

Soutenance de thèse de Stéphanie Chevalier

29 septembre 2022

Inférence logique de réseaux booléens à partir de connaissances et d'observations de processus de différenciation cellulaire

Encadrement :

Christine FROIDEVAUX, Loïc PAULEVÉ, Andrei ZINOVYEV



Thèse préparée au

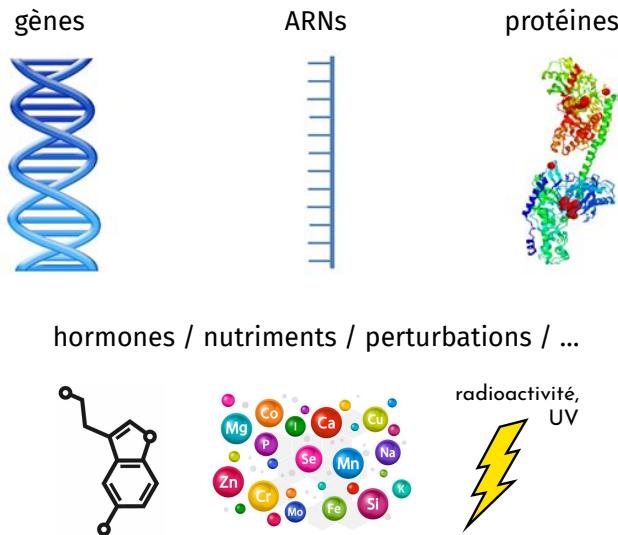


en collaboration avec



Comprendre les mécanismes de régulation d'un processus cellulaire

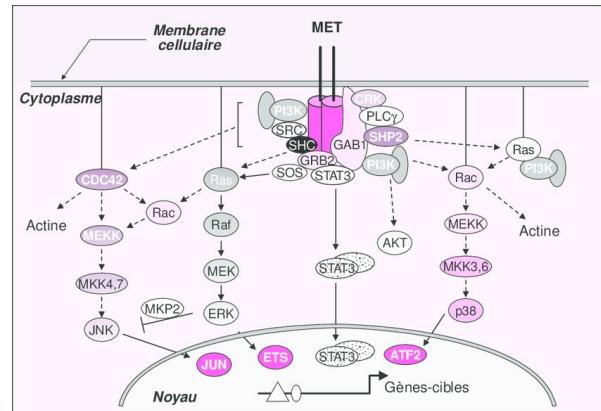
Résultat d'un **enchaînement d'interactions**
entre composants biologiques :



Réseaux d'interactions à la base
des comportements cellulaires

⇨ voies métaboliques, voies de signalisation ...

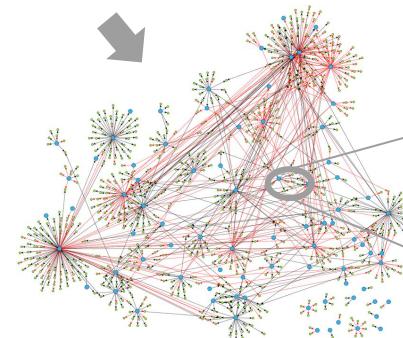
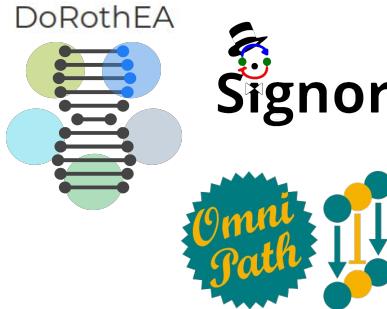
Exemple :



D. Tulasne et al., *L'hépatocyte growth factor/scatter factor et son récepteur MET*, Med Sci (Paris), 2001

Connaissances sur les interactions concernant un processus cellulaire

- Bases de données d'interactions**



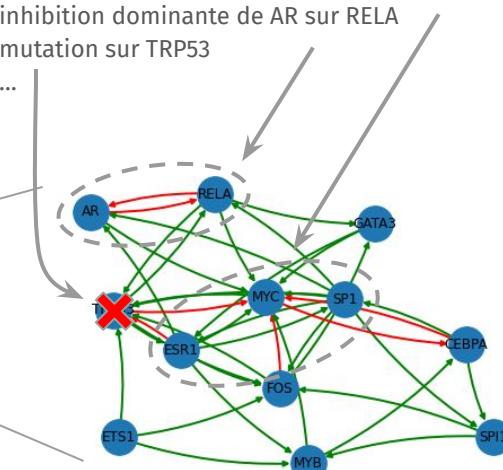
Graphe d'interactions (GI),
Prior Knowledge Network (PKN)

- Expertises**

Spécificité du contexte biologique ,
compétitions et coopérations connues , ...

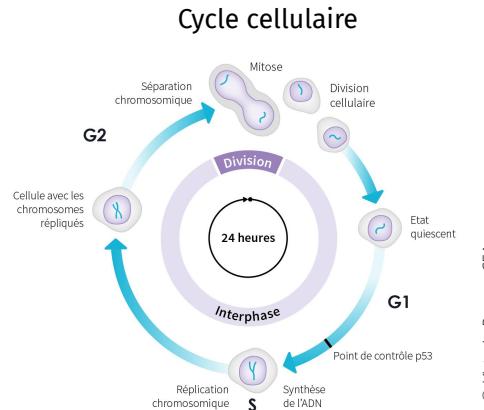
Exemple :

- coopération ESR1 et SP1 pour l'activation de MYC
- inhibition dominante de AR sur RELA
- mutation sur TRP53
- ...

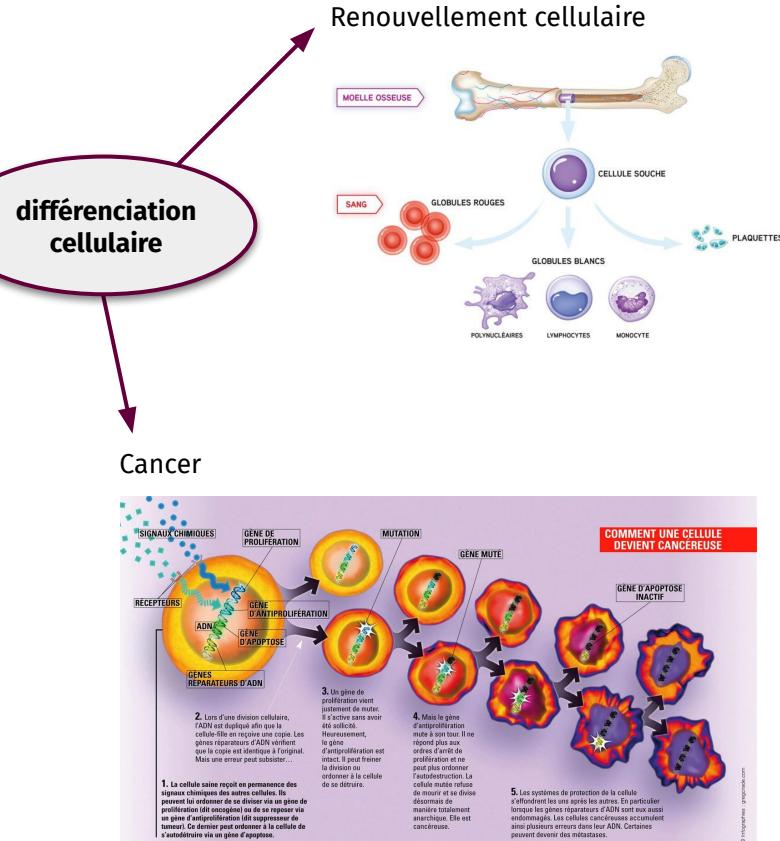
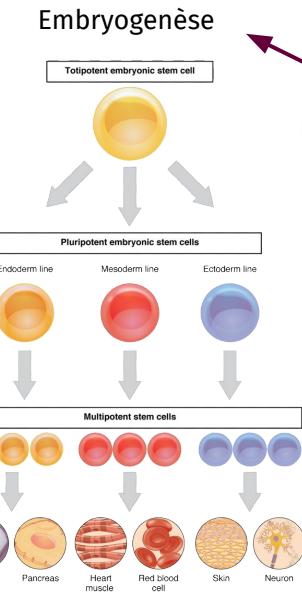
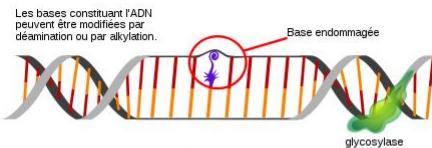


Exemples de processus cellulaires

Exemples de processus cellulaires :

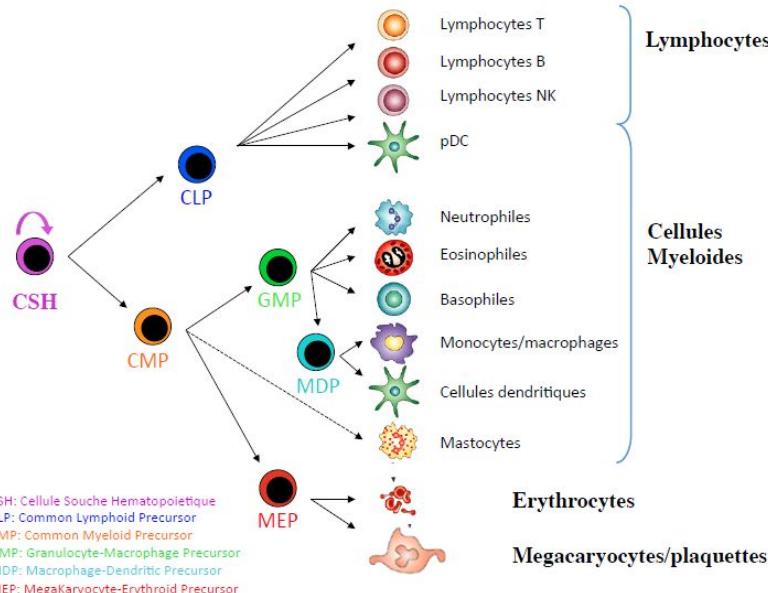


Réparation de l'ADN



Données contextuelles sur un processus cellulaire

- Propriétés dynamiques

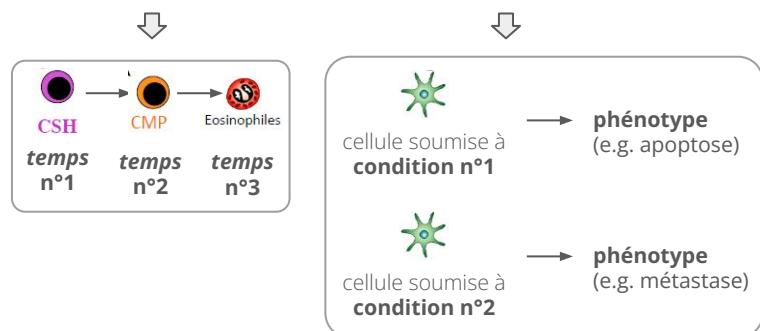


Processus de différenciation cellulaire
à partir d'une cellule souche hématopoïétique (CSH)

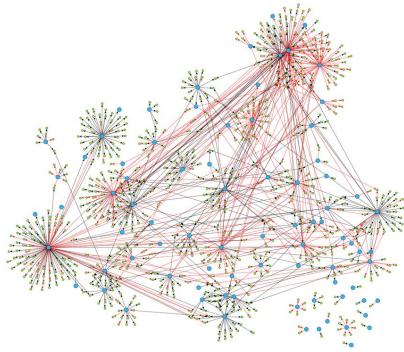
- Mesures expérimentales

CSH	gène A : <i>silencieux</i> gène B : <i>exprimé</i> gène C : <i>exprimé</i>	CMP	gène A : <i>silencieux</i> gène B : <i>silencieux</i> gène C : ?	Eosinophiles	gène A : exprimé gène B : exprimé gène C : <i>silencieux</i>
-----	--	-----	--	--------------	--

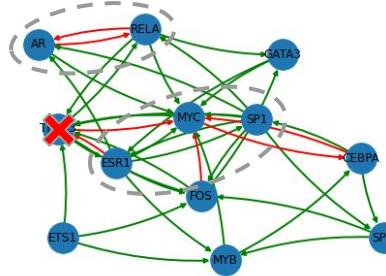
Données issues du séquençage des ARN,
sous la forme de séries temporelles ou de données de perturbation



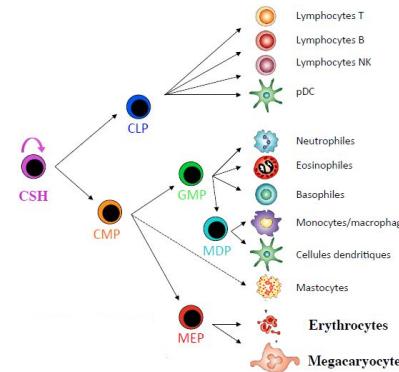
Comprendre les mécanismes de régulation d'un processus cellulaire



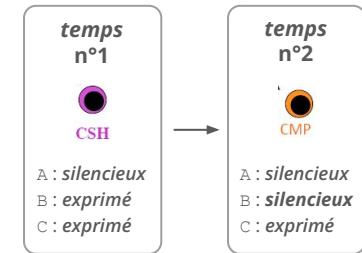
Prior Knowledge Network



Expertise



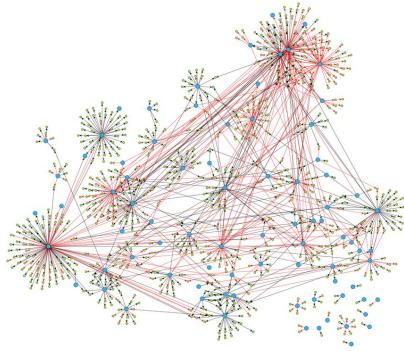
Propriétés dynamiques



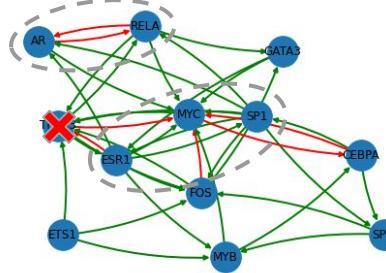
Observations

Quels sont les composants impliqués dans la régulation du processus observé, et comment interagissent-ils les uns avec les autres ?

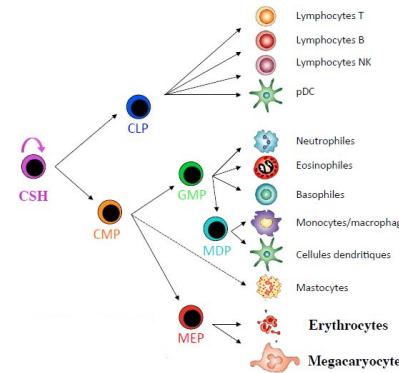
Comprendre les mécanismes de régulation d'un processus cellulaire



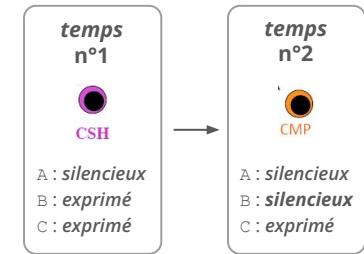
Prior Knowledge Network



Expertise

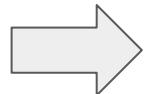


Propriétés dynamiques



Observations

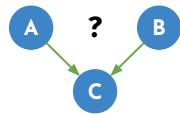
Quels sont les composants impliqués dans la régulation du processus observé, et comment interagissent-ils les uns avec les autres ?



Modélisation dynamique

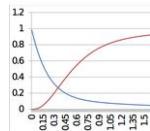
Créer un modèle d'un processus biologique

Reproduire *in silico* un processus :



exploration des mécanismes de régulation

tests d'hypothèses, découverte de nouveaux acteurs, ...



simulation du comportement cellulaire face à des perturbations

prédictions, recherche de cibles thérapeutiques, ...

2 familles de modèles dynamiques :

- **quantitatifs :**

représentation détaillée d'un système quantitatif
(évolution de quantité au cours du temps)

$$\frac{d(mRNA)}{dT} = k_{transcription} - (k_{mRNA\ degradation} \cdot mRNA)$$

- **qualitatifs :**

représentation abstraite des quantités
(changement de phénotypes / d'activité de gènes)

$$f_1(x) = \neg x_4$$

$$f_2(x) = \neg x_1 \wedge \neg x_3$$

$$f_3(x) = \neg x_2$$

Modélisation dynamique qualitative : Réseau booléen (RB)

Réseaux booléens très utilisés :

- adaptés à la **granularité** des connaissances
- adaptés à une grande **taille** de données
- facilement **intelligibles**
- peuvent modéliser des comportements cellulaires **complexes**
- **reproduisent** avec succès des résultats expérimentaux
- **prédisent** avec succès des comportements cellulaires

$$f_A(x) = x_B \wedge x_C$$

$$f_B(x) = x_A \vee x_C$$

$$f_C(x) = \neg x_B$$

Exemple de réseau booléen

Modélisation dynamique qualitative : Réseau booléen (RB)

Un réseau booléen (RB) de dimension n
 est une fonction $f: \{0,1\}^n \rightarrow \{0,1\}^n$
 telle que $\forall i \in \{1, \dots, n\}, f_i: \{0,1\}^n \rightarrow \{0,1\}$

Exemple d'un RB de dimension 3 :

$$f = (f_1, f_2, f_3)$$

$$\text{avec } f_1(x) := \neg x_2$$

$$f_2(x) := \neg x_1$$

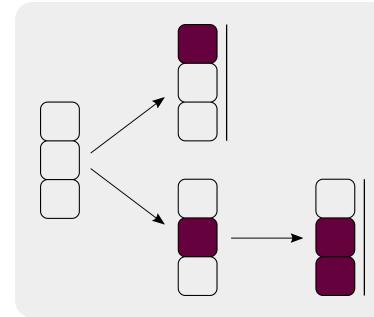
$$f_3(x) := \neg x_1 \wedge x_2$$

Une **configuration** est un vecteur $x \in \{0,1\}^n$

Exemples de configurations de taille 3 :

000, 001, 110, 010, etc...

Sémantique : manière de calculer l'évolution de l'état des composants



dynamique **asynchrone** de f

Trajectoire : séquence finie de configurations telles qu'on passe de l'une à l'autre par l'application d'une seule transition.

Modélisation dynamique qualitative : Réseau booléen (RB)

Un réseau booléen (RB) de dimension n est une fonction $f: \{0,1\}^n \rightarrow \{0,1\}^n$ telle que $\forall i \in \{1, \dots, n\}, f_i: \{0,1\} \rightarrow \{0,1\}$

Sémantique : manière de calculer l'évolution de l'état des composants

Exemple

$f = (f_1, \dots, f_n)$
avec

Comment créer un réseau booléen qui modélise un comportement cellulaire ?

dynamique asynchrone de l'

Une configuration est un vecteur $x \in \{0,1\}^n$

Exemples de configurations de taille 3 :

000, 001, 110, 010, etc...

Trajectoire : séquence finie de configurations telles qu'on passe de l'une à l'autre par l'application d'une seule transition.

Méthodes d'inférence de réseaux booléens modèles de comportements biologiques

Démarche usuelle : construction manuelle (raffinage successif par cycles essai-simulation)

Outils d'inférence automatique :

	CGA-BNI [Trinh and Kwon, 2021]	SgpNet [Gao et al., 2022]	TaBooN [Aghamiri and Delaplace, 2021]	CellINOpt-ILP [Gjerga et al., 2020]	BONITA [Palli et al., 2019]	ASKEed [Vaginay et al., 2021]	Caspo-ts [Ostrowski et al., 2016]	GAPORE [Liu et al., 2021]	BRE:IN [Goldfeder and Kugler, 2019]
<i>Données en entrée :</i>									
états stables	X	X	X						X
séries temporelles					X	X	X	X	(comportements descriptibles en CTL et LTL)
perturbations				X	X				
graphe d'interaction			X	X	X	X	X		X
<i>Propriétés dyn. RB :</i>									
points fixes	X	X	X	X	X				X
transition						X (asynchrone)		X (synchrone)	(synchrone et asynchrone)
atteignabilité				X (synchrone)	X (syn, asynchrone)		X (asynchrone)		
<i>Algorithme :</i>									
PLNE				X					
génétique	{ optimisation décision } X	X			X			X	
recherche tabou				X (étape 2, inférence du réseau booléen)					
problème de satisfiabilité	{ optimisation décision }			X (étape 1, inférence des fonctions locales)		X (ASP)	X (ASP)		X (SMT)

Applicabilité des méthodes d'inférence de l'état de l'art

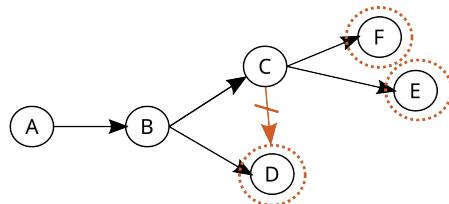
Comportements modélisables selon les outils :

- ensemble d'états stables
- séries temporelles, selon différentes interprétations
- tout comportement modélisable en logique temporelle (LTL et CTL) (BRE:IN)

Échelle des réseaux considérée : < 40 composants

Limites des méthodes d'inférence de l'état de l'art

Complexité du problème :



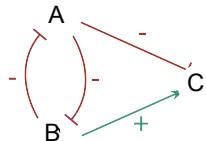
B est-il **atteignable** depuis A ?
(sémantique asynchrone)

problème PSPACE-complet

→ réponse très difficile à obtenir

Nombre de réseaux booléens à explorer :

Exemple :



Composant C

⇒ 6 règles possibles

$$\begin{array}{ll} f_C(x) = 0 & ; \quad f_C(x) = x_2 \\ f_C(x) = 1 & ; \quad f_C(x) = \neg x_1 \wedge x_2 \\ f_C(x) = \neg x_1 & ; \quad f_C(x) = \neg x_1 \vee x_2 \end{array}$$

Graphe d'interactions

⇒ $6 \times 3 \times 3 = 54$ RB possibles.

degré entrant	# fonctions booléennes monotones
0	2
1	3
2	6
3	20
4	168
5	7 581
6	7 828 354
7	2 414 682 040 998
8	56 130 437 228 687 557 907 788

→ **explosion combinatoire**

Limites des méthodes d'inférence de l'état de l'art



Asynchrone : perte de comportements possibles dans un système respectant la même logique de régulation

RB f :

$$f_A(x) = \neg x_B$$

$$f_B(x) = \neg x_A$$

$$f_C(x) = \neg x_A \wedge x_B$$

Raffinement multivalué de f :

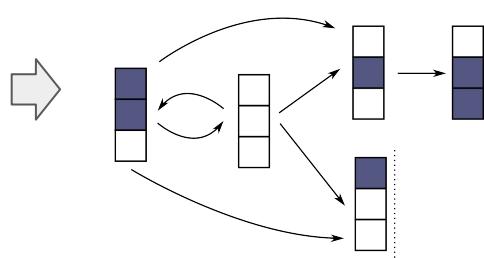
$$F_A(x) = +1 \text{ si } x_B < 2, \text{ sinon } -1$$

$$F_B(x) = +1 \text{ si } x_A < 2, \text{ sinon } -1$$

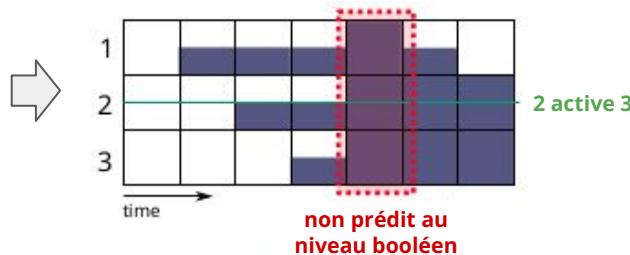
$$F_C(x) = +1 \text{ si } x_A < 2 \text{ et } x_B \geq 1, \text{ sinon } -1$$

GI de f :

Dynamique asynchrone de f :



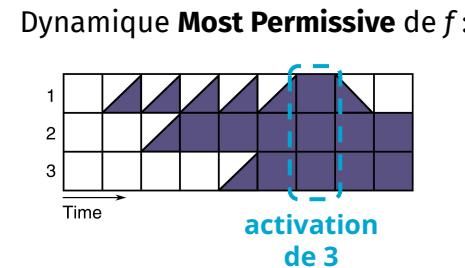
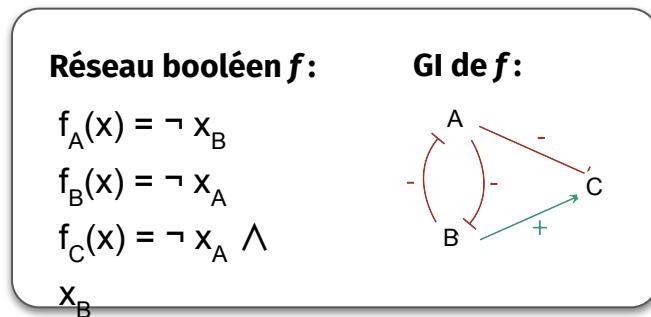
Dynamique asynchrone de F (multivalué):



Solution : Sémantique *Most Permissive* (MP)

Asynchrone : perte de comportements possibles dans un système respectant la même logique de régulation

⇒ **problème dû à la sémantique choisie et non à la nature binaire du formalisme**



- **Garanties formelles** sur les comportements prédits vis-à-vis de modèles multivalués et continus
- **Complexité réduite** pour vérifier des propriétés dynamiques jusqu'alors difficilement abordables



Challenge

Inférence automatique de réseaux booléens avec :

- observations et connaissances sur
plusieurs milliers de gènes

- ensemble complexe des
propriétés dynamiques d'une différenciation

- dynamique du réseau booléen en
sémantique *Most Permissive*

Mes contributions

Problème 1 :

Propriétés d'un réseau booléen modèle de différenciation cellulaire ?

Problème 2 :

Méthode pour inférer des modèles à partir de données de différenciation ?

Problème 3 :

Applicabilité sur données réelles ?

Problème 1 :

Propriétés d'un réseau booléen modèle de différenciation cellulaire ?

- Modélisation logique de propriétés dynamiques élémentaires
- Formalisation de la compatibilité réseau booléen \leftrightarrow propriétés

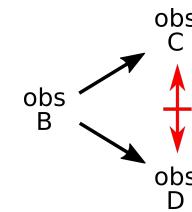
Propriétés dynamiques des observations

Propriétés élémentaires :

- **trajectoire** : ordre entre les observations (série temporelle)

$$\text{obs}_A \longrightarrow \text{obs}_B$$

- **bifurcation** : évolutions différentes et irréversibles à partir d'un état initial



- **stabilité (totale ou partielle)** : états cellulaires stables
(de l'ensemble ou d'une partie des composants)



Combinaison : **arbre de différenciation** pour représenter le processus de différenciation cellulaire

Exemples d'objets définis

Définition :
Observation

Système à n composants, une observation o avec $1 \leq k \leq n$ composants :
 $o = \{ (i_1, v_1), (i_2, v_2), \dots, (i_k, v_k) \}$ avec $1 \leq i \leq n$, $v \in \{0,1\}$, $\nexists i : \{ (i,0), (i,1) \} \subseteq o$.

Exemple :

- **observation** d'un système à 4 composants :
 $o = \{(2,1), (3,1)\}$

composant
 n° 1 2 3 4
 ||||
 11

- **configurations de taille 4 compatibles avec o** :
 $0110, 0111, 1110, 1111$

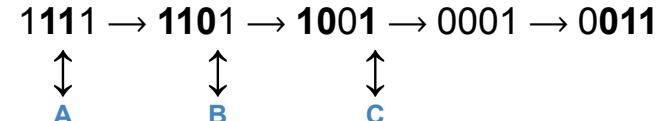
Définition : **Trajectoire compatible avec une liste d'observations**

f un RB avec sémantique σ , $S = (o_1, \dots, o_m)$ une liste d'observations.
Une trajectoire $T_{\sigma}^f = (x_1, \dots, x_l)$ **est compatible avec S** $\Leftrightarrow \exists i_1, \dots, i_m \in \{1, \dots, l\}$
 avec $i_1 \leq i_2 \leq \dots \leq i_m : \forall k \in [1, \dots, m], x_{ik} \text{ est compatible avec } o_k$.

Exemple : • liste d'observations S :



• trajectoire compatible avec S :



Compatibilité existentielle avec une liste d'observations

Existentielle :

observation d'un comportement possible du système.

Définition existentielle compatibilité réseau booléen avec liste d'observations :

RB existuellement compatible

avec $S = (o_1, o_2, o_3, o_4)$ ssi

$\exists c_1$ compatible avec o_1 et

$\exists T_\sigma^f$ commençant par $c_1 : T_\sigma^f$ compatible avec S .

$$S = (A, B, C, D)$$

A) composant
n° 1 2 3 4

11

B) composant
n° 1 2 3 4

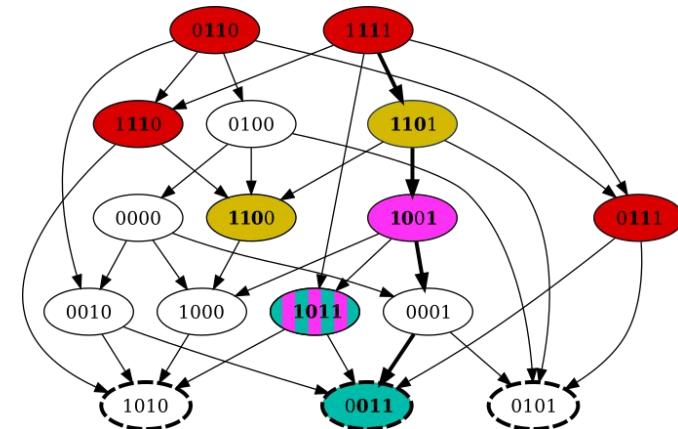
110

C) composant
n° 1 2 3 4

10 1

D) composant
n° 1 2 3 4

011



graphe de transition $f_1(x) = \neg x_4$
 $\text{def } f: f_2(x) = \neg x_1 \wedge \neg x_3 \wedge x_4$
 $f_3(x) = \neg x_2$
 $f_4(x) = \neg x_1$

f existuellement compatible avec L :

À partir de 1111 compatible avec A,

\exists au moins une trajectoire compatible avec S.

Compatibilité universelle avec une liste d'observations

Universelle :

observation du seul comportement possible du système.

Définition universelle compatibilité réseau booléen avec liste d'observations :

RB universellement compatible

avec $S = (o_1, o_2, o_3, o_4)$ ssi

$\exists c_1$ compatible avec o_1 :

$\forall T_\sigma^f$ commençant par c_1 ,

- soit T_σ^f compatible avec S ,
- soit $\exists T'_\sigma^f$ suffixe de T_σ^f compatible avec S .

$$S = (A, B, C, D)$$

A) composant
n° 1 2 3 4

$$\begin{matrix} & 1 \\ & 1 \end{matrix}$$

B) composant
n° 1 2 3 4

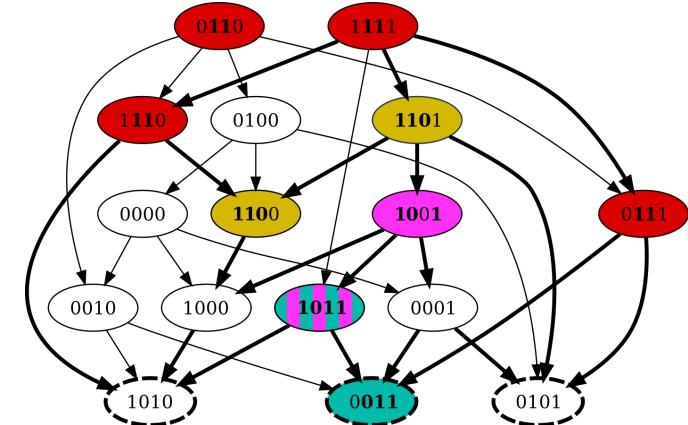
$$\begin{matrix} & 1 \\ & 1 \\ & 0 \end{matrix}$$

C) composant
n° 1 2 3 4

$$\begin{matrix} & 1 \\ & 0 \\ & 1 \end{matrix}$$

D) composant
n° 1 2 3 4

$$\begin{matrix} & 0 \\ & 1 \\ & 1 \end{matrix}$$



graphe de transition $f_1(x) = \neg x_4$
 $\text{def } f: \quad f_2(x) = \neg x_1 \wedge \neg x_3 \wedge x_4$
 $f_3(x) = \neg x_2$
 $f_4(x) = \neg x_1$

f n'est pas universellement compatible avec S :

\forall configuration compatible avec A,

\exists une trajectoire ni compatible avec S,
ni préfixe d'une trajectoire compatible avec S.

Compatibilité universelle avec une liste d'observations

Universelle :

observation du seul comportement possible du système.

Définition universelle compatibilité réseau booléen avec liste d'observations :

RB universellement compatible

avec $S = (o_1, o_2, o_3, o_4)$ ssi

$\exists c_1$ compatible avec o_1 :

$\forall T_\sigma^f$ commençant par c_1 ,

- soit T_σ^f compatible avec S ,
- soit $\exists T'_\sigma^f$ suffixe de T_σ^f compatible avec S .

$$S = (A, B, C, D)$$

A) composant
n° 1 2 3 4

11

B) composant
n° 1 2 3 4

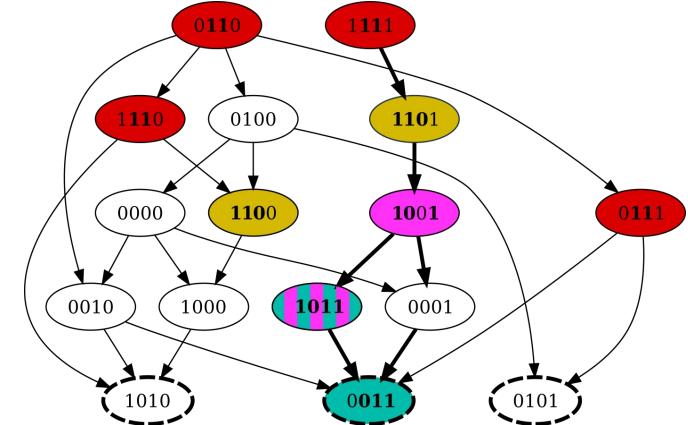
110

C) composant
n° 1 2 3 4

10 1

D) composant
n° 1 2 3 4

011



graphe de transition d'un RB f'

f' universellement compatible avec S :

À partir de 1111 compatible avec A, chaque trajectoire est compatible avec S.

Compatibilités définies

Réseau booléen compatible :

- existentiellement / universellement avec **une liste d'observations**
- existentiellement / universellement avec **une bifurcation**
- existentiellement / universellement avec **un (ensemble de) marqueur(s) de stabilité partielle / totale**
- universellement avec **un ensemble de marqueurs de stabilité partielle / totale atteignable depuis une observation**
- avec **un arbre de différenciation**

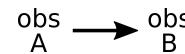
Problème 2 :

Méthode pour inférer des modèles à partir de données de différenciation ?

→ Conception et implémentation de contraintes logiques

Atteignabilit  positive existentielle (trajectoire)

\exists un chemin d'une configuration compatible avec une observation initiale
  une configuration compatible avec une observation cible.



Point fixe existentielle (marqueur de stabilit  totale)

Une configuration compatible avec l'observation donn e est un point fixe.



Confinement existentielle (marqueur de stabilit  partielle)

  partir d'une configuration compatible avec l'observation donn e, il est impossible
d'atteindre une configuration ayant une valeur diff rente pour l'un des marqueurs.



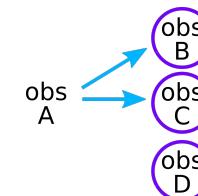
Atteignabilit  n gative existentielle (bifurcation)

\nexists chemin entre une configuration compatible avec une observation initiale
et une configuration compatible avec une observation cible.



Point fixe (atteignable) universelle

Tous les points fixes (atteignables) sont compatibles avec un ensemble donn e d'observations.



Programmation logique

Exprime un problème dans un système de logique formelle via :

- un ensemble de **faits** élémentaires : détermine le domaine du problème
- un ensemble de **règles logiques** : associent des conséquences plus ou moins directes aux faits

Langage de programmation logique choisi : **Answer-Set Programming (ASP)**

- adapté à la résolution de problèmes combinatoires
- calcul de modèles stables [Gelfond and Lifschitz, 1988] \Leftrightarrow minimalité des modèles obtenus

aperçu de la
syntaxe ASP

Fait : H .

Règle : $H \leftarrow L_1, \dots, L_n$.

Contrainte d'intégrité : $\leftarrow L_1, \dots, L_n$.

L_i ($1 \leq i \leq n$) : **littéral** (de la forme A ou $\text{not } A$)

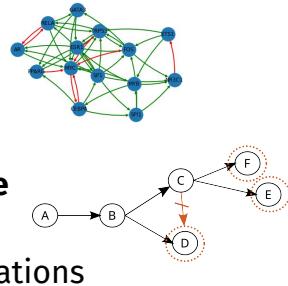
→ **A : atome** (construction élémentaire de la connaissance)

→ **not** : connecteur logique, désigne la négation par défaut

Implémentation de l'inférence de modèle

Le programme logique en ASP contient :

- **formalisme du réseau booléen**
- **manière de calculer sa dynamique**
sémantique Most Permissive
- **domaine de ses fonctions**
PKN
- **propriétés de sa dynamique**
observations,
contraintes liant les observations



Encodage du réseau booléen :

fonction booléenne exprimée
en **forme normale disjonctive**

Prédicat clause(*N,C,L,S*)

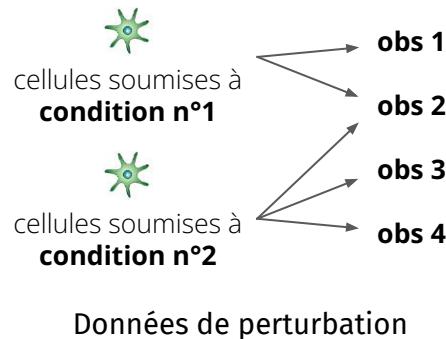
- atome *L*
- avec le signe *S* (-1, 1)
- est inclus dans la clause n°*C*
- pour $f_{\textcolor{red}{N}}$

Exemple : $f_{\textcolor{red}{a}}(x) = \textcolor{blue}{x_c} \vee (\neg x_{\textcolor{green}{a}} \wedge \textcolor{magenta}{x_b})$ → clause(*a,1,c,1*).
clause(*a,2,a,-1*).
clause(*a,2,b,1*).

Encodage de la contrainte de point fixe existentielle :

← is_fp(*X*) , cfg(*X,N,V*) , eval(*X,N,-V*) .

Challenge de la contrainte universelle



Réseau booléen souhaité :
universellement compatible avec un ensemble de marqueurs de stabilité totale atteignable depuis une observation



Formule booléenne à deux niveaux de quantificateur :

Étant donné y (configuration), $\exists f(\text{RB}) \forall x$ (configuration) :

$f(x) = x$ si et seulement si :

- soit x est compatible avec un marqueur de stabilité totale
- soit $y \not\rightarrow^* x$ (x non atteignable depuis y)

Contrainte universelle sur les points fixes

2QBF : abordable en ASP (en programmation logique disjonctive)

⇒ **technique de saturation** = règle disjonctive + saturation + modèles minimaux

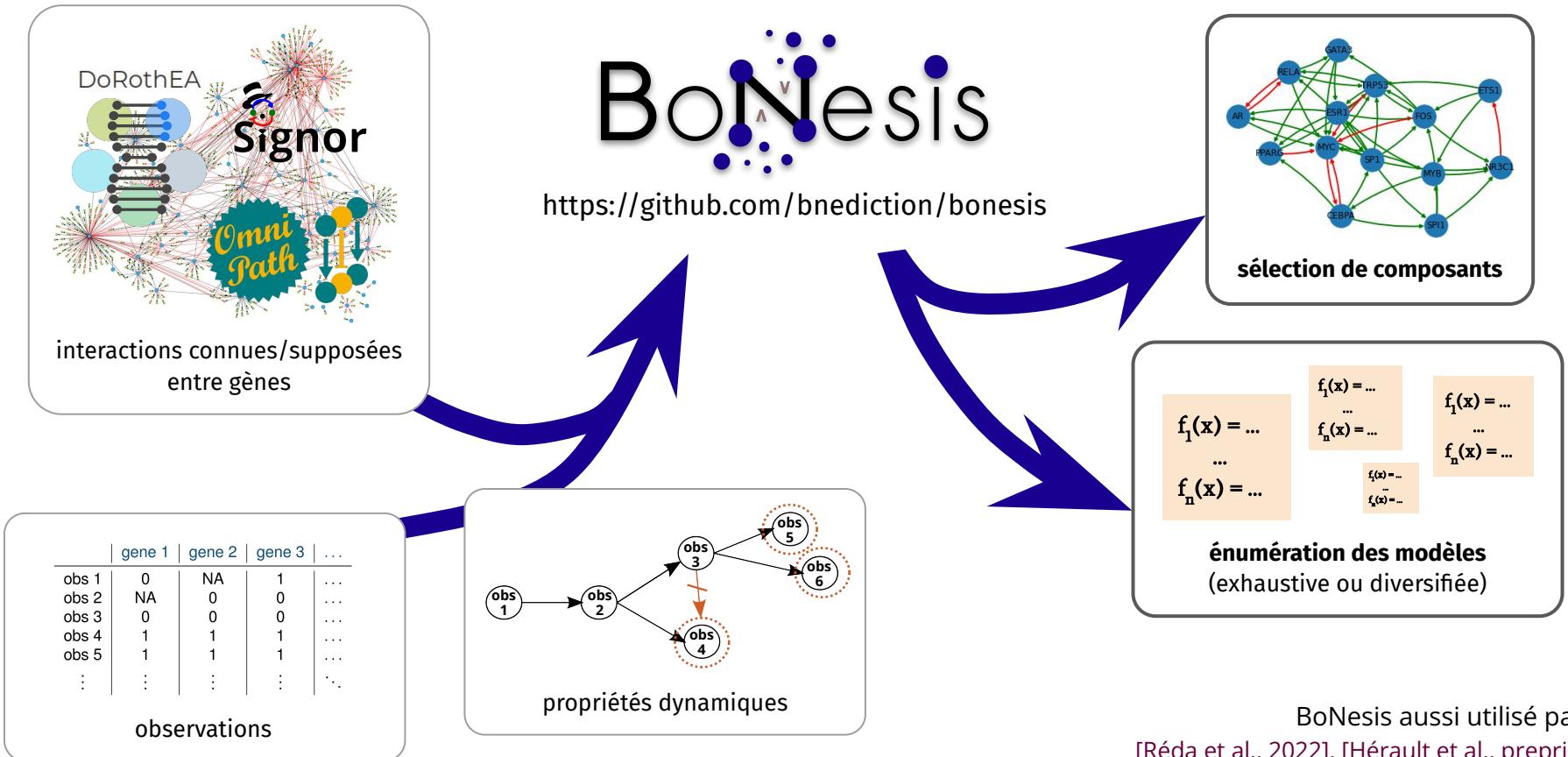
Principe de la contrainte universelle sur les points fixes atteignables :

1. Déduction d'une configuration z (règle disjonctive)
2. Validation et saturation sur la disjonction si :
 - z n'est pas un point fixe
 - z est compatible avec un marqueur de stabilité totale
 - z n'est pas atteignable depuis la configuration initiale
3. Solution rejetée si absence de validation

Solveur forcé de considérer toutes les configurations possibles (modèles minimaux)



Contraintes ASP intégrées dans *BoNesis* (*Boolean Network synthesis*)



BoNesis aussi utilisé par :
 [Réda et al., 2022], [Héault et al., preprint]

Problème 3 :

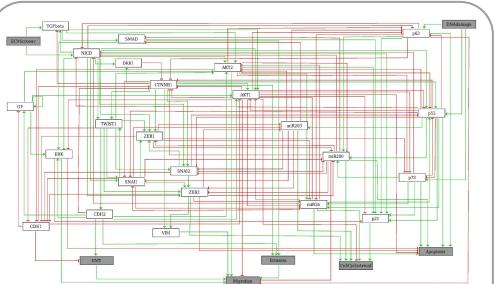
Applicabilité de la méthode sur données réelles ?

**→ Ensemble de modèles :
impact de la variabilité des modèles sur les prédictions**

**→ Données de grande taille :
impact sur la démarche de modélisation**

Impact de la variabilité des modèles sur les prédictions

Modélisation de la régulation du destin cellulaire dans la progression du cancer



PKN: graphe d'interactions du modèle publié [Cohen et al. 2015]

	miR34	dommage ADN	invasion	...
cond. initiale 1	1	0	0	...
cond. initiale 2	1	1	0	...
apoptose	0	0	0	...
métastase	0	1	1	...
⋮	⋮	⋮	⋮	⋮

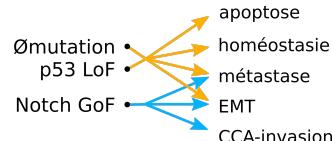
Observations: conditions initiales et phénotypes cellulaires [Cohen et al. 2015]

BonEsis

énumération des réseaux booléens compatibles

ensemble de 1000 modèles

En collaboration avec l'Institut Curie



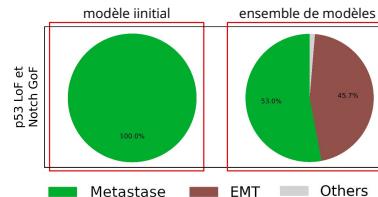
Propriétés dynamiques du modèle publié [Cohen et al. 2015]

- universalité des points fixes
- universalité des points fixes atteignables depuis les 3 conditions



Simulation du double mutant p53 LoF + Notch GoF:

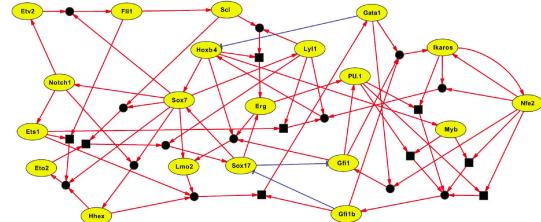
⇒ nuance son potentiel métastatique



Utilisation de données de grande taille

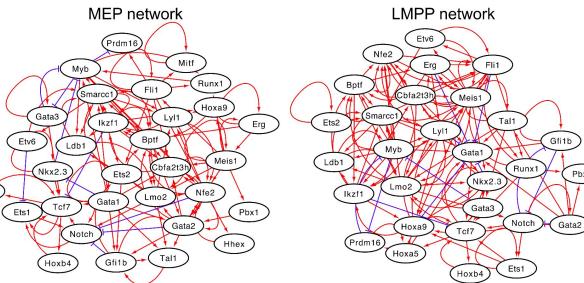
Modélisation de la régulation de l'hématopoïèse

Comparaison de modèles experts :



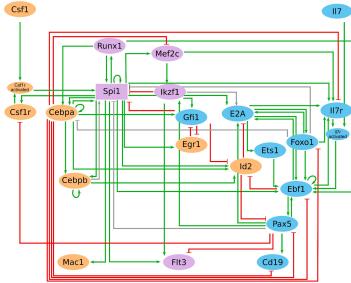
Graphe d'interactions du RB de [Moignard al., 2015]

→ 20 gènes



Graphes d'interactions des 2 RB de [Hamey et al., 2017],
1 RB par séries temporelles : a) HSC vers MEP, b) HSC vers LMPP.

➡ 31 gènes



Graphe d'interactions du RB de [Collombet al., 2017]

→ 20 gènes

seulement 2 gènes communs aux 3 modèles

Quel est le résultat d'une inférence automatique de modèles

à partir de mesures d'expression de gènes *single-cell*

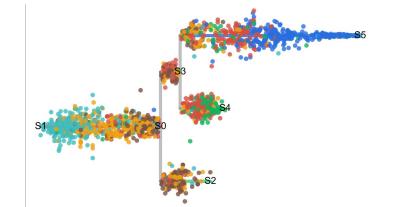
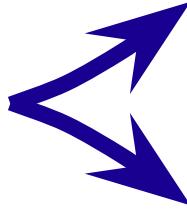
sans a priori sur les gènes régulant ce processus ?

Utilisation de données de grande taille

Modélisation de la régulation de l'hématopoïèse : 1. Préparation des données

	gene 1	gene 2	gene 3	...
cellule 1	4.891	6.877	0.0	...
cellule 2	1.423	0.0	6.913	...
cellule 3	0.0	0.0	0.0	...
cellule 4	0.0	2.423	0.0	...
cellule 5	2.599	1.804	6.419	...
⋮	⋮	⋮	⋮	⋮

Mesures d'expression de gènes single-cell

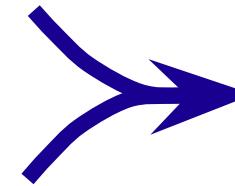


Reconstruction de trajectoire
logiciel STREAM [Chen et al., 2019]

	gene 1	gene 2	gene 3	...
cellule 1	1	1	NA	...
cellule 2	NA	0	1	...
cellule 3	0	0	0	...
cellule 4	NA	1	NA	...
cellule 5	NA	0	1	...
⋮	⋮	⋮	⋮	⋮

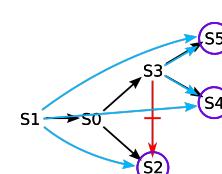
Binarisation des mesures d'expression

méthode PROFILE [Béal et al., 2019]




	gene 1	gene 2	gene 3	...
S0	0	NA	1	...
S1	NA	1	0	...
S2	0	0	0	...
S3	1	1	1	...
S4	1	1	0	...
S5	1	0	1	...

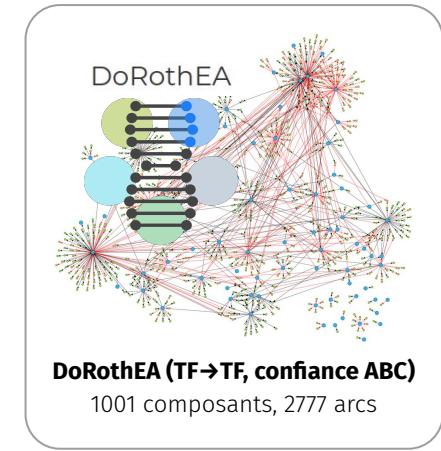
Création des observations



Description du comportement
(contraintes existentielles et universelles)

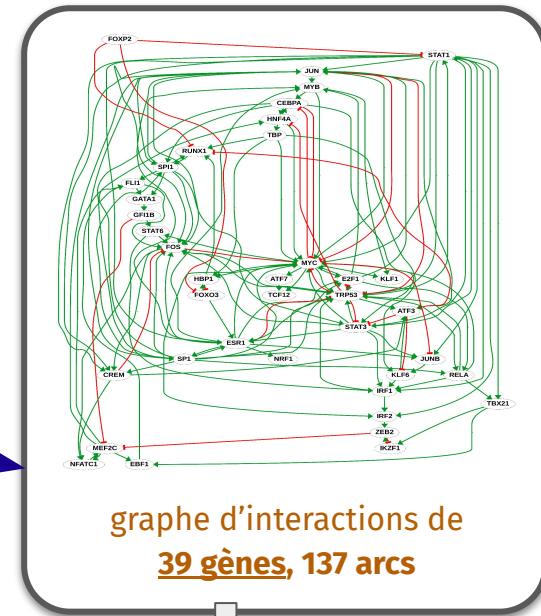
Utilisation de données de grande taille

Modélisation de la régulation de l'hématopoïèse : 2. Construction du PKN



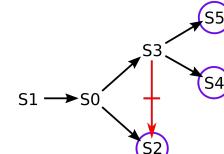
Bon
e
n
e
s
i
s

sélection de
composants



	gene 1	gene 2	gene 3	...
S0	0	NA	1	...
S1	NA	1	0	...
S2	0	0	0	...
S3	1	1	1	...
S4	1	1	0	...
S5	1	0	1	...

Observations S0, S1, S2, S3, S4, S5



Propriétés dynamiques

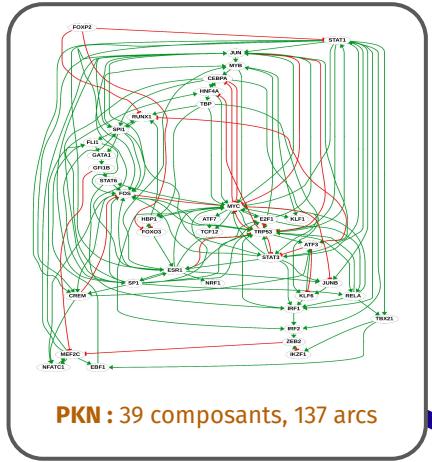
- 5 atteignabilités positives
- 1 atteignabilité négative
- 3 points fixes

30 gènes fonctionnellement liés à la **différenciation**
dont **19** spécifiquement à l'**hématopoïèse**

6 gènes en commun avec **chacun des 3 modèles experts**

Utilisation de données de grande taille

Modélisation de la régulation de l'hématopoïèse : 3. Inférence de modèles



Bon
e
n
e
s
i
s

énumération des réseaux booléens compatibles

RB n°1 :

```

1 ATF3 <- JUN
2 ATF2 <- JUN
3 CEBPA <- IRF4
4 CREM <- 0
5 E2F1 <- TRP53
6 EBF1 <- TBX21
7 EGR1 <- ATF2
8 FLI1 <- EBF1
9 FOS <- JUN
10 FOXO3 <- 1
11 FOXP2 <- 1
12 GATA1 <- FLI1
13 GF11B <- GATA1
14 HBP1 <- FOXO3
15 IRF4 <- ATF2
16 IKZF1 <- TBX21
17 IRF1 <- STAT1|STAT6|(RELA&KLF6)|(STAT3&KLF6)
18 IRF2 <- MYC
19 IRF4 <- STAT1|E2F1&MEF2C)|(SP1&MEF2C)
20 JUND <- 1
21 KLF1 <- MYB
22 KLF9 <- TBP
23 KLF9 <- GF11B|ZEB2
24 MYB <- E2F1
25 MYC <- (TRP53&TBP)|(FO5&CEBP&STAT3
26 NFATC1 <- CREM
27 NFATC1 <- E2F1
28 RELA <- E2F1
29 RUNX1 <- ATF3
30 SP1 <- CEBPA
31 SP1 <- EGR1|RUNX1
32 STAT1 <- FOXP2
33 STAT3 <- SP1|TRP53
34 STAT6 <- MYC
35 TBX21 <- RELA
36 TCF12 <- ATF7
37 TRP53 <- MYC
38 TRP53 <- SP1
39 ZEB2 <- IRF2

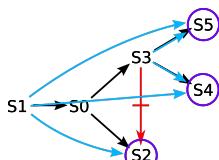
```

RB n°1000 :

```

1 ATF3 <- JUN|(CREMKLF6)|(CREM&TRP53)|(TRP53&KLF6)
2 ATF2 <- JUN
3 CEBPA <- IRF4&JUN44
4 CREM <- (FO5&JUN&STAT1)|(JUN&SP1&STAT1)
5 E2F1 <- 0
6 EBF1 <- TBX21
7 EGR1 <- (SP1&EGR1)|(SP1&EGR1&KLF6)|(STAT3&EGR1)
8 ESR1 <- (SP1&EGR1)|(SP1&EGR1&KLF6)|(STAT3&EGR1)
9 FOS <- (JUN&EGR1)|(JUN&EGR1)|(SP1&EGR1)|(SP1&EGR1)|(SP1&EGR1)
10 FOXO3 <- (ATF2&KLF6)|(ATF2&KLF6)|(ATF2&KLF6)
11 FOXP2 <- 1
12 GATA1 <- FLI1
13 GF11B <- GATA1
14 HBP1 <- FOXO3
15 IRF4 <- (SP1&CEBPA)|(TRP53&CEBPA)
16 IKZF1 <- TBX21
17 IRF1 <- (KLF6&STAT1)|(KLF6&STAT1)|(STAT3&KLF6)|(RELA&STAT1)|(STAT3&STAT1)
18 IRF2 <- (MYC&STAT1)|(STAT1&IRF1)|(MYC&STAT1)
19 IRF4 <- (ATF2&E2F1)|(E2F1&MEF2C)|(SP1&E2F1)|(SP1&E2F1)|(ESR1&STAT1)|(SP1&E2F2)
20 JUND <- STAT1|STAT3
21 KLF1 <- MYB
22 KLF9 <- TBP
23 KLF9 <- GF11B|ZEB2
24 MYB <- E2F1
25 MYC <- ((FO5&JUN&STAT3)|(FO5&JUN&E2F1&HNF4A))|((FO5&HNF4&E2F1&STAT1))
|((FO5&E2F1&STAT3&HNF4A))|((CEBP&AP1&SP1&GSE1&STAT1)|(TBX&YB&HNF4&ESR1&E2F1|FO5))
|((CEBP&AP1&SP1&GSE1&STAT1)|(TBX&YB&HNF4&ESR1&E2F1|FO5))
26 NFATC1 <- (CREM&EGR1)|(CREM&EGR1)
27 NF1 <- ESR1
28 RUNX1 <- ((ATF3&FOXP2)|(HBP1|ATF3)|(SP1&TBP)|(HBP1|FOXP2|TBP)
|((CEBP&AP1&SP1&GSE1&STAT1)|(TBX&YB&HNF4&ESR1&E2F1|FO5))
29 ATF3 <- (SP1&KLF6)
30 ATF2 <- 0
31 ATF4 <- 0
32 STAT1 <- FOXP2
33 STAT3 <- SP1|TRP53
34 STAT6 <- MYC
35 TBP <- IRF4
36 TBX21 <- RELA
37 TCF12 <- ATF7
38 TRP53 <- MYC
39 ZEB2 <- IRF2

```



Propriétés dynamiques

- 5 atteignabilités positives
- 1 atteignabilité négative
- 3 points fixes
- universalité des points fixes depuis S1 et S3

Observations S0, S1, S2, S3, S4, S5

ensemble de 1000 modèles

Utilisation de données de grande taille

Modélisation de la régulation de l'hématopoïèse : 4. Analyse de l'ensemble de modèles

Aperçu de la variabilité sur 1000 modèles

Attracteurs des modèles :

- aucun attracteur cyclique
- uniquement des points fixes compatibles avec S2, S4 et S5

Variabilité sur les 1000 modèles :

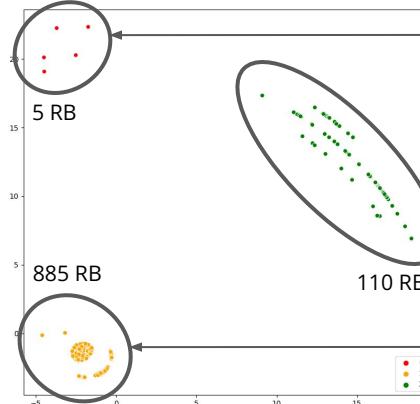
# fonctions différentes	# gènes concernés
1	12
2	9
3	5
4	4
5	5
6	1 (RUNX1)
16	1 (MYC)
44	1 (TRP53)
682	1 (FOS)



12 gènes à la fonction conservée au travers des 1000 modèles (dont 9 absents de l'état de l'art)

} variabilité concentrée sur quelques gènes

3 groupes de modèles de structure similaire :



Nuage de points résultant d'un MDS

Particularités :

- fonctions de peu de clauses
- fonctions indépendantes de FOS
- fonction constante sur CREM

fonctions avec nombreuses clauses

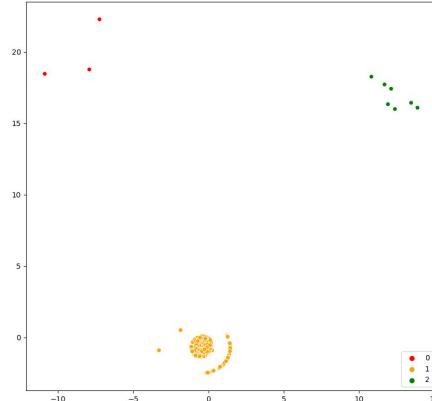
- grande homogénéité (seulement 4 gènes variables)
- fonctions avec nombreuses clauses
- seul groupe où IKZF1 est variable

Perspectives :

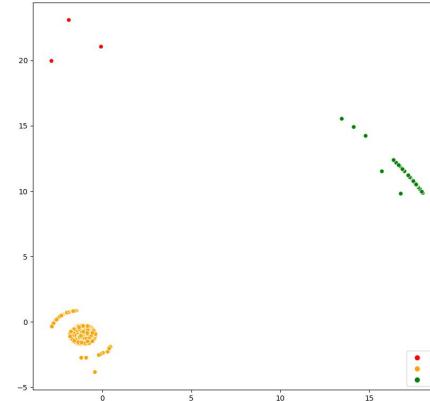
Poursuivre l'exploration des modèles de l'hématopoïèse

via un regard expert sur l'ensemble des modèles

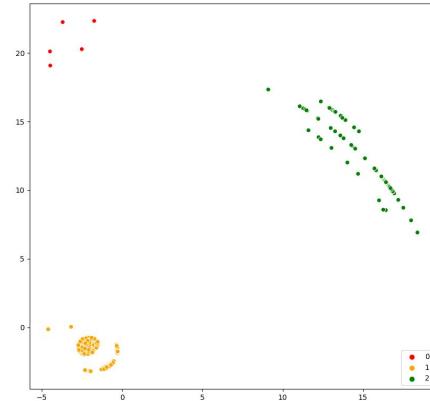
Critère pour estimer le nombre de modèles suffisant pour explorer la diversité des solutions possibles en comparant l'évolution de la diversité des modèles ?



MDS sur les **250** premiers RB



MDS sur les **500** premiers RB



MDS sur les **1000** RB

Perspectives :

Impact des paramètres de construction des observations sur les modèles obtenus

Single-cell, comparaison des ensembles de modèles obtenus selon :

- binarisation des mêmes données avec deux méthodes différentes
- changement du nombre de cellules incluses dans les observations

Concevoir des contraintes supplémentaires

Contrainte universelle
sur les **points fixes**



Contrainte universelle
sur les **attracteurs**

3QBF → ~~ASP~~

Faciliter la démarche de modélisation



Conclusion

Inférence automatique de modèles dynamiques à partir de données de différenciation

Modélisation logique de la différenciation

- Propriétés dynamiques élémentaires : trajectoire, bifurcation, stabilité
- Compatibilité du réseau booléen avec ces propriétés

Implémentation de contraintes garantissant les propriétés de la différenciation

- À disposition de la communauté : logiciel BoNesis
- Richesse des comportements modélisables

Applications biologiques

- Variabilité des modèles impactent les prédictions
- Inférence automatique possible à partir de données de très grande taille (> 1000 gènes)
- Sélection de gènes d'intérêt

Publications :

- Article de conférence

Synthesis and Simulation of Ensembles of Boolean Networks for Cell Fate Decision. CMSB 2020. Stéphanie Chevalier, Vincent Noël, Laurence Calzone, Andrei Zinovyev, Loïc Paulevé.

- Article de conférence

Synthesis of Boolean Networks from Biological Dynamical Constraints using Answer-Set Programming. ICTAI 2019. Stéphanie Chevalier, Christine Froidevaux, Andrei Zinovyev, Loïc Paulevé

Article en préparation sur les applications.