

# Analyse des données transcriptômes

Anne Siegel

# Le problème

On dispose de...

- Un réseau biologique qualitatif: sommets  $\mathcal{X}$ , arêtes  $f(X, Y) = + (-)$  ssi une augmentation de  $X$  provoque une augmentation (diminution) de  $Y$ .
- Des données expérimentales sur un ensemble de sommets  $\mathcal{S}: xp(S_i) = +, -$ .

But: repérer ce qui est contradictoire entre le réseau qualitatif et les données expérimentales.

# Le problème

$$\begin{array}{ccccc} & - & & + & \\ AA & \longrightarrow & LXR & \longrightarrow & FAS \\ + & & + & & + \end{array}$$

Une erreur de  $A$  vers  $LXR$   
Une erreur de  $A$  vers  $FAS$

Si la première erreur est résolue,  
la second erreur aussi

$$\begin{array}{ccccc} & - & & + & \\ AA & \longrightarrow & LXR & \longrightarrow & FAS \\ + & & + & & - \end{array}$$

Une erreur de  $A$  vers  $LXR$   
Une erreur de  $LXR$  vers  $FAS$

Les deux erreurs sont  
indépendantes

Séparer les contradictions véritablement indépendantes de celle qui sont déduites des autres.

# Démarche

- Construire un graphe d'influence sur les cibles de l'expérimentation.
  - **Sommets**: différentes formes des cibles  $S$  de l'expérimentation (exemple:  $LXR$ -ARN et  $LXR$ -actif).
  - $infl_+(S, T)$ : nombre de chemins dont le produit des signes est positif qui vont de  $S$  vers  $T$  dans le réseaux qualitatif et qui ne passent par aucun sommet de  $S$ .
  - $infl_-(S, T)$  = nombre de chemins de signe  $-$  qui vont de  $S$  vers  $T$ ...
- **Contradiction entre  $A$  et  $B$** :  
 $xp(A) * \text{signe d'un chemin de } A \text{ vers } B \neq xp(B)$

# Dénombrer les contradictions

- $contr(B)$  nombre de chemins contradictoires arrivant en  $A$ .
- Contradictions transmises d'un sommet  $A$  à un de ses successeurs  $B$  :

$$contr(A) * infl_{xp(A)*xp(B)}(A, B)$$

- $B$  est un sommet porteur d'une contradiction ssi

$$contr(B) \neq \sum_{A \rightarrow B} contr(A) * infl_{xp(A)*xp(B)}(A, B)$$

## Exemple

	−		+	
<i>AA</i>	→	<i>LXR</i>	→	<i>FAS</i>
+		+		+
0		1		1

$$\text{infl}_{xp(LXR)*xp(FAS)}(LXR, FAS) = 1$$

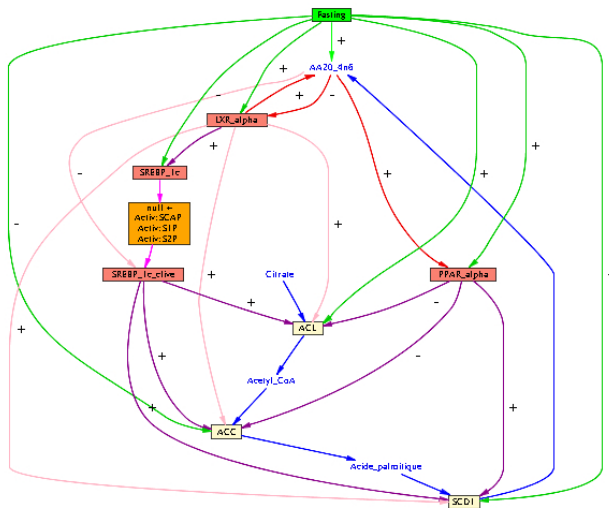
*LXR* est seul porteur d'une contradiction

	−		+	
<i>AA</i>	→	<i>LXR</i>	→	<i>FAS</i>
+		+		−
0		1		1

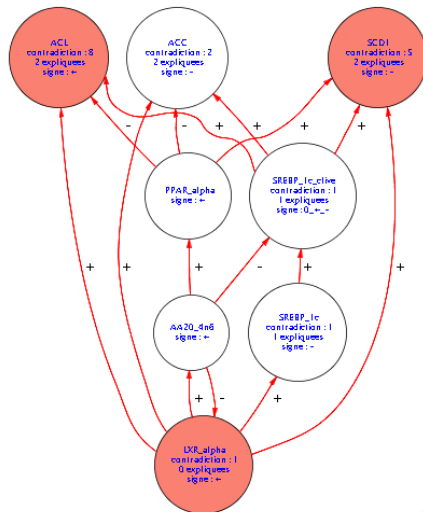
$$\text{infl}_{xp(LXR)*xp(FAS)}(LXR, FAS) = 0$$

*LXR* et *FAS* portent deux contradictions.

## Application: lipogénèse



# Application: lipogénèse





## Ce qu'on obtient

Les sommets pour lesquels une **contradiction** existe entre les données expérimentales et la bibliographie.

- Données fausses (ex: ACL)
- arêtes manquantes (ex:  $LXR \rightarrow LXR - \text{actif}$ )
- lieux de balances: toujours repérés.

### Restrictions

- Il faut connaître le signe des régulations  
(à calculer automatiquement en fonction des équations ?)
- Que se passe-t-il en cas de boucles ?
- Il faut suffisamment de données mais pas trop dans le réseau.  
(Identifier les variables à déterminer absolument, plan d'expérience ?)

## A éclaircir...

Utiliser les données pour **retrouver l'état du réseau le plus probable**:  
retrouver le signe des variables XP et éventuellement de leurs  
voisins en utilisant les connaissances sur la dynamique et sur l'état  
du réseau ?

