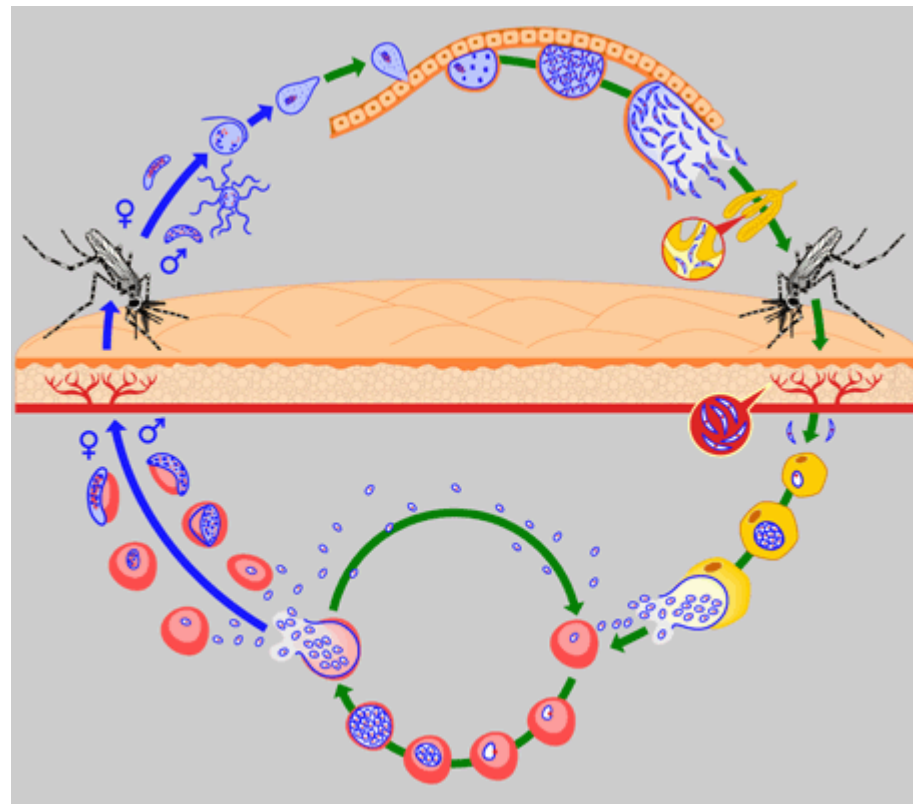
- For each target gene, rank the couples of all genes by their estimated mutual information and sample size;
- When two mutual information are equal, the one estimated from a larger sample comes first;
- Choose the best couples;
- Design the interaction graph



# Layout

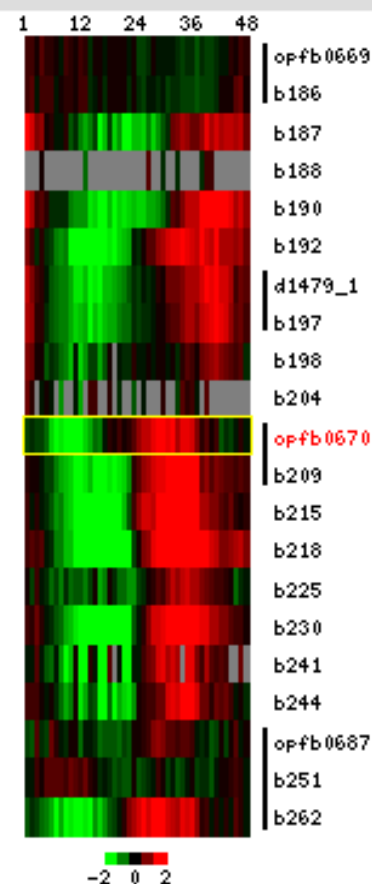
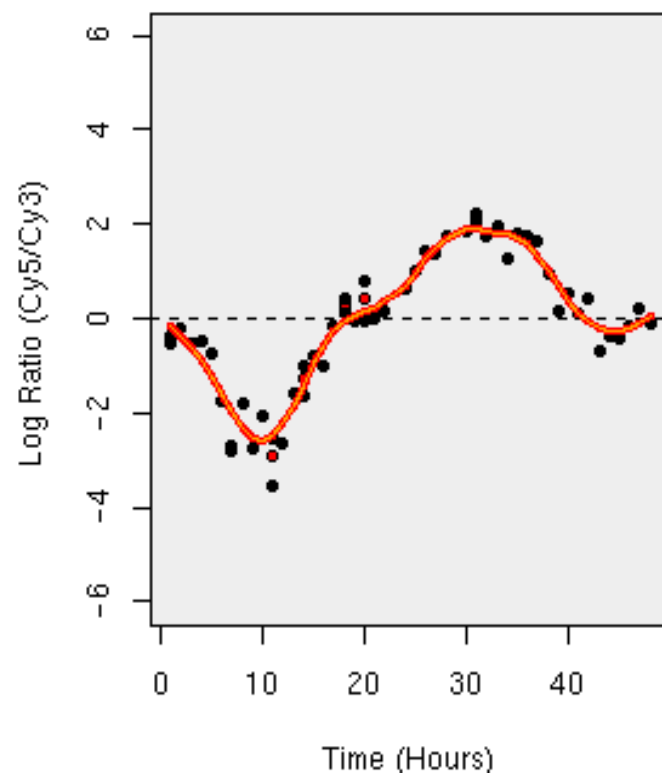
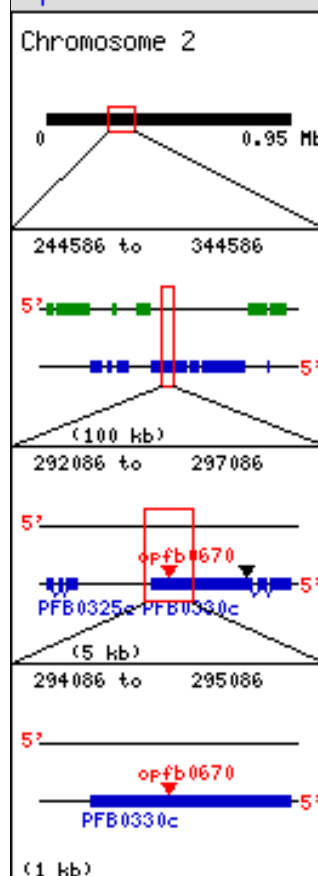
- Introduction
- Probabilistic genetic network (PGN)
- Cell cycle regulation
- Design of PGN from data
- **Malaria parasite**

# The life cycle of the malaria parasite





OligoID	Status	Maximum Hour	Minimum Hour	Amplitude Score (log2)	Score (%)	Phase (-Pi to +Pi)	CGH %3D7	Avg. Med. Intensity
<a href="#">opfb0670</a>	UNIQUE	30	10	4.5	87	0.06	89	3211.57



[← OLIGO →](#)

PlasmoDB ID Description

[PFB0330c](#) cysteine protease, putative

Oligo Sequence

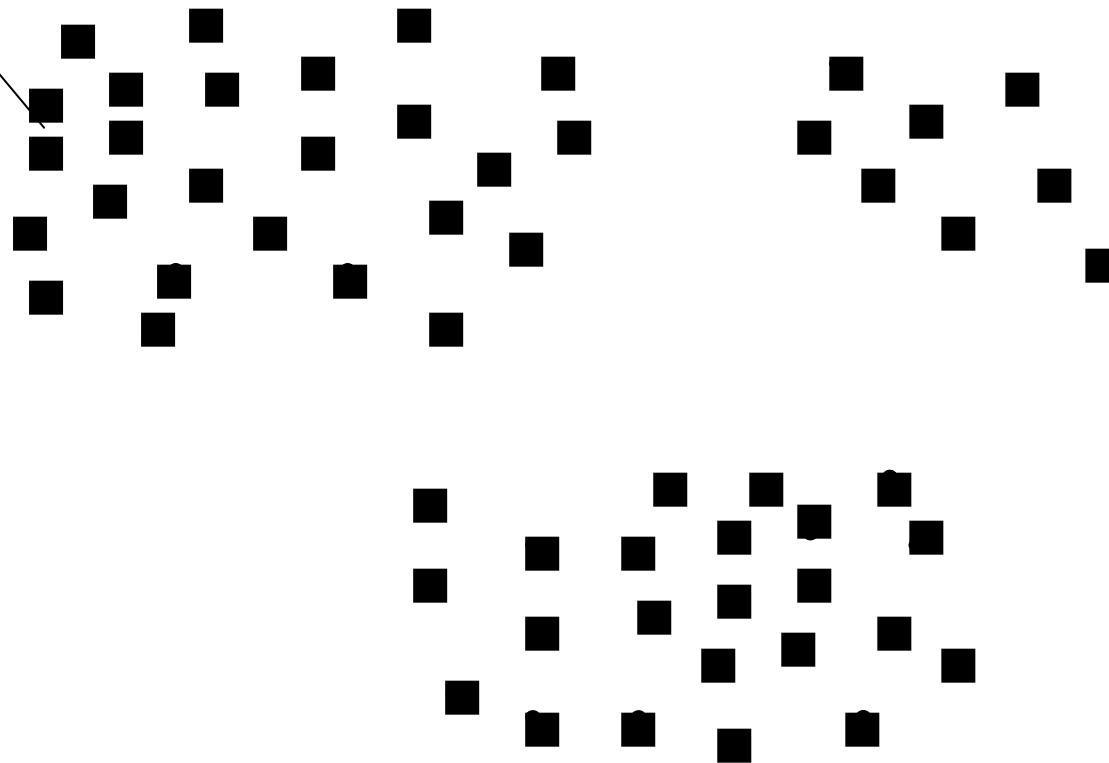
[BLAST @ PlasmoDB](#)

5'

CTGCCCAAGATGAGCCACCTACTGATAATGTAGAATCACAAGCAGAAAATAACAAAAAACAGAAATTTA

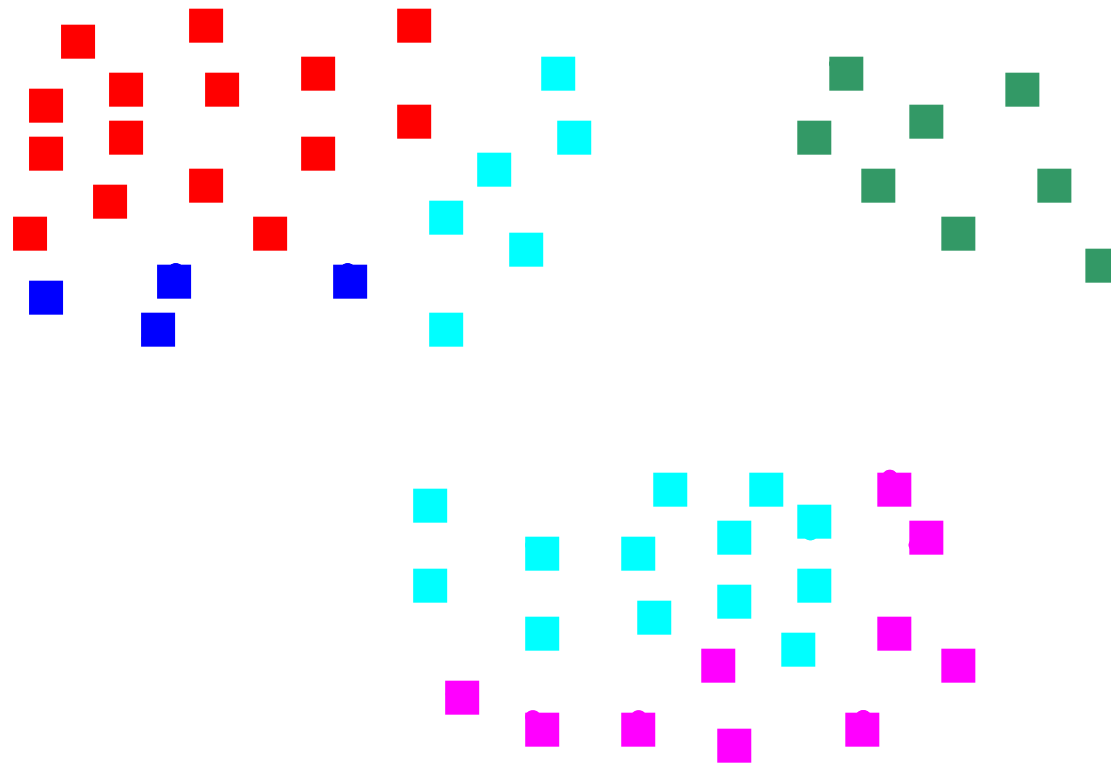
# Malaria parasite genes with almost sinusoidal signals

Sinuousoidal  
signals



DeRisi, 2003.

# Functional Classification

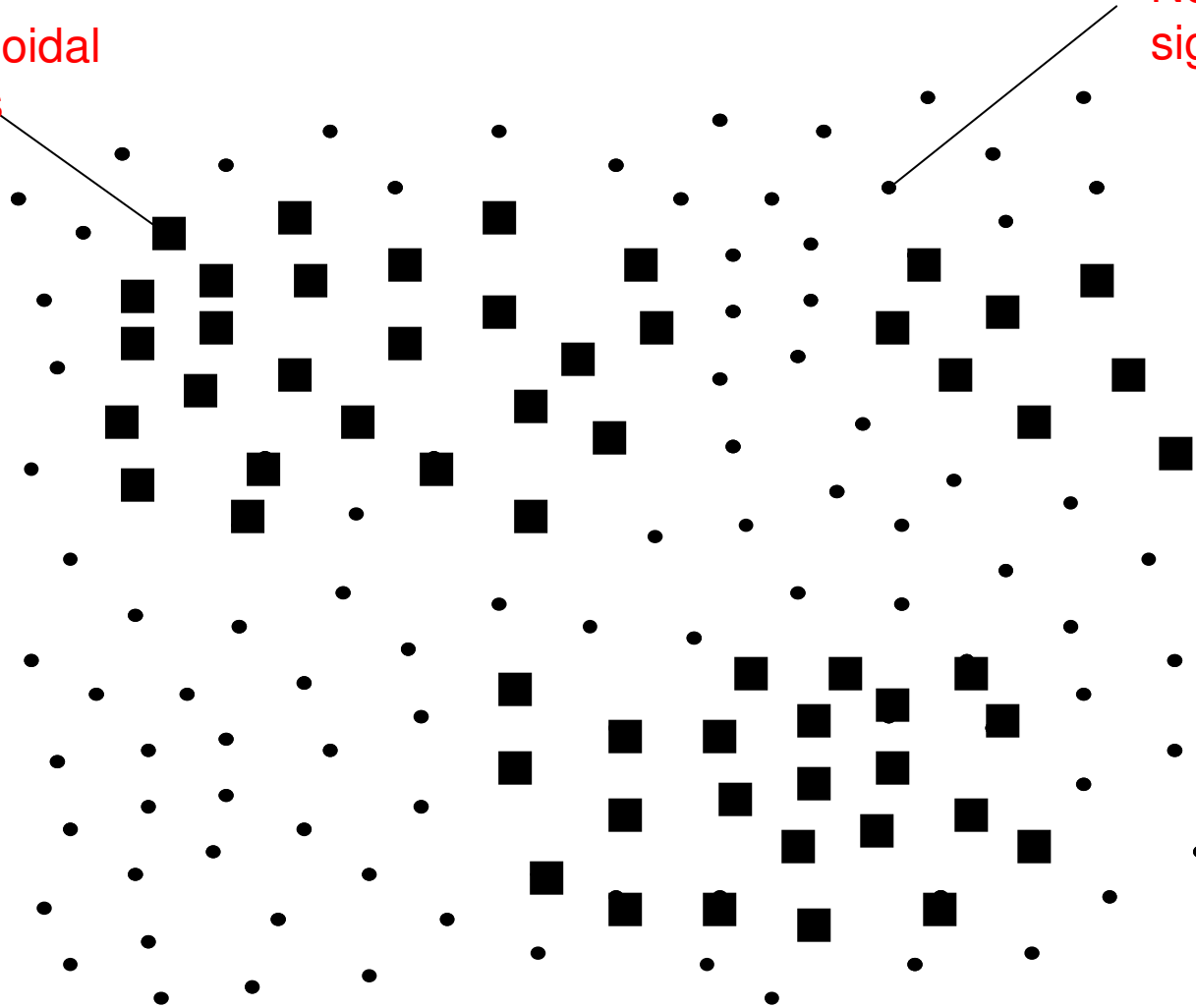


DeRisi, 2003.

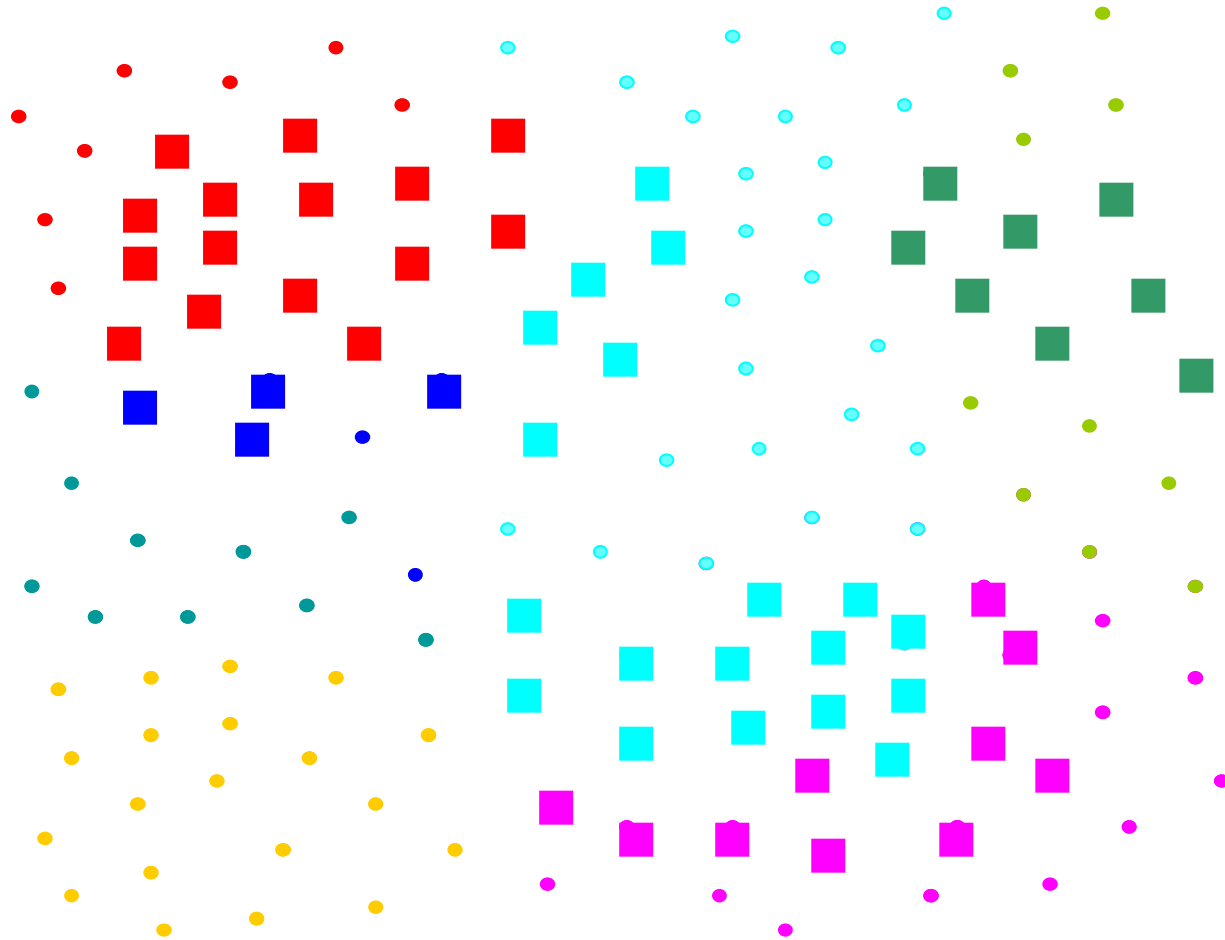
# Malaria parasite genes

Sinuousoidal signals

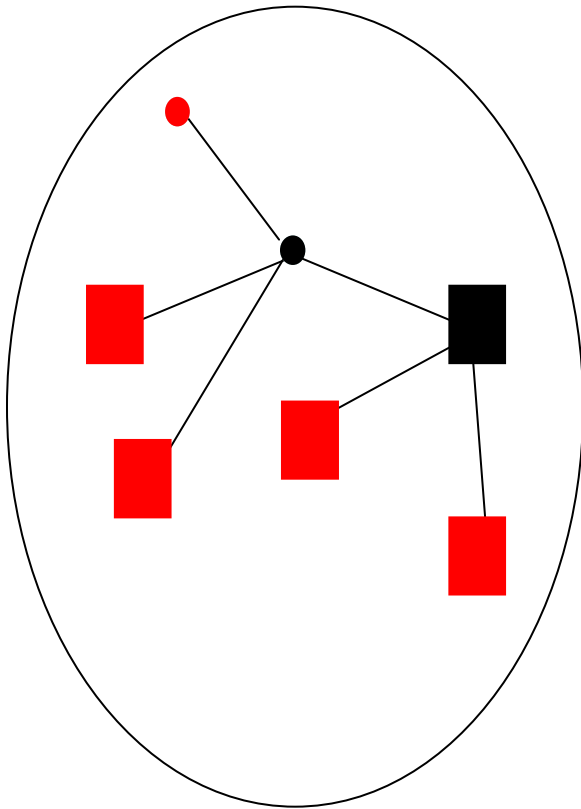
Non sinuousoidal signals



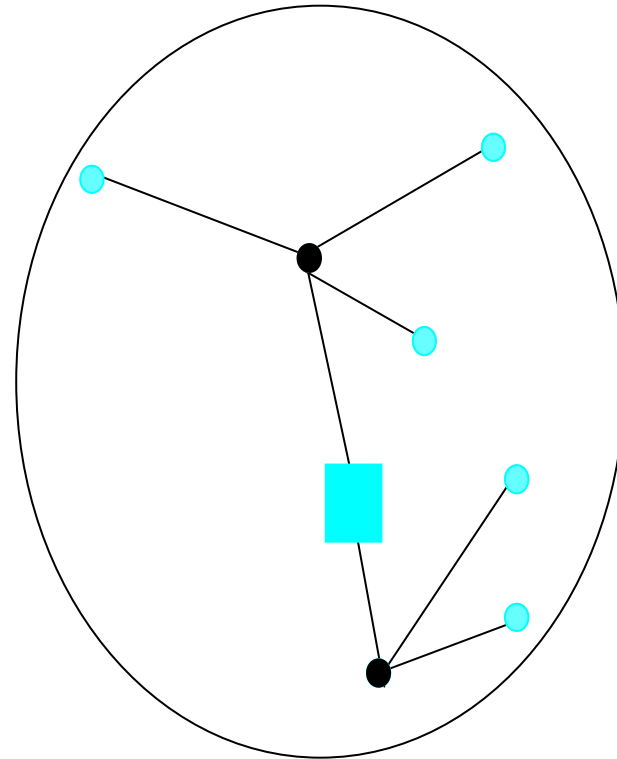
# Functional Classification



# Interaction Graph

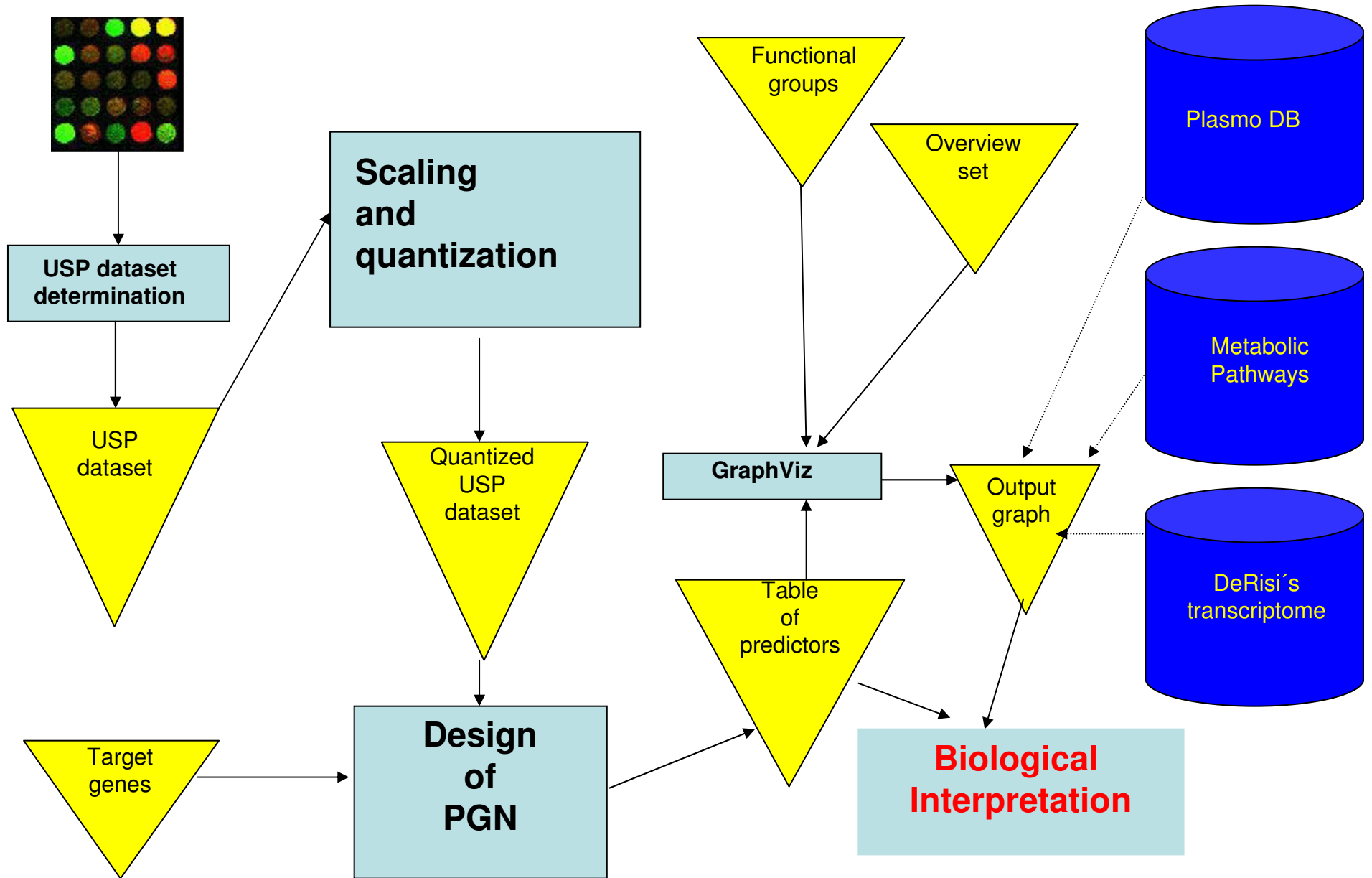


Glycolysis

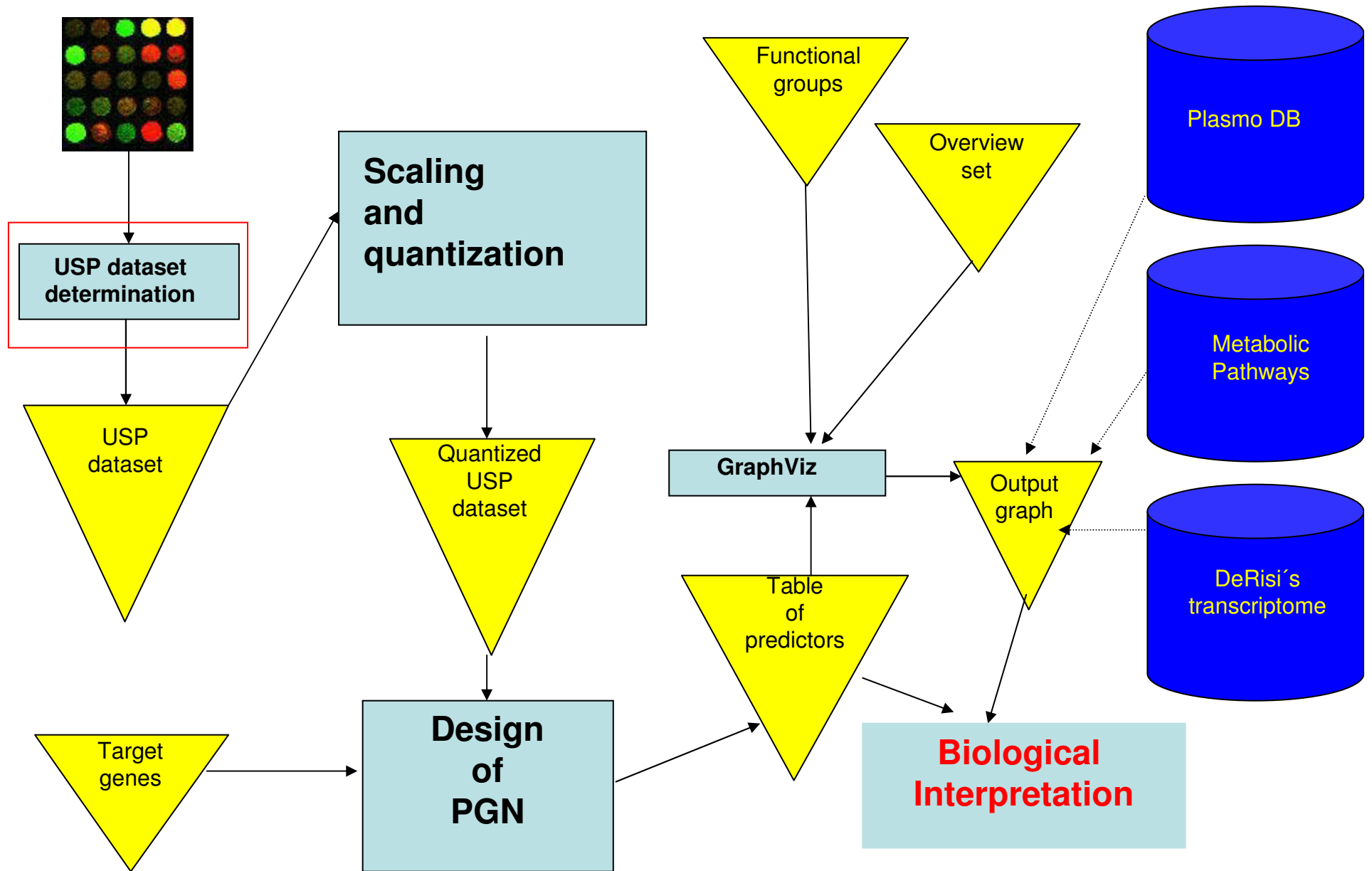


Apicoplast

# System architecture

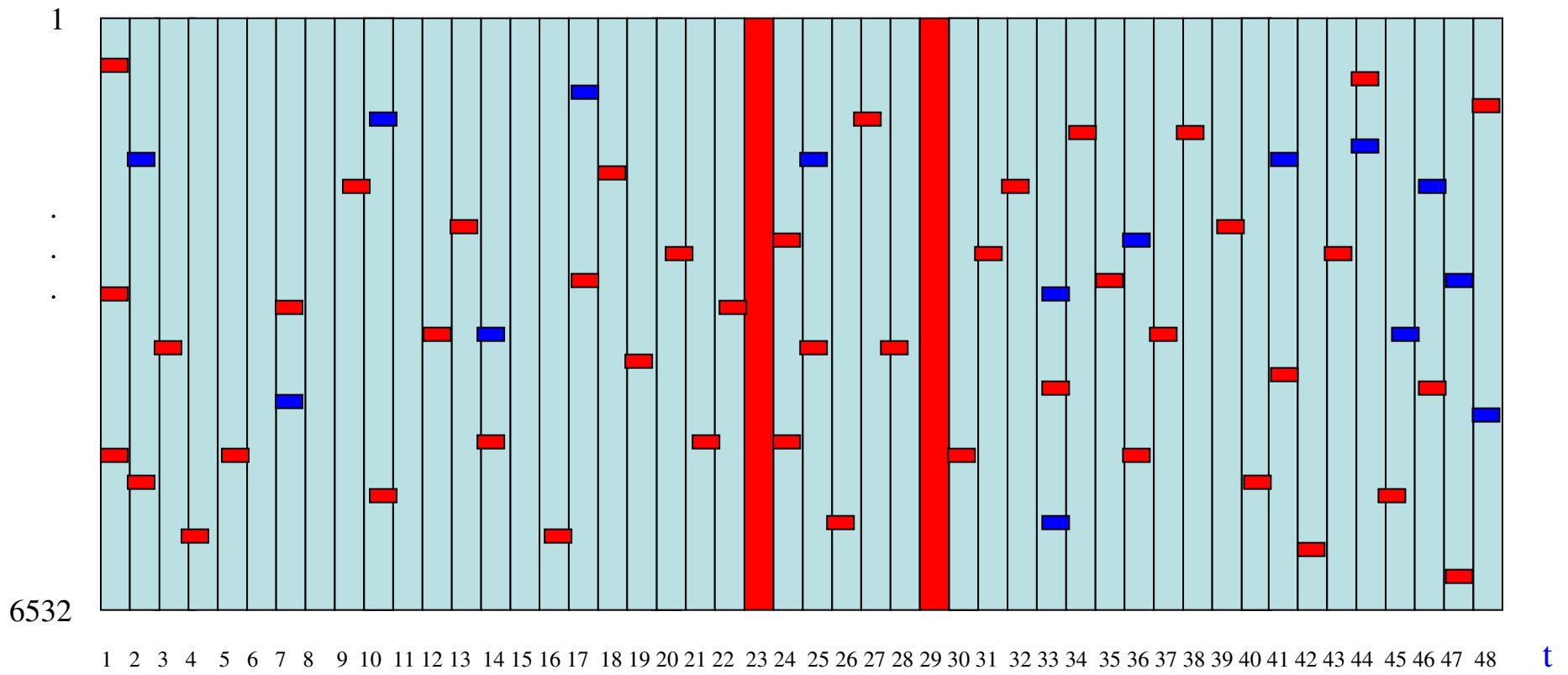


# System architecture





# Genes



Good spots



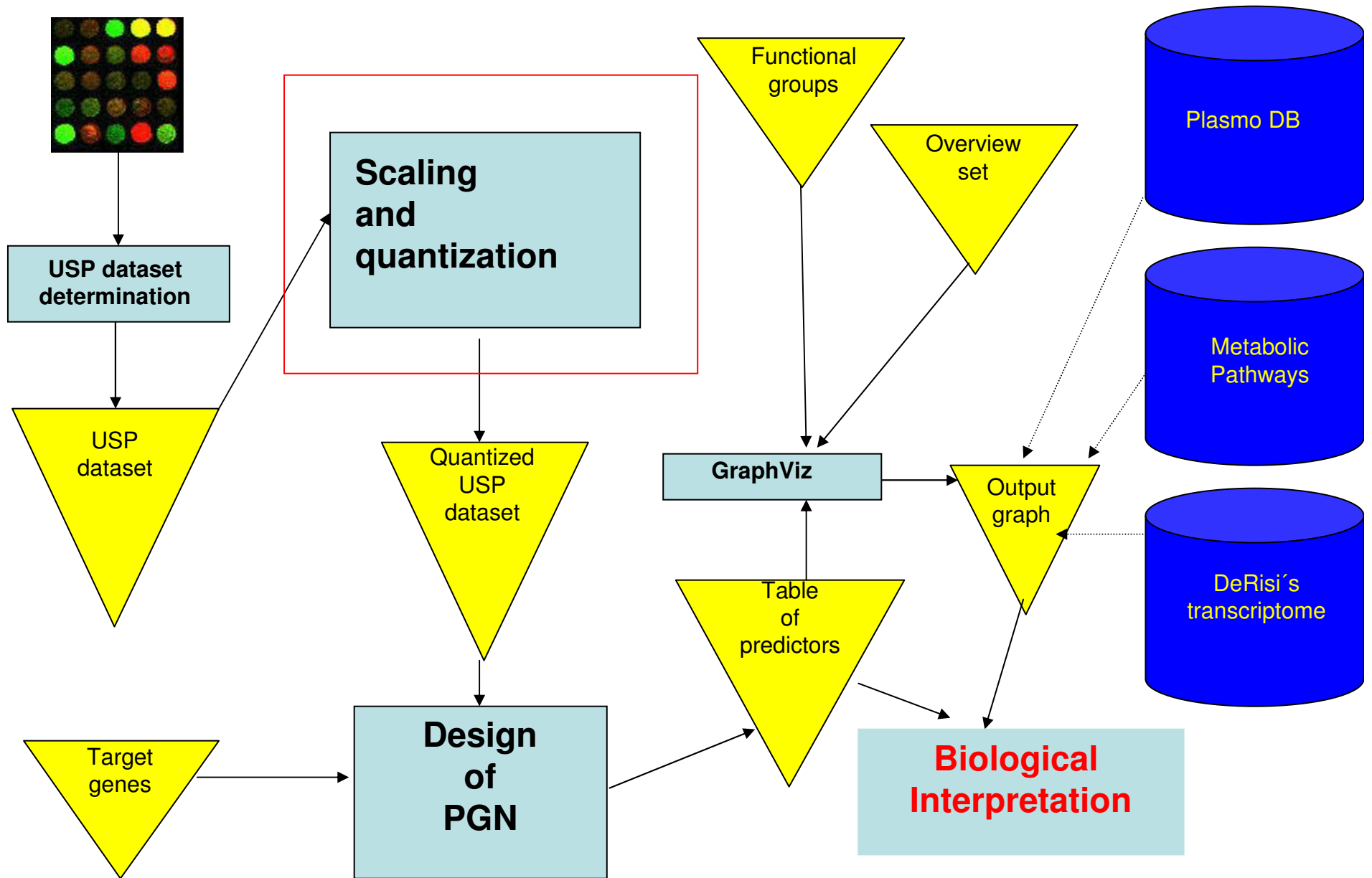
Weak spots



Bad spots

**NO INTERPOLATION**

# System architecture



# Scaling

For each  $i$ , estimate the mean  $\hat{E}[x_i[t]]$   
and standard deviation  $\hat{\sigma}[x_i[t]]$   
of the spots

Scale normalization of the spots

$$n_i[t] = \frac{x_i[t] - \hat{E}[x_i[t]]}{\hat{\sigma}[x_i[t]]}$$

# Quantization

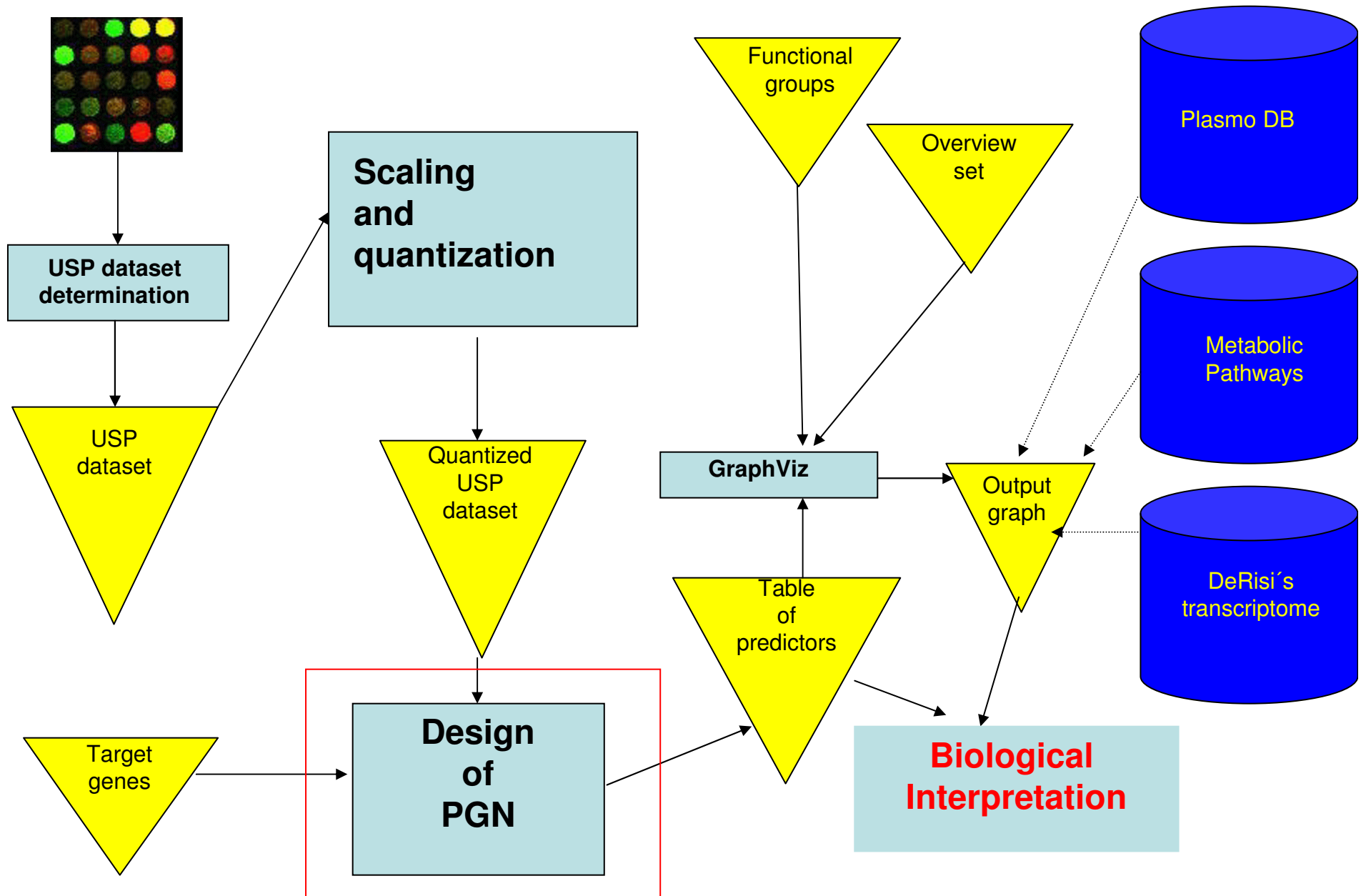
Let  $n_i^+[t]$  and  $n_i^-[t]$  denote, respectively, the normalized signals greater and lower than zero at  $t$ .

If  $n_i^+[t] > \hat{E}[n_i^+[t]]$ , then  $x_i[t] = +1$

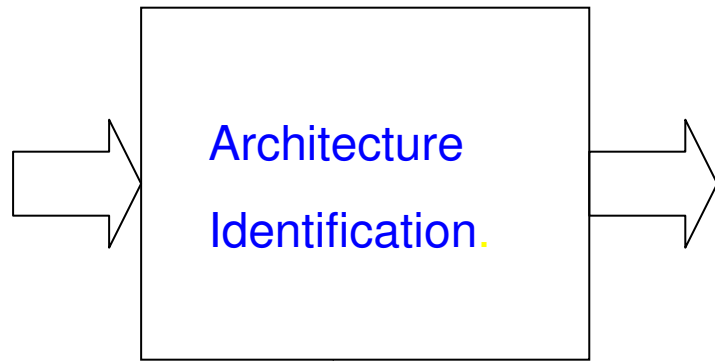
If  $n_i^-[t] \geq \hat{E}[n_i^-[t]]$  and  $n_i^+[t] \leq \hat{E}[n_i^+[t]]$ , then  $x_i[t] = 0$

If  $n_i^-[t] < \hat{E}[n_i^-[t]]$ , then  $x_i[t] = -1$

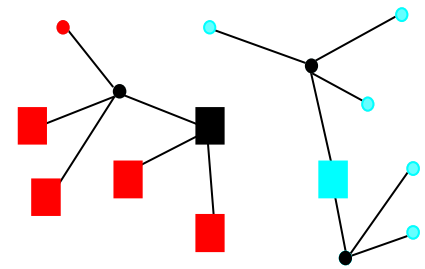
# System architecture



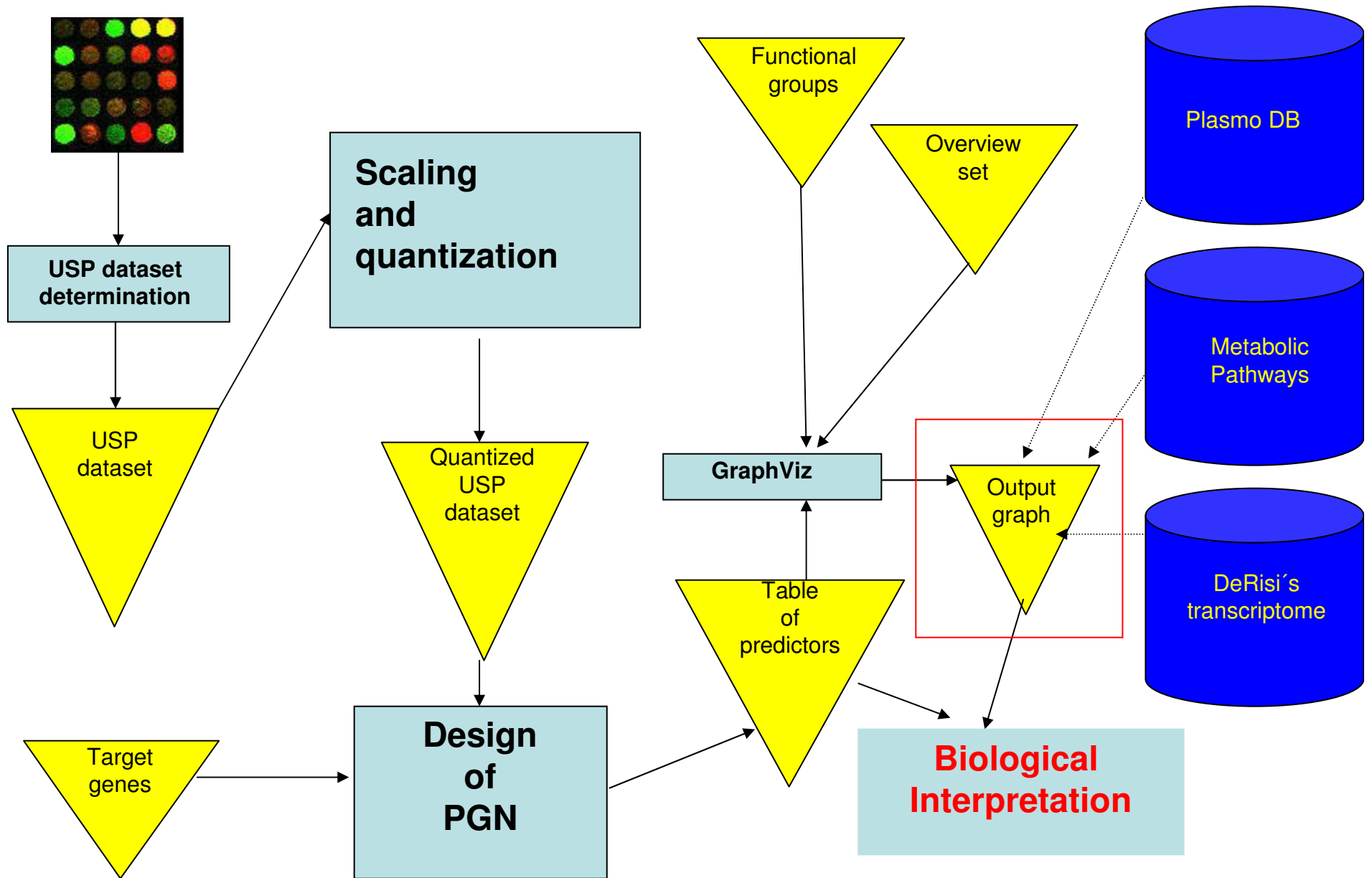
$x[1], x[2], \dots, x[48]$



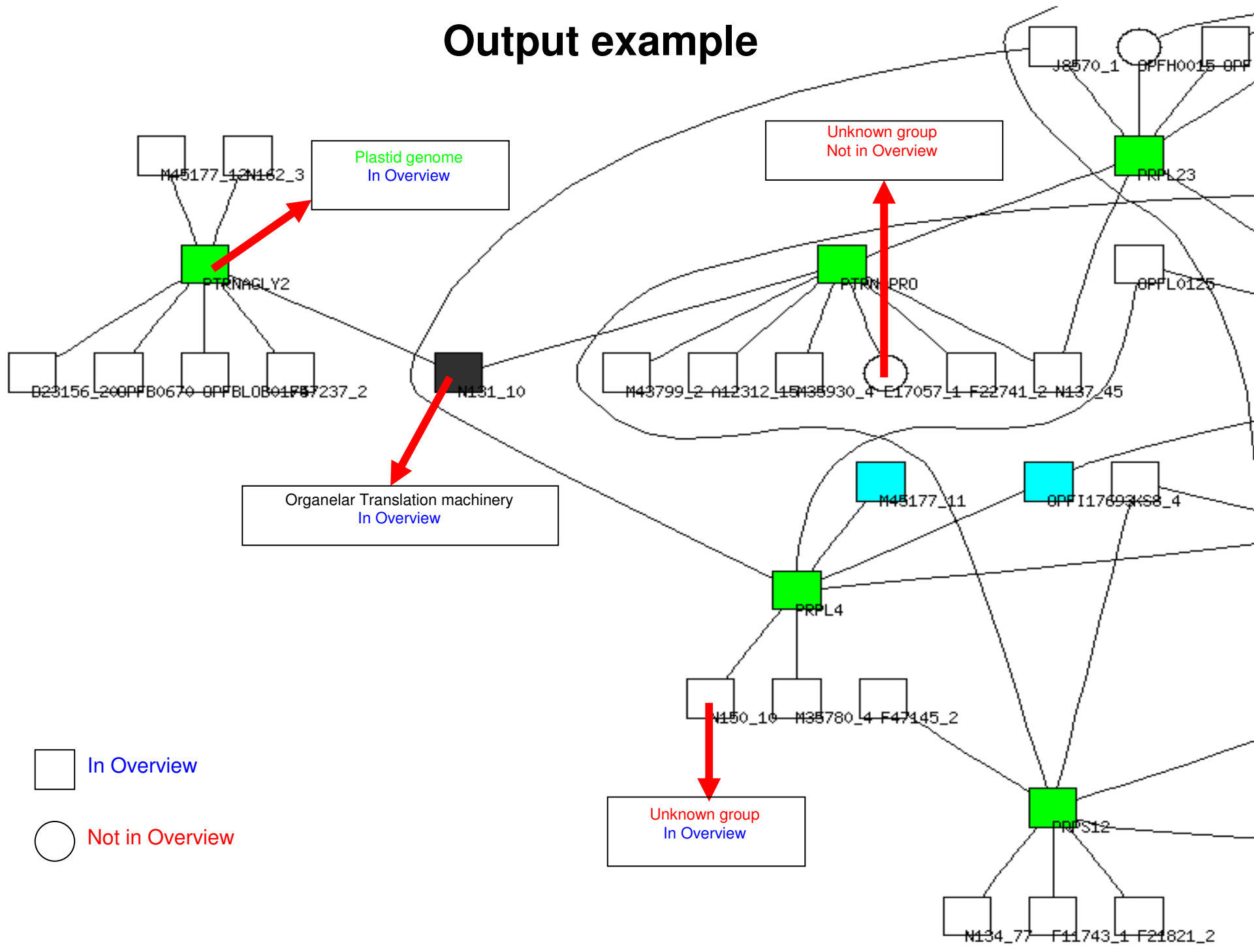
target genes



# System architecture



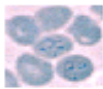
# Output example



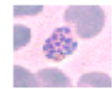


# OPFB0670

<a href="#">PlasmoDB</a>	<a href="#">Metabolic Pathway</a>	<a href="#">Derisi Lab</a>
--------------------------	-----------------------------------	----------------------------



# *P. falciparum* PFB0330c



[Back](#)

[Home](#) [Downloads](#) [Tools](#) [Queries](#) [BLAST](#) [History](#) [CDs & Links](#) [Browse](#) [Data Sources](#) [SRT](#) [Help](#)

## *Plasmodium falciparum* / CHR 2 / PFB0330c

cysteine protease, putative

### Summary view

[Add this gene to your History](#)

#### [Annotation](#)

#### [Protein](#)

#### [Expression](#)

#### [Sequence](#)

[Curated Annotation](#)

[UserComments](#)

[GO Process](#)

[GO Component](#)

[GO Function](#)

[EC number](#)

[RefSeqs](#)

[Metabolic Pathways](#)

[MR4 Reagents](#)

[Ortholog Group](#)

[Ortholog Views](#)

[Orthologs](#)

[BLASTP non-Pf](#)

[\(graphic\)](#)

[BLASTP other](#)

[\(graphic\)](#)

[BLASTP NRDB](#)

[PDB structures](#)

[Structural Models](#)

[Features \(graphic\)](#)

[Pfam](#)

[PROSITE](#)

[TM domains](#)

[SignalP](#)

[PlasmoAP](#)

[Motifs \(graphic\)](#)

[Motifs](#)

[Proteomics \(graphic\)](#)

[Mass spec. data](#)

[Microarrays](#)

[Developmental series](#)

[\(clone array\)](#)

[Developmental series](#)

[\(Affy array\)](#)

[Developmental series](#)

[\(glass slide array\)](#)

[Proteomics \(graphic\)](#)

[Mass spec. data](#)

[DNA \(graphic\)](#)

[Exons](#)

[SNPs](#)

[mRNA/RNA sequence](#)

[Protein sequence](#)

## Annotation

[back to top](#)

### Curated Annotation

\*\*\* None \*\*\*

# *P. falciparum* Gene: [PFB0330c](#)

[Back](#)

ID: PFB0330C

Comment:

This gene was predicted and reviewed manually for the Oct. 3, 2002 Nature publication by Gardner et al. This gene has at least one intron

Superclasses: [Genes](#) -> [UNCLASSIFIED](#)

Chromosome: Chromosome 2

Map Position (centisomes): [31.287 \[click to view in chromosome browser\]](#)

Map Position (nucleotides): 296,317 -> 297,583

Products: [cysteine protease, putative](#)

Gene-Reaction Schematic: [?](#)



---

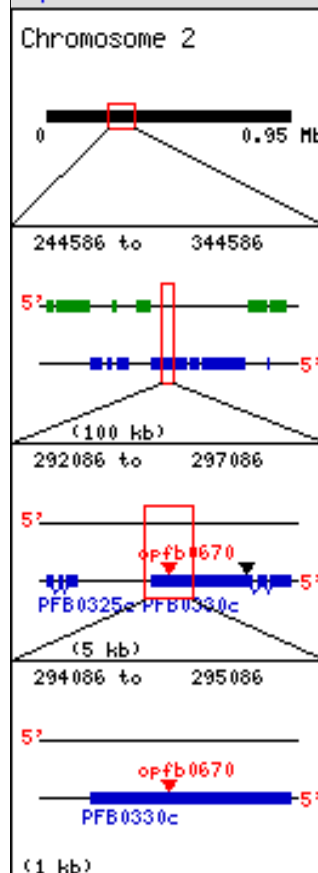
[Query Page](#)

[Advanced Query Page](#)

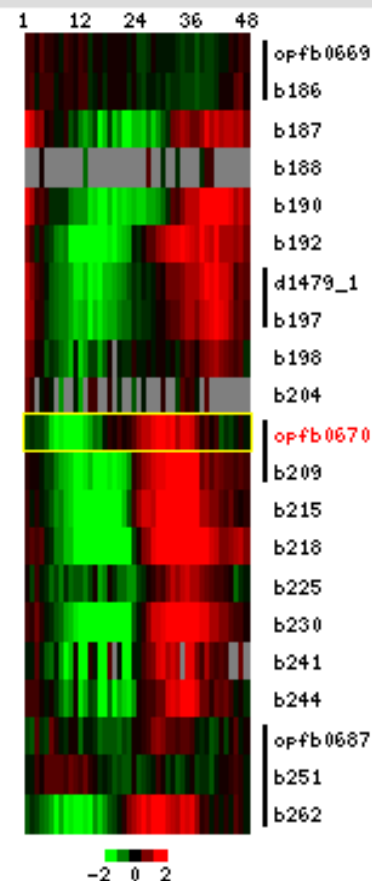
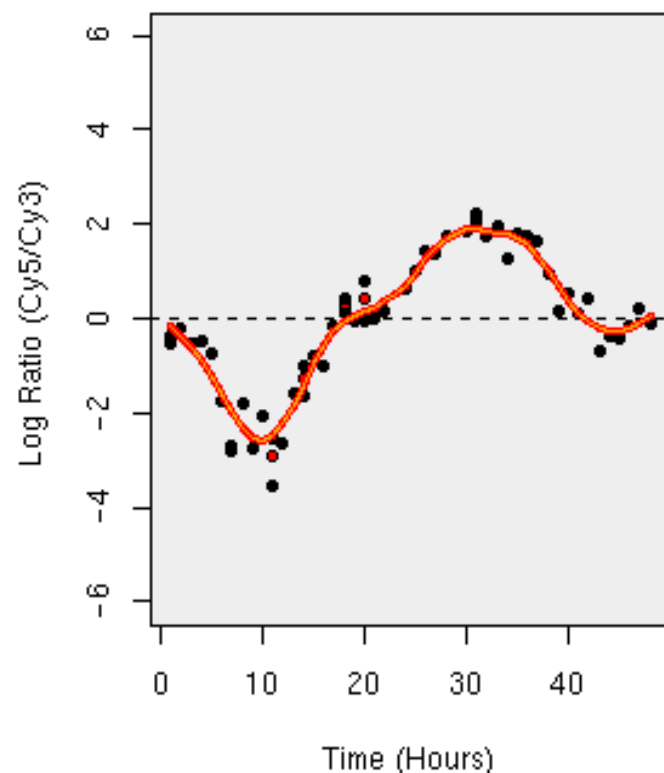
[BioCyc Home](#)

[Report Errors or Provide Feedback](#)

OligoID	Status	Maximum Hour	Minimum Hour	Amplitude Score (log2)	Score (%)	Phase (-Pi to +Pi)	CGH %3D7	Avg. Med. Intensity
<a href="#">opfb0670</a>	UNIQUE	30	10	4.5	87	0.06	89	3211.57



[← OLIGO →](#)



PlasmoDB ID Description

[PFB0330c](#) cysteine protease, putative

Oligo Sequence

[BLAST @ PlasmoDB](#)

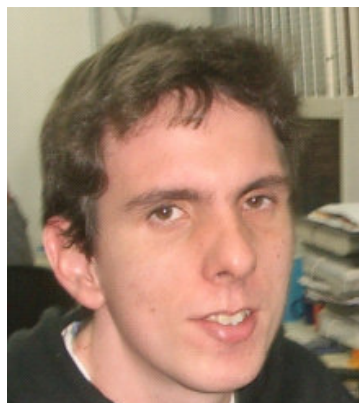
5'

CTGCCCAAGATGAGCCACCTACTGATAATGTAGAATCACAAGCAGAAAATAACAAAAAACAGAAATTTA









J. Barrera, R.M. Cesar Jr., C. P. Pereira, D. Martins,  
R. Z. Vencio, E. F. Merino, M. M. Yamamoto

[www.bioinfo.usp.br](http://www.bioinfo.usp.br)

