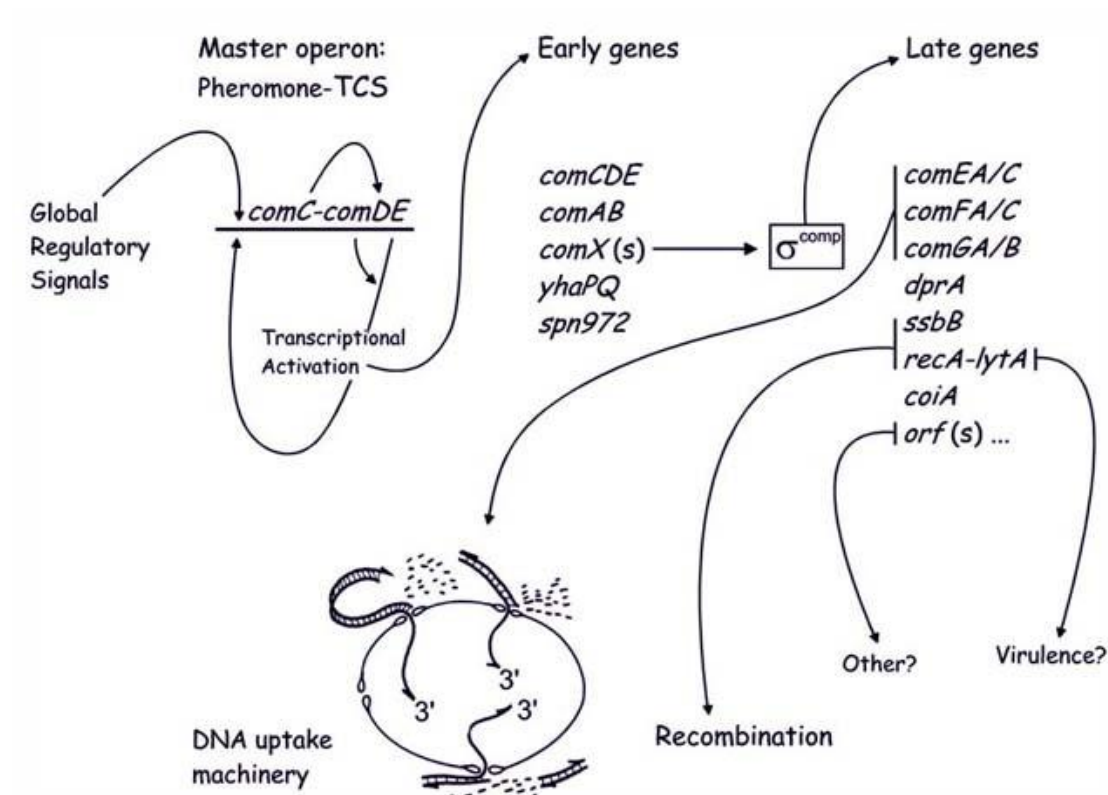


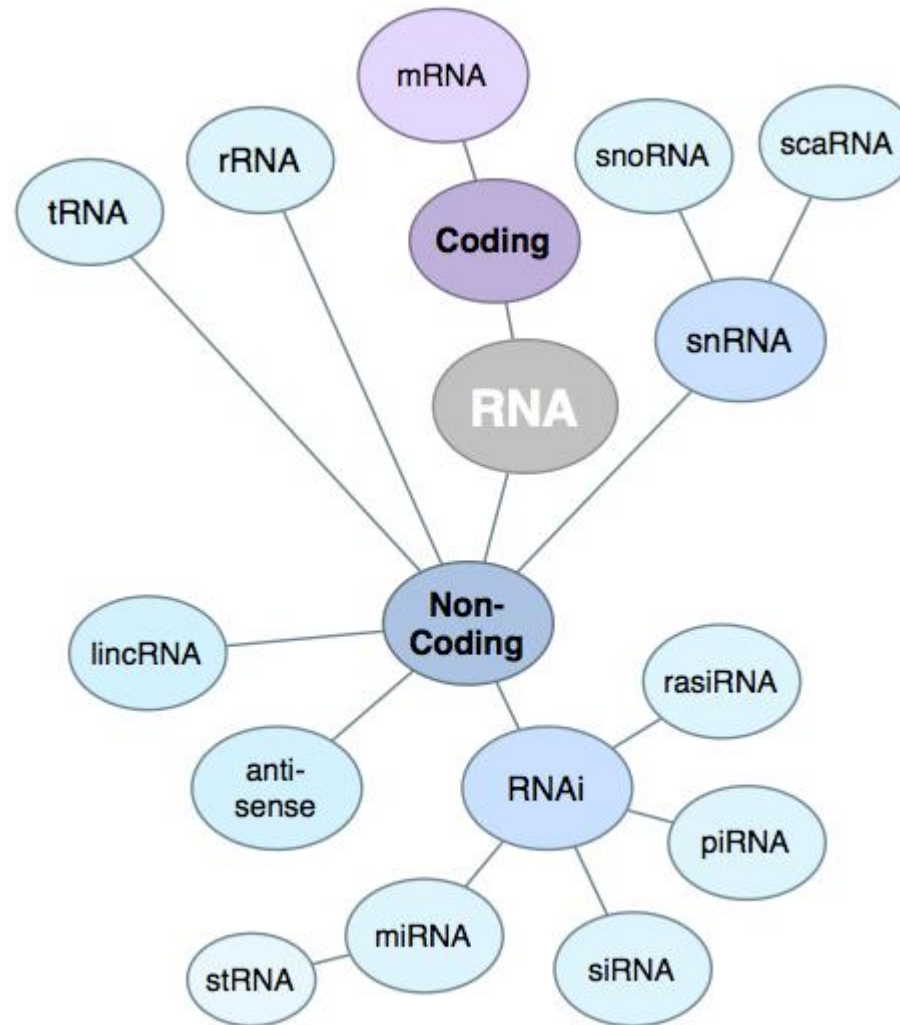
Réseaux de régulation

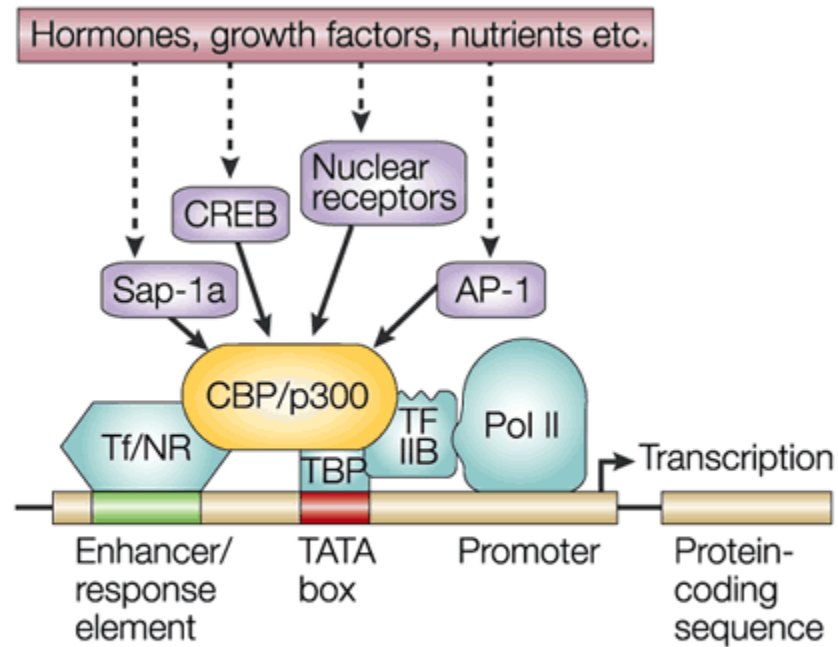
Master 1 MABS

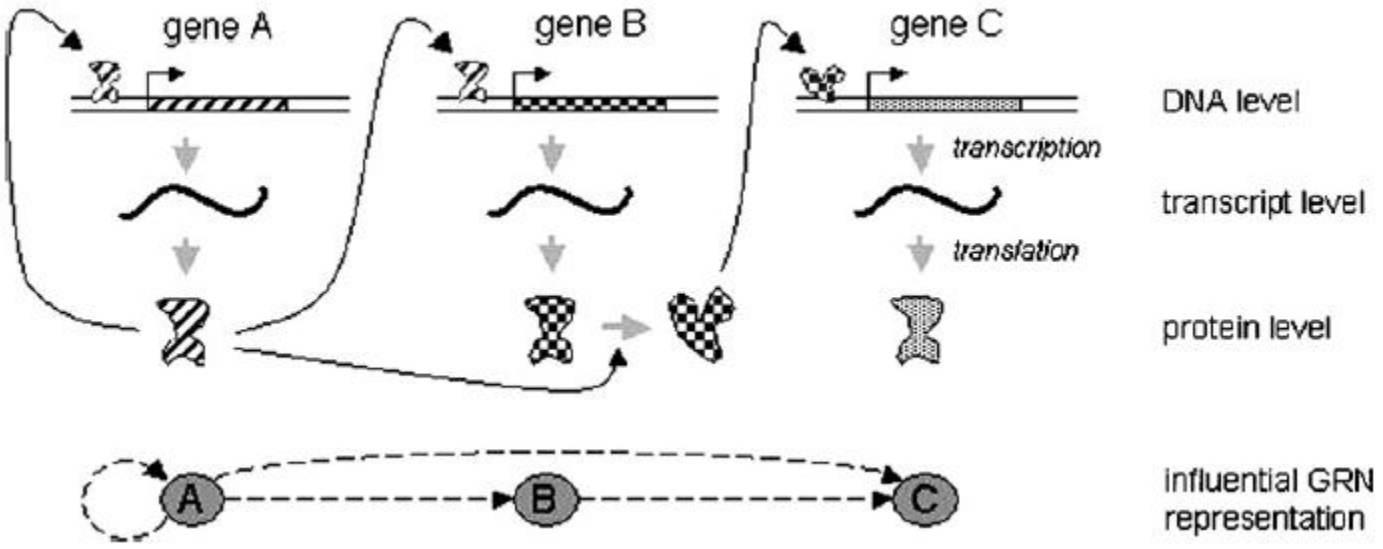


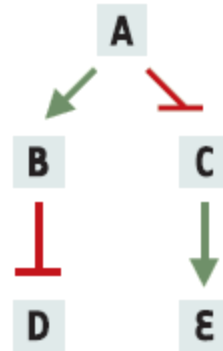
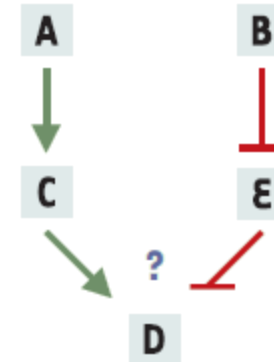
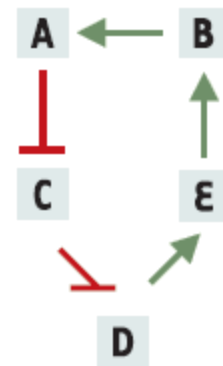
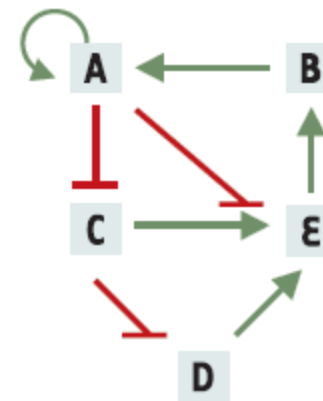
- Compréhension de propriétés du vivant conférées par les modalités d'expression du génome
- Expression
 - ◆ ARN codant et non codant
- Relation génotype/phénotype
 - ◆ et épigénétique : étude des changements, héritable au cours des divisions cellulaires, qui affectent la fonction des génomes sans altération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

RNA World





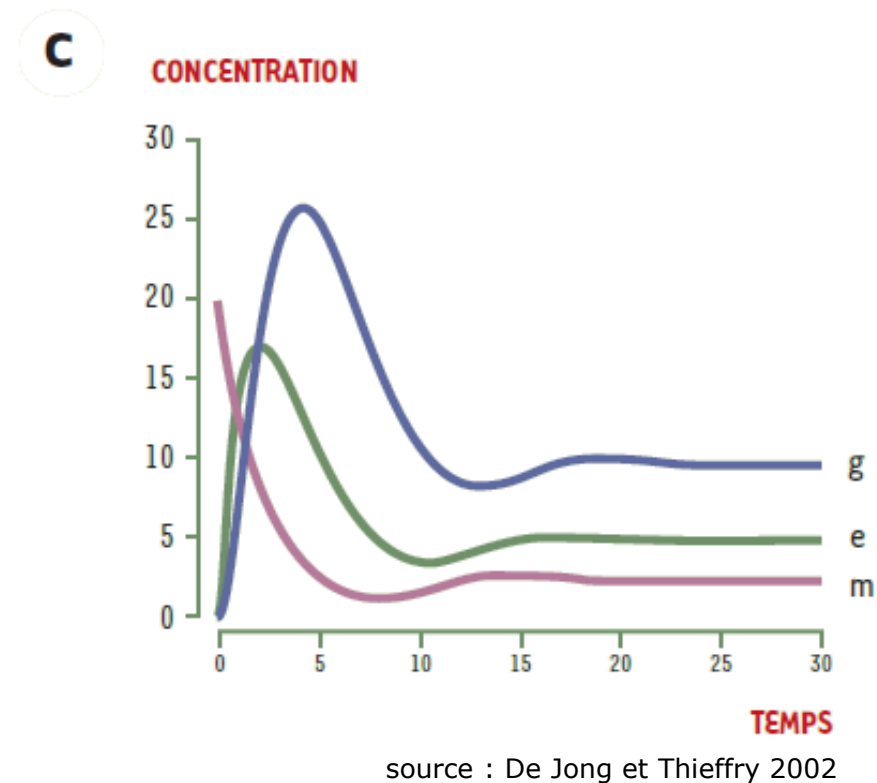
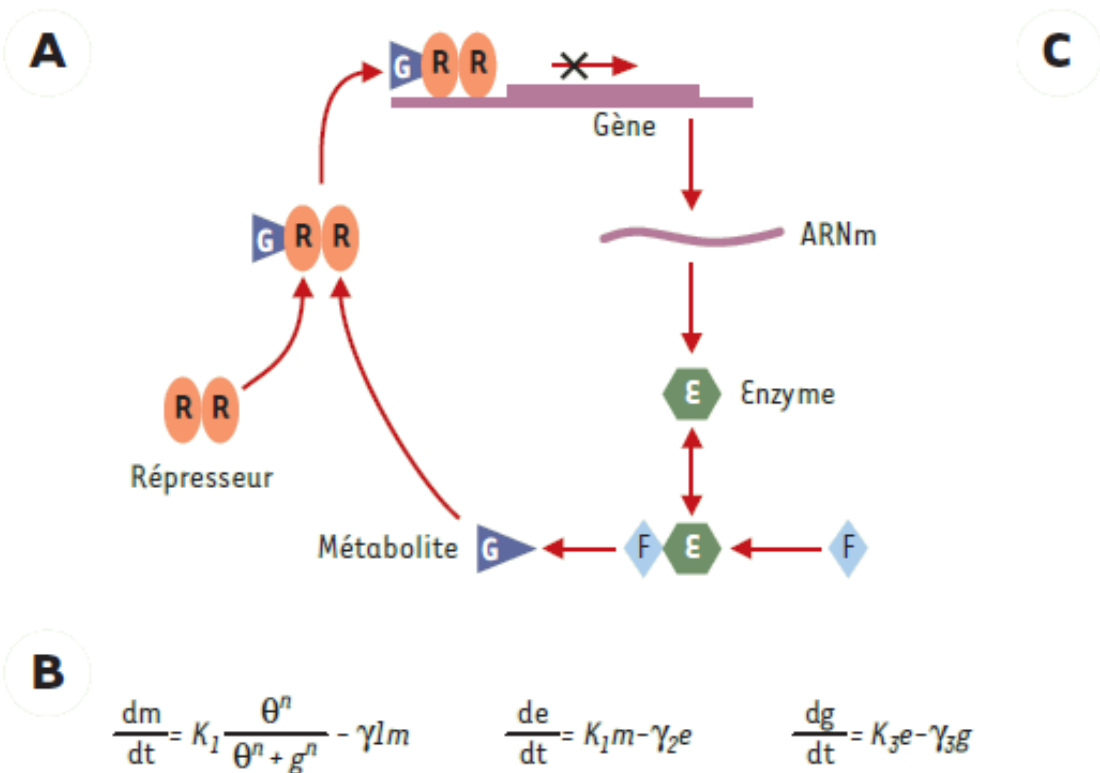


A**B****C****D****E**

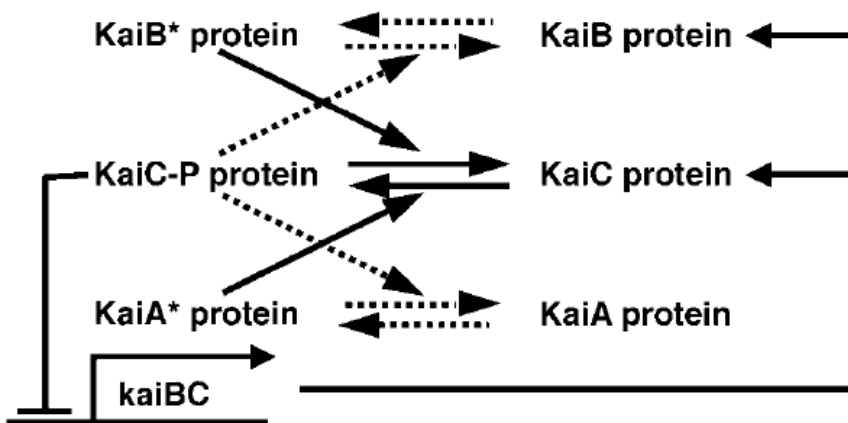
- Données d'expression
 - ◆ perturbations sur le phénomène d'intérêt
 - ◆ ensemble de conditions (>60) capturant un maximum de variations
- Données ChIP et ChIP-seq
 - ◆ localisation des sites de fixation des facteurs de transcription
- Facteurs de transcription connus
 - ◆ ex : RegulonDB, TransFac,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

- Systèmes d'équations
- Réseaux booléens
- Réseaux bayésiens
- Réseaux bayésiens dynamiques
- Réseaux d'influence

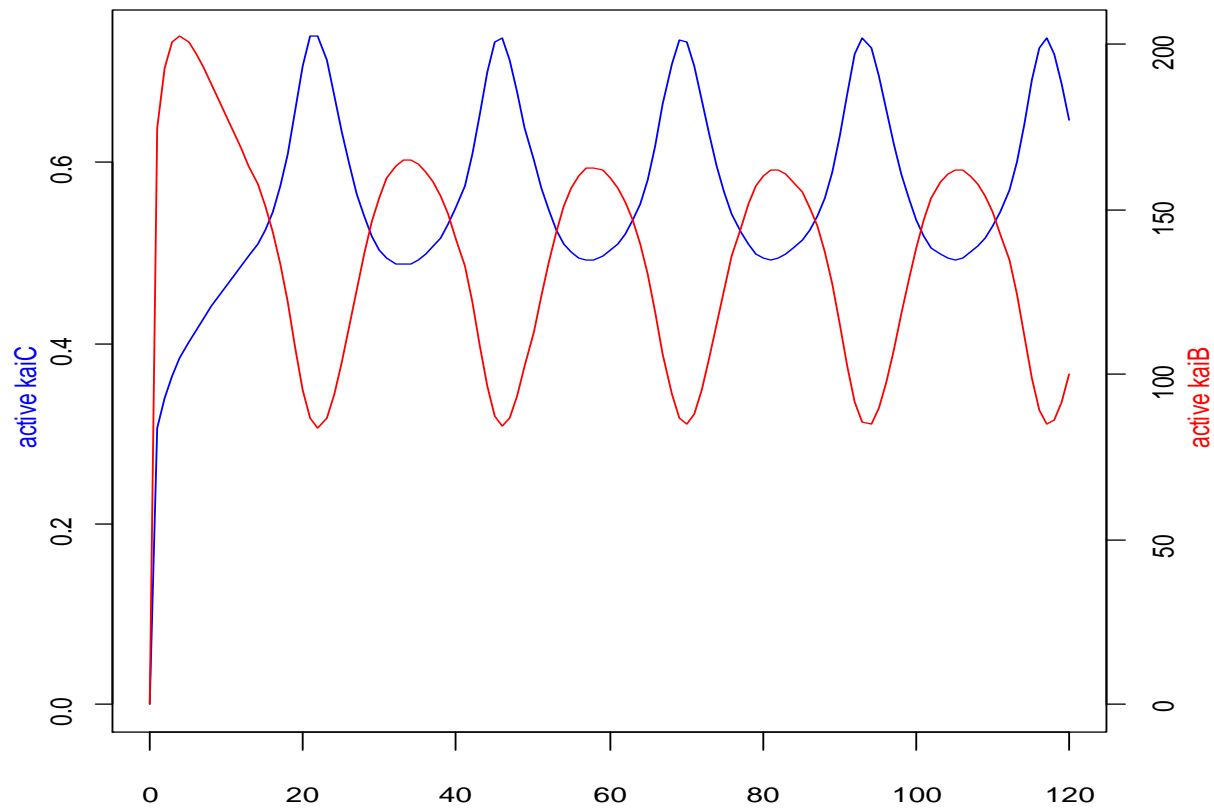
- Concentrations/activités des molécules
- Vitesse de production, dégradation
- Constante de seuil, association, dissociation, coopération



Exemple : oscillations circadiennes chez les cyanobactéries



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



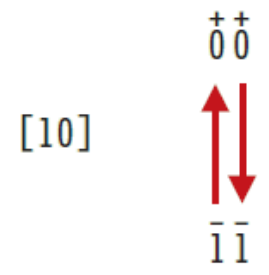
Synchrone

$$\begin{cases} x_{t+1} = \bar{y}_t \\ y_{t+1} = \bar{x}_t \end{cases}$$

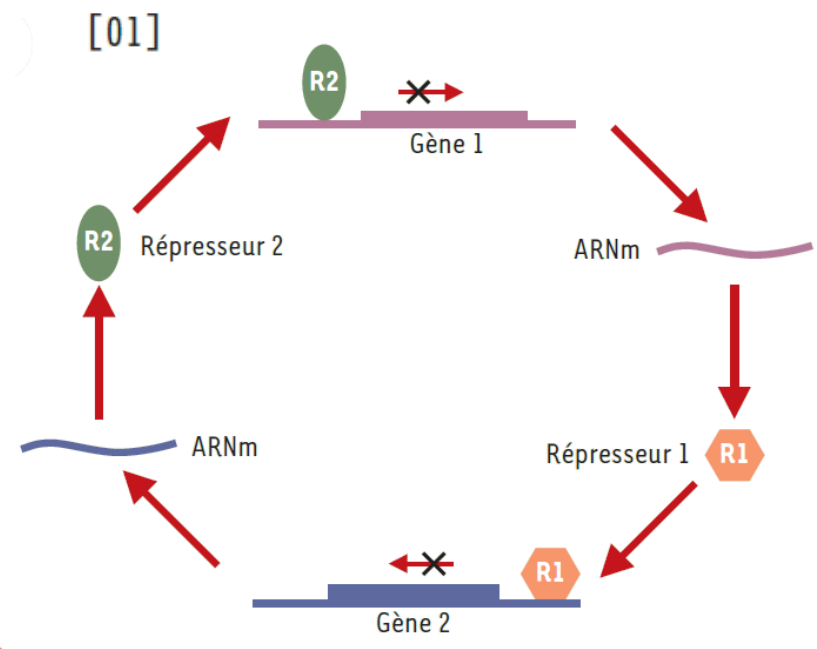
$(xy)_t$	$(xy)_{t+1}$
00	11
01	01
10	10
11	00

équations

table des états



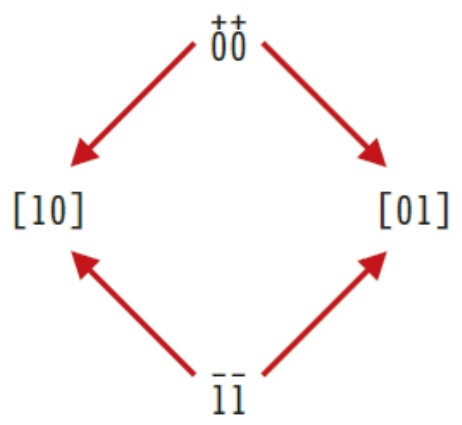
graphe des séquences d'états

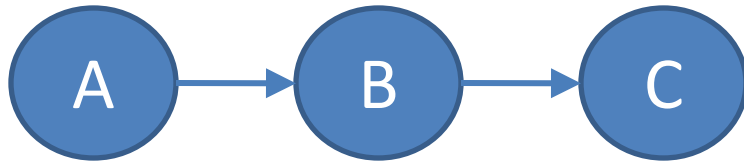


Asynchrone

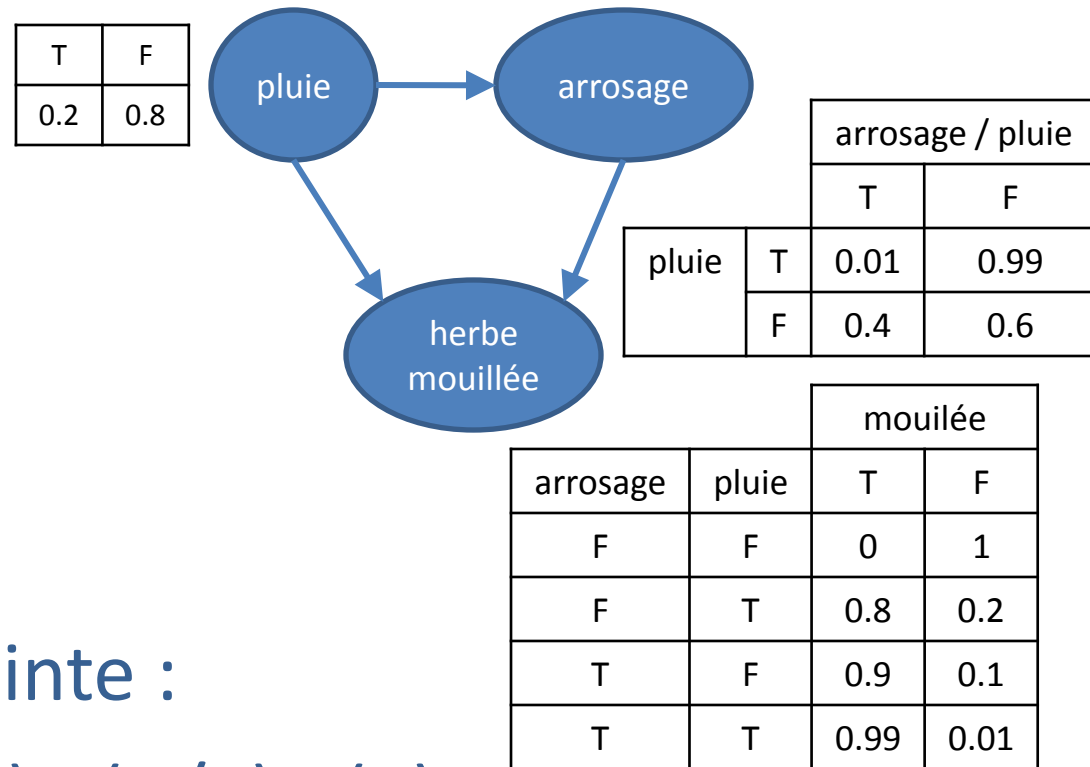
$$\begin{cases} x = \bar{y} \\ y = \bar{x} \end{cases}$$

xy	XY
00	11
01	01
10	10
11	00





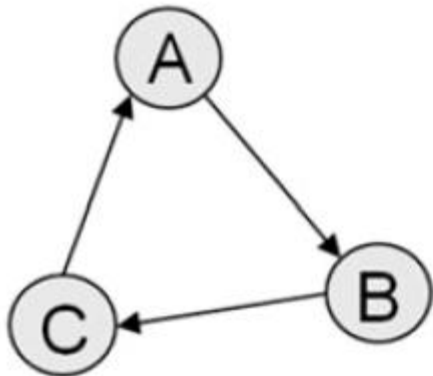
- Niveau d'expression de B dépend de celui de A (influence)
 - ♦ $P(B/A)$
- Le niveau d'expression de C
 - ♦ dépend de celui de B
 - ♦ est indépendant de celui de A sachant B
- Avantages :
 - ♦ capture l'aspect stochastique de la régulation
 - ♦ possibilité d'intégrer des régulations connues
 - ♦ peu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



- Loi de probabilité jointe :
 - ◆ $P(M,A,P) = P(M/A).P(A/P).P(P)$

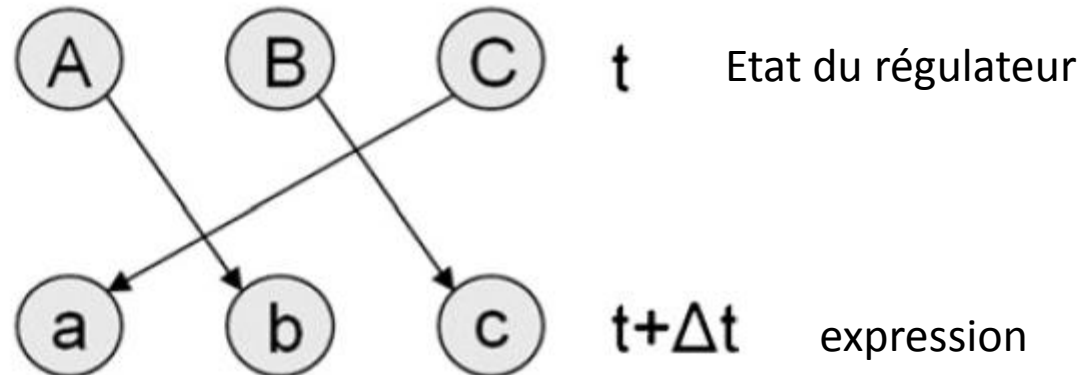
- Information temporelle (temps discret)
 - ◆ variable indicée par son pas de temps
 - ◆ distribution des probabilités d'une variable dépend de l'état de ses prédécesseurs au pas de temps précédent

Static BN



Not allowed !

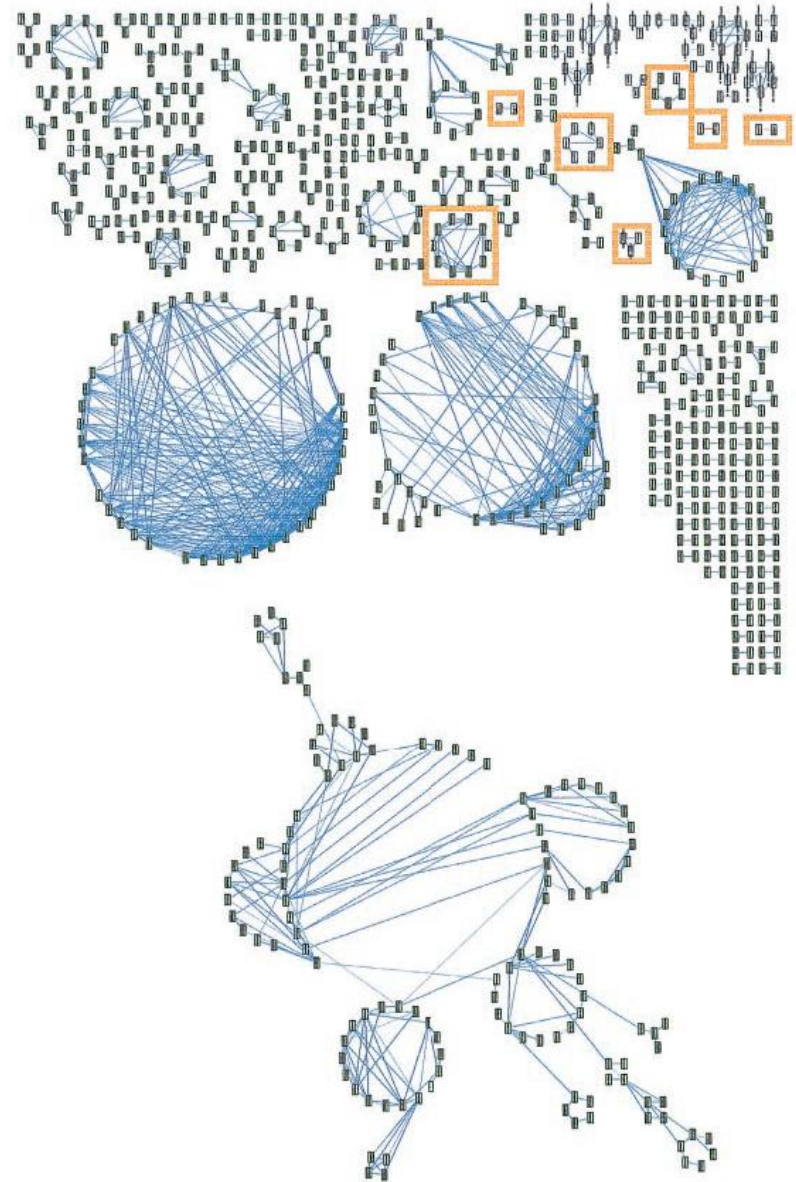
Dynamic BN



- 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

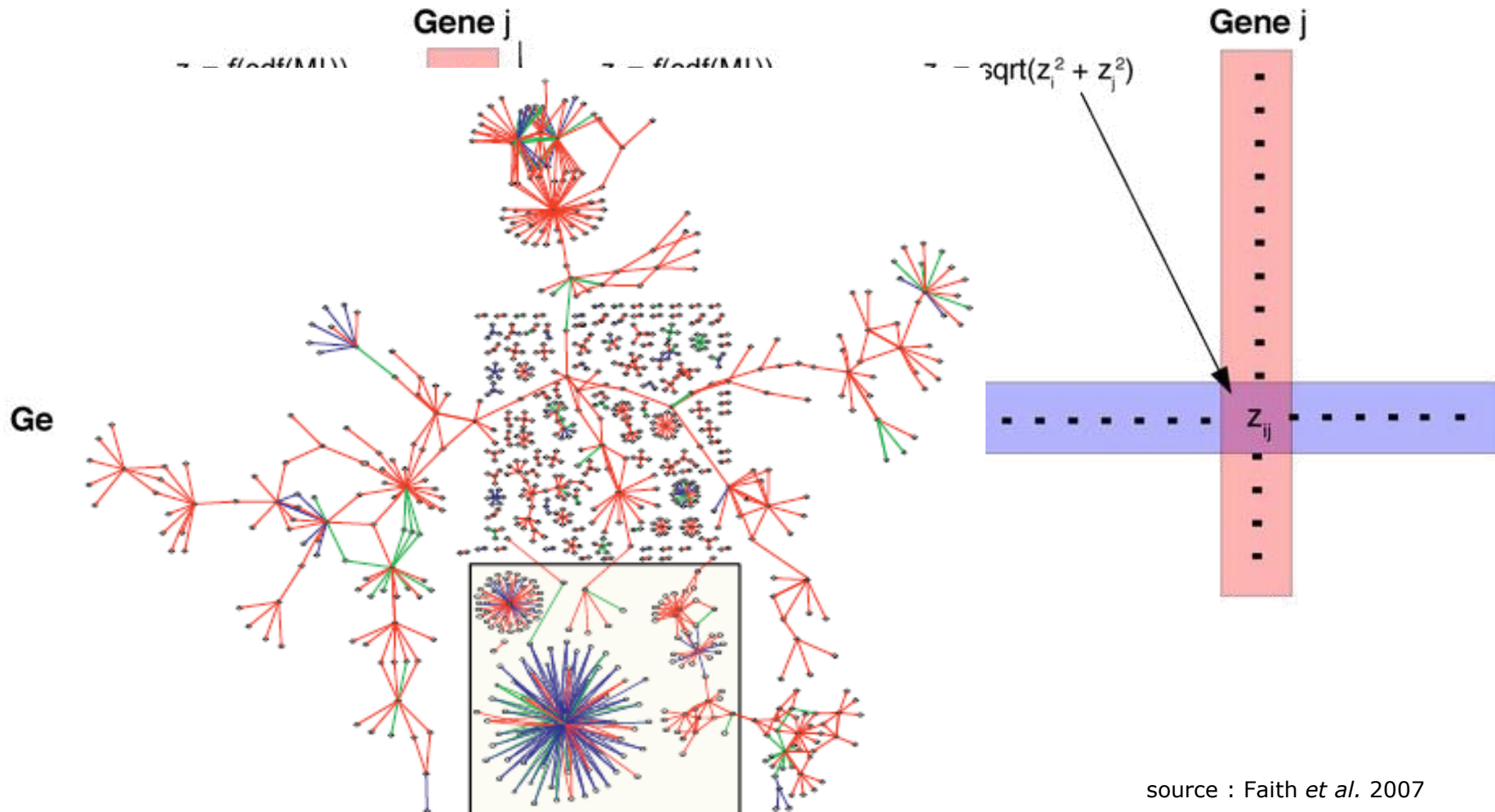


- fit entre observations et modèle
- Optimisation
 - ♦ de la structure ($2^n - 1$ connexions pour chaque sommet, BIC)
 - feature selection : considérer uniquement les gènes différentiellement exprimés
 - feature mapping : agréger les ensembles tels que les opérons ou les gènes co-exprimés impliqués dans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

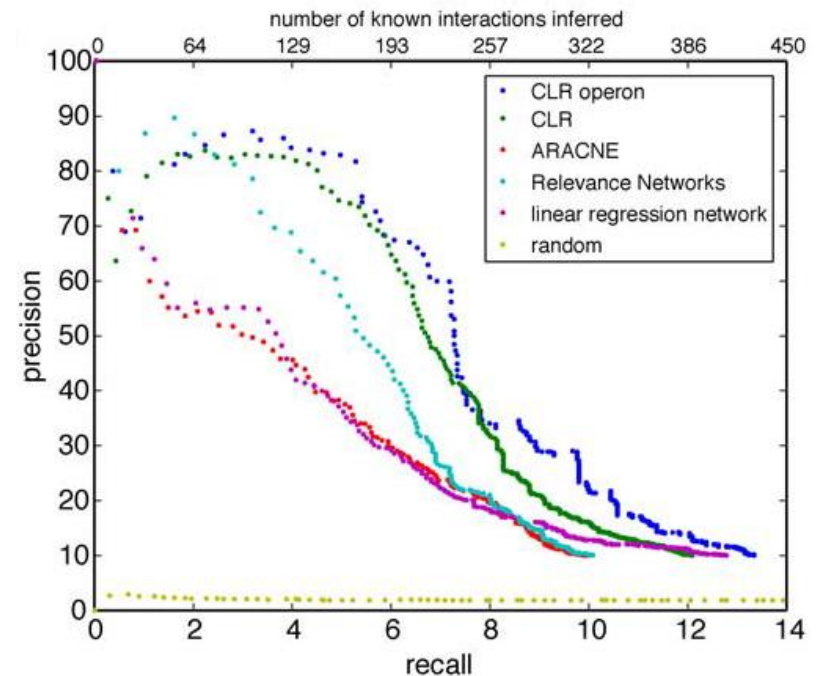
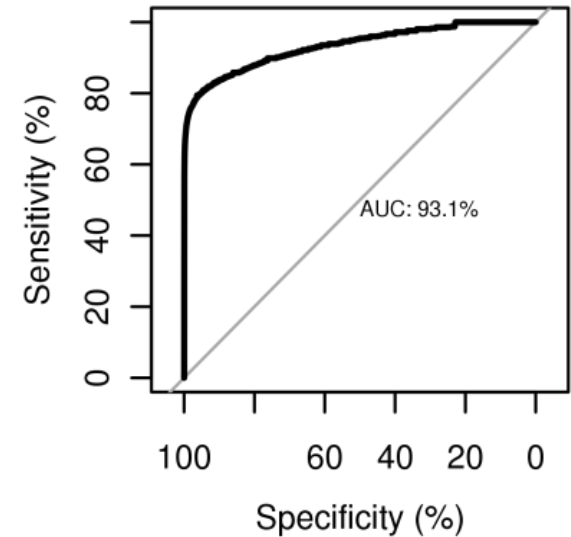
- Information Mutuelle

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

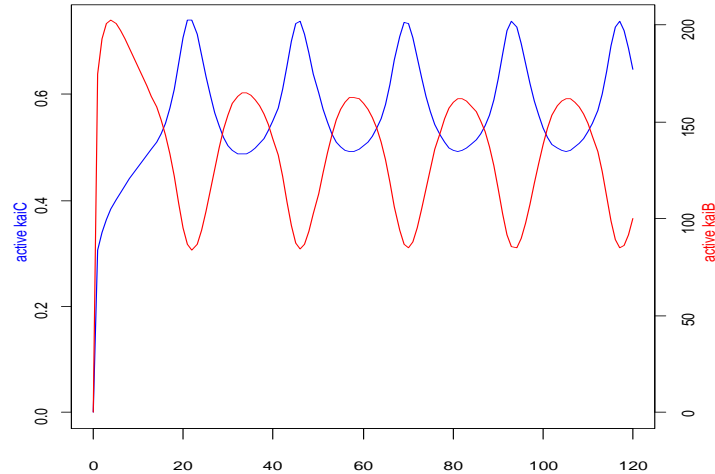
- Context Likelihood Ration (CLR)



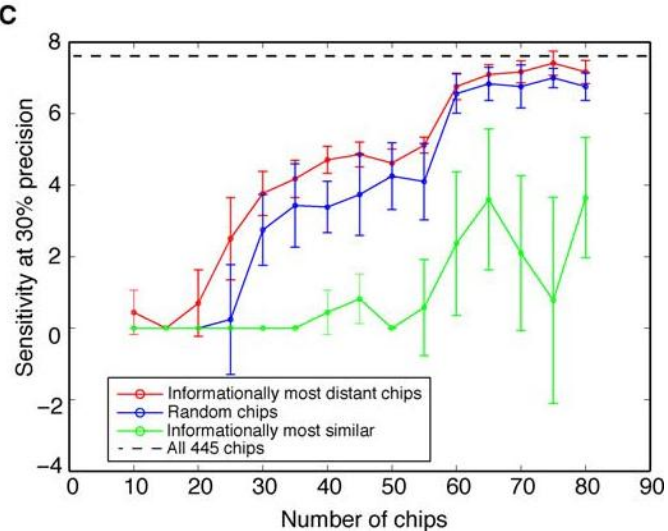
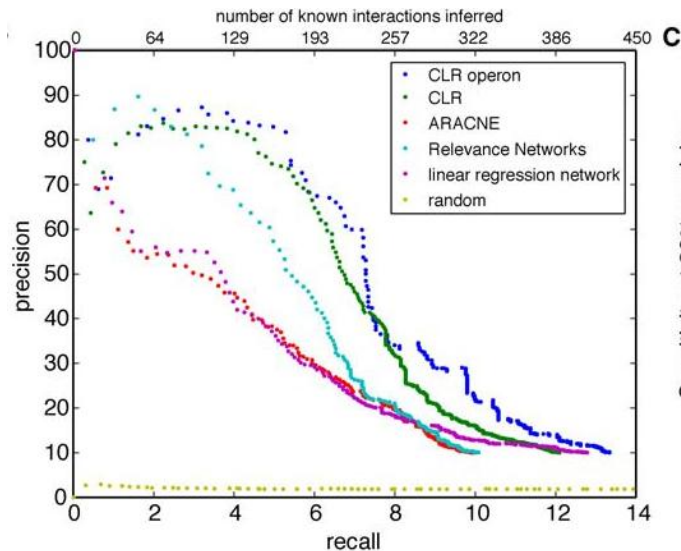
- Besoin d'un jeu de données de référence ou bien validation expérimentale des prédictions
 - ◆ souvent appel à des données synthétiques
 - ◆ RegulonDB pour *E. coli* ou SGD pour *S. cerevisiae*
- True Positive, False Positive, True Negative, False Negative links
 - ◆ Recall = sensibilité = $TP / (TP+FN)$
 - ◆ Précision = $TP / (TP+FP)$
 - ◆ 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	
SCALE		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	
SCALE		1 25 50 75 100 125

A conserved motif is found in
6 out of 8 promoters