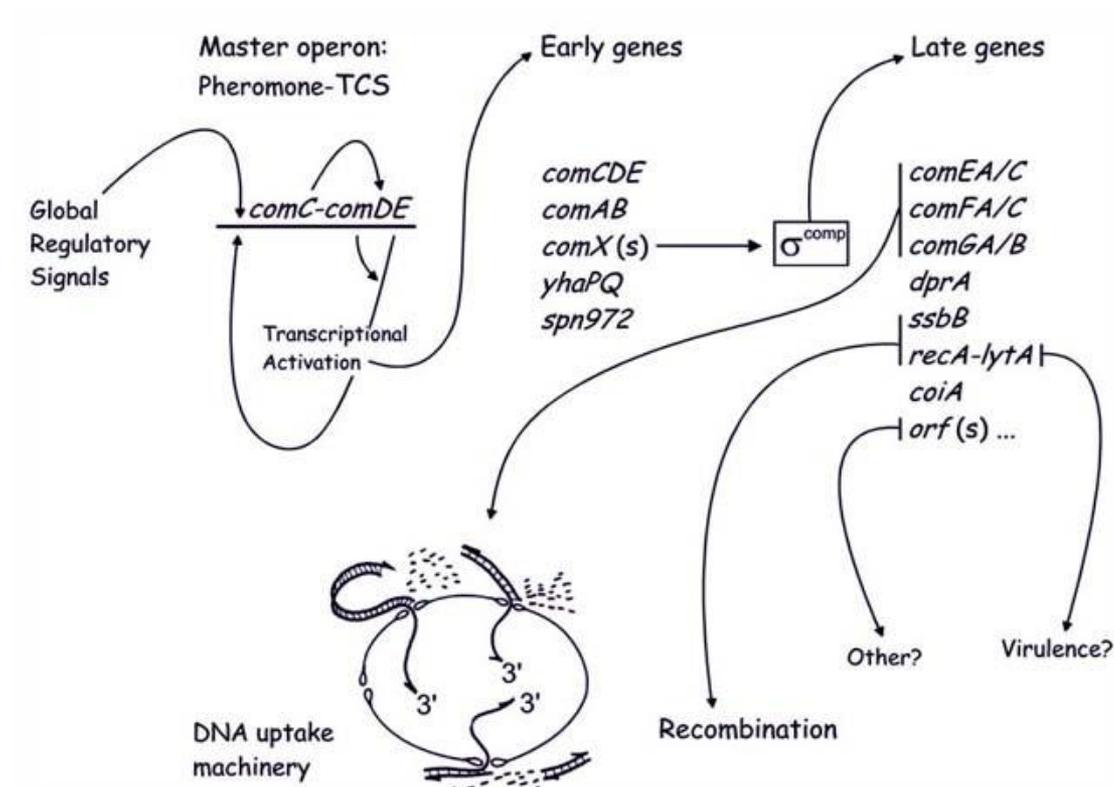


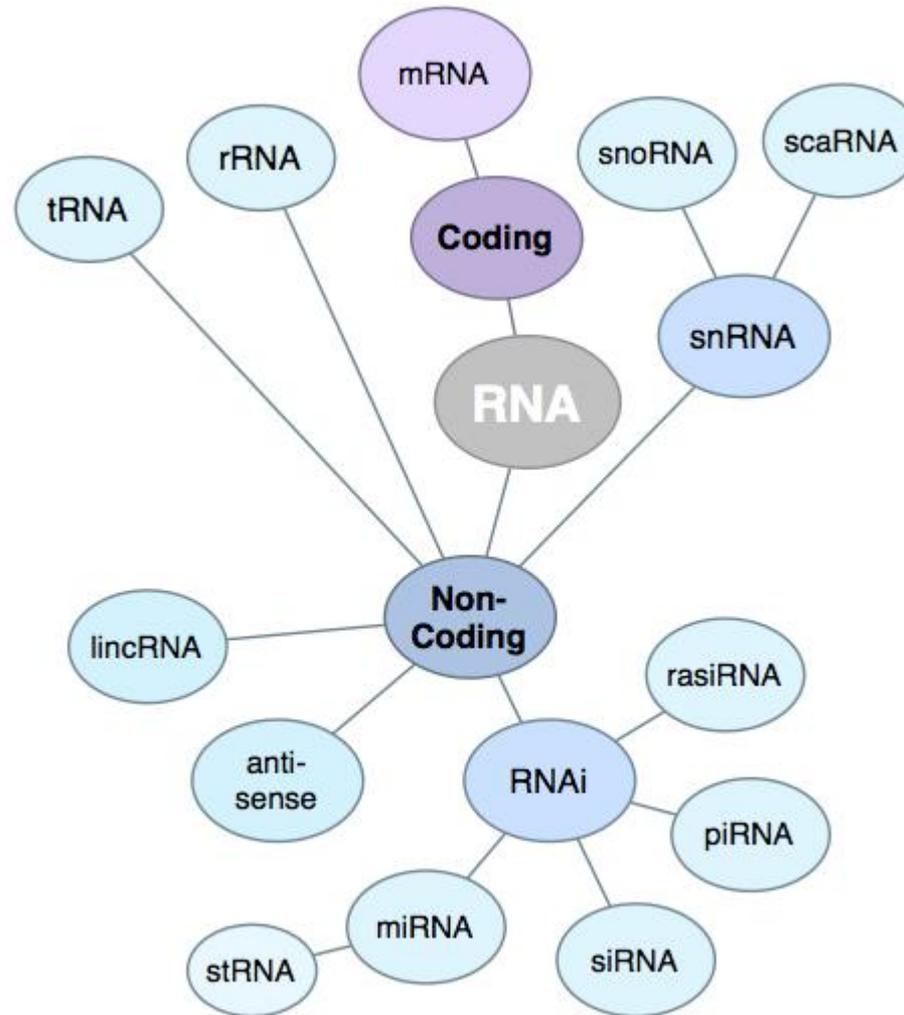
# Réseaux de régulation

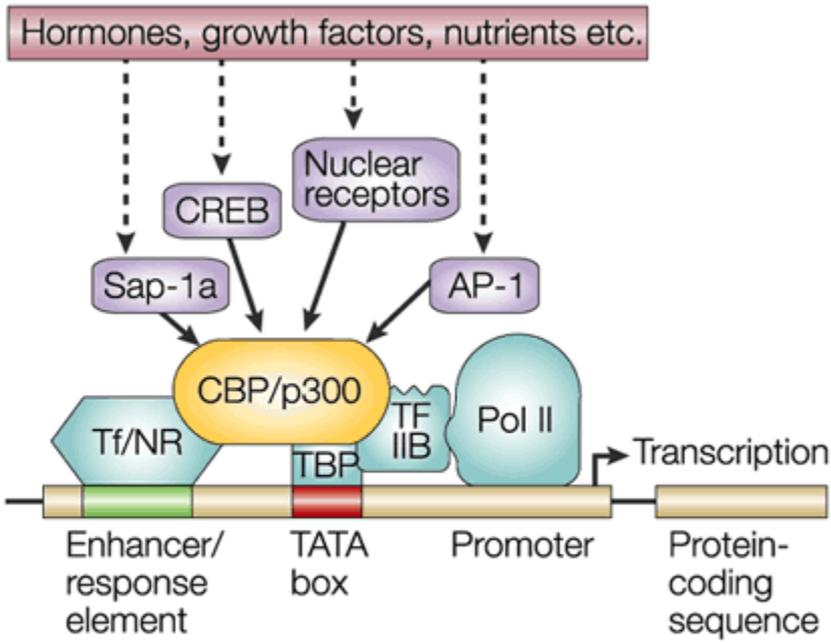
## Master 1 MABS

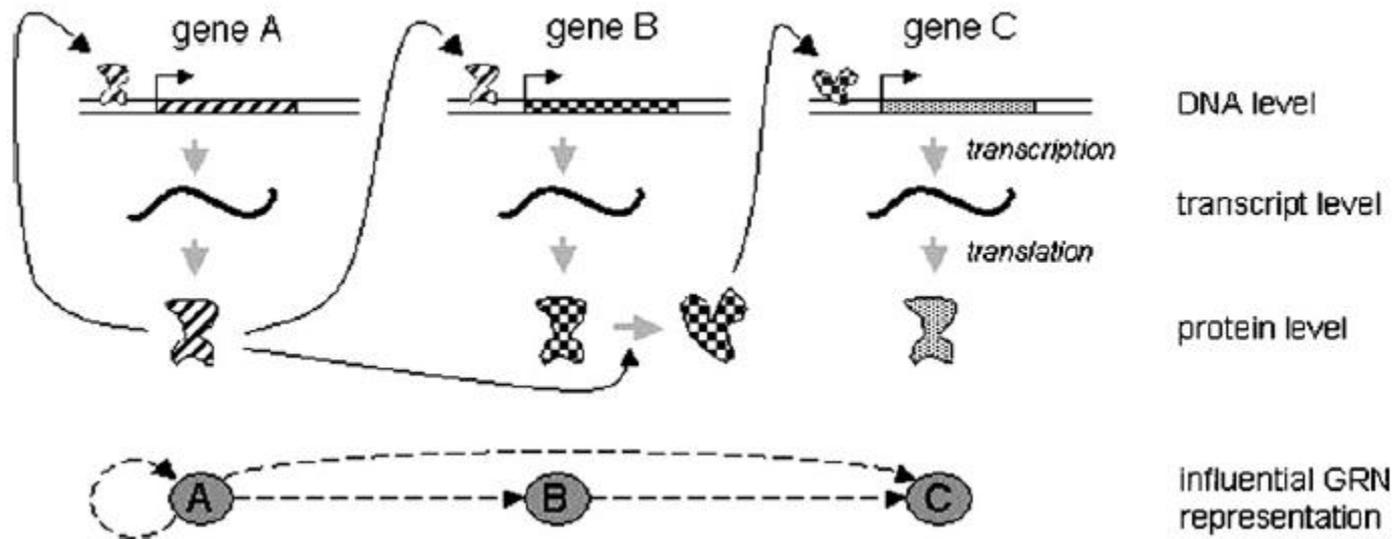


- Compréhension de propriétés du vivant conférées par les modalités d'expression du génome
- Expression
  - ◆ ARN codant et non codant
- Relation génotype/phénotype
  - ◆ et épigénétique : étude des changements, héritable au cours des divisions cellulaires, qui affectent la fonction des génomes sans altération de la séquence ADN.
- Phénomènes complexes :
  - ◆ différenciation
  - ◆ développement
  - ◆ vieillissement
  - ◆ adaptation
  - ◆ effet de l'environnement

## RNA World





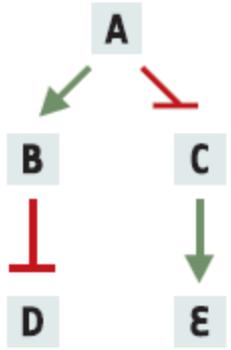


# Représentations et modèles

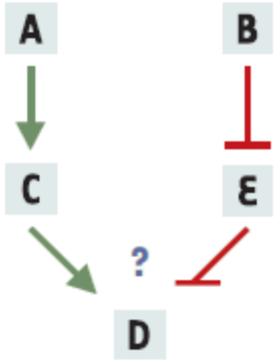
**A**



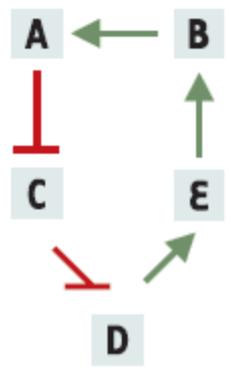
**B**



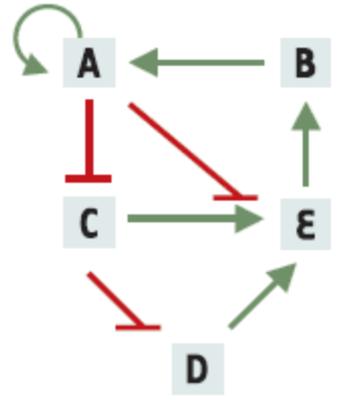
**C**



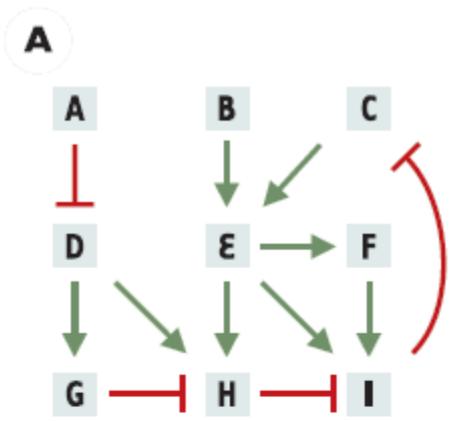
**D**



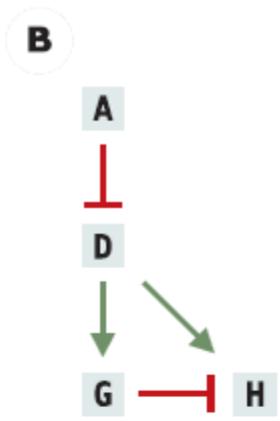
**E**



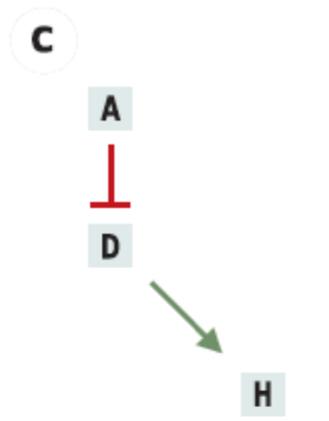
# Décomposition en modules



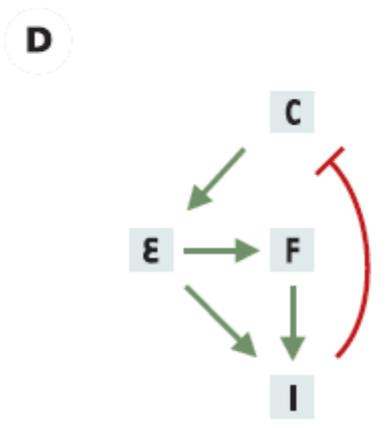
Graphe d'interaction



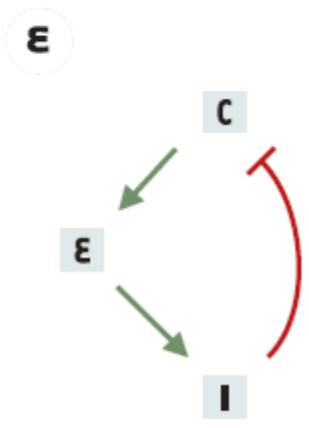
Ordre partiel



Plus court chemin entre les sommets A et H



Composante fortement connexe

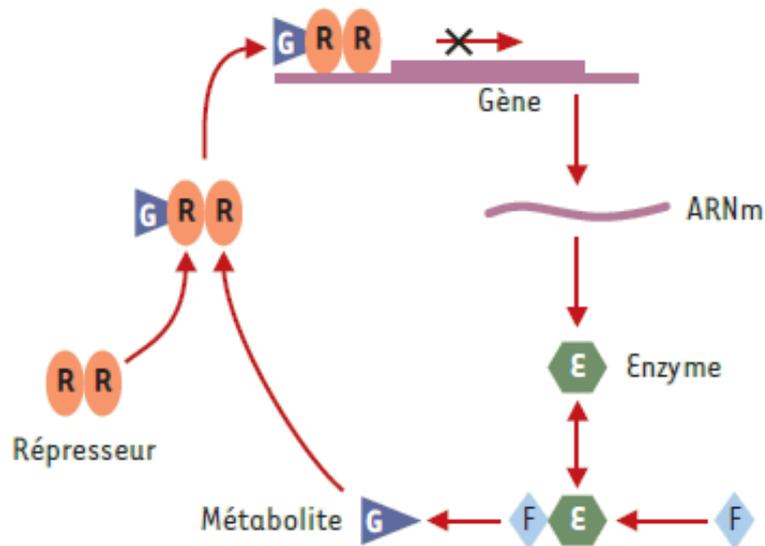


Circuits

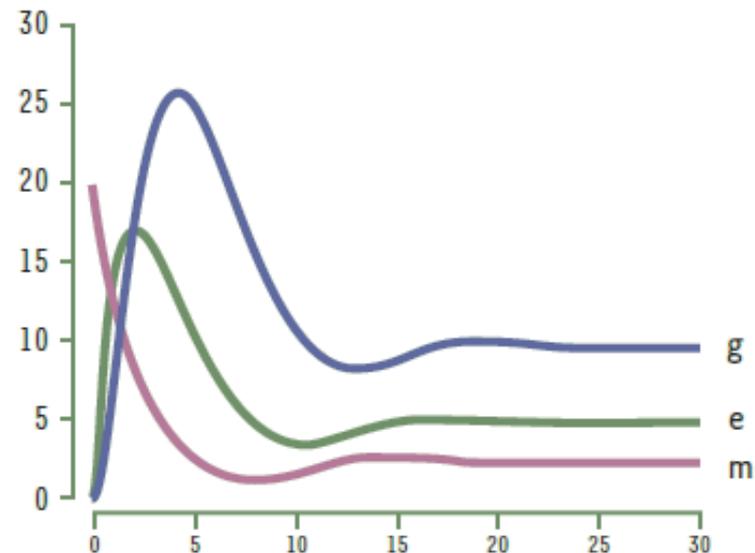
- Données d'expression
  - ◆ perturbations sur le phénomène d'intérêt
  - ◆ ensemble de conditions (>60) capturant un maximum de variations
- Données ChIP et ChIP-seq
  - ◆ localisation des sites de fixation des facteurs de transcription
- Facteurs de transcription connus
  - ◆ ex : RegulonDB, TransFac, JASPAR
- Prédiction de régions régulatrices
  - ◆ co-expression observée puis recherche des promoteurs
- Littérature et annotations

- Systèmes d'équations
- Réseaux booléens
- Réseau de petri
- Réseaux bayésiens
- Réseaux bayésiens dynamiques
- Réseaux d'influence

- Concentrations/activités des molécules
- Vitesse de production, dégradation
- Constante de seuil, association, dissociation, coopération



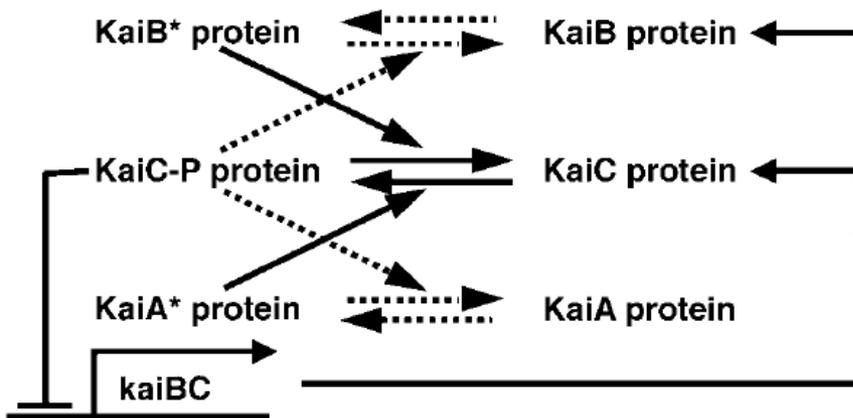
CONCENTRATION



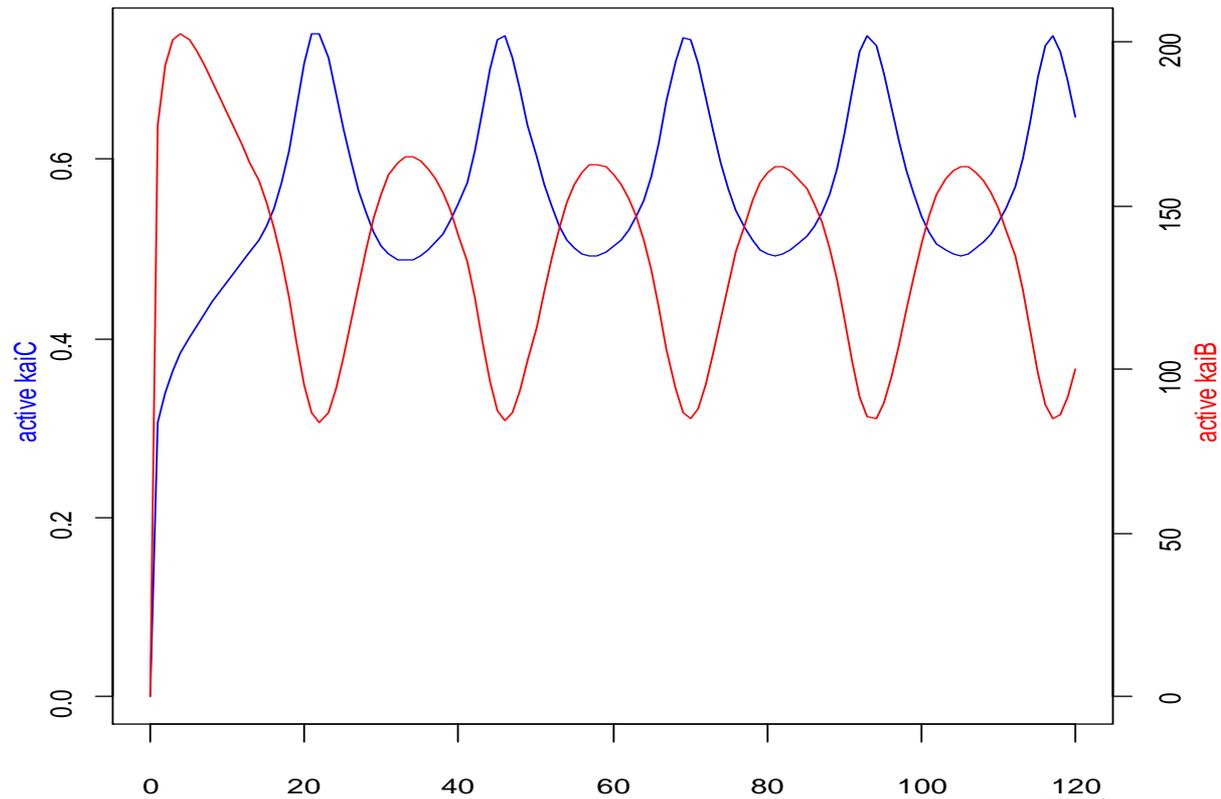
$$\frac{dm}{dt} = K_1 \frac{\theta^n}{\theta^n + g^n} - \gamma_1 m \quad \frac{de}{dt} = K_1 m - \gamma_2 e \quad \frac{dg}{dt} = K_3 e - \gamma_3 g$$

TEMPS

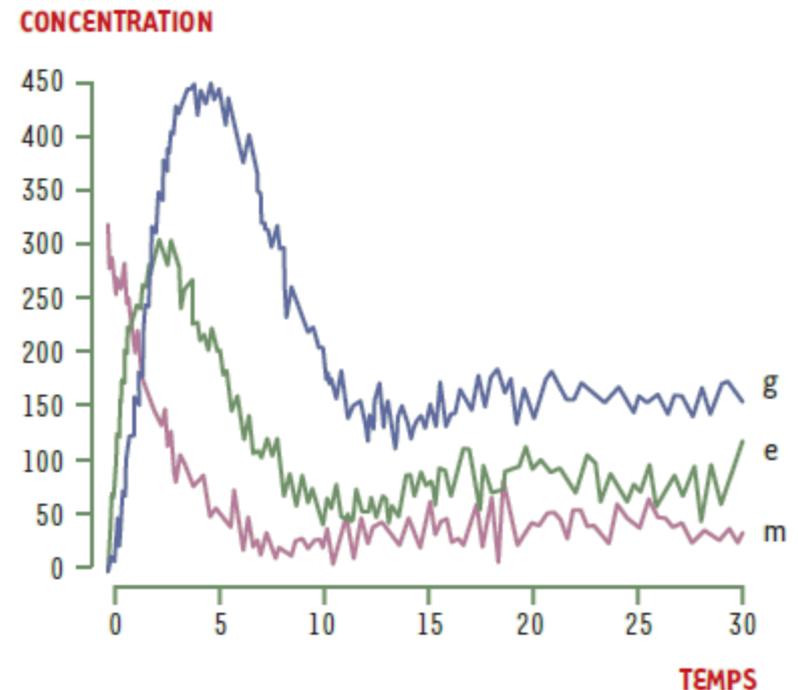
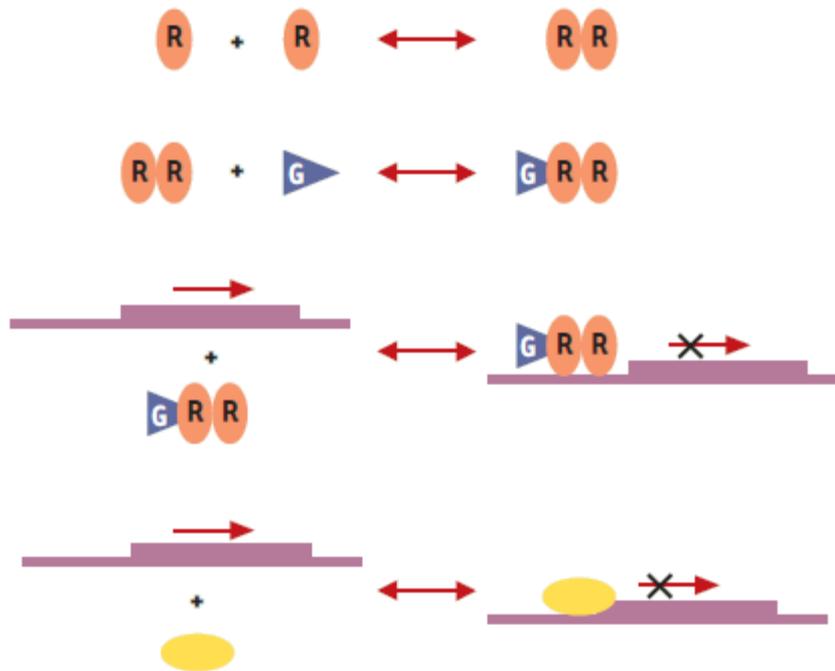
# Exemple : oscillations circadiennes chez les cyanobactéries



$$\begin{aligned} dx/dt &= pa(C_0s - x) - bx(y + f), \\ dy/dt &= g(B_0s - y) - k_1yx^n/(q^n + x^n), \\ dB/dt &= \varepsilon_1\{B_0\lambda/(1 + h_1x^m) - \mu B\}, \\ dC/dt &= \varepsilon_1\{C_0\lambda/(1 + h_1x^m) - \mu C\}. \end{aligned}$$



- Concentrations très faibles
- Probabilités associées aux réactions/transitions
- En général, trop complexe pour une solution analytique → simulations



## Synchrone

$(xy)_t$	$(xy)_{t+1}$
$\begin{matrix} ++ \\ 00 \end{matrix}$	11
[01]	01
[10]	10
$\begin{matrix} -- \\ 11 \end{matrix}$	00

$$x_{t+1} = \bar{y}_t$$

$$y_{t+1} = \bar{x}_t$$

équations

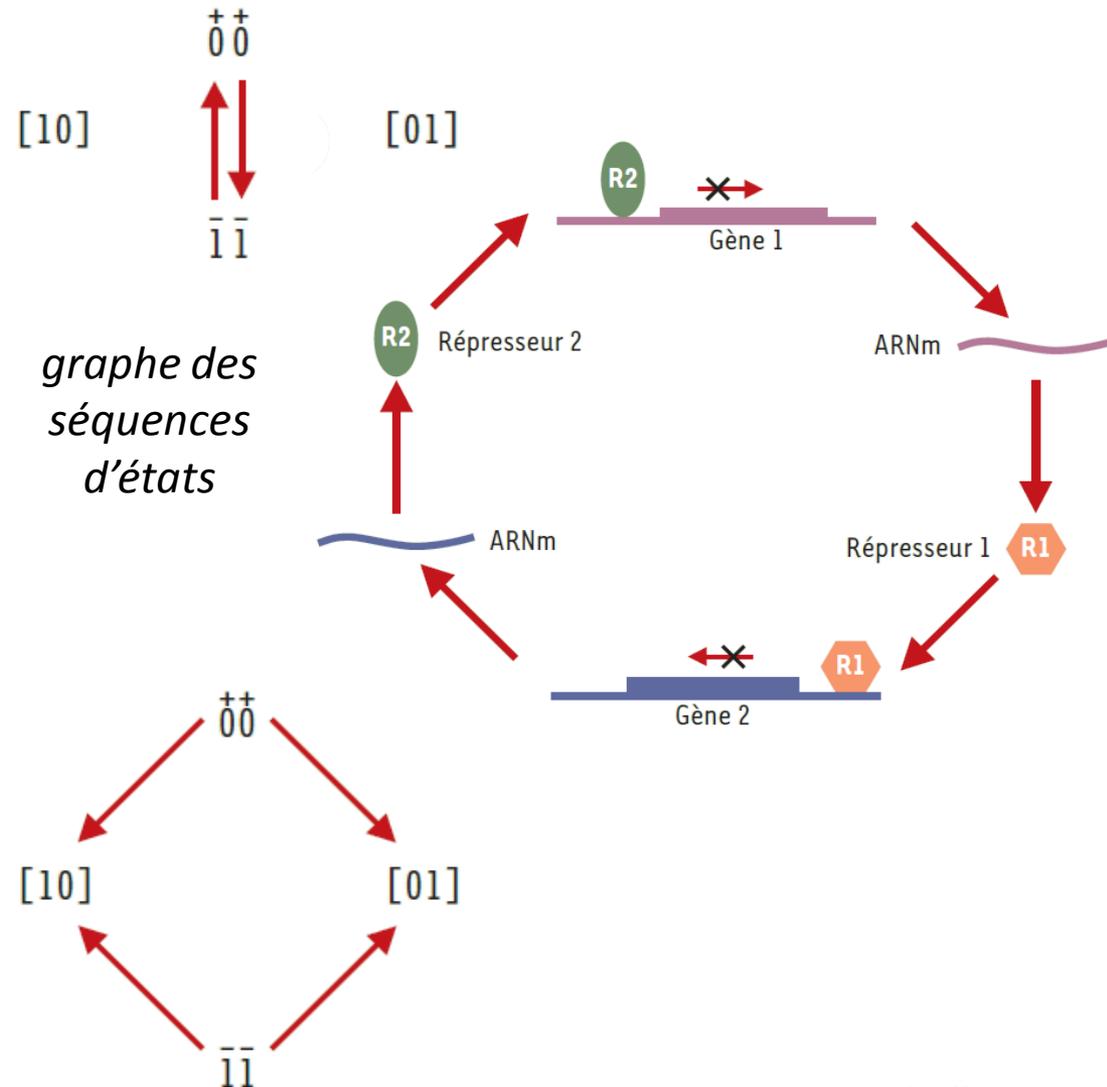
table des états

## Asynchrone

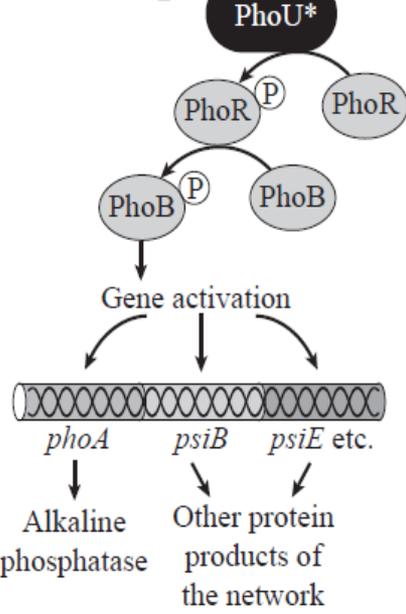
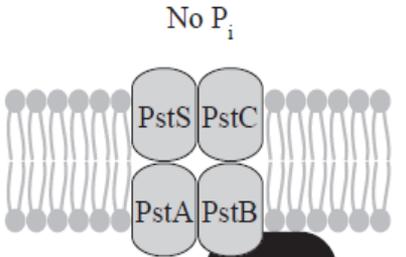
xy	XY
$\begin{matrix} ++ \\ 00 \end{matrix}$	11
[01]	01
[10]	10
$\begin{matrix} -- \\ 11 \end{matrix}$	00

$$x = \bar{y}$$

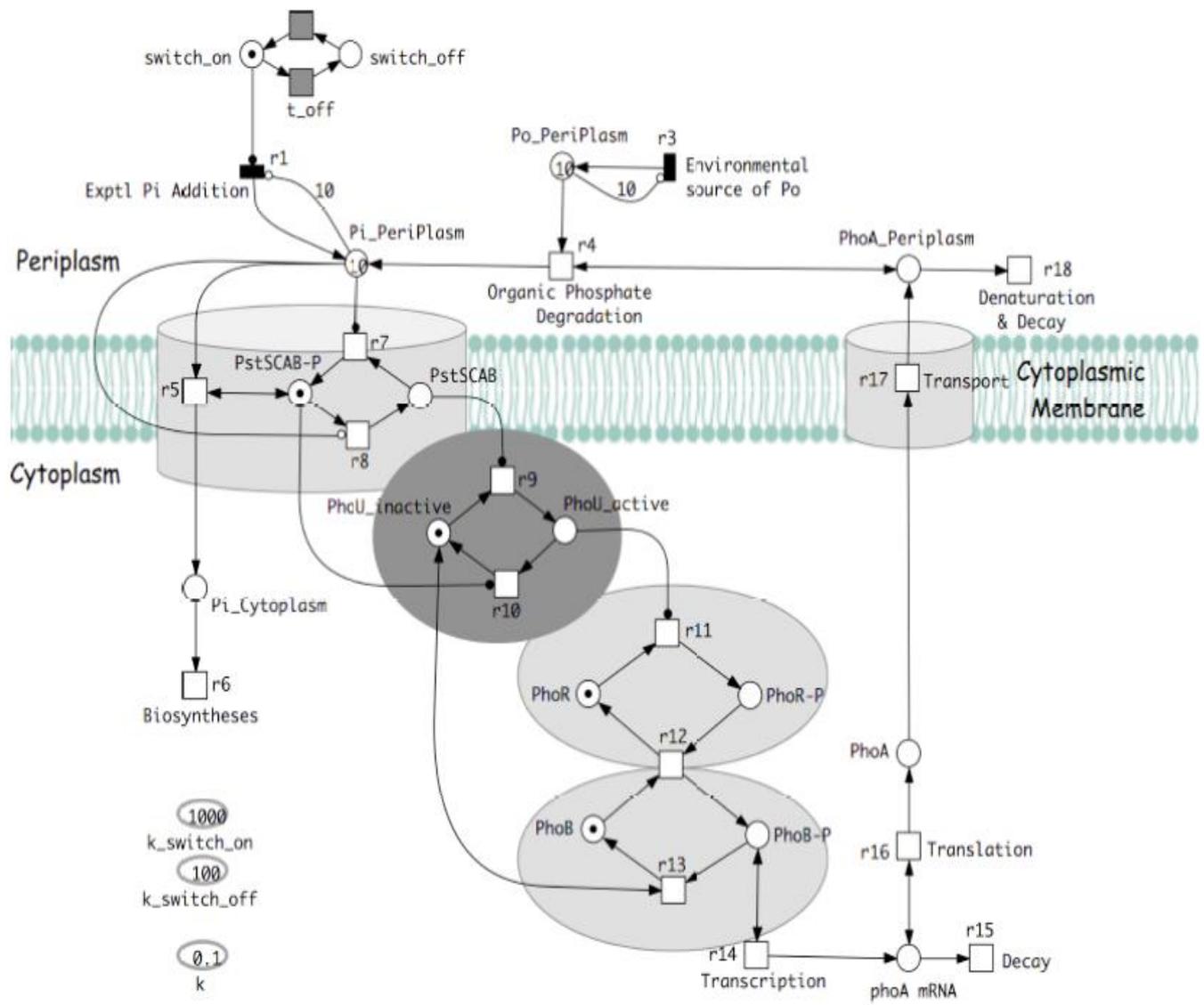
$$y = \bar{x}$$



# Réseau de petri



Neidhardt *et al.* 1990



Durzinsky *et al.*, 2011



- Niveau d'expression de B dépend de celui de A (influence)
  - ♦  $P(B/A)$
- Le niveau d'expression de C
  - ♦ dépend de celui de B
  - ♦ est indépendant de celui de A sachant B

	cond1	cond2	cond3	cond4	cond5
A	off	low	high	low	low
B	off	low	high	high	low
C	off	off	low	high	low

observations



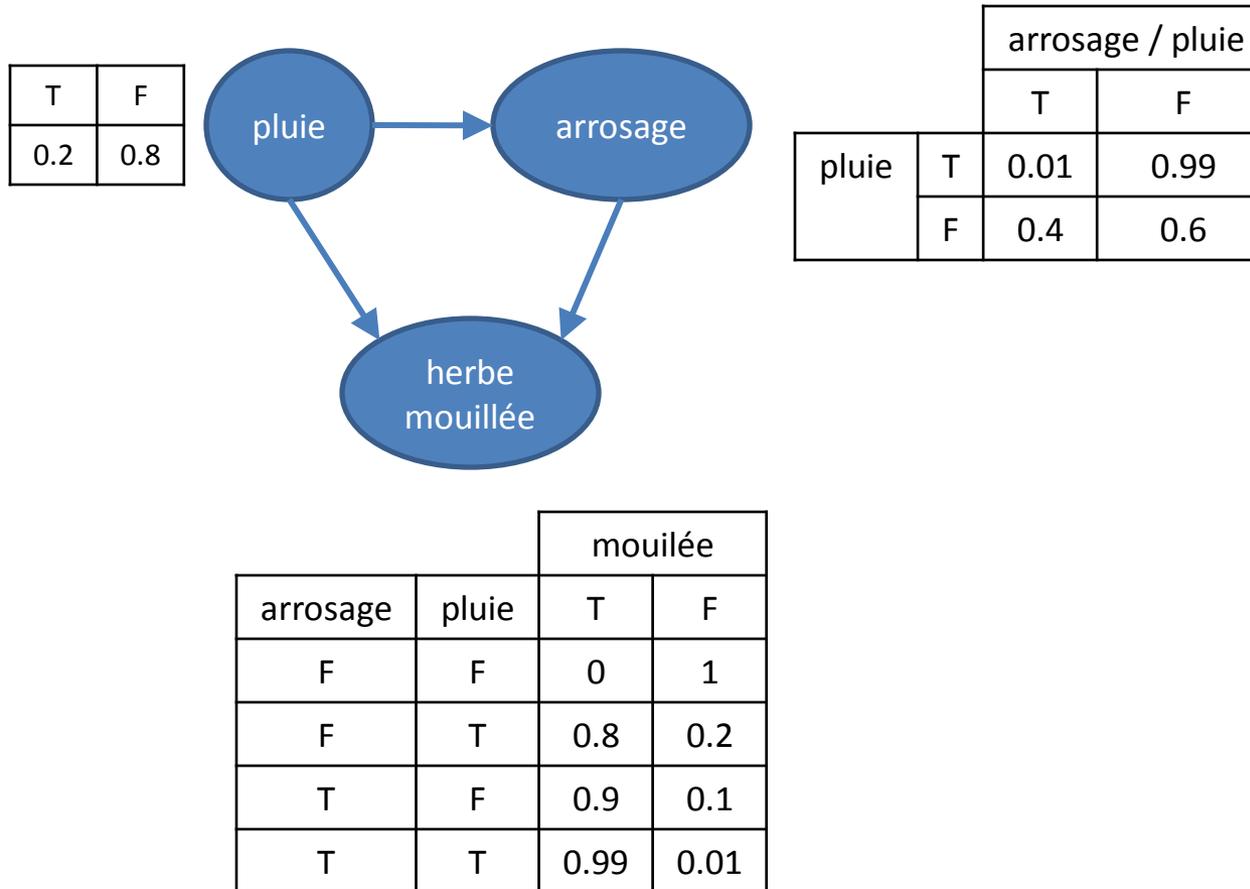
$P(B/A)$

B\A	off (1)	low (3)	high (1)
off	1/1	0/3	0/1
low	0/1	2/3	0/1
high	0/1	1/3	1/1



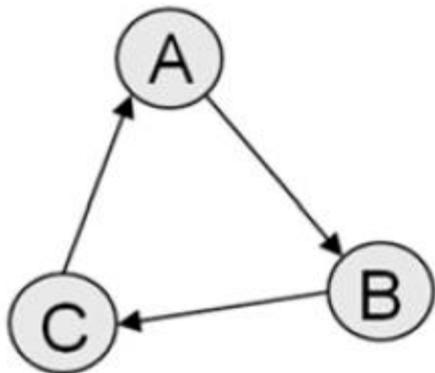
- Niveau d'expression de B dépend de celui de A (influence)
  - ♦  $P(B/A)$
- Le niveau d'expression de C
  - ♦ dépend de celui de B
  - ♦ est indépendant de celui de A sachant B
- Avantages :
  - ♦ capture l'aspect stochastique de la régulation
  - ♦ possibilité d'intégrer des régulations connues
  - ♦ peu de sur-apprentissage et robustesse
  - ♦ quantitatif (niveau d'expression) ou qualitatif (on/off)
- Inconvénient : Pas de cycle donc pas de boucle d'auto-régulation

# Réseaux bayésiens : illustration



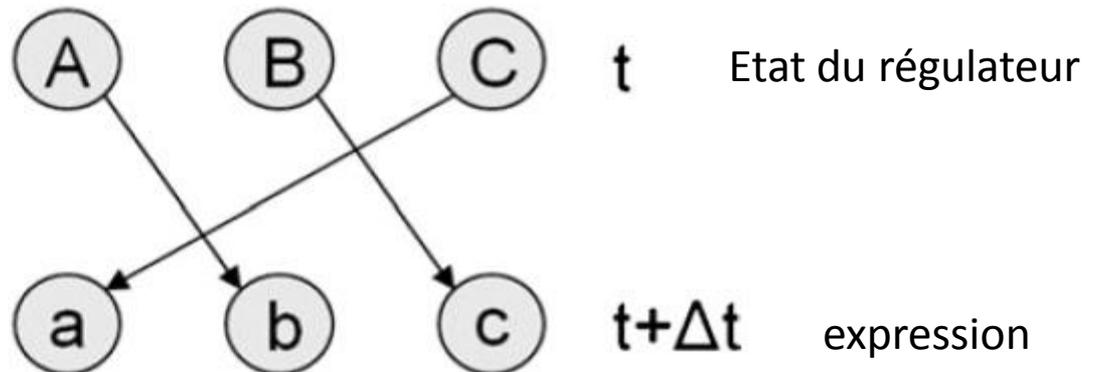
- Information temporelle (temps discret)
  - ◆ variable indicée par son pas de temps
  - ◆ distribution des probabilités d'une variable dépend de l'état de ses prédécesseurs au pas de temps précédent

Static BN

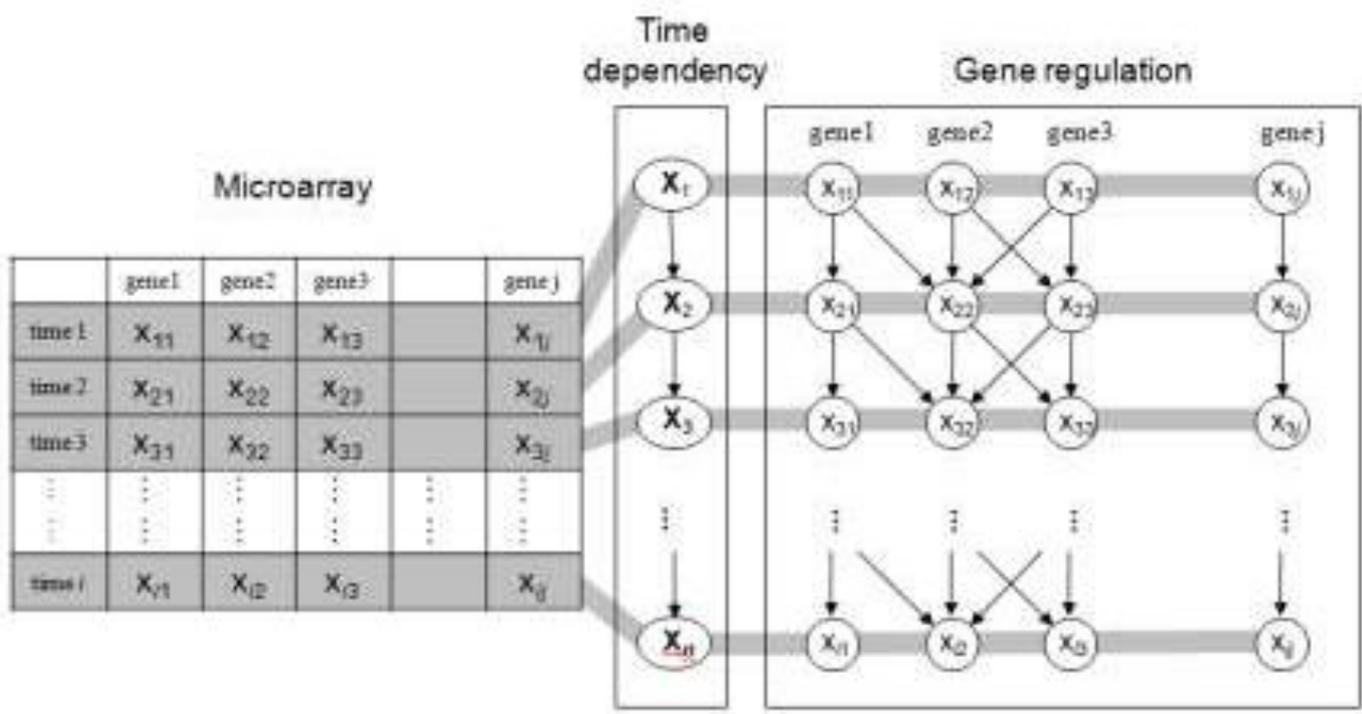
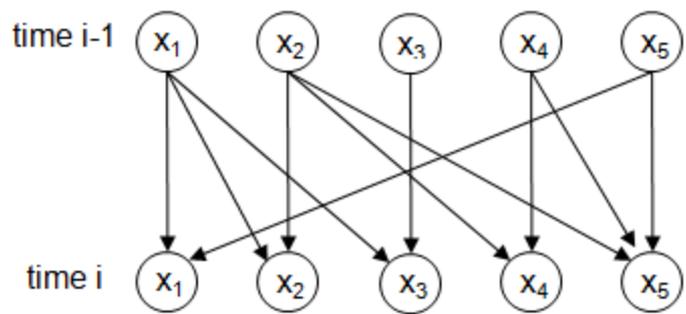
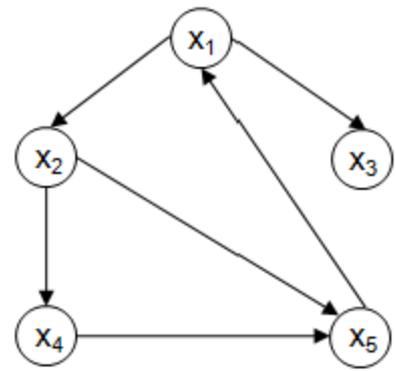


Not allowed !

Dynamic BN



# Réseaux bayésiens dynamiques

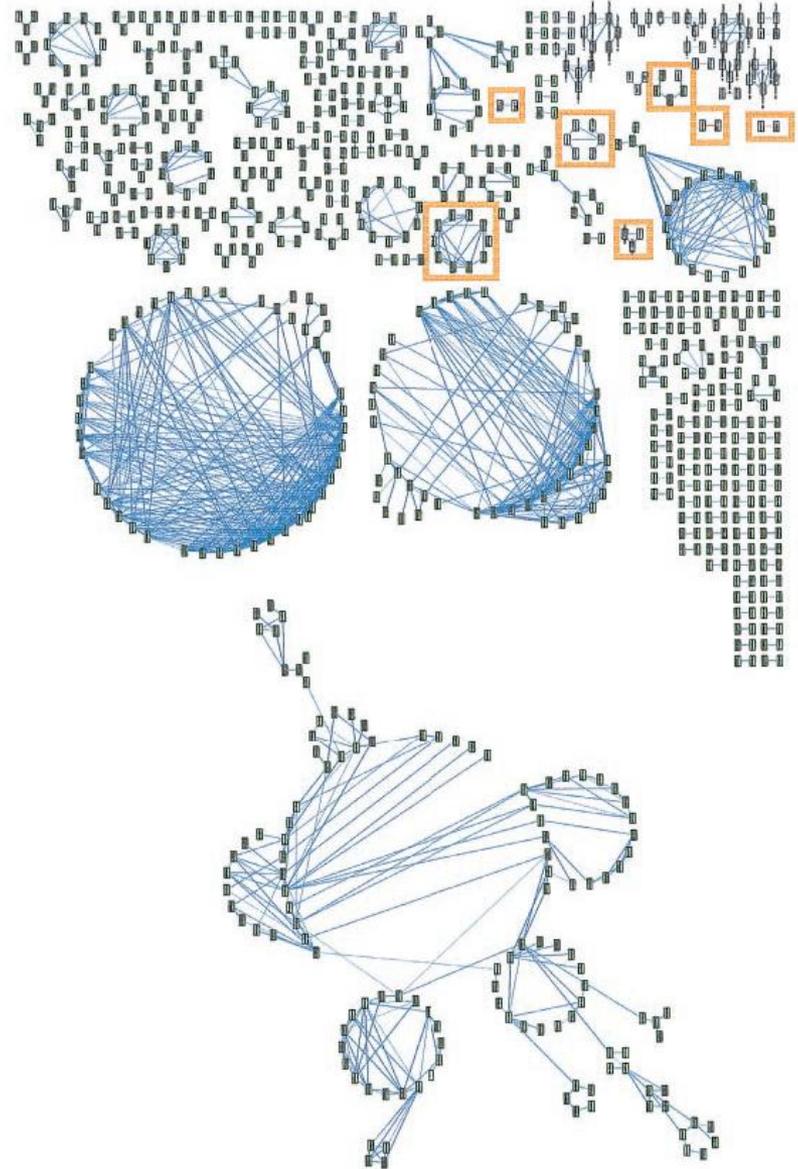


- Adéquation entre observations et modèle
- Généralement en 3 étapes :
  - ◆ inférence de(s) la structure
  - ◆ détermination des paramètres
  - ◆ sélection du meilleur modèle
- Optimisation
  - ◆ de la structure ( $2^n - 1$  connexions pour chaque sommet, BIC)
    - feature selection : considérer uniquement les gènes différentiellement exprimés
    - feature mapping : agréger les ensembles tels que les opérons ou les gènes co-exprimés impliqués dans un même processus biologique
  - ◆ des paramètres : fonction de score
    - moindres carrés
    - maximum de vraisemblance
- Contraintes et incorporation de connaissances

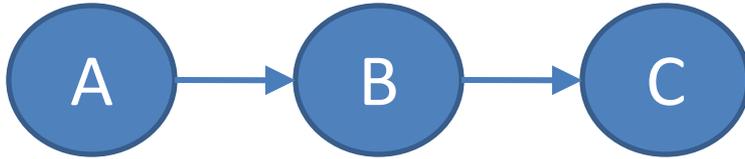
- Mesure de similarité entre profils
  - ◆ Basée sur la corrélation entre l'activité de 2 éléments

$$\hat{r}^2 = \frac{r}{abs(r)} r^2$$

- ◆ seuil pour l'inférence d'un lien entre les éléments



- Information Mutuelle



$$I(X; Y) = \sum_{i,j} P(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}$$

observations

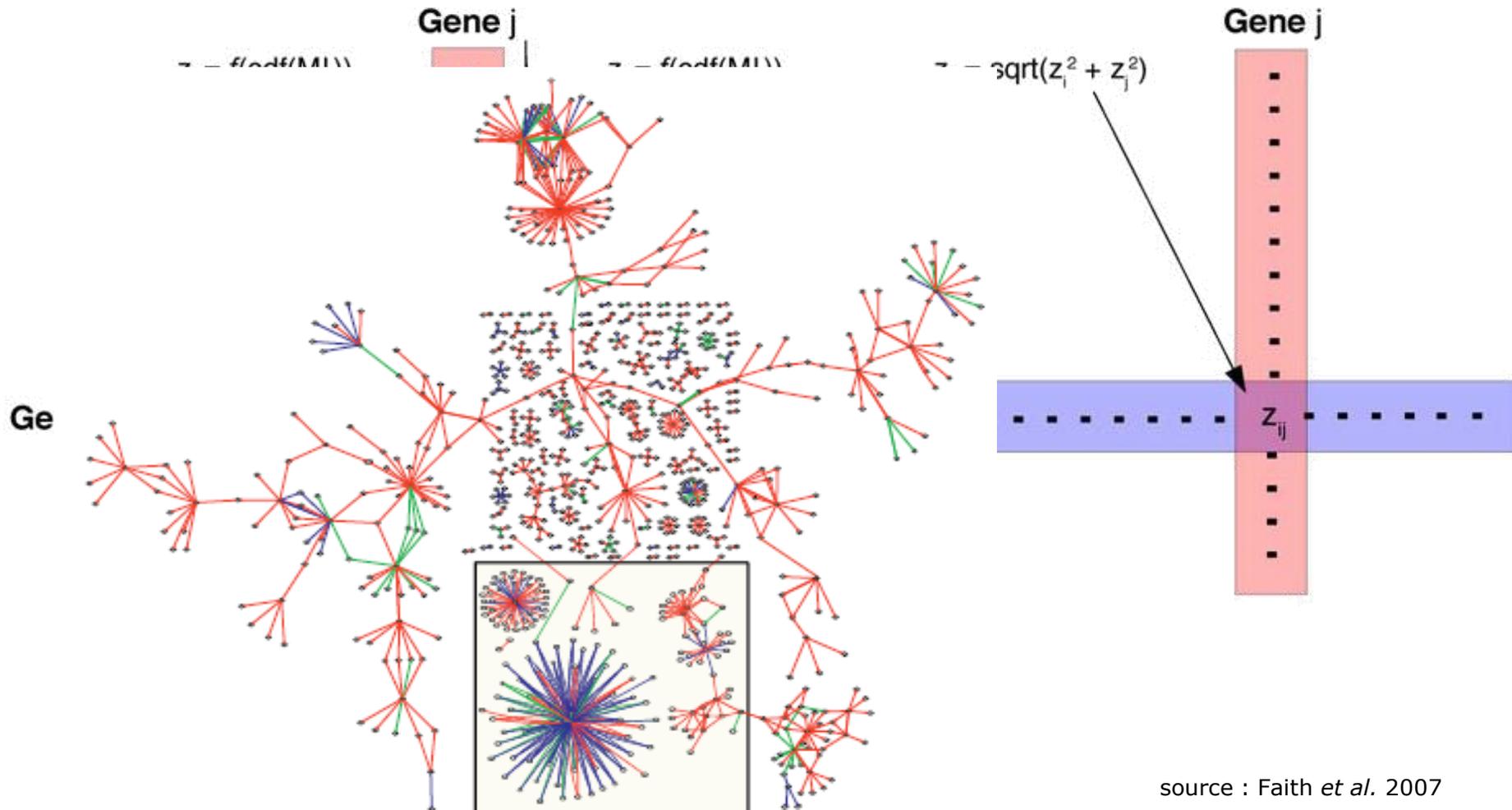
	cond1	cond2	cond3	cond4	cond5
A	high	low	high	low	low
B	high	low	high	high	low

	B = low	B = high	total
A = low	2 / 5	1 / 5	3 / 5
A = high	0	2 / 5	2 / 5
total	2 / 5	3 / 5	1

- Information Mutuelle

$$I(X; Y) = \sum_{i,j} P(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}$$

- Context Likelihood Ration (CLR)



## BMC Bioinformatics



Proceedings

Open Access

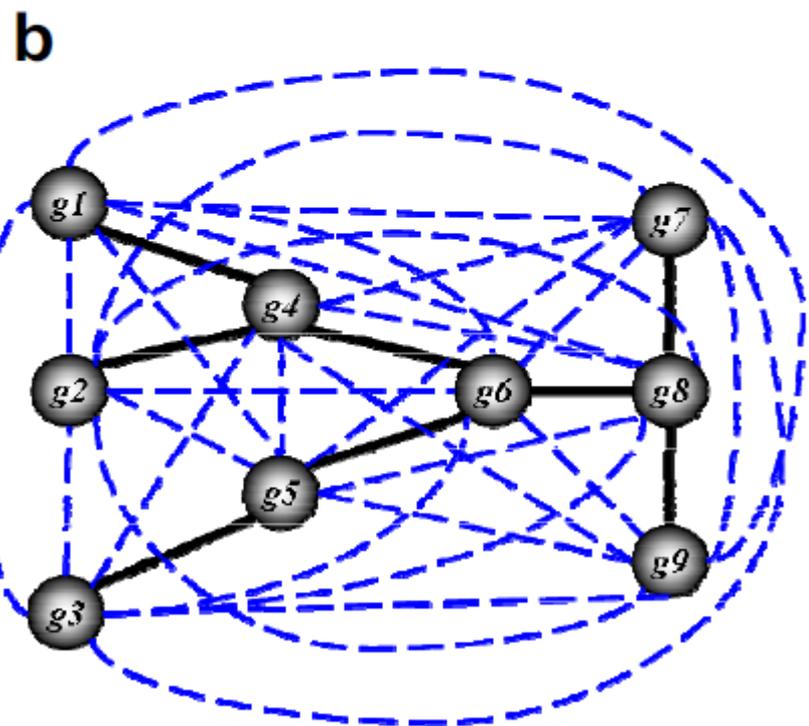
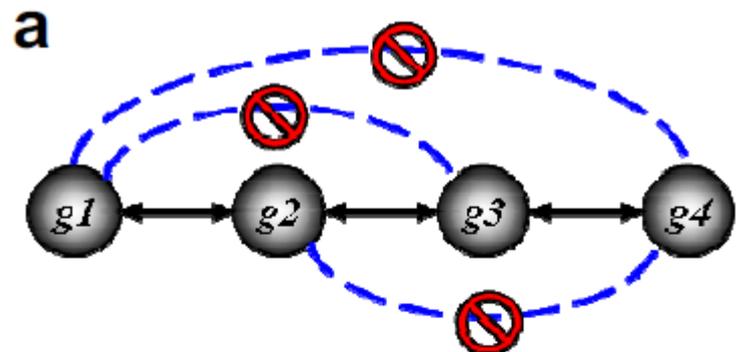
### ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context

Adam A Margolin<sup>1,2</sup>, Ilya Nemenman<sup>2</sup>, Katia Basso<sup>3</sup>, Chris Wiggins<sup>2,4</sup>, Gustavo Stolovitzky<sup>5</sup>, Riccardo Dalla Favera<sup>3</sup> and Andrea Califano<sup>\*1,2</sup>

#### Data Processing Inequality

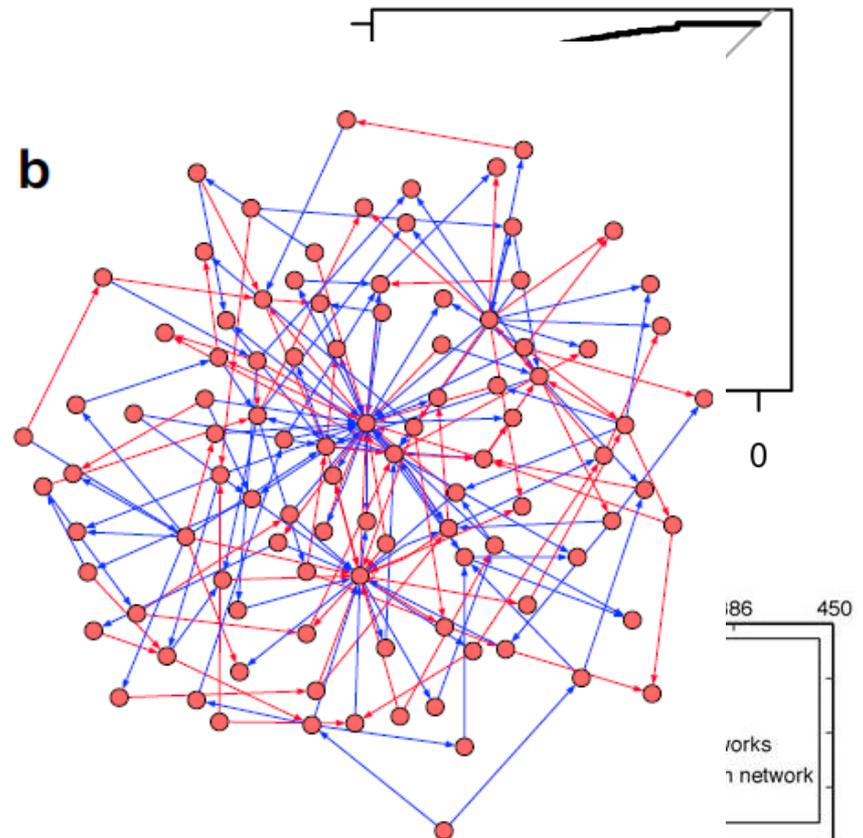
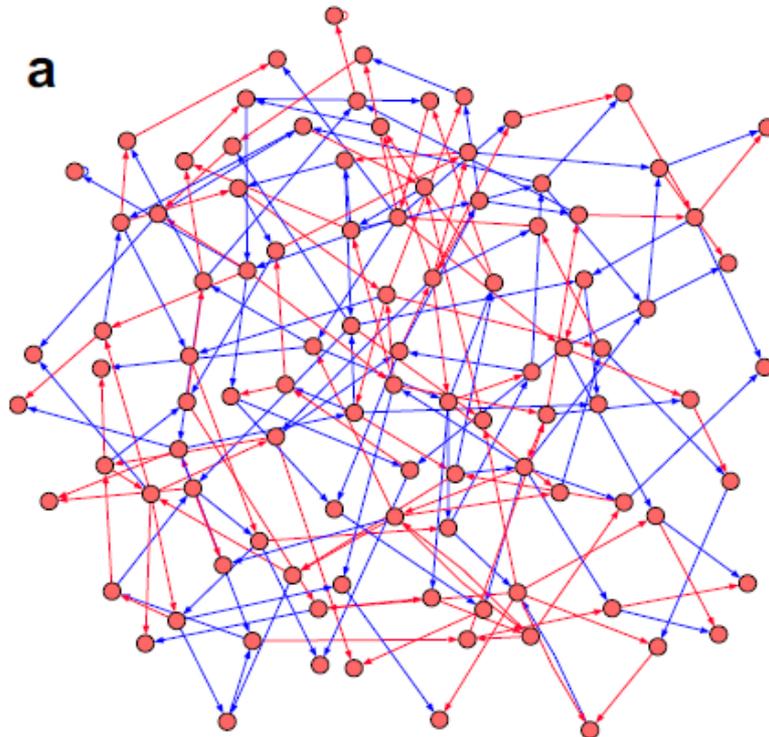
The DPI (Figure 2) [14] states that if genes  $g_1$  and  $g_3$  interact only through a third gene,  $g_2$ , (i.e., if the interaction network is  $g_1 \leftrightarrow \dots \leftrightarrow g_2 \leftrightarrow \dots \leftrightarrow g_3$  and no alternative path exists between  $g_1$  and  $g_3$ ), then

$$I(g_1, g_3) \leq \min [I(g_1, g_2); I(g_2, g_3)]. \quad (3)$$

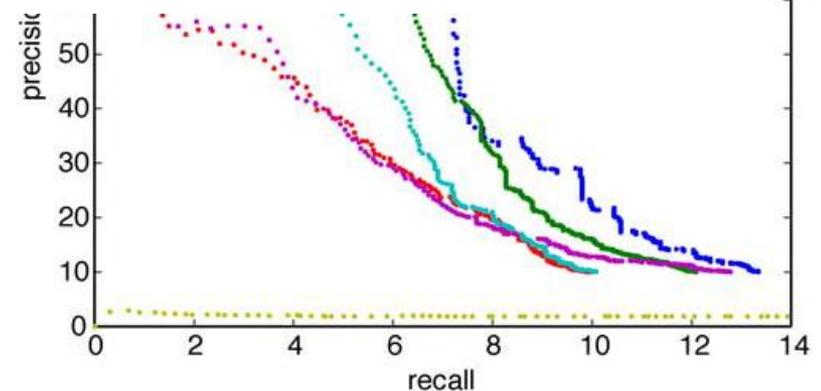


- Besoin d'un jeu de données de référence

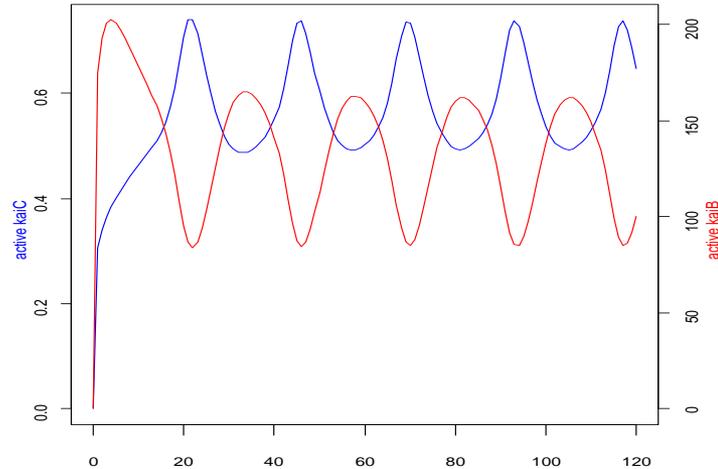
• T  
T  
li



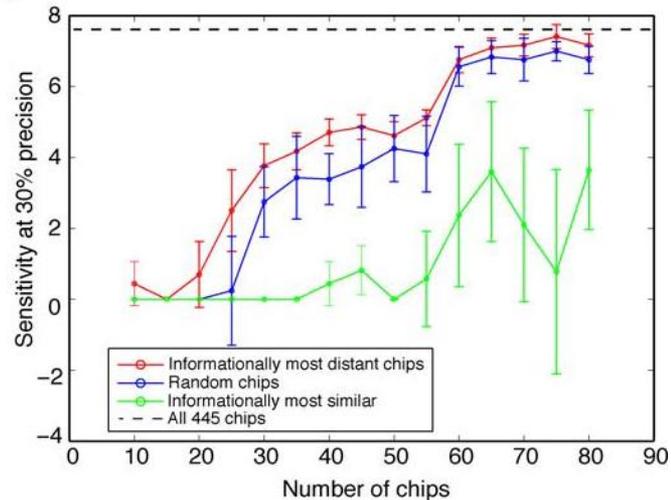
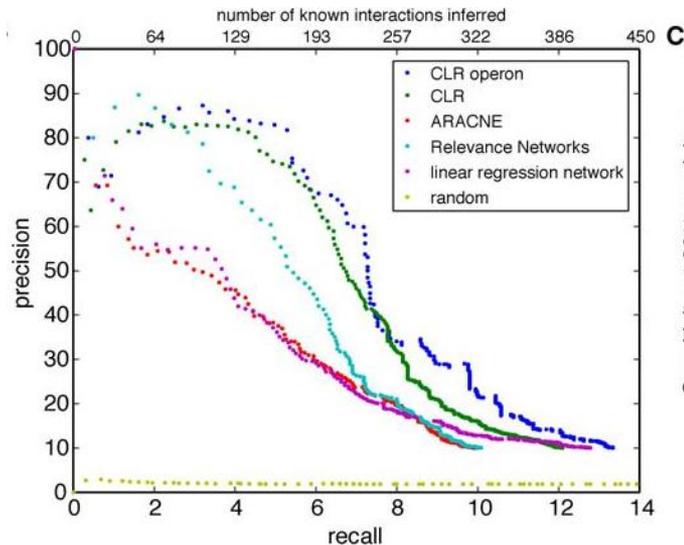
- ◆ Precision =  $IP / (IP+FP)$
- ◆ Spécificité =  $TN / (TN+FP)$
- ◆ FPR = 1 - spécificité
- ◆ FDR = 1 - précision



- Validation d'un modèle, d'un comportement observé



- Simulation et prédiction



## Controls

A  
LexA

	p-value	Motif location
recN	4.4e-09	
yebG	6.4e-10	
lexA	8.3e-10	
uvrA	1e-08	
sulA	1.7e-08	
dinI	2e-08	
dinP	4.4e-08	
recA	6.5e-08	
<b>SCALE</b>		1   25   50   75   100   125

The known motif is found in  
8 out of 13 promoters

## Putative novel regulons

C  
YnaE

	p-value	Motif location
cspB	3.9e-10	
cspG	6.3e-09	
b1374_s	1.1e-08	
cspH	3.7e-08	
b1459	2.9e-07	
rhsE	6.1e-07	
<b>SCALE</b>		1   25   50   75   100   125

A conserved motif is found in  
6 out of 8 promoters