

Étude de la régulation anti-sens par l'analyse différentielle de données transcriptomiques dans le domaine végétal

Marc LEGEAY

LERIA

Université d'Angers
Laboratoire d'Étude et de Recherche en Informatique d'Angers
Institut de Recherche en Horticulture et Semences



Rapporteurs :

Mme Céline ROUVEIROL

Professeure des universités

Université Paris-Nord

Mme Fariza TAHI

Maître de conférences - HDR

Université d'Evry

Examineur :

M. Jérémie BOURDON

Professeur des universités

Université de Nantes

Direction de thèse :

Mme Béatrice DUVAL

Professeure des universités

Université d'Angers

M. Jean-Pierre RENOÜ

Directeur de recherche

INRA d'Angers

12 décembre 2017

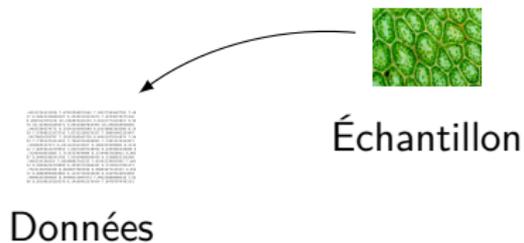


Biologie des systèmes

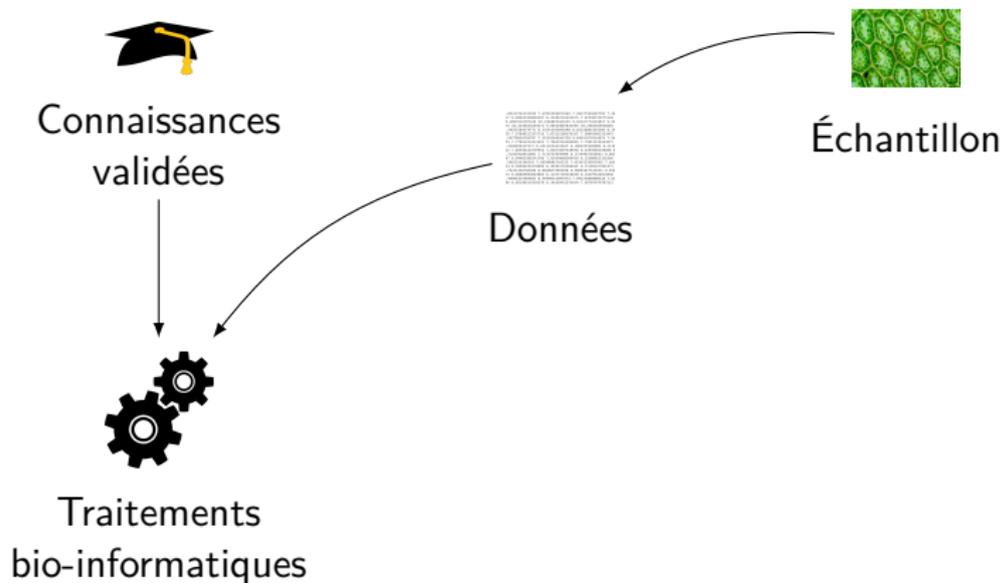


Échantillon

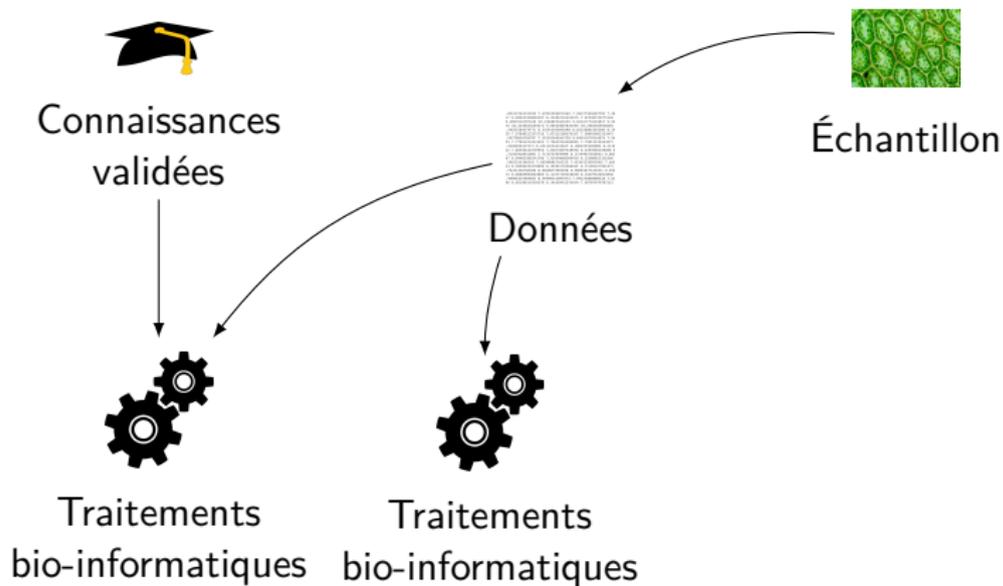
Biologie des systèmes



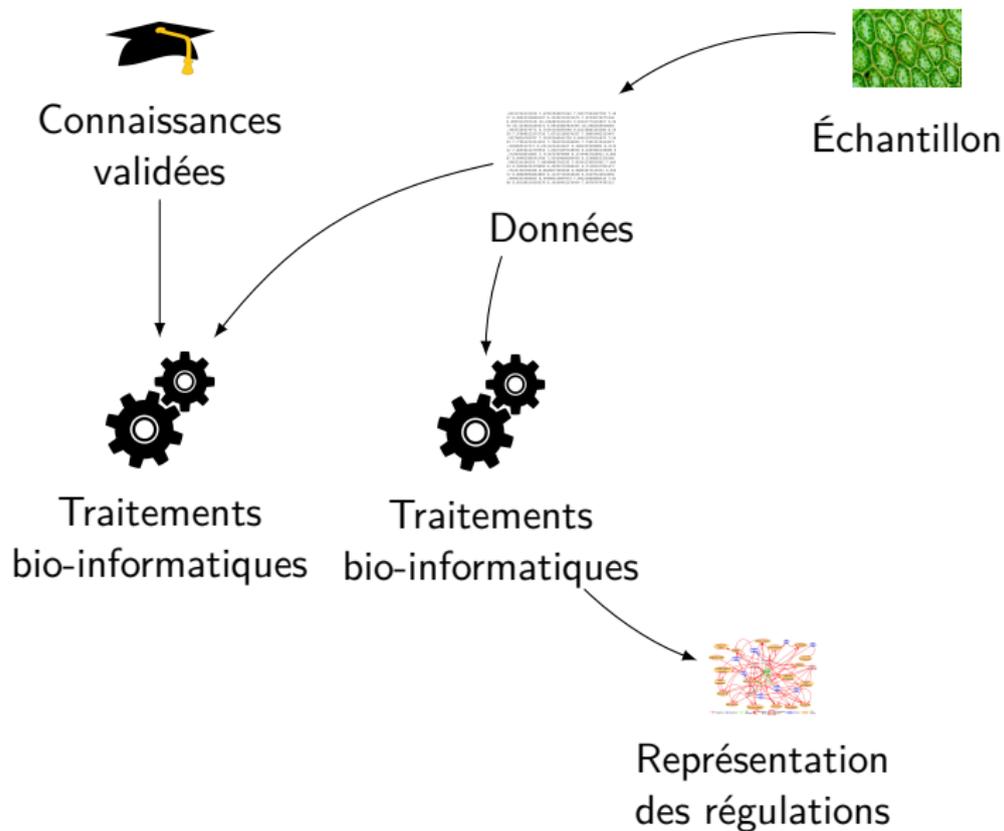
Biologie des systèmes



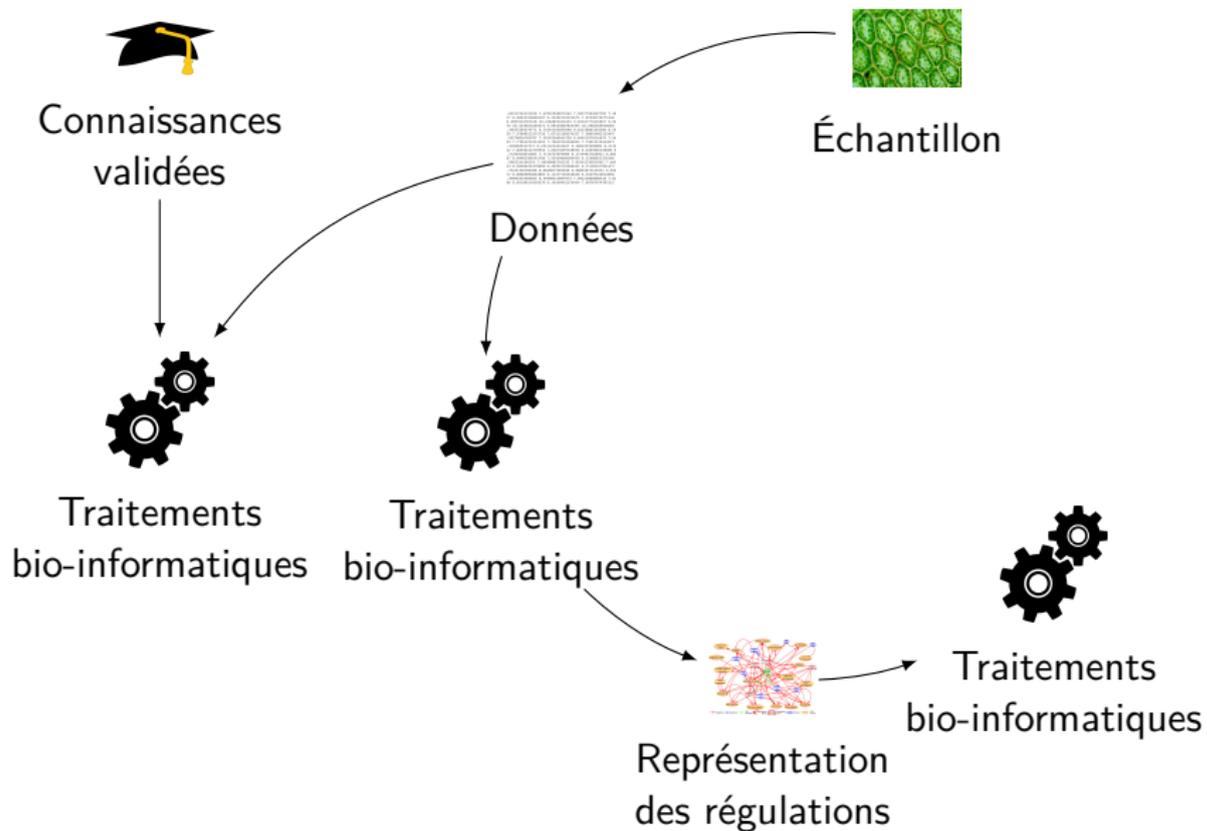
Biologie des systèmes



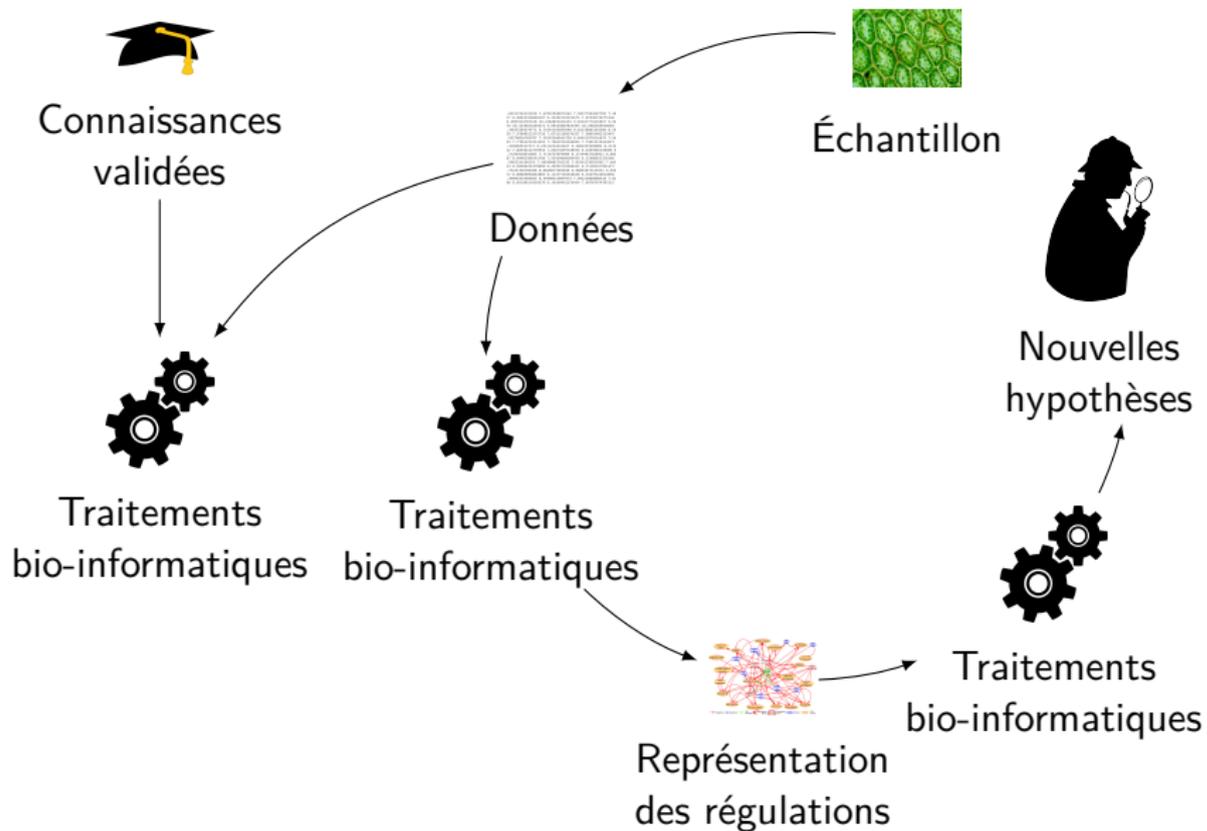
Biologie des systèmes



Biologie des systèmes



Biologie des systèmes

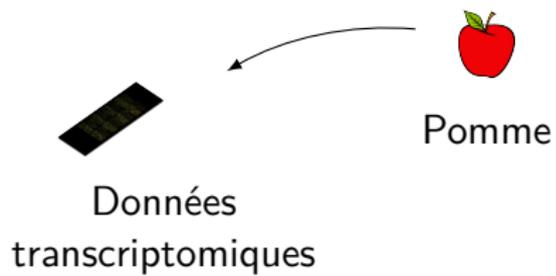


Contexte de la thèse

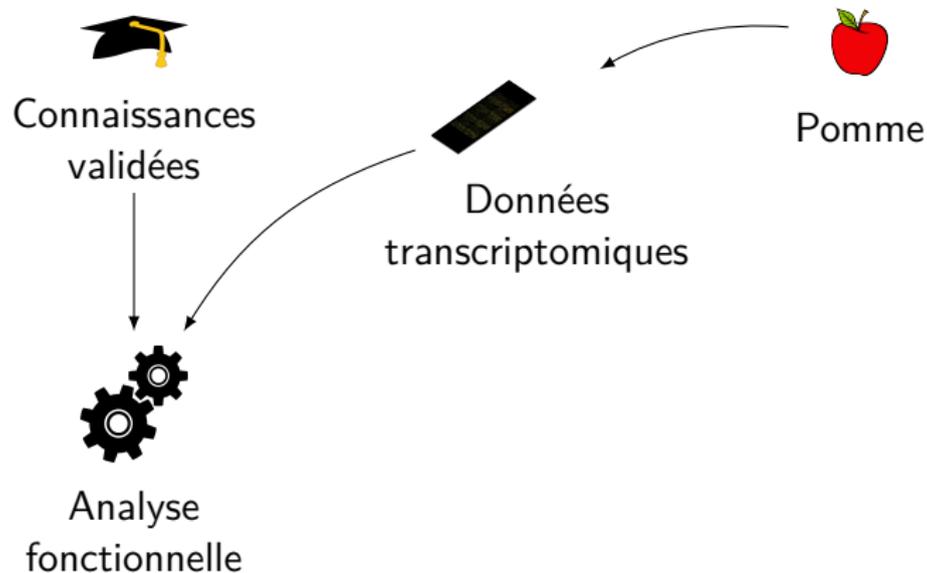


Pomme

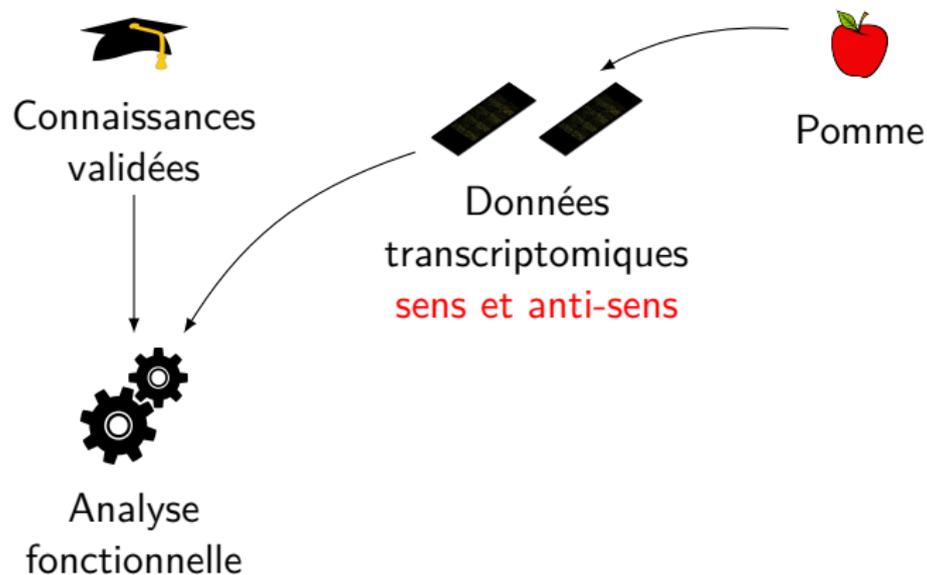
Contexte de la thèse



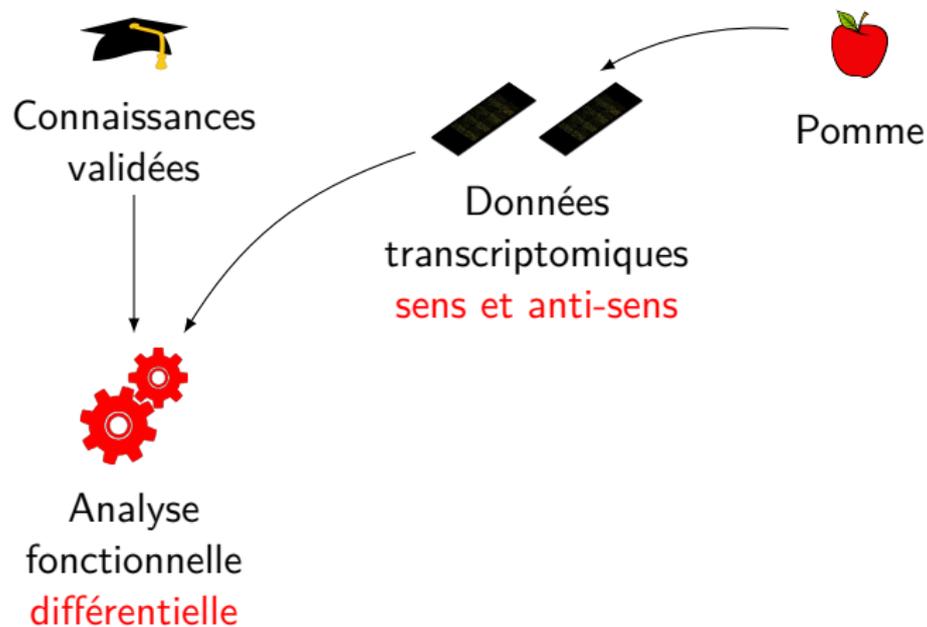
Contexte de la thèse



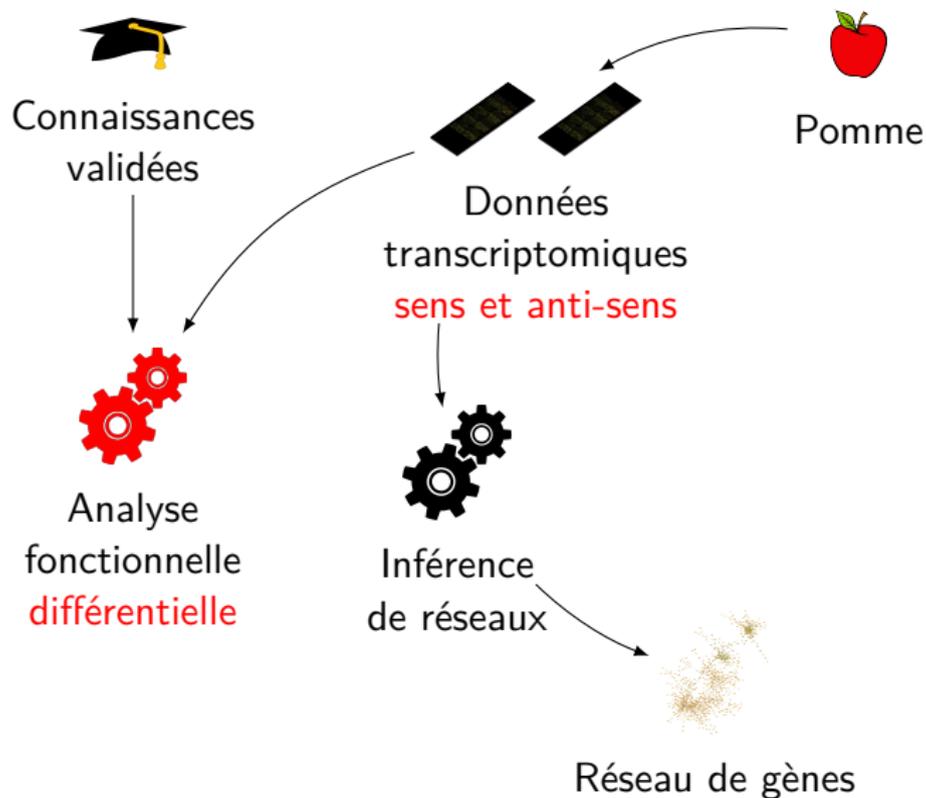
Contexte de la thèse



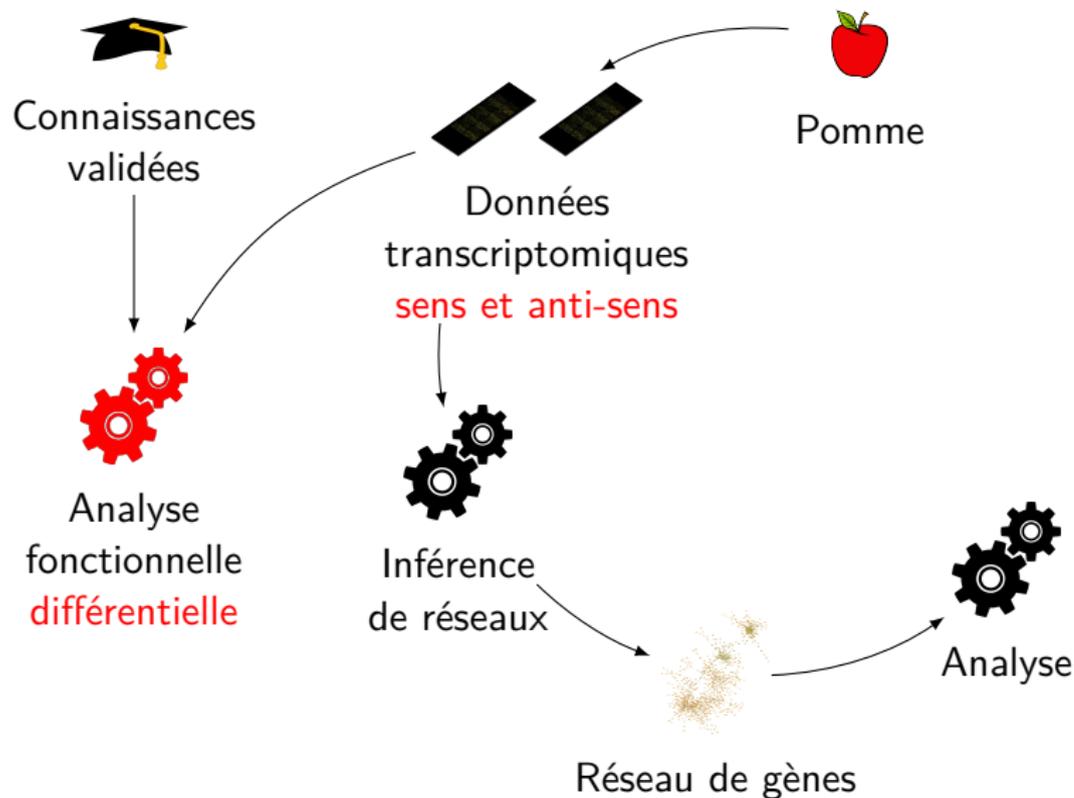
Contexte de la thèse



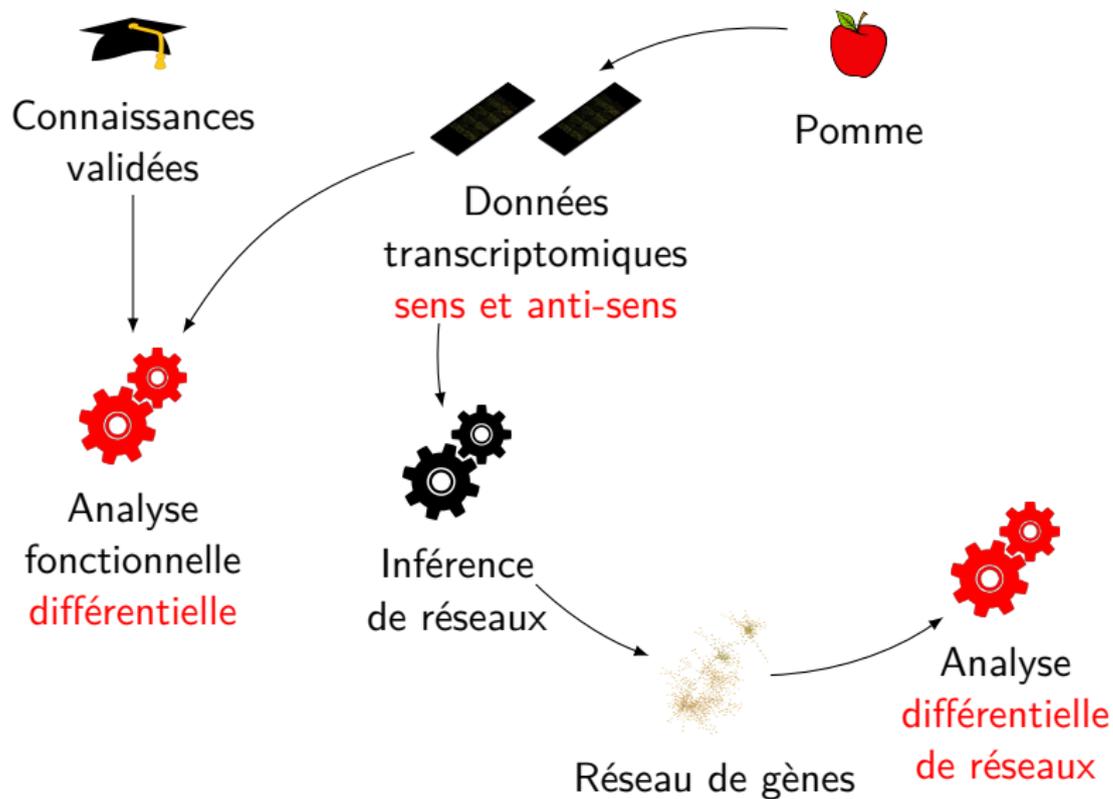
Contexte de la thèse



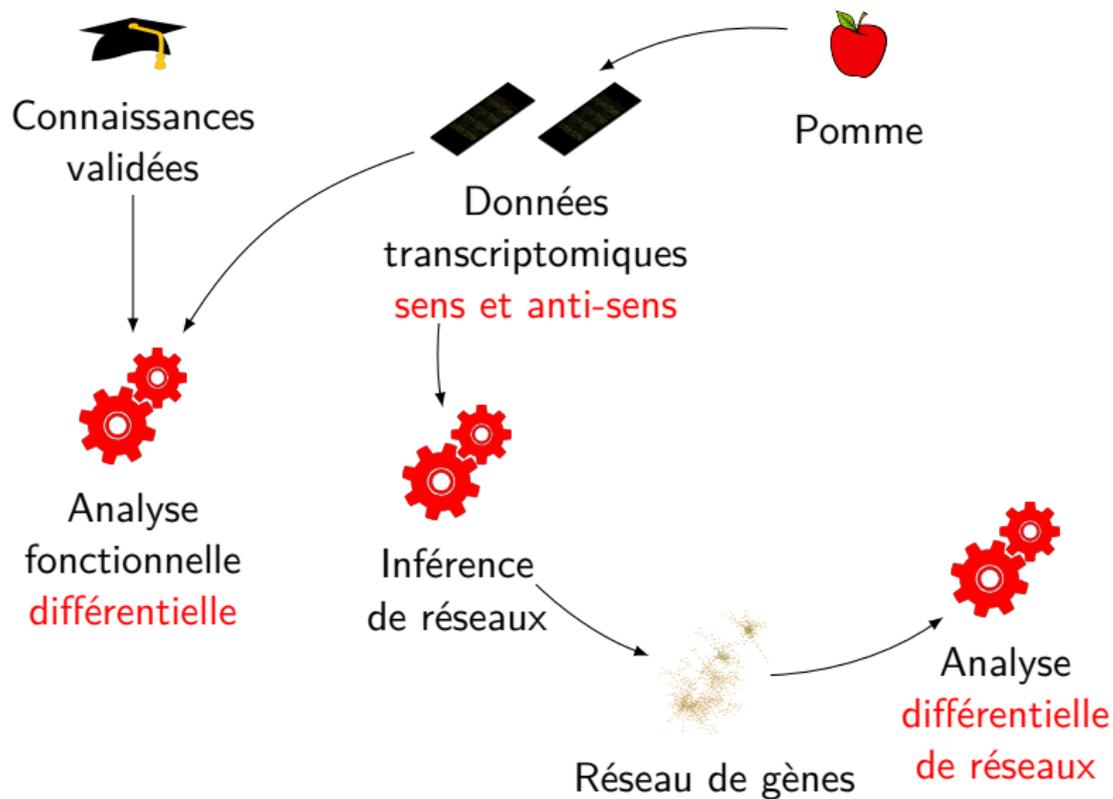
Contexte de la thèse



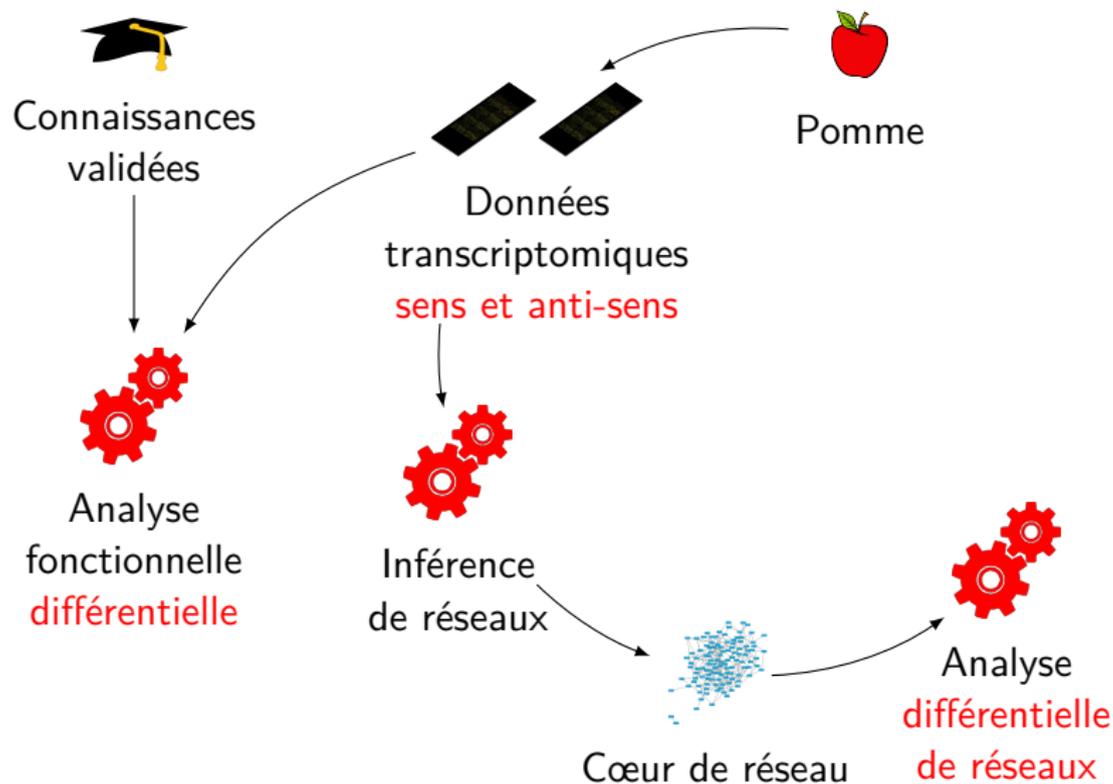
Contexte de la thèse



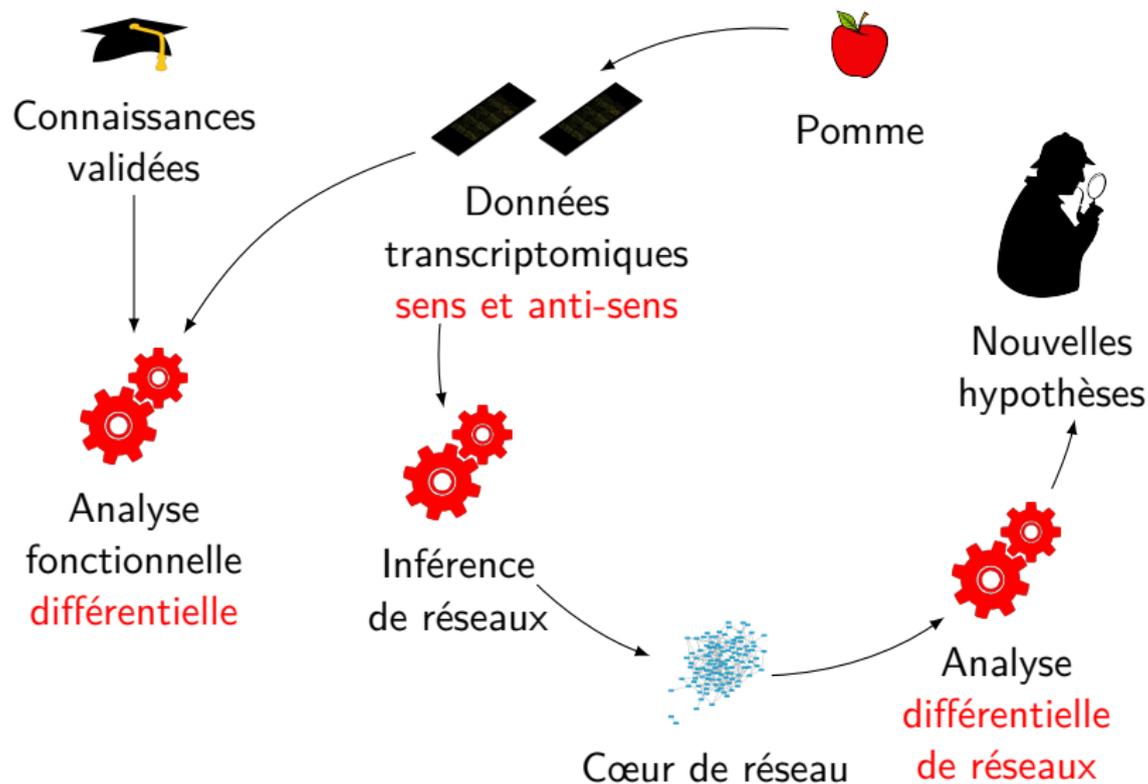
Contexte de la thèse



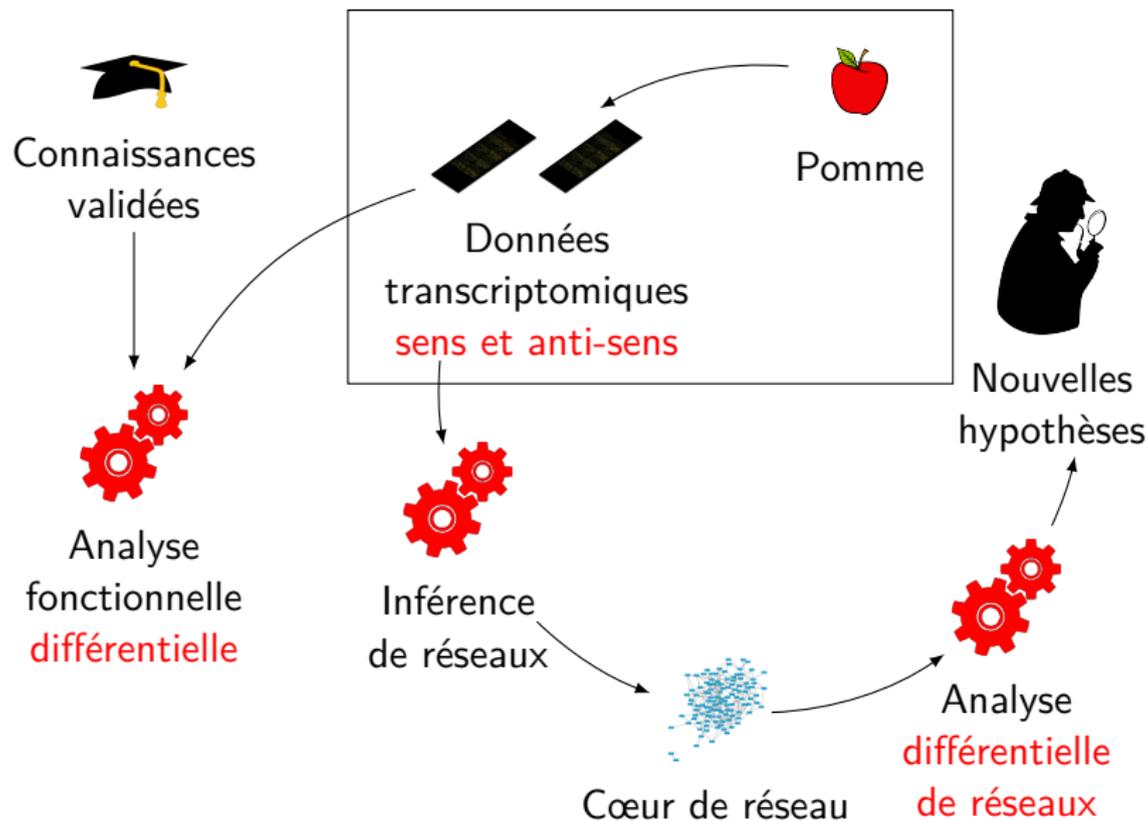
Contexte de la thèse



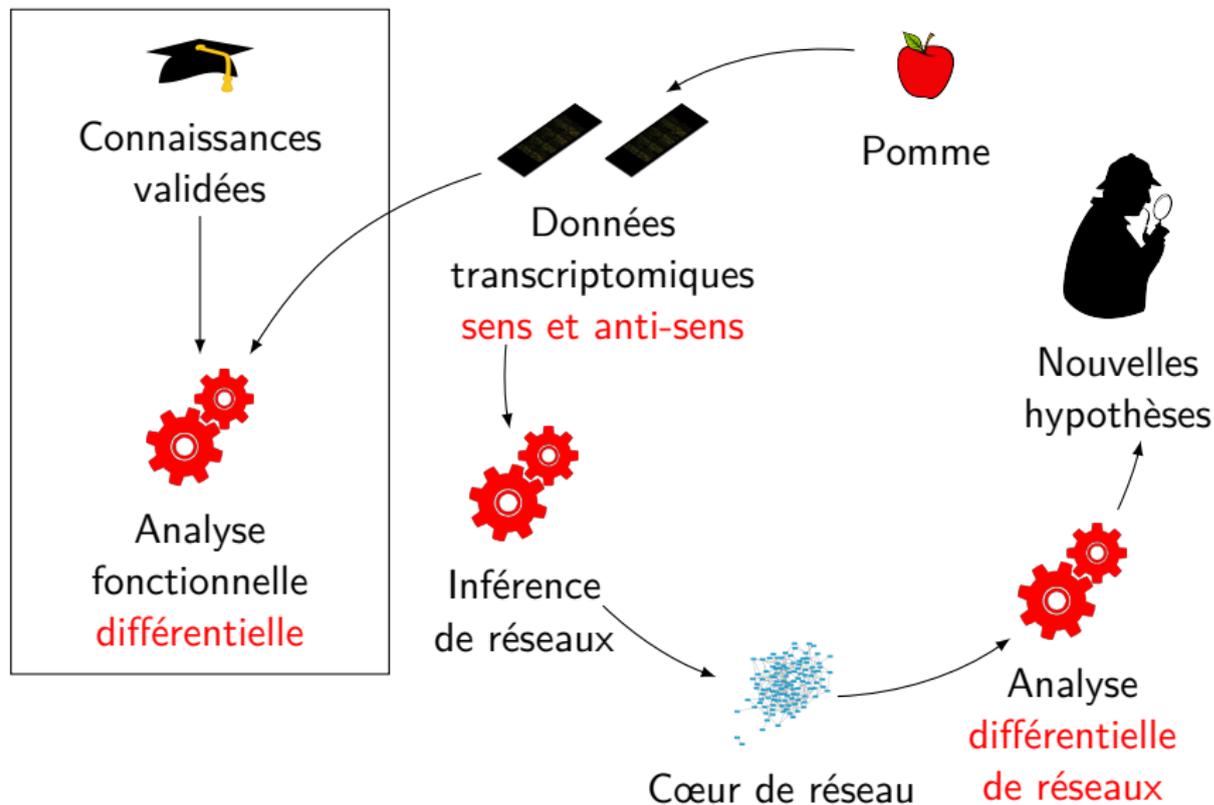
Contexte de la thèse



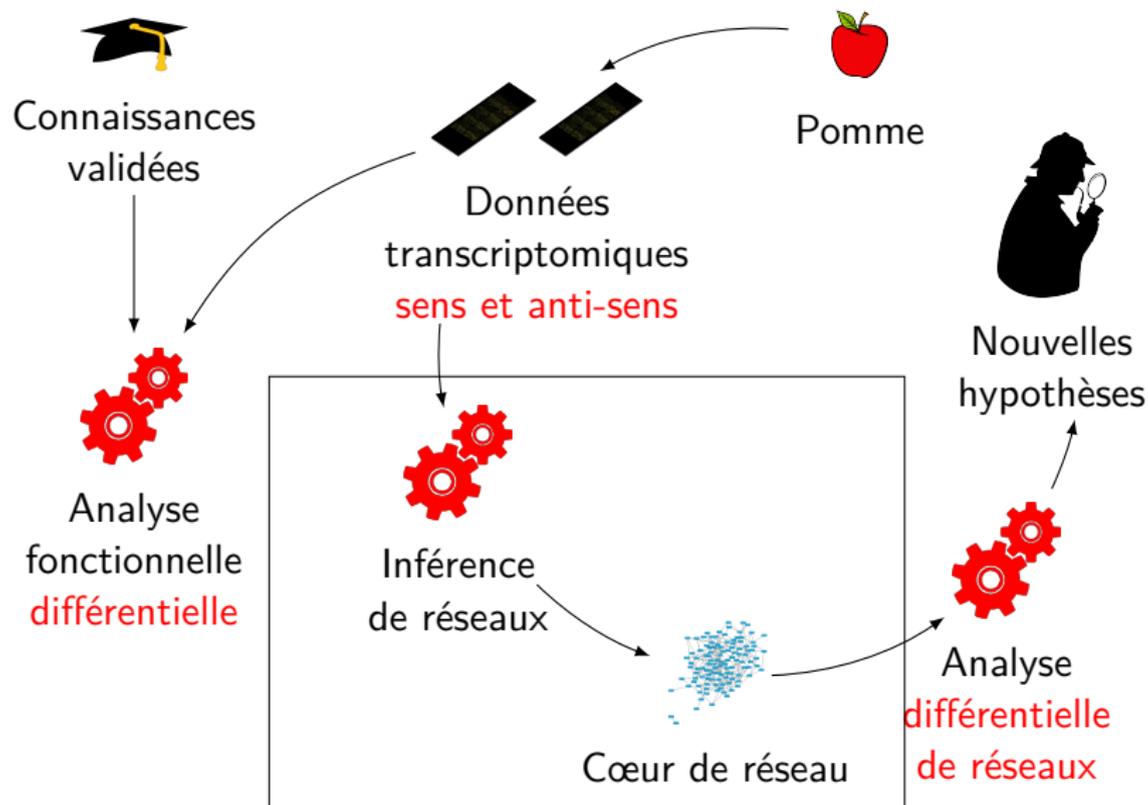
Contexte de la thèse



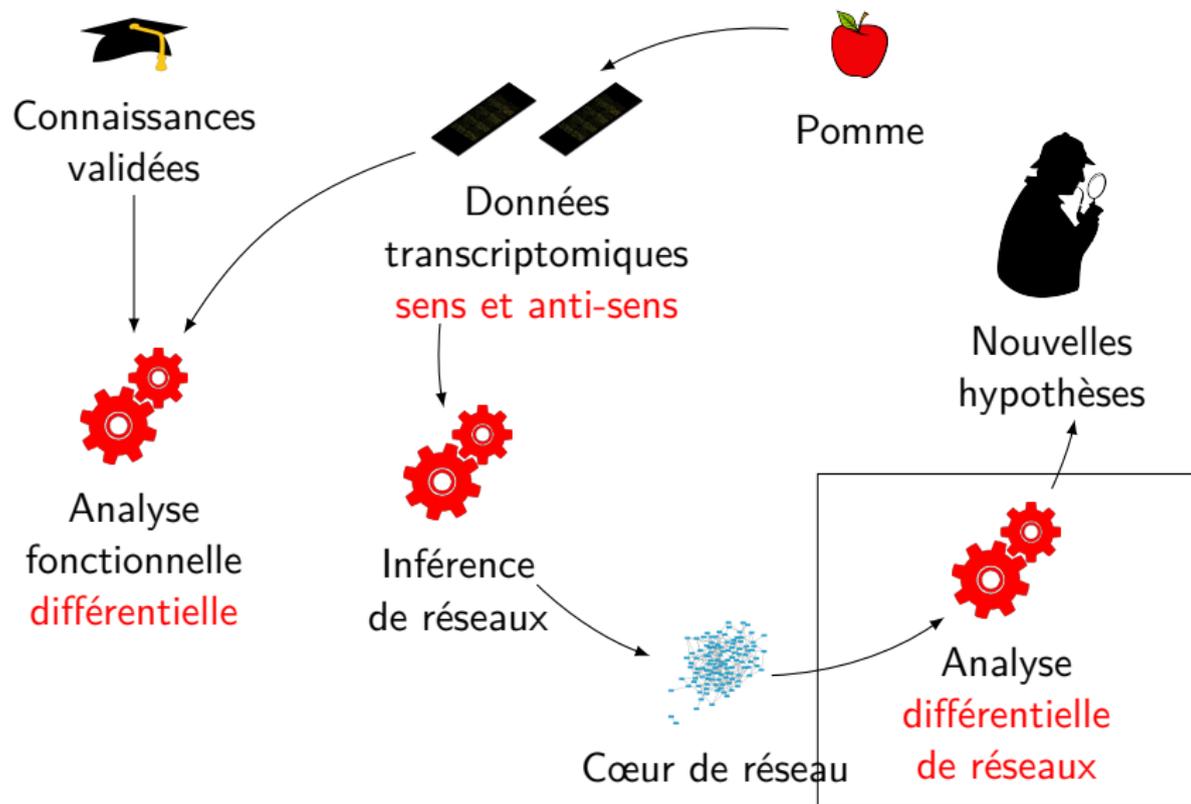
Contexte de la thèse



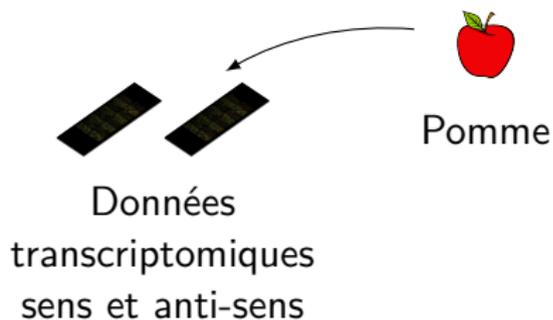
Contexte de la thèse



Contexte de la thèse



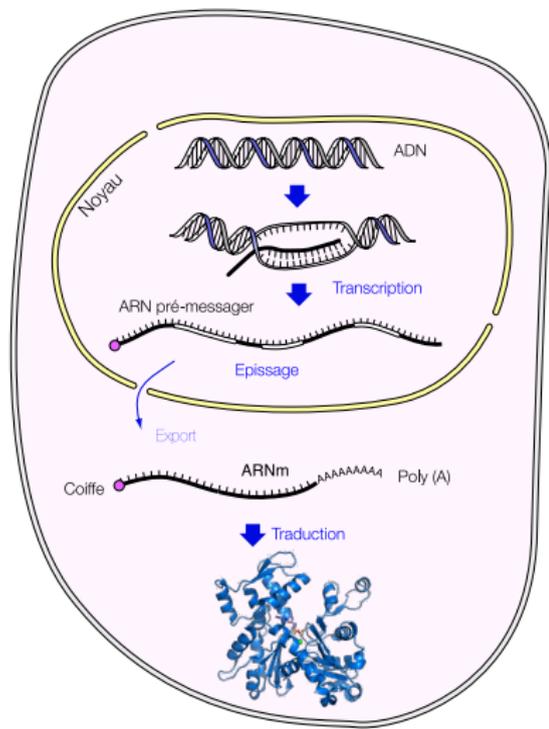
Données transcriptomiques



Transcriptomique

Dogme :

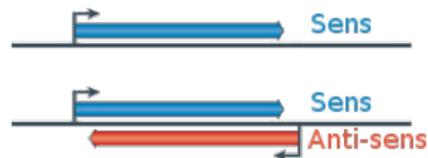
- Acide Désoxyribonucléique (ADN)
- ↓ Transcription d'un brin
- Acide Ribonucléique (ARN)
- ↓ Traduction
- Protéine



Transcription anti-sens

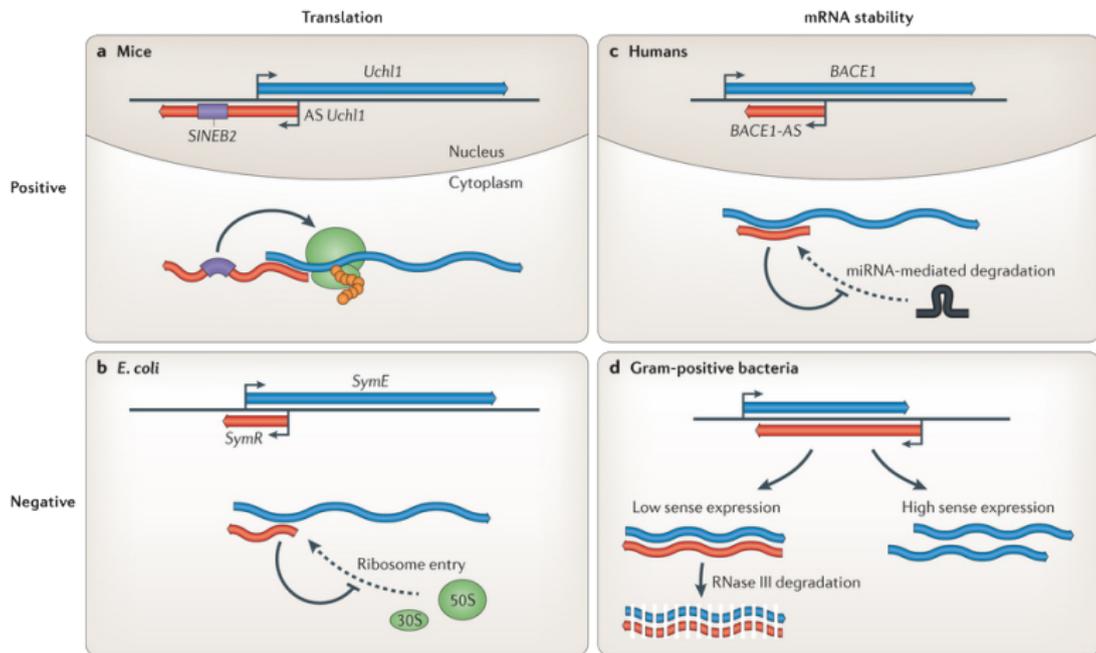
ARN anti-sens

- Transcrit du brin opposé au gène
- Molécule d'ARN complémentaire avec d'autres transcrits



Mécanisme peu étudié, quelques pistes ...

Contrôle post-transcriptionnel

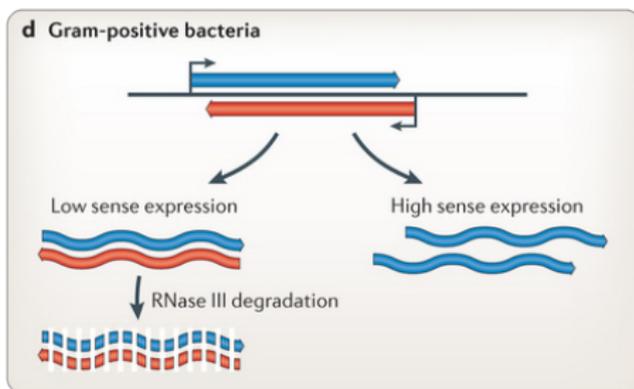


Nature Reviews | Genetics

PTGS

Post-Transcriptional Gene Silencing

- Extinction de l'expression du gène après la transcription
- Deux effets qui se cumulent :
 - Dégradation de l'ARNm
 - Production de petits ARN interférents (siRNA)



Mesure des transcrits

Puce à ADN

- Méthode permettant de quantifier l'expression des gènes
- Sonde : courte séquence d'oligonucléotides (ADN simple brin)
- Chaque séquence d'une sonde est sélectionnée pour identifier un gène

Puce AryANE v1.0

- Puce à ADN développée à l'IRHS pour le Pommier
- Basée sur le séquençage de 2010 : 750 Mb ; 57 386 gènes putatifs
- 126 022 sondes d'oligonucléotides 60-mer
- 63 011 sondes sens ; 63 011 sondes anti-sens

Motivations

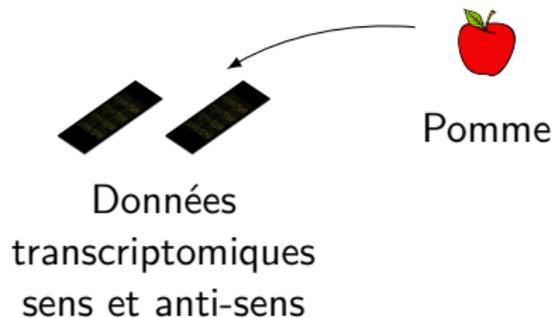
- Étude de l'IRHS en 2014 par Celton *et al.* sur le transcriptome
- Mesure de l'expression sens et anti-sens dans 8 organes du pommier
- ① 65% transcrits anti-sens exprimés dans au moins un organe lorsque le transcrit sens est exprimé (30% chez *Arabidopsis thaliana*)
- ② Le niveau d'expression anti-sens varie selon les fonctions biologiques
- ③ Le nombre de transcrits anti-sens exprimés varie selon les organes
- ④ Corrélation entre transcription anti-sens et présence de petits ARN

Quel est le rôle de la transcription anti-sens ?

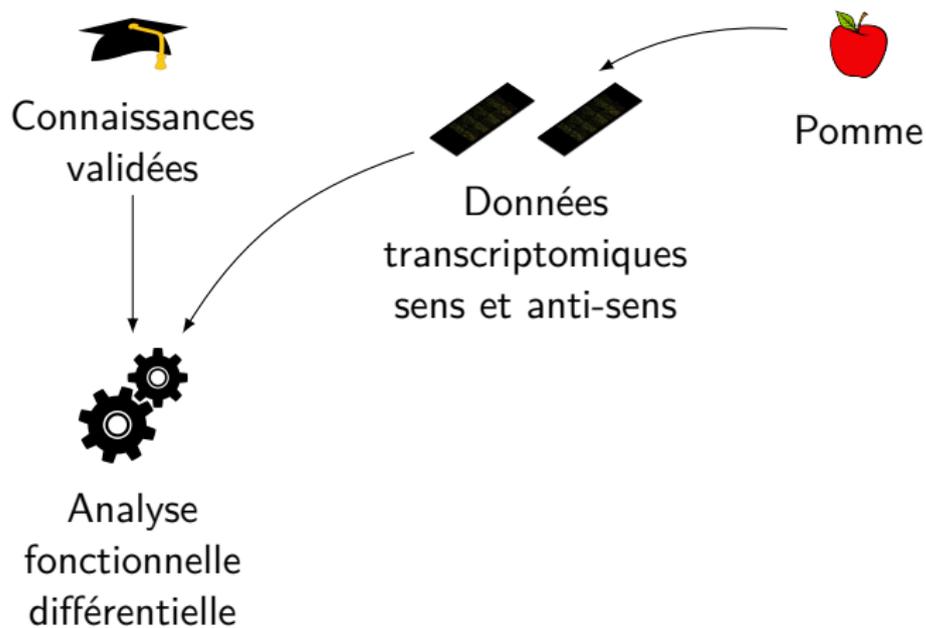
Données biologiques

- Contexte : Maturation du fruit stocké en chambre froide
 - Deux expériences : récolte (H) et 60 jours après récolte (60DAH)
 - 22 échantillons in H \longleftrightarrow 22 échantillons dans 60DAH
- Normalisation des données par la méthode des quantiles
- Sélection des sondes les plus spécifiques : 96 120 sondes
- Sélection des transcrits qui évoluent (différentiellement exprimées)
- 1 625 transcrits d'intérêt :
 - 931 S
 - 694 AS
 - 1 425 couples dont 200 ont S et AS qui évoluent

Analyse fonctionnelle différentielle



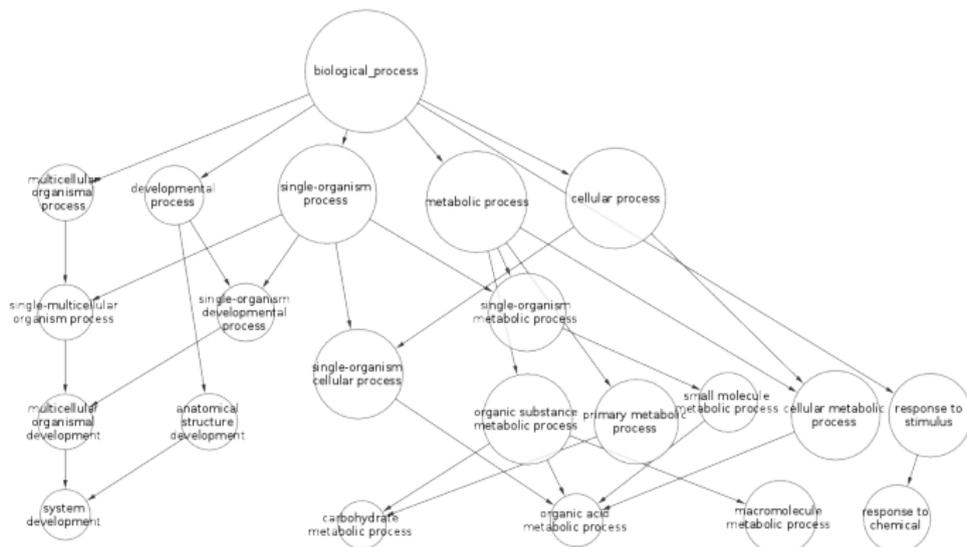
Analyse fonctionnelle différentielle



Gene Ontology



- Vocabulaire hiérarchisé pour décrire des fonctions biologiques (Biological process)
- Associe un gène à un ensemble de termes (fonctions)
- Disponible pour des organismes de référence
- GO slim : Ontologie allégée (15 catégories contre 3 163)



Analyse fonctionnelle

Soient :

- Un génome de taille G
- Un échantillon de taille g

Grâce à la GO, on connaît :

- Les fonctions biologiques associées aux gènes de l'échantillon
- Le nombre E de gènes du génome associés à une fonction donnée f
- Le nombre e de gènes de l'échantillon associés à f

Quelles sont les fonctions significativement sur-représentées par l'échantillon de gènes ?

Test d'enrichissement fonctionnel

Test hypergéométrique

Quelle est la probabilité qu'un échantillon de taille g tiré aléatoirement dans le génome contienne au moins e gènes étiquetés par f ?

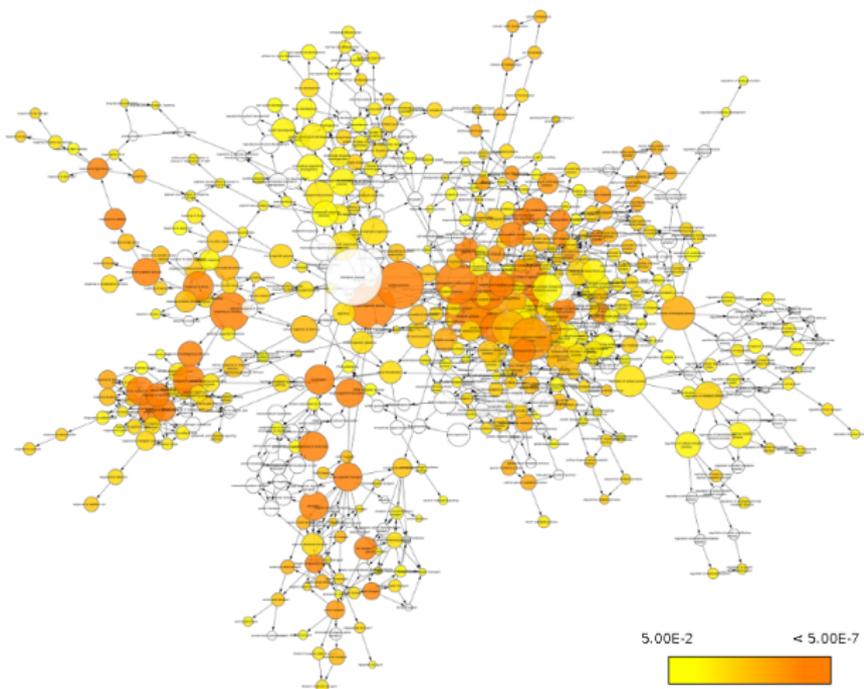
$$P(X \geq e) = 1 - \sum_{k=1}^{e-1} \frac{\binom{E}{k} \binom{G-E}{g-k}}{\binom{G}{g}}$$

Seules les p-valeurs inférieures à 0.05 sont considérées.

Test d'enrichissement fonctionnel

Outils : BiNGO, App de Cytoscape

Ontologie des fonctions sur-représentées par les données sens :

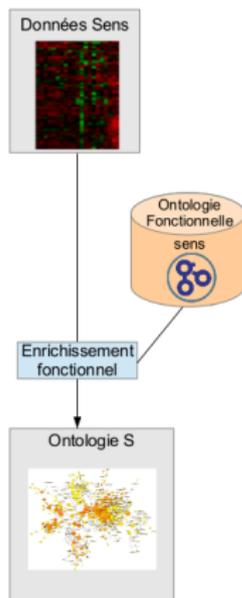


Analyse fonctionnelle différentielle

- F_S
- F_{SAS}
- *termes révélés par les anti-sens*
 $= F_{SAS} \setminus F_S$

Termes révélés par les anti-sens

Fonctions biologiques
significativement sur-représentées
uniquement en ajoutant les données
anti-sens aux données sens

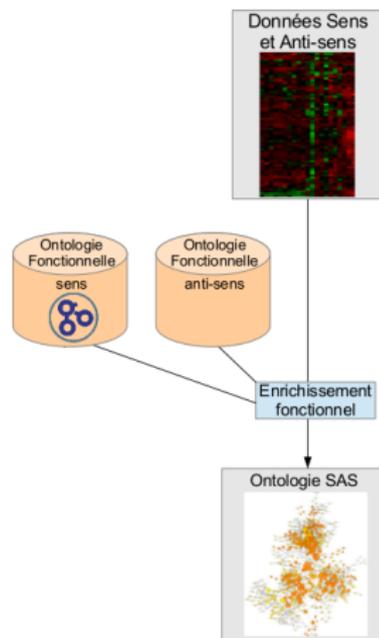


Analyse fonctionnelle différentielle

- F_S
- F_{SAS}
- *termes révélés par les anti-sens*
 $= F_{SAS} \setminus F_S$

Termes révélés par les anti-sens

Fonctions biologiques
significativement sur-représentées
uniquement en ajoutant les données
anti-sens aux données sens

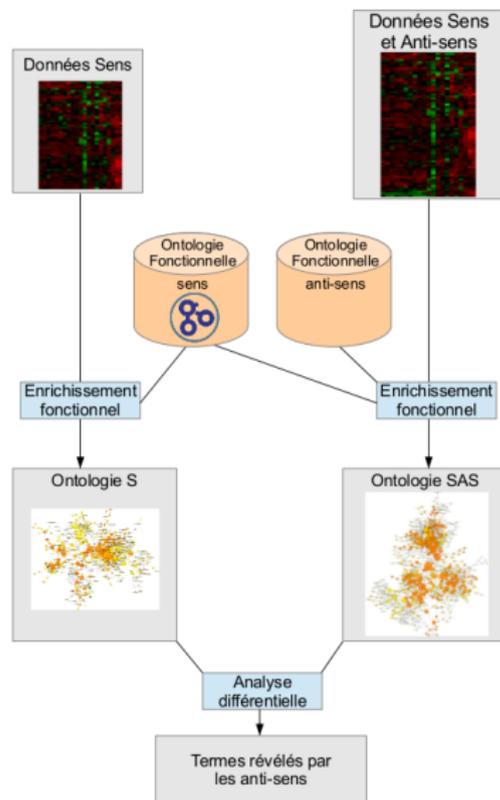


Analyse fonctionnelle différentielle

- F_S
- F_{SAS}
- *termes révélés par les anti-sens*
 $= F_{SAS} \setminus F_S$

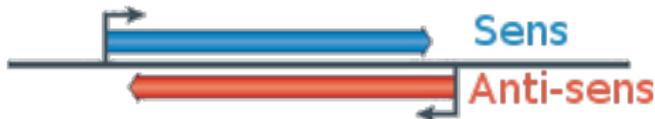
Termes révélés par les anti-sens

Fonctions biologiques
 significativement sur-représentées
 uniquement en ajoutant les données
 anti-sens aux données sens



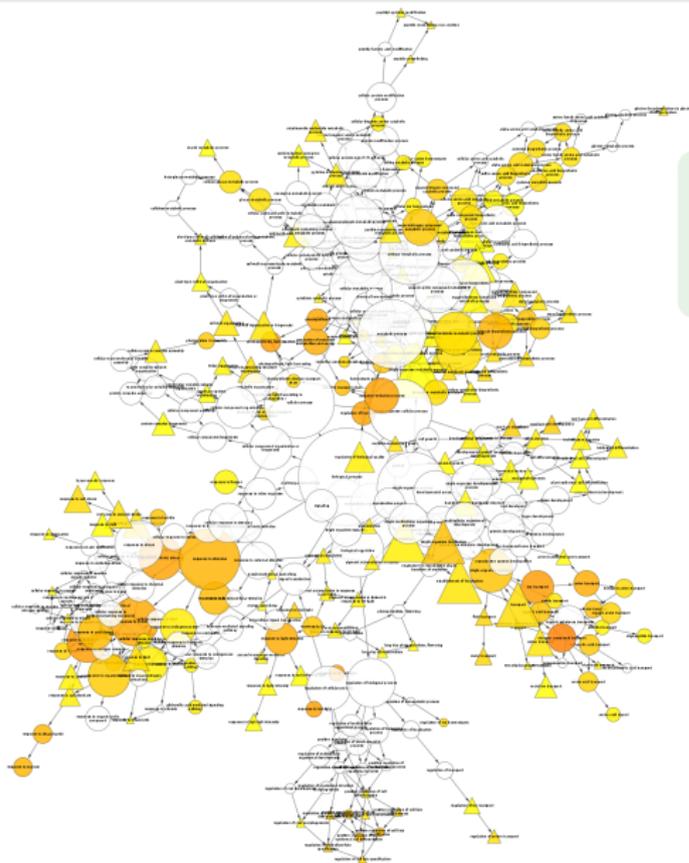
Gene Ontology Pommier

- Pas d'annotation Pommier dans la GO
- ⇒ Association des gènes du pommier avec les catégories des orthologues d'Arabidopsis
- Les gènes sans orthologues sont étiquetés « fonction inconnue »



- Pas d'annotation pour les anti-sens
- ⇒ Association des anti-sens avec la catégorie du sens complémentaire

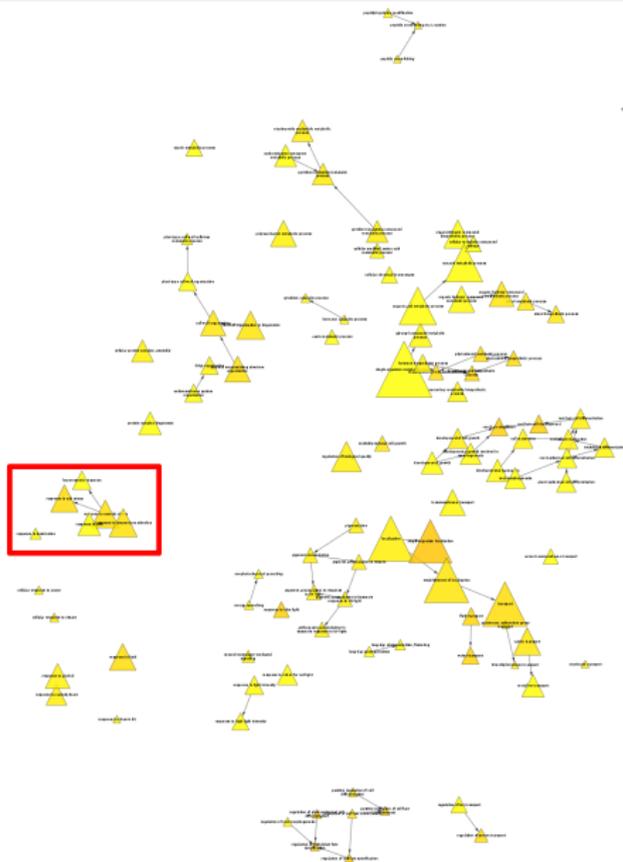
Résultats graphiques



Ontologie SAS avec les termes révélés par les anti-sens mis en valeur

- 104 termes révélés par les anti-sens
- Les termes révélés par les anti-sens sont répartis dans l'ontologie
- Les termes révélés par les anti-sens sont relatifs à la condition expérimentale

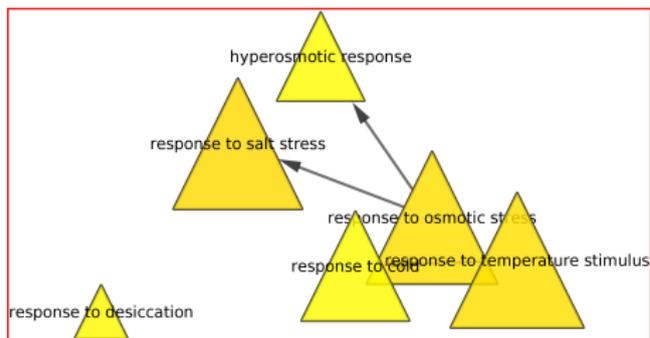
Résultats graphiques



Ontologie SAS avec les termes révélés par les anti-sens mis en valeur

- 104 termes révélés par les anti-sens
- Les termes révélés par les anti-sens sont répartis dans l'ontologie
- Les termes révélés par les anti-sens sont relatifs à la condition expérimentale

Résultats graphiques



Ontologie SAS avec les termes révélés par les anti-sens mis en valeur

- 104 termes révélés par les anti-sens
- Les termes révélés par les anti-sens sont répartis dans l'ontologie
- Les termes révélés par les anti-sens sont relatifs à la condition expérimentale

Interprétation biologique

- Analyse fonctionnelle sens :
 - termes liés aux hormones
 - termes liés au transport
 - ⇒ résultats cohérents avec la condition expérimentale
 - Termes révélés par les anti-sens :
 - paroi cellulaire
 - réponse osmotique
 - réponse au froid
 - ⇒ résultats cohérents avec la condition expérimentale
- ⇒ Les transcrits anti-sens jouent un rôle particulier dans le processus

Évolution des transcrits associés aux termes révélés par les anti-sens

hyperosmotic response

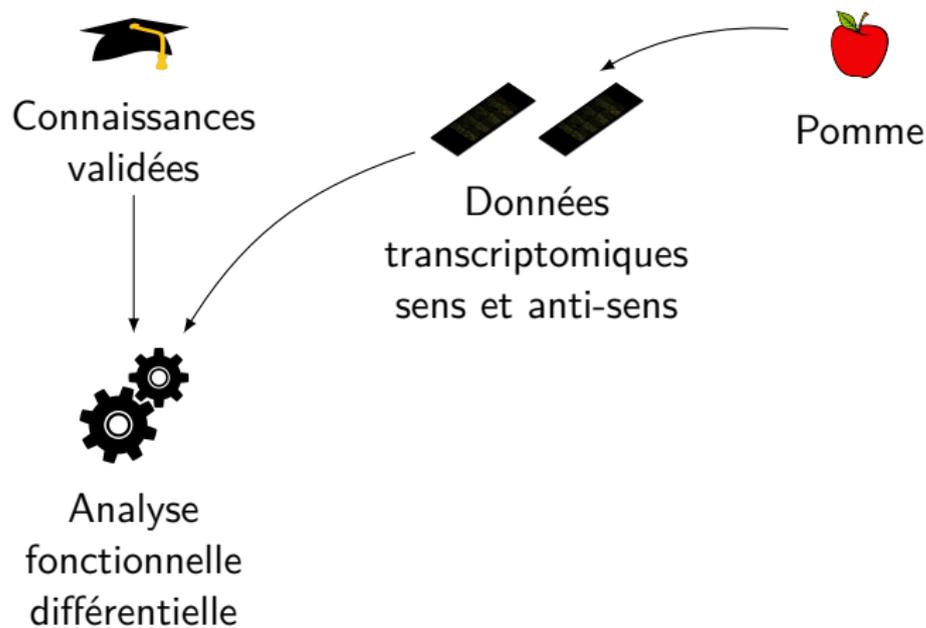
S	e(S)	e(AS)	AS
MDP0000398010_R	-1	-1	MDP0000398010
MDP0000598183_R	-1	-1	MDP0000598183
MDP0000237978_R	-1	-1	MDP0000237978
MDP0000561026_R	-1	-1	MDP0000561026
MDP0000151721_R	-1	-1	MDP0000151721
MDP0000272802_R	-1	-1	MDP0000272802
MDP0000283052_R	-1	-1	MDP0000283052
MDP0000920069_R	-1	-1	MDP0000920069
MDP0000797759_R	-1	-1	MDP0000797759
MDP0000323330_R	-1	-1	MDP0000323330
MDP0000187714_R	-1	-1	MDP0000187714
MDP0000952079_R	-1	-1	MDP0000952079
MDP0000250138_R	-1	-1	MDP0000250138
MDP0000324254_R	-1	-1	MDP0000324254
MDP0000632173_R	-1	-1	MDP0000632173
MDP0000505439_R	-1	-1	MDP0000505439
MDP0000165604_R	-1	-1	MDP0000165604
MDP0000094767_R	-1	-1	MDP0000094767
MDP0000360864_R	-1	-1	MDP0000360864
MDP0000258530_R	-1	-1	MDP0000258530
MDP0000595826_R	-1	-1	MDP0000595826
MDP0000216726_R	-1	-1	MDP0000216726
MDP0000938969_R	-1	-1	MDP0000938969
MDP0000389181_R	-1	-1	MDP0000389181
MDP0000841405_R	-1	-1	MDP0000841405
MDP0000331536_R	-1	-1	MDP0000331536
MDP0000263161_R	1	-1	MDP0000263161
MDP0000813339_R	1	-1	MDP0000813339
MDP0000293432_R	-1	1	MDP0000293432
MDP0000766240_R	-1	1	MDP0000766240
MDP0000464276_R	1	1	MDP0000464276

- 1 625 transcrits d'intérêts : 1 425 couples sens/anti-sens
- Observation de l'évolution des transcrits

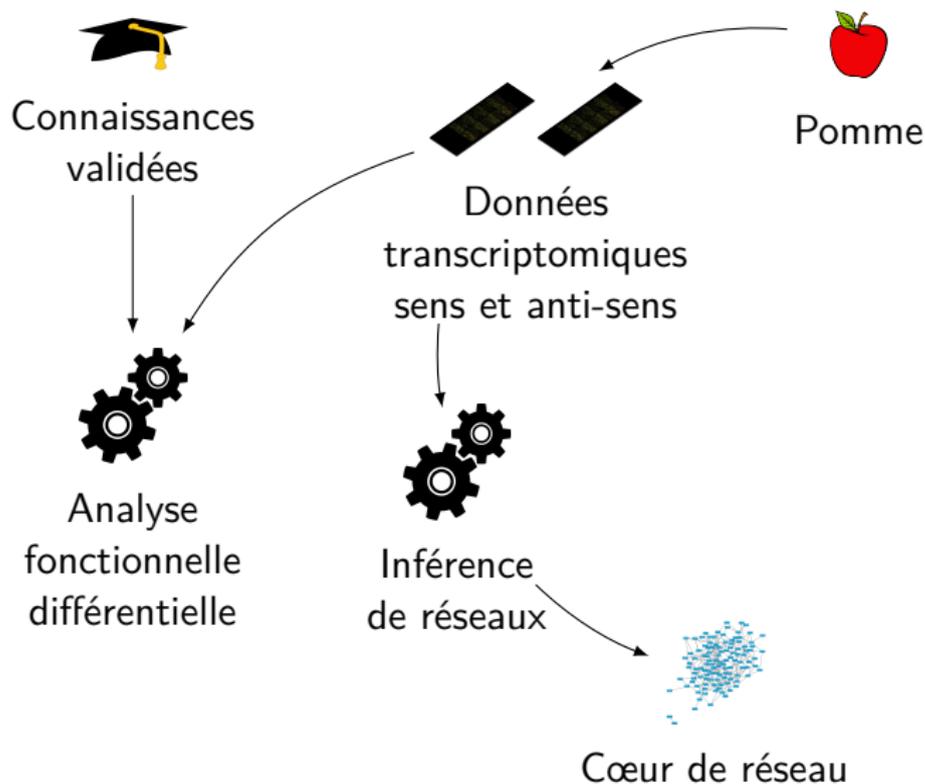
Recherche de motifs dans les séquences promotrices

- Analyse des séquences promotrices des sens et des anti-sens du terme *response to cold*
- Estimation de la région promotrice de l'anti-sens
- Recherche de séquences de facteurs de transcription connus
- Recherche de motifs particuliers dans la région promotrice

Inférence et Réseaux de gènes

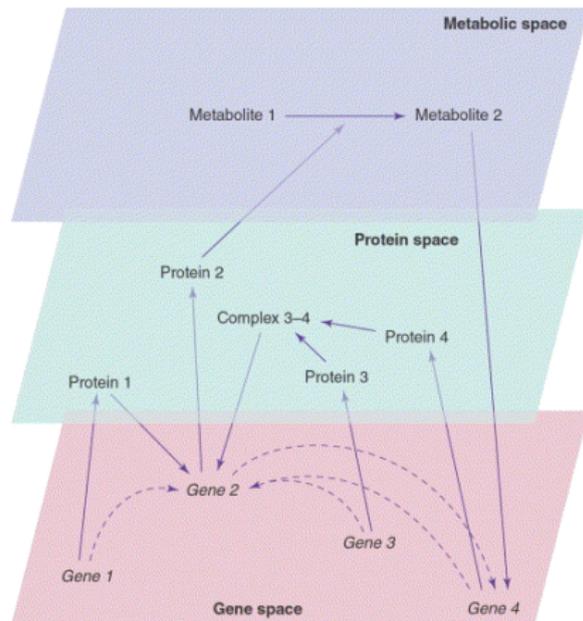


Inférence et Réseaux de gènes



Réseaux de gènes

- Type de modélisation en *biologie des systèmes*
- Projection des interactions au niveau génomique
- Modèle graphique



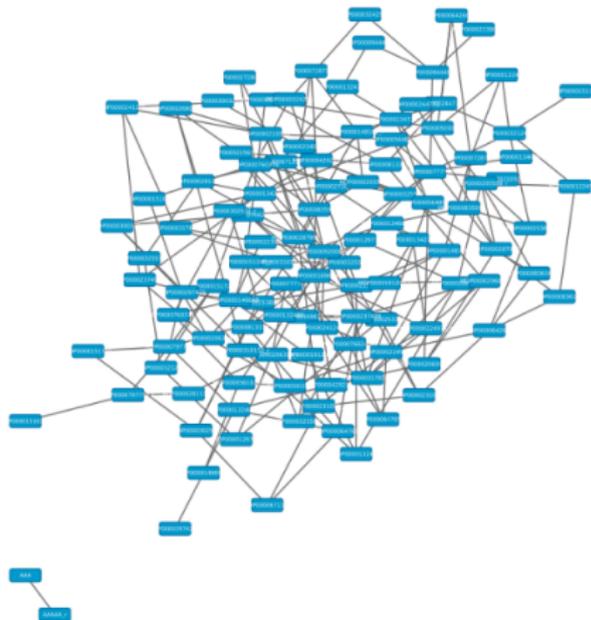
TRENDS in Biotechnology

Brazhnik, P., Fuente, A. d. I., et Mendes, P. (2002). Gene networks : how to put the function in genomics. Trends in Biotechnology, 20(11) :467-472

Inférence de réseaux

P échantillons de N gènes

$$X = \begin{pmatrix} X_{1,1} & \cdots & X_{1,P} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & X_{i,j} & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ X_{N,1} & \cdots & X_{N,P} \end{pmatrix}$$



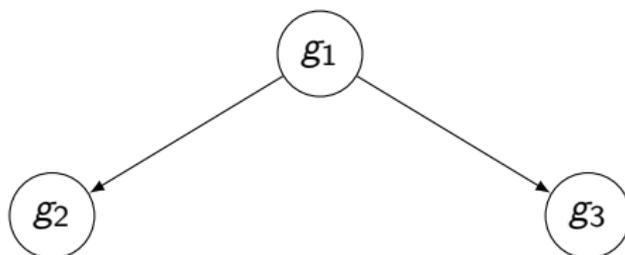
Nœud : gène

Lien : interaction **significative** entre deux gènes

Méthodes d'inférence de réseaux

De nombreuses méthodes d'inférence de réseaux de gènes ont été proposées. Elles diffèrent selon :

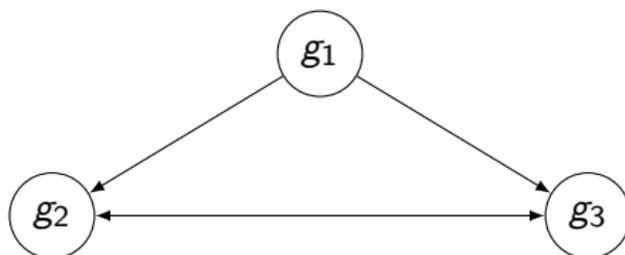
- Détermination des interactions
- Suppression des redondances



Méthodes d'inférence de réseaux

De nombreuses méthodes d'inférence de réseaux de gènes ont été proposées. Elles diffèrent selon :

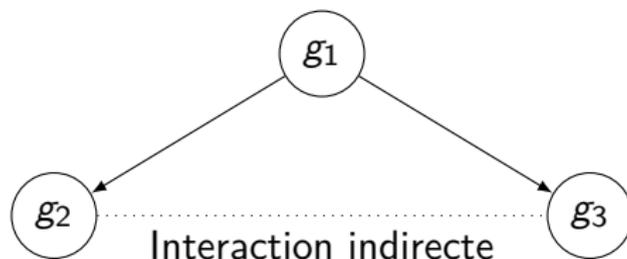
- Détermination des interactions
- Suppression des redondances



Méthodes d'inférence de réseaux

De nombreuses méthodes d'inférence de réseaux de gènes ont été proposées. Elles diffèrent selon :

- Détermination des interactions
- Suppression des redondances



Extended Core Network

Méthode d'inférence proposée

- Utilise l'*information mutuelle* : identifie les dépendances paire à paire
- Construit un « cœur de réseau étendu » : les meilleures interactions pour chaque gène
- Méthode inspirée de C3NET (Altay et Emmert-Streib, 2010)
- Permet d'identifier les interactions significatives du réseau pour l'analyse différentielle

ECN

L'algorithme

- Utilisation d'une matrice d'information mutuelle IM
- Détermination des meilleures interactions (ECN_BI) avec un taux d'acceptation $r = 0.05$

IM	g_1	g_2	g_3	g_4
g_1		0.70	1.10	0.80
g_2	0.70		0.69	0.50
g_3	1.10	0.69		0.70
g_4	0.80	0.50	0.70	

ECN

L'algorithme

- Utilisation d'une matrice d'information mutuelle IM
- Détermination des meilleures interactions (ECN_BI) avec un taux d'acceptation $r = 0.05$

IM	g_1	g_2	g_3	g_4
g_1		0.70	1.10	0.80
g_2	0.70		0.69	0.50
g_3	1.10	0.69		0.70
g_4	0.80	0.50	0.70	

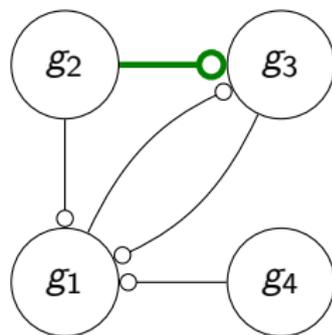
ECN

L'algorithme

- Utilisation d'une matrice d'information mutuelle IM
- Détermination des meilleures interactions (ECN_BI) avec un taux d'acceptation $r = 0.05$
- ECN_BI : $g_u \text{ --- } g_v$ indique que, pour g_u , l'interaction $IM(g_u, g_v)$ est une valeur maximale

IM	g_1	g_2	g_3	g_4
g_1			1.10	
g_2	0.70		0.69	
g_3	1.10			
g_4	0.80			

⇒



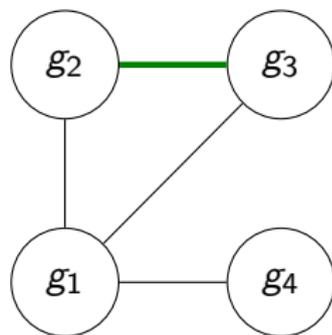
ECN

L'algorithme

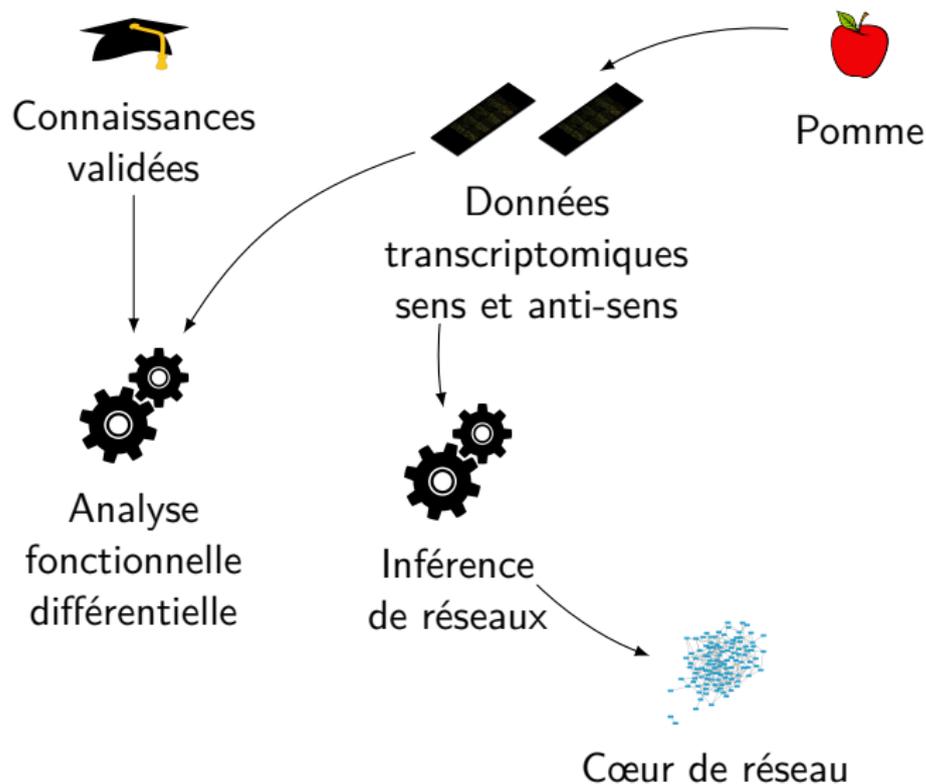
- Utilisation d'une matrice d'information mutuelle IM
- Détermination des meilleures interactions (ECN_BI) avec un taux d'acceptation $r = 0.05$
- ECN_BI : $g_u - o g_v$ indique que, pour g_u , l'interaction $IM(g_u, g_v)$ est une valeur maximale
- ECN : Cœur de réseau, non-orienté

IM	g_1	g_2	g_3	g_4
g_1		0.70	1.10	0.80
g_2	0.70		0.69	
g_3	1.10	0.69		
g_4	0.80			

⇒

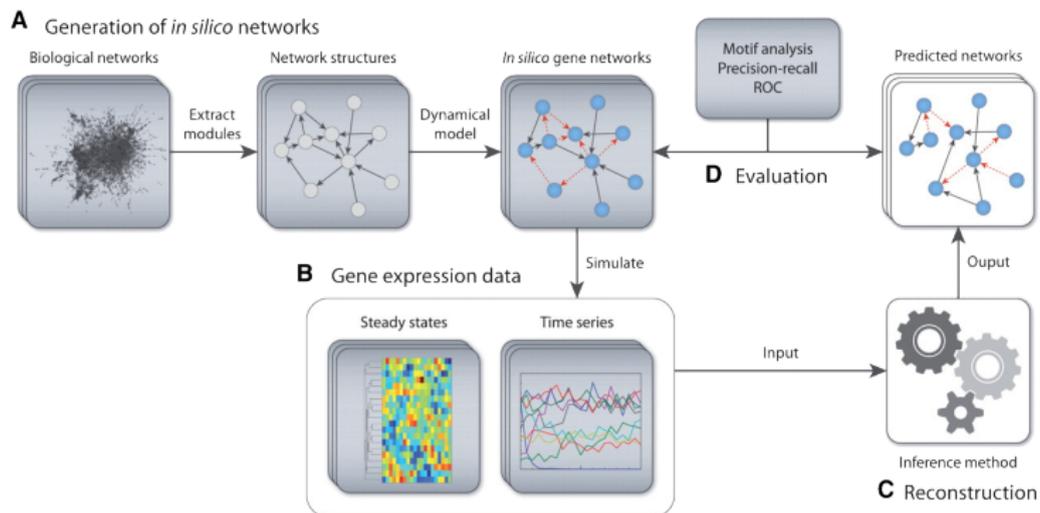


Évaluation



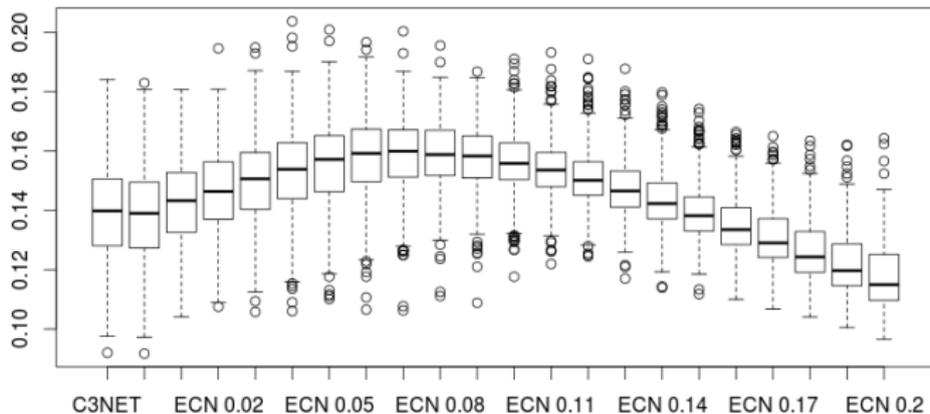
Validation des méthodes d'inférence

- Étape classique dans l'apprentissage
- Évaluer les méthodes d'inférence avec différents indicateurs
- Utilisation de générateur de données *in silico*



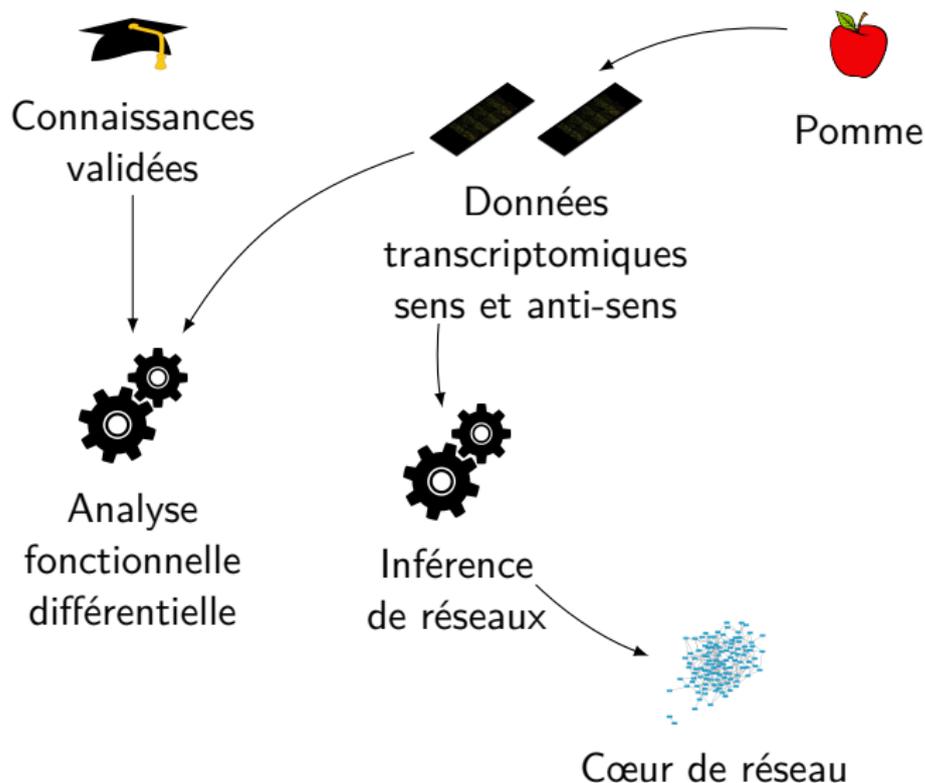
Schaffter, T., Marbach, D., et Floreano, D. (2011). GeneNetWeaver : *in silico* benchmark generation and performance profiling of network inference methods. *Bioinformatics*, 27(16) :2263–2270.

Évaluation d'ECN par rapport à C3NET

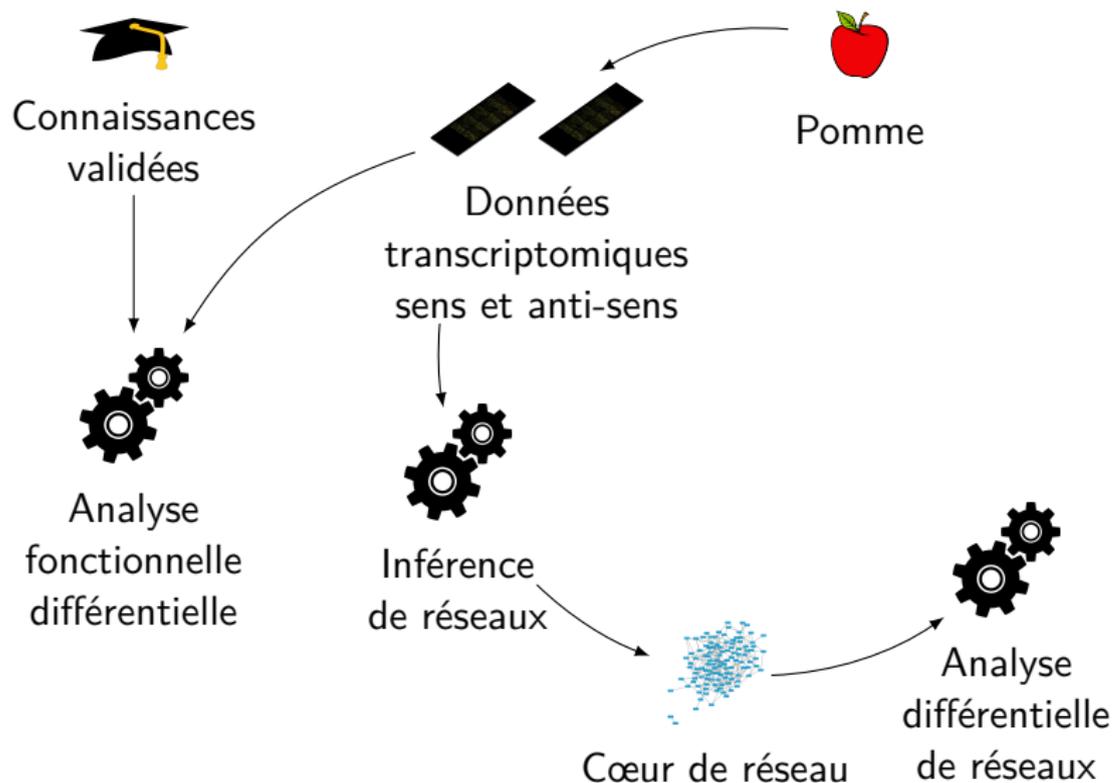


- Évaluation sur des données simulées à partir de réseaux de bactérie et de levure
- Indicateur utilisé : F -score
- Amélioration par rapport à C3NET pour un faible taux (entre 0.05 et 0.10)

Analyse différentielle de réseaux

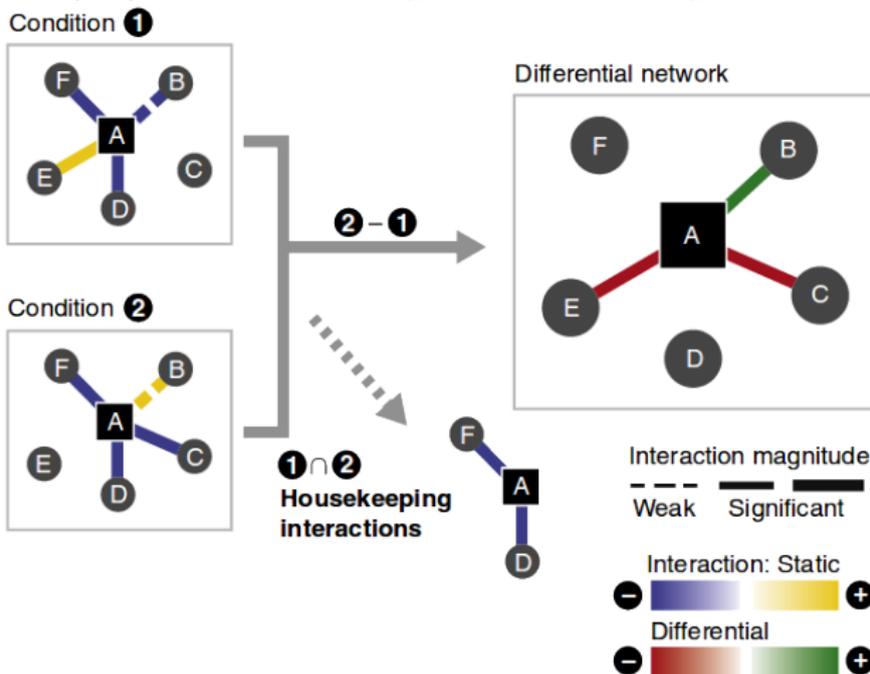


Analyse différentielle de réseaux



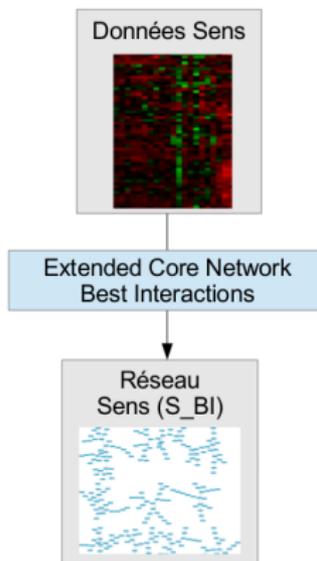
Analyse différentielle de réseaux

Ideker, T. et Krogan, N. J. (2012). Differential network biology. *Molecular Systems Biology*, 8 :565.

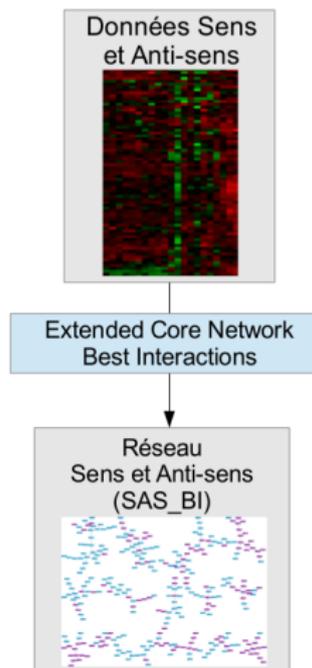


Toujours un même ensemble de gènes dans différentes conditions

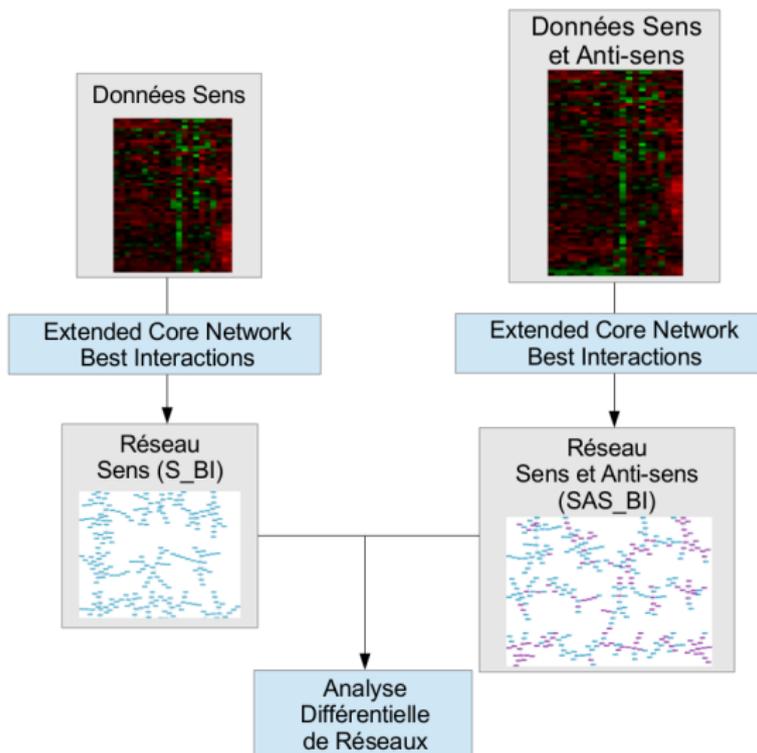
Analyse différentielle de réseaux



Analyse différentielle de réseaux



Analyse différentielle de réseaux

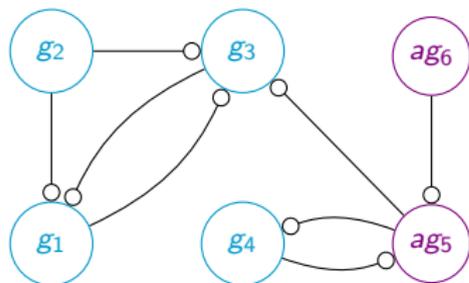
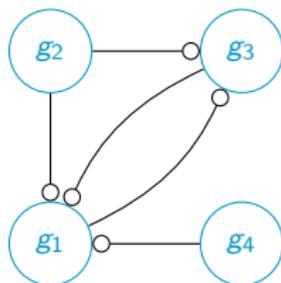


Comparaison des réseaux

Construction des réseaux S et SAS

IM	g_1	g_2	g_3	g_4
g_1			1.10	
g_2	0.70		0.69	
g_3	1.10			
g_4	0.80			

IM	g_1	g_2	g_3	g_4	ag_5	ag_6
g_1			1.10			
g_2	0.70		0.69			
g_3	1.10				1.10	
g_4					1.05	
ag_5			1.10	1.05		
ag_6					0.90	

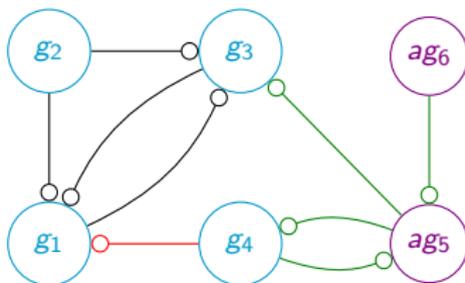


Comparaison des réseaux

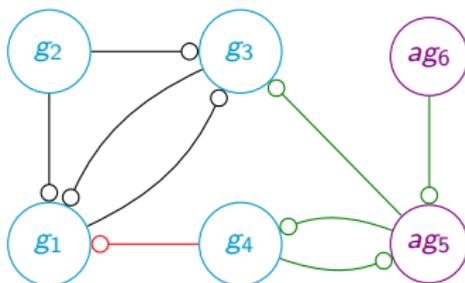
Construction d'un graphe de changements

IM	g_1	g_2	g_3	g_4
g_1			1.10	
g_2	0.70		0.69	
g_3	1.10			
g_4	0.80			

IM	g_1	g_2	g_3	g_4	ag_5	ag_6
g_1			1.10			
g_2	0.70		0.69			
g_3	1.10				1.10	
g_4					1.05	
ag_5			1.10	1.05		
ag_6					0.90	



Gènes AS-impactés



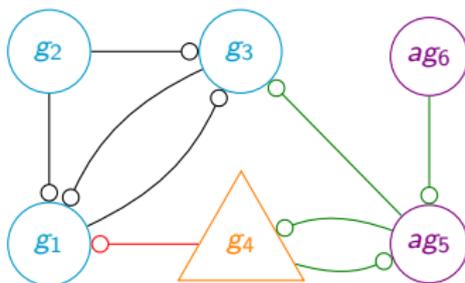
Gène AS-impacté

Soient $S = (V_S, A_S)$ un cœur de réseau orienté inféré par ECN_BI à partir de G_S et $SAS = (V_{SAS}, A_{SAS})$ un cœur de réseau orienté inféré par ECN_BI à partir de $G_S \cup G_{AS}$.

Un gène $g \in G_S$ est un *gène AS-impacté* si, et seulement si :

$$N_S(g) \cap N_{SAS}(g) = \emptyset$$

Gènes AS-impactés



Gène AS-impacté

Soient $S = (V_S, A_S)$ un cœur de réseau orienté inféré par ECN_BI à partir de G_S et $SAS = (V_{SAS}, A_{SAS})$ un cœur de réseau orienté inféré par ECN_BI à partir de $G_S \cup G_{AS}$.

Un gène $g \in G_S$ est un *gène AS-impacté* si, et seulement si :

$$N_S(g) \cap N_{SAS}(g) = \emptyset$$

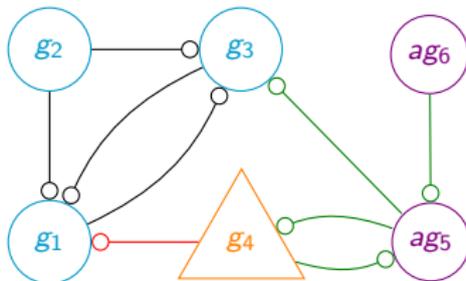
Motifs de changements

Motif de changements

Soient $S = (V_S, A_S)$ un cœur de réseau orienté inféré par ECN_BI à partir de G_S et $SAS = (V_{SAS}, A_{SAS})$ un cœur de réseau orienté inféré par ECN_BI à partir de $G_S \cup G_{AS}$. Soit $g \in G_S$ un gène AS-impacté.

Le *motif de changements* autour de g est un graphe $M = (V_M, A_M)$ tel que :

- $V_M = \{g\} \cup N_S(g) \cup N_{SAS}(g) \cup \bigcup_{v \in N_{SAS}(g)} N_{SAS}(v)$
- $A_M = \{(v_1, v_2) \in A_S \mid v_1, v_2 \in V_M\} \cup \{(v_1, v_2) \in A_{SAS} \mid v_1, v_2 \in V_M\}$



Motifs de changements

Motif de changements

Soient $S = (V_S, A_S)$ un cœur de réseau orienté inféré par ECN_BI à partir de G_S et $SAS = (V_{SAS}, A_{SAS})$ un cœur de réseau orienté inféré par ECN_BI à partir de $G_S \cup G_{AS}$. Soit $g \in G_S$ un gène AS-impacté.

Le *motif de changements* autour de g est un graphe $M = (V_M, A_M)$ tel que :

- $V_M = \{g\} \cup N_S(g) \cup N_{SAS}(g) \cup \bigcup_{v \in N_{SAS}(g)} N_{SAS}(v)$
- $A_M = \{(v_1, v_2) \in A_S \mid v_1, v_2 \in V_M\} \cup \{(v_1, v_2) \in A_{SAS} \mid v_1, v_2 \in V_M\}$



Motifs de changements

Motif de changements

Soient $S = (V_S, A_S)$ un cœur de réseau orienté inféré par ECN_BI à partir de G_S et $SAS = (V_{SAS}, A_{SAS})$ un cœur de réseau orienté inféré par ECN_BI à partir de $G_S \cup G_{AS}$. Soit $g \in G_S$ un gène AS-impacté.

Le *motif de changements* autour de g est un graphe $M = (V_M, A_M)$ tel que :

- $V_M = \{g\} \cup N_S(g) \cup N_{SAS}(g) \cup \bigcup_{v \in N_{SAS}(g)} N_{SAS}(v)$
- $A_M = \{(v_1, v_2) \in A_S \mid v_1, v_2 \in V_M\} \cup \{(v_1, v_2) \in A_{SAS} \mid v_1, v_2 \in V_M\}$



Motifs de changements

Motif de changements

Soient $S = (V_S, A_S)$ un cœur de réseau orienté inféré par ECN_BI à partir de G_S et $SAS = (V_{SAS}, A_{SAS})$ un cœur de réseau orienté inféré par ECN_BI à partir de $G_S \cup G_{AS}$. Soit $g \in G_S$ un gène AS-impacté.

Le *motif de changements* autour de g est un graphe $M = (V_M, A_M)$ tel que :

- $V_M = \{g\} \cup N_S(g) \cup N_{SAS}(g) \cup \bigcup_{v \in N_{SAS}(g)} N_{SAS}(v)$
- $A_M = \{(v_1, v_2) \in A_S \mid v_1, v_2 \in V_M\} \cup \{(v_1, v_2) \in A_{SAS} \mid v_1, v_2 \in V_M\}$



Motifs de changements

Motif de changements

Soient $S = (V_S, A_S)$ un cœur de réseau orienté inféré par ECN_BI à partir de G_S et $SAS = (V_{SAS}, A_{SAS})$ un cœur de réseau orienté inféré par ECN_BI à partir de $G_S \cup G_{AS}$. Soit $g \in G_S$ un gène AS-impacté.

Le *motif de changements* autour de g est un graphe $M = (V_M, A_M)$ tel que :

- $V_M = \{g\} \cup N_S(g) \cup N_{SAS}(g) \cup \bigcup_{v \in N_{SAS}(g)} N_{SAS}(v)$
- $A_M = \{(v_1, v_2) \in A_S \mid v_1, v_2 \in V_M\} \cup \{(v_1, v_2) \in A_{SAS} \mid v_1, v_2 \in V_M\}$



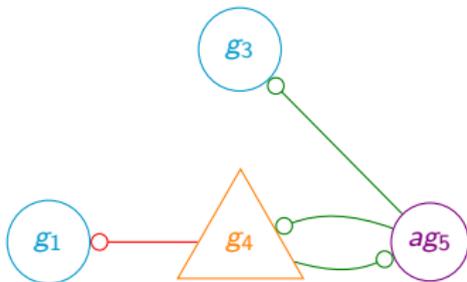
Motifs de changements

Motif de changements

Soient $S = (V_S, A_S)$ un cœur de réseau orienté inféré par ECN_BI à partir de G_S et $SAS = (V_{SAS}, A_{SAS})$ un cœur de réseau orienté inféré par ECN_BI à partir de $G_S \cup G_{AS}$. Soit $g \in G_S$ un gène AS-impacté.

Le *motif de changements* autour de g est un graphe $M = (V_M, A_M)$ tel que :

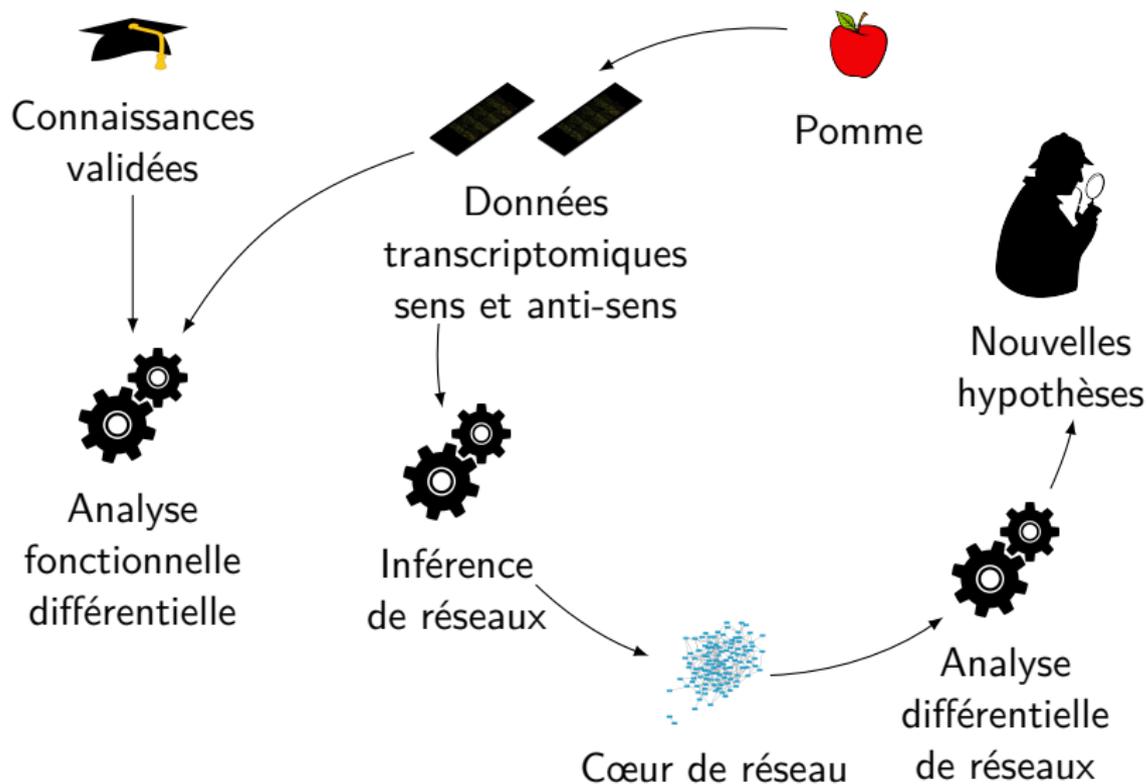
- $V_M = \{g\} \cup N_S(g) \cup N_{SAS}(g) \cup \bigcup_{v \in N_{SAS}(g)} N_{SAS}(v)$
- $A_M = \{(v_1, v_2) \in A_S \mid v_1, v_2 \in V_M\} \cup \{(v_1, v_2) \in A_{SAS} \mid v_1, v_2 \in V_M\}$



Résultats

- Utilisation d'ECN à 0.05
- Rappel : 931 sens, 624 anti-sens : 1 625 nœuds SAS
- 308 gènes AS-impactés ($\approx 30\%$ des sens) donc 308 motifs de changements

Nouvelles hypothèses



Interprétation biologique

Enrichissement des motifs avec la GO slim :

- Sur 308 motifs, 72 sont enrichis pour un ensemble de 291 gènes
 - Processus identifiés (127 gènes – 64 motifs) :
 - maturation
 - stress oxydatif
 - stress au froid
 - Voie biologique peut-être sous-estimée :
 - brassinostéroïdes
- ⇒ L'intégration des transcrits anti-sens permet de mettre en lumière de nouvelles régulations

Reconfiguration des interactions entre gènes AS-impactés



- Recherche de connexions du cœur de réseau S qui ne sont pas représentées dans le cœur de réseau SAS : composante connexe de sens AS-impactés
 - Recherche des connexions des gènes AS-impactés dans le cœur de réseau SAS
- ⇒ Permet de voir si des connexions du réseau S se retrouvent dans le cœur de réseau SAS grâce à des anti-sens

Reconfiguration des interactions entre gènes AS-impactés



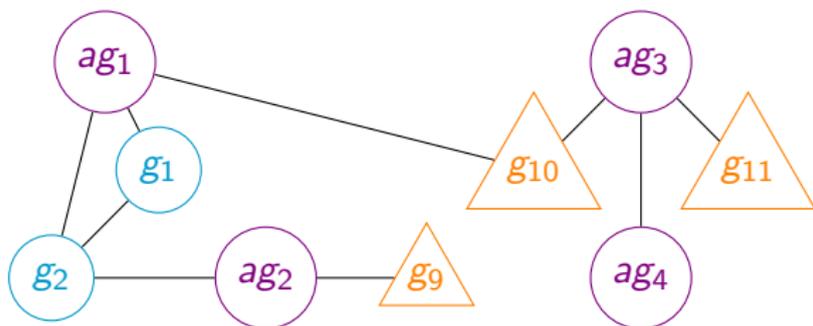
- Recherche de connexions du cœur de réseau S qui ne sont pas représentées dans le cœur de réseau SAS : composante connexe de sens AS-impactés
 - Recherche des connexions des gènes AS-impactés dans le cœur de réseau SAS
- ⇒ Permet de voir si des connexions du réseau S se retrouvent dans le cœur de réseau SAS grâce à des anti-sens

Reconfiguration des interactions entre gènes AS-impactés



- Recherche de connexions du cœur de réseau S qui ne sont pas représentées dans le cœur de réseau SAS : composante connexe de sens AS-impactés
 - Recherche des connexions des gènes AS-impactés dans le cœur de réseau SAS
- ⇒ Permet de voir si des connexions du réseau S se retrouvent dans le cœur de réseau SAS grâce à des anti-sens

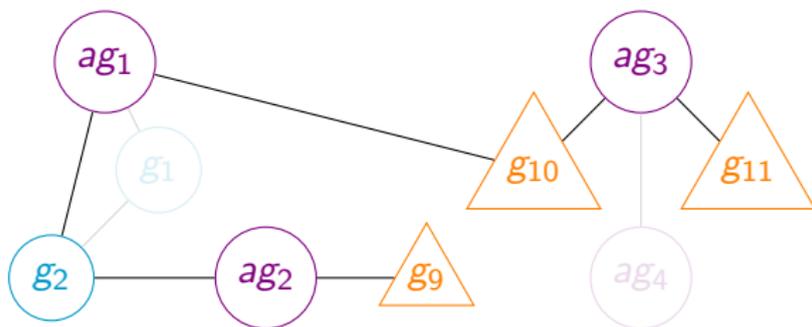
Reconfiguration des interactions entre gènes AS-impactés



- Recherche de connexions du cœur de réseau S qui ne sont pas représentées dans le cœur de réseau SAS : composante connexe de sens AS-impactés
- Recherche des connexions des gènes AS-impactés dans le cœur de réseau SAS

⇒ Permet de voir si des connexions du réseau S se retrouvent dans le cœur de réseau SAS grâce à des anti-sens

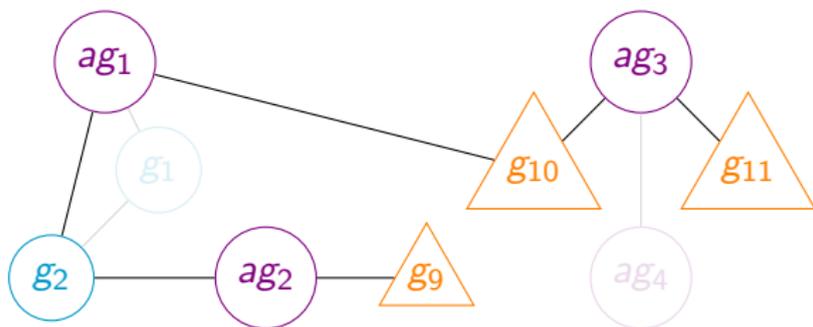
Reconfiguration des interactions entre gènes AS-impactés



- Recherche de connexions du cœur de réseau S qui ne sont pas représentées dans le cœur de réseau SAS : composante connexe de sens AS-impactés
- Recherche des connexions des gènes AS-impactés dans le cœur de réseau SAS

⇒ Permet de voir si des connexions du réseau S se retrouvent dans le cœur de réseau SAS grâce à des anti-sens

Reconfiguration des interactions entre gènes AS-impactés



- Recherche de connexions du cœur de réseau S qui ne sont pas représentées dans le cœur de réseau SAS : composante connexe de sens AS-impactés
 - Recherche des connexions des gènes AS-impactés dans le cœur de réseau SAS
- ⇒ Permet de voir si des connexions du réseau S se retrouvent dans le cœur de réseau SAS grâce à des anti-sens

Recherche de reconnections

Arbres de Steiner

Problème de l'arbre de Steiner minimal

Soit un graphe $G = (V, A)$ et un ensemble de nœuds T tel que $T \subseteq V$.
Un *arbre de Steiner* $G' = (V', A')$ est un sous-graphe de G tel que :

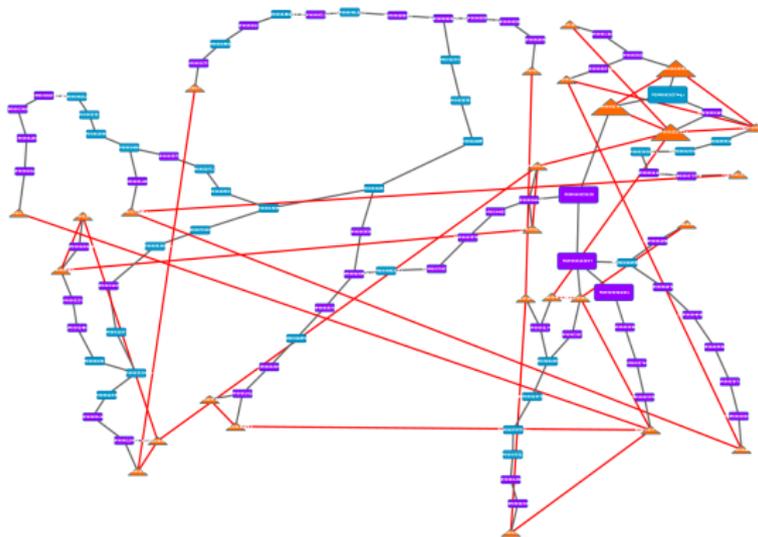
- $T \subseteq V' \subseteq V$,
- $A' \subseteq A$,
- il existe un chemin entre chaque paire de T dans G' .

G' est un *arbre de Steiner minimal* si $|A'|$ est minimal.

- Problème NP-complet
- Pas de méthode exacte utilisable
- Approche par la sélection des plus courts chemins
- Deux versions :
 - Nombre d'arêtes
 - Distance ontologique

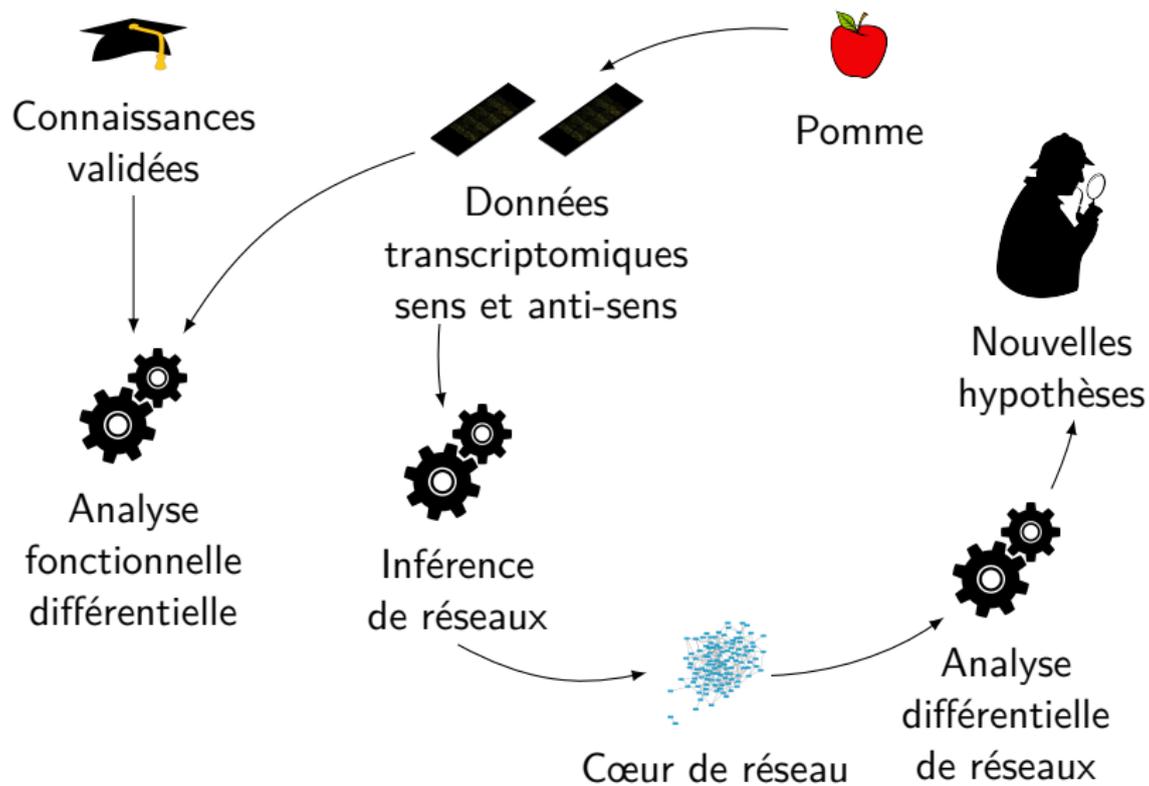
Résultats et interprétations

- 29 sous-graphes AS-impactés donnent 35 arbres de Steiner
- Enrichissement des gènes AS-impactés et des arbres
- Certains arbres sont composés de motifs de changements



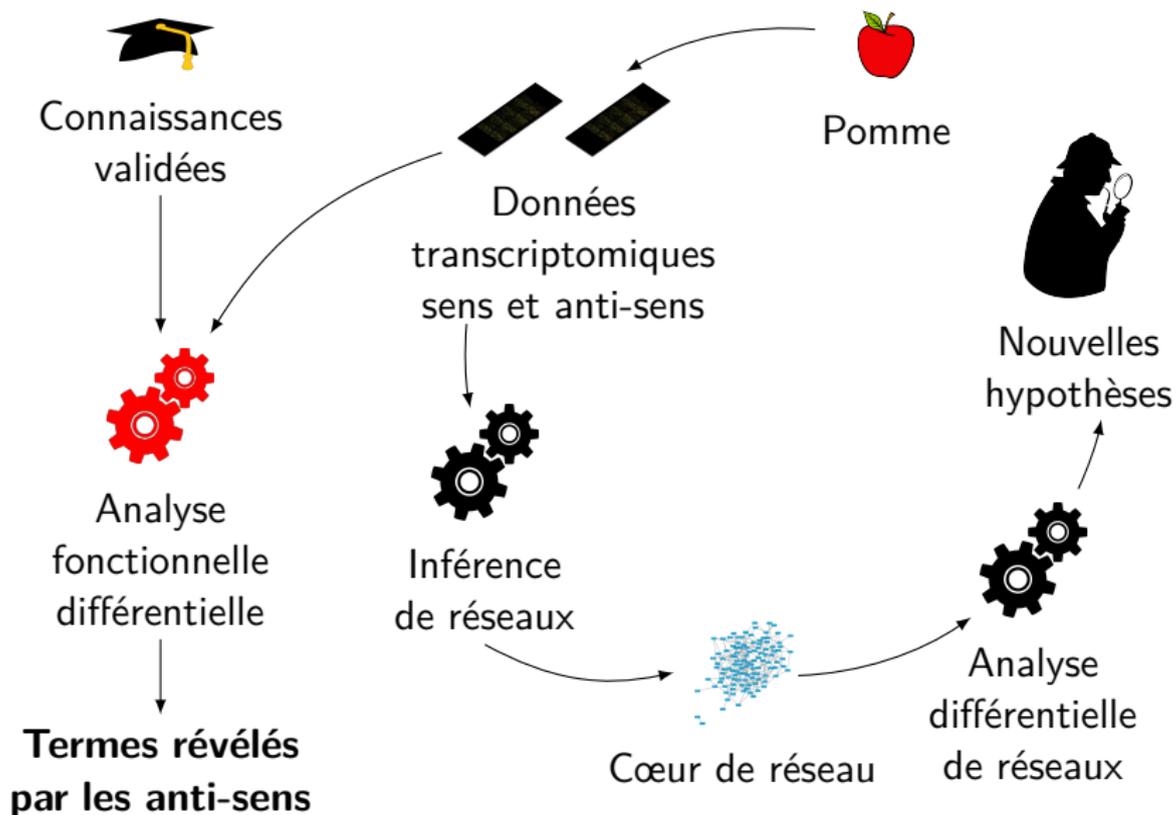
Étude exploratoire à large-échelle

de l'impact de la transcription anti-sens

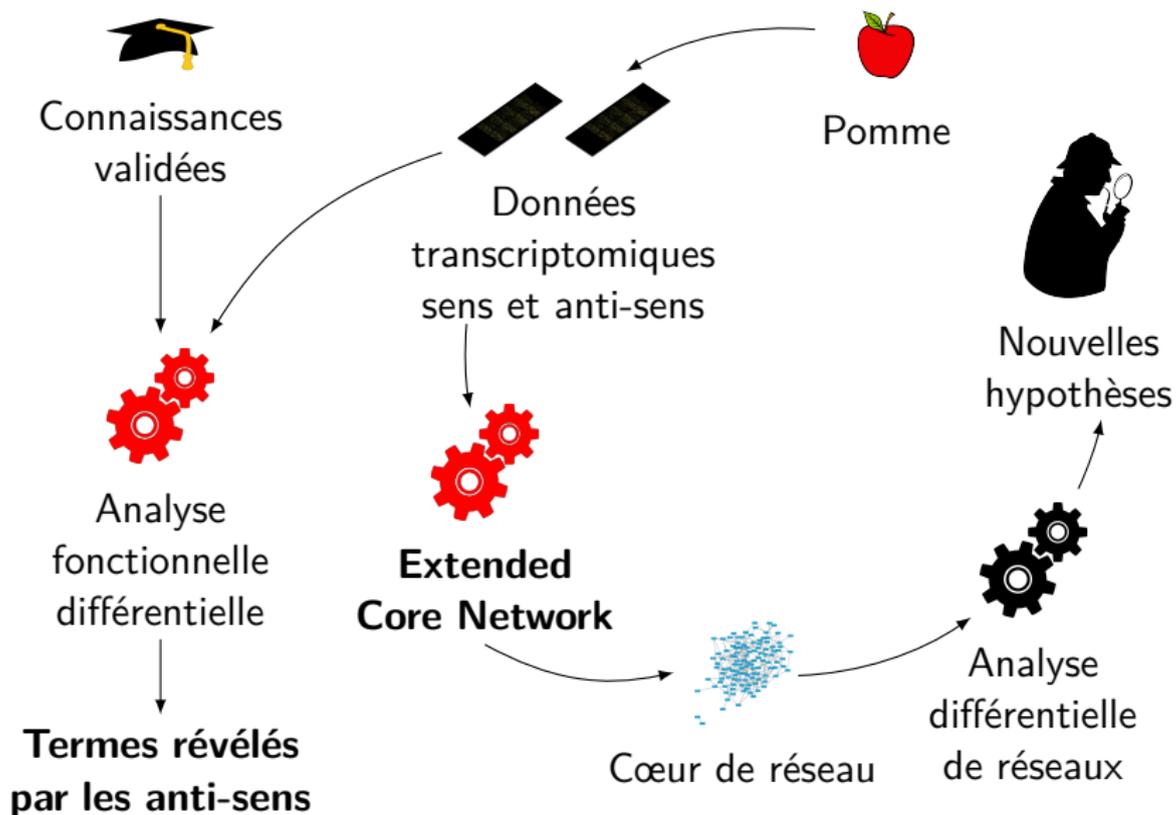


Étude exploratoire à large-échelle

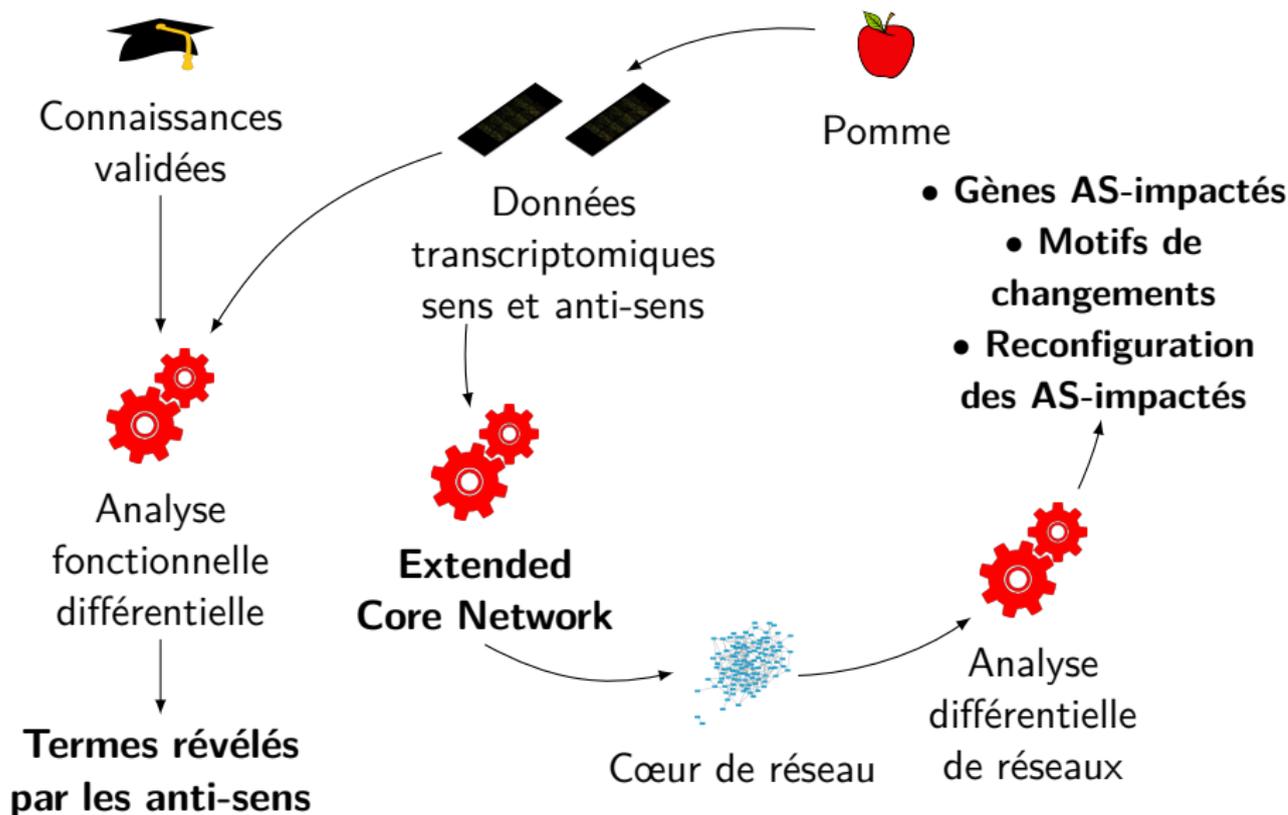
de l'impact de la transcription anti-sens



Étude exploratoire à large-échelle de l'impact de la transcription anti-sens



Étude exploratoire à large-échelle de l'impact de la transcription anti-sens



Perspectives

- Juillet 2017 : nouveau séquençage à l'IRHS
- 651 Mb ; 44 105 gènes putatifs
- Utilisation du nouveau séquençage
 - Nouvelles orthologie avec Arabidopsis
 - Meilleure définition du TSS (analyse des promoteurs)
 - ⇒ Interprétations biologiques plus précises
- Traitements développés en R utilisables avec ce nouveau séquençage

Perspectives

- Analyse différentielle de réseaux
 - Domaine en plein essor en médecine
 - Peu utilisé en biologie végétale
 - Nécessite des réseaux d'interactions robustes
- Constitution d'un réseau d'interactions robustes
 - Méthode d'ensemble learning
 - Réseau consensus entre plusieurs méthodes d'inférence

Article de revue internationale en cours de soumission

1. Large scale study of anti-sense regulation by differential network analysis.
Article « accept after revisions » pour une publication dans le journal *BMC Systems Biology*.

Conférences internationales

2. Differential network analysis of anti-sense regulation.
Bioinformatics and Biomedical Engineering - 5th International Work-Conference, IWBBIO 2017, de *Lecture Notes in Computer Science*, 2017.
3. Inference and differential analysis of extended core networks : A way to study anti-sense regulation.
IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2016, IEEE Computer Society, 2016.
4. Differential functional analysis and change motifs in gene networks to explore the role of anti-sense transcription.
Bioinformatics Research and Applications - 12th International Symposium, ISBRA 2016, de *Lecture Notes in Computer Science*. Springer, 2016.

Étude de la régulation anti-sens par l'analyse différentielle de données transcriptomiques dans le domaine végétal

Marc LEGEAY

LERIA

Université d'Angers
Laboratoire d'Étude et de Recherche en Informatique d'Angers
Institut de Recherche en Horticulture et Semences



Rapporteurs :

Mme Céline ROUVEIROL

Professeure des universités

Université Paris-Nord

Mme Fariza TAHI

Maître de conférences - HDR

Université d'Evry

Examineur :

M. Jérémie BOURDON

Professeur des universités

Université de Nantes

Direction de thèse :

Mme Béatrice DUVAL

Professeure des universités

Université d'Angers

M. Jean-Pierre RENO

Directeur de recherche

INRA d'Angers

12 décembre 2017

