

# Méthode d'évaluation de la pertinence d'un réseau

## Application à un jeu de données protéomiques

Elodie Marchadier

NETBIO

14 & 15 novembre 2023

# Plan de l'exposé

- Présentation du jeu de données
- Etude du jeu de données
- Construction de réseaux, méthodes disponibles
- Méthodes d'évaluation de la pertinence des réseaux
- Comparaison de méthodes de construction de réseaux

# Plan de l'exposé

- Présentation du jeu de données

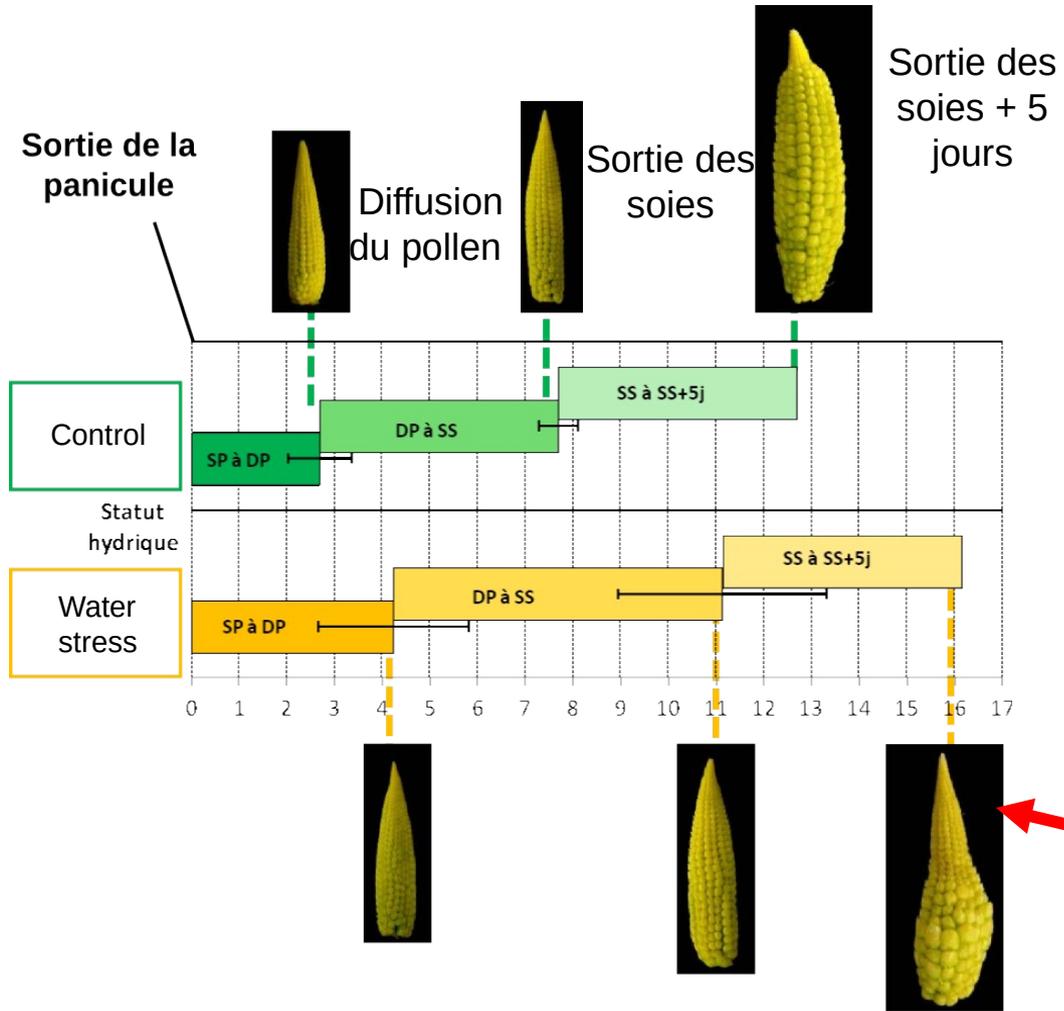
Etude du jeu de données

Construction de réseaux, méthodes disponibles

Méthodes d'évaluation de la pertinence des réseaux

Comparaison de méthodes de construction de réseaux

# Impact du stress hydrique sur le protéome des ovules de maïs



# Design expérimental

- **Conditions**

- water stress à l'apparition de la panicule
- control

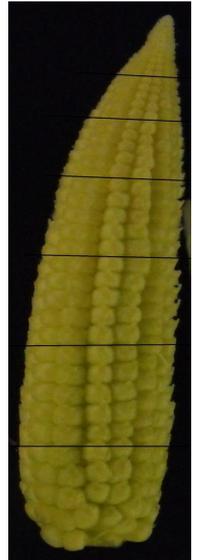
- **Temps**

- date de sortie des soies
- date de sortie des soies + 5 jours

- **Zones**

- zones 1 à 5 (4 zones en condition stress)

- **3 réplicats biologiques**



# Plan de l'exposé

Présentation du jeu de données

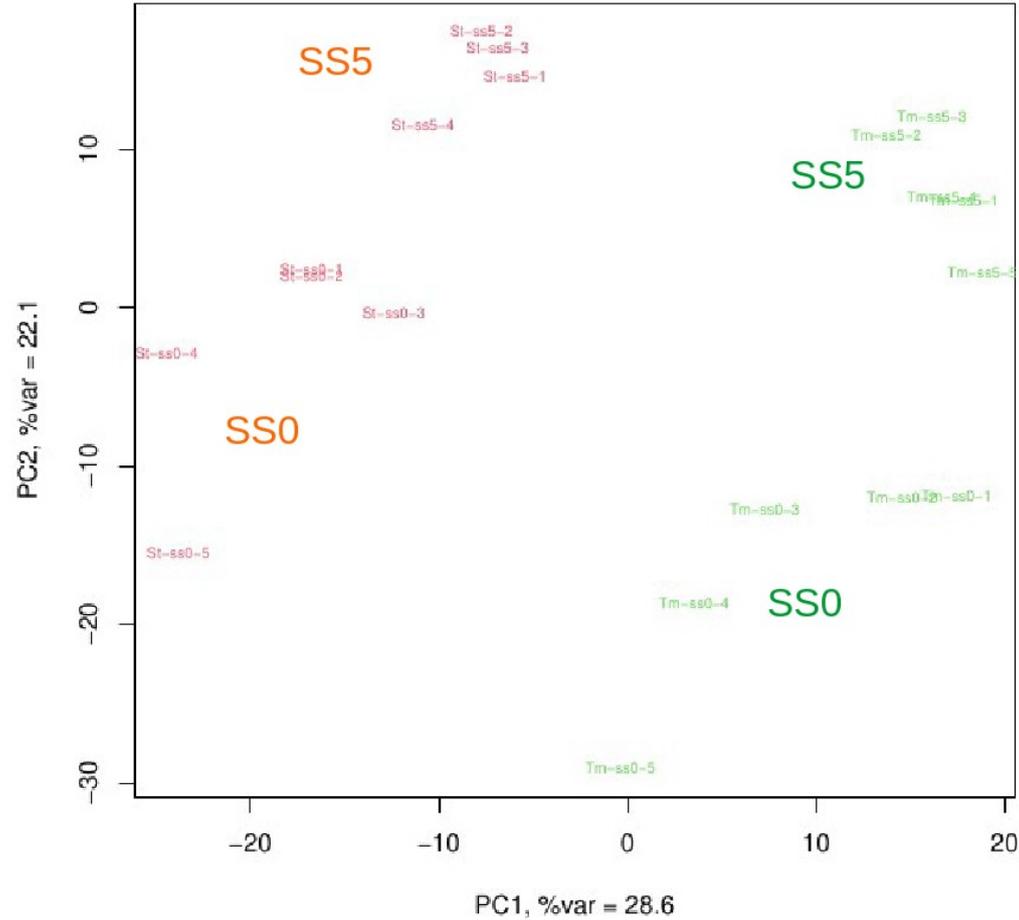
- Etude du jeu de données

Construction de réseaux, méthodes disponibles

Méthodes d'évaluation de la pertinence des réseaux

Comparaison de méthodes de construction de réseaux

# Analyse en composantes principales



# Plan de l'exposé

Présentation du jeu de données

Etude du jeu de données

- **Construction de réseaux, méthodes disponibles**

Méthodes d'évaluation de la pertinence des réseaux

Comparaison de méthodes de construction de réseaux

# Méthodes de construction de réseaux

V.A. Huynh-Thu and G. Sanguinetti.  
*Gene regulatory network inference: an  
introductory survey*  
Methods Mol Biol. 2019

## Méthodes basées sur les données

- corrélation ([Pearson](#), WGCNA)
- information mutuelle (ARACNE, CLR, MRNET)
- régression (directed networks, time series, ...)

## Méthodes probabilistes

- gaussian graphical models ([GeneNet](#))
- réseaux bayésiens (CatNet, Banjo)

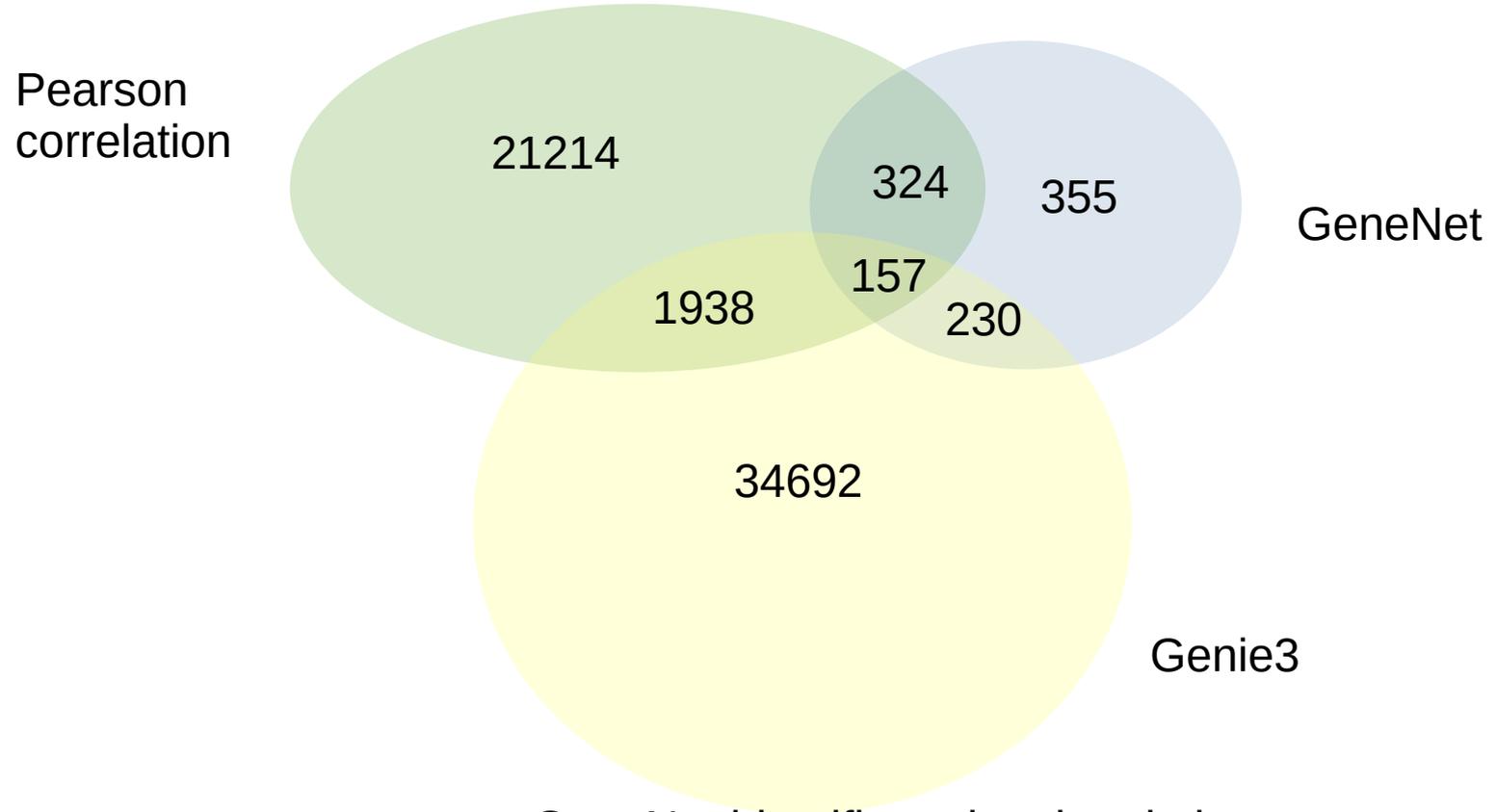
## Méthodes basées sur le machine learning

- Random forest ([Genie3](#))

## Modèles dynamiques (time series data)

- réseaux dynamiques bayésiens (G1DBN, GRENITS)
- méthodes basées sur les équations différentielles (TSNI, Inferelator)

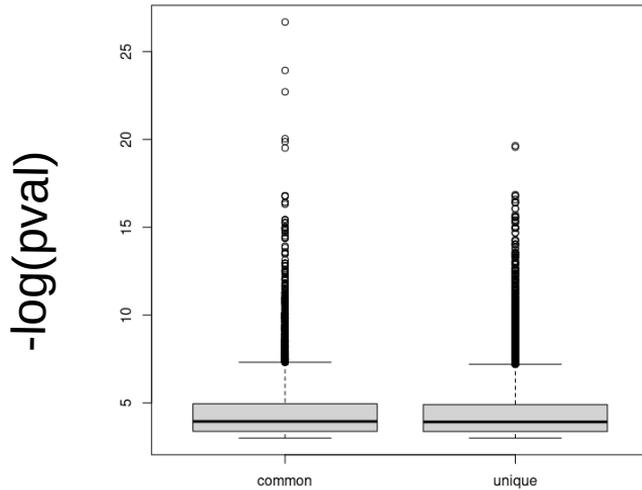
# Les relations identifiées par chacune des méthodes sont elles les mêmes ?



GeneNet identifie moins de relations  
mais elles semblent plus robustes car confortées par les autres méthodes

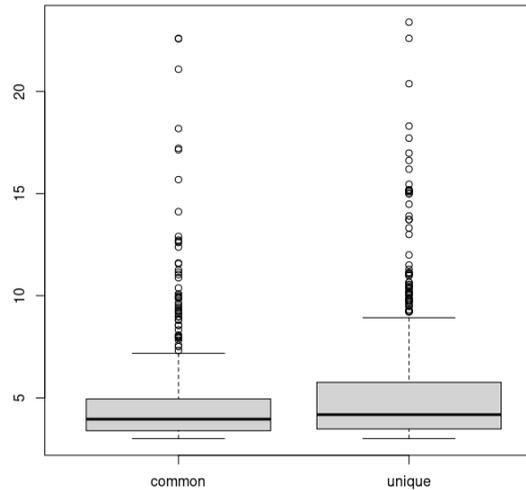
# Les relations identifiées par chacune des méthodes sont elles les mêmes ?

## Pearson correlation



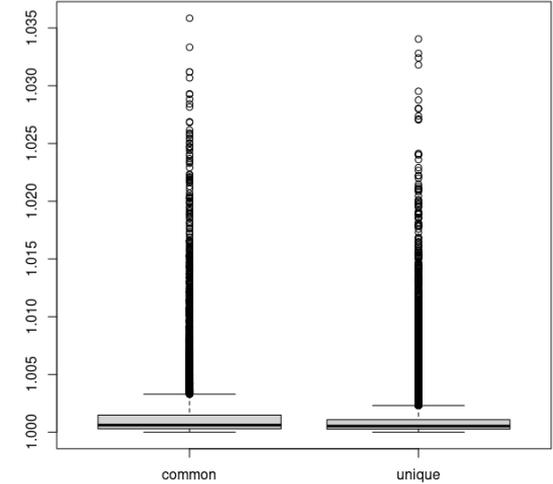
Les relations identifiées par les autres méthodes ne sont pas plus significatives

## GeneNet



Les relations identifiées par les autres méthodes sont les moins significatives

## Genie3



Les relations identifiées par les autres méthodes sont les plus significatives

# Plan de l'exposé

Présentation du jeu de données

Etude du jeu de données

Construction de réseaux, méthodes disponibles

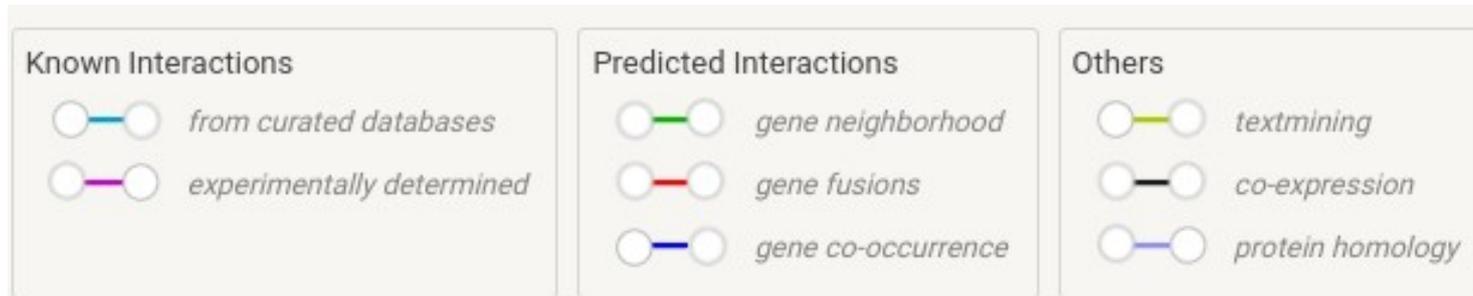
- Méthodes d'évaluation de la pertinence des réseaux

Comparaison de méthodes de construction de réseaux

# Evaluer la pertinence des réseaux à partir de connaissances indépendantes

Base de données STRING

regroupe des liens fonctionnels et contextuels entre protéines



Utilisation du réseau « STRING » pour contruire un réseau de référence

# Evaluer la pertinence des réseaux à partir de connaissances indépendantes

Base de données STRING

regroupe des liens fonctionnels et contextuels entre protéines

Your Input:

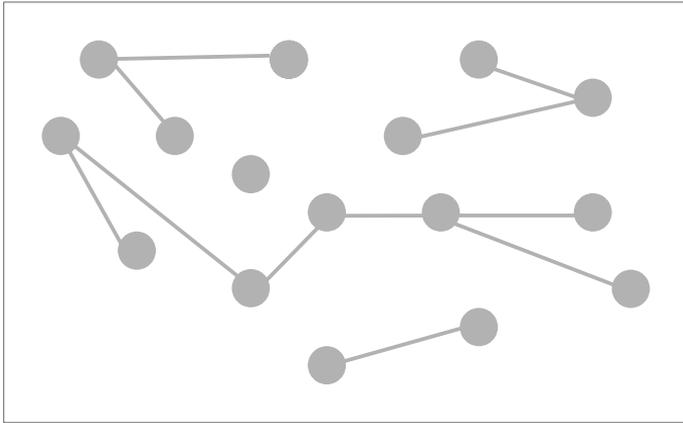
● ADK1 *Adenylate kinase, chloroplastic; Catalyzes the reversible transfer of the terminal phosphate group between ATP and AMP. Plays an important role in cellular energy homeostasis and in adenine nucleotide metabolism. The maize enzyme also works with CMP, albeit with 10% of the activity with AMP. (222 aa)*

Predicted Functional Partners:

|               |   | Neighborhood | Gene Fusion | Cooccurrence | Coexpression | Experiments | Databases | Textmining | [Homology] | Score |
|---------------|---|--------------|-------------|--------------|--------------|-------------|-----------|------------|------------|-------|
| ● Adk1        | <i>Adenylate kinase 1; Belongs to the adenylate kinase family.</i>  |              |             | ●            | ●            |             |           |            | ●          | 0.900 |
| ● atpF        | <i>ATP synthase subunit b, chloroplastic; F(1)F(0) ATP synthase produces ATP from ADP in the presence of a proton or sod...</i> |              |             |              | ●            |             | ●         | ●          |            | 0.791 |
| ● Si486014c02 | <i>Adenylate kinase; Belongs to the adenylate kinase family.</i>  |              |             | ●            |              |             |           | ●          |            | 0.754 |
| ● A0A1D6K135  | <i>Guanylate kinase.</i>  |              |             |              | ●            | ●           | ●         | ●          |            | 0.728 |
| ● Cl30467_1   | <i>Guanylate kinase 1.</i>  |              |             |              | ●            | ●           | ●         | ●          |            | 0.728 |
| ● A0A1D6IW63  | <i>Nucleoside diphosphate kinase.</i>   |              | ●           |              | ●            |             | ●         | ●          |            | 0.710 |
| ● A0A1D6EH51  | <i>Nucleoside diphosphate kinase.</i>   |              | ●           |              | ●            |             | ●         | ●          |            | 0.709 |
| ● A0A1D6IBV3  | <i>Nucleoside diphosphate kinase.</i>   |              | ●           |              | ●            |             | ●         | ●          |            | 0.709 |
| ● A0A1D6IZB9  | <i>Nucleoside diphosphate kinase.</i>   |              | ●           |              | ●            |             | ●         | ●          |            | 0.709 |
| ● A0A1D6Q3H8  | <i>Nucleoside diphosphate kinase.</i>   |              | ●           |              | ●            |             | ●         | ●          |            | 0.709 |

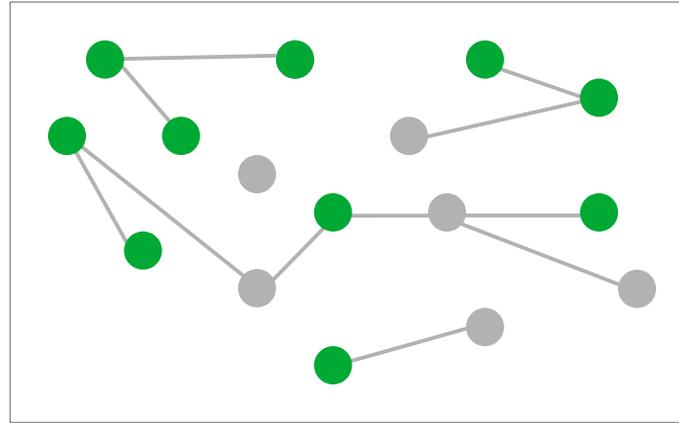
Utilisation du réseau « STRING » pour contruire un réseau de référence

# Construction d'un réseau « référence »



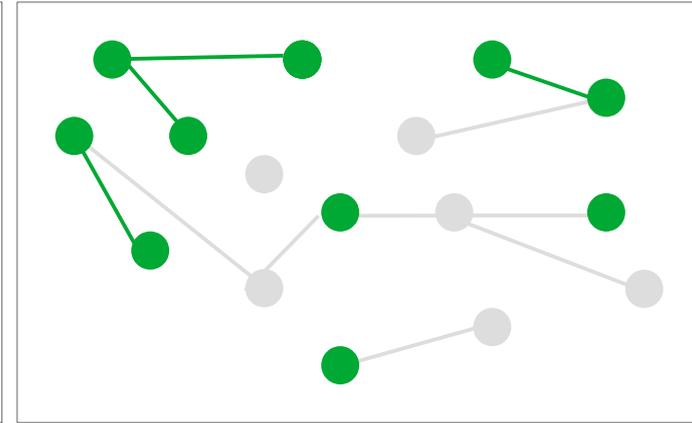
Complete reference network

29416 protéines  
14.158.036 relations



Proteins of our dataset

673 / 800 protéines  
présentes dans STRING

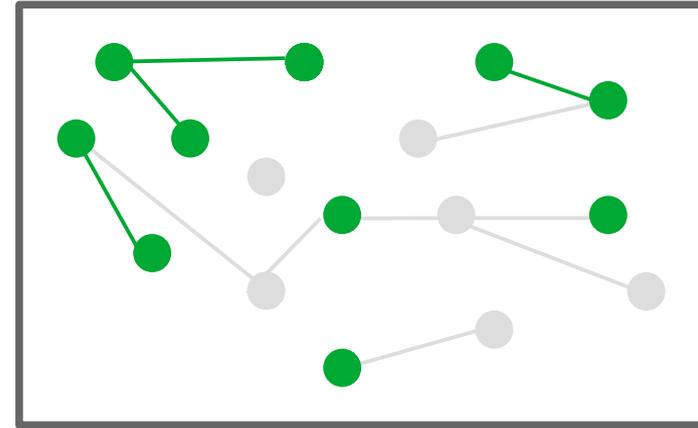


Dataset reference network

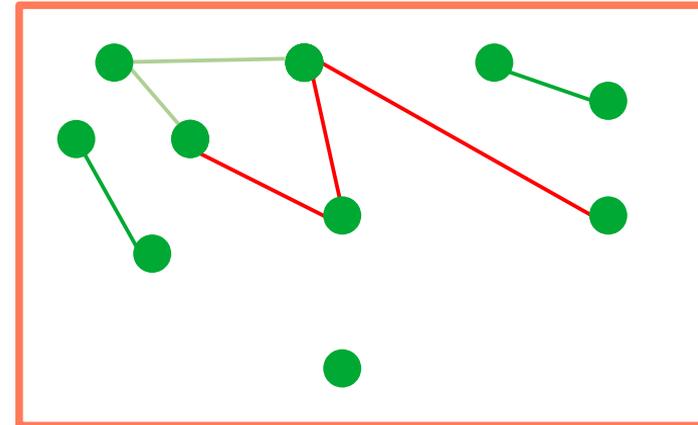
673 protéines  
41.939 relations

# Comparaison du réseau inferré au réseau « référence »

|                   | Relation présente | Relation absente | Total (observé) |
|-------------------|-------------------|------------------|-----------------|
| Relation présente | 2                 | 3                | 5               |
| Relation absente  | 2                 | 38               | 40              |
| Total (reference) | 4                 | 41               | 45              |



Dataset reference network

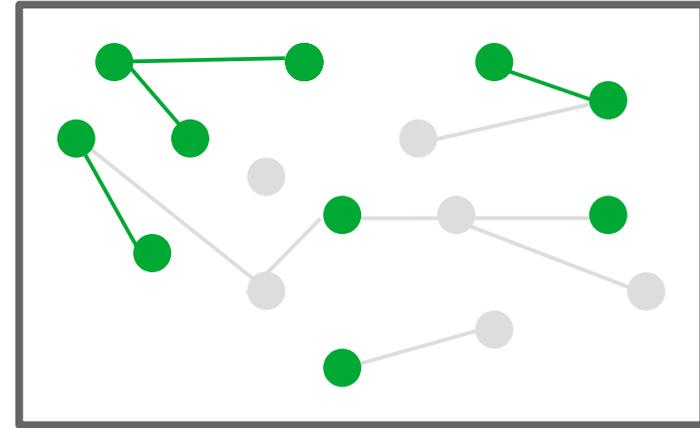


Dataset inferred network

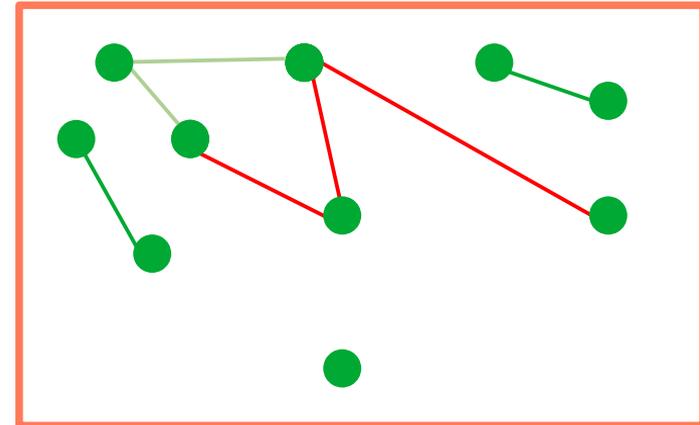
# Comparaison du réseau inferré au réseau « référence »

|                   | Relation présente | Relation absente | Total (observé) |
|-------------------|-------------------|------------------|-----------------|
| Relation présente | +++<br>VP         | —<br>FP          |                 |
| Relation absente  | —<br>FN           | ++<br>VN         |                 |

Total (reference)



Dataset reference network



Dataset inferred network

# Plan de l'exposé

Présentation du jeu de données

Etude du jeu de données

Construction de réseaux, méthodes disponibles

Méthodes d'évaluation de la pertinence des réseaux

- Comparaison de méthodes de construction de réseaux

# Différences dans les capacité des méthodes à identifier des interactions du réseau de référence

|               | Pearson correlation | GeneNet | Genie3 |
|---------------|---------------------|---------|--------|
| Ref Inter.    | 3330                | 177     | 4571   |
| Ref No inter. | 20303               | 889     | 19062  |

% des relations du réseau  
de référence identifiées

14.1 %

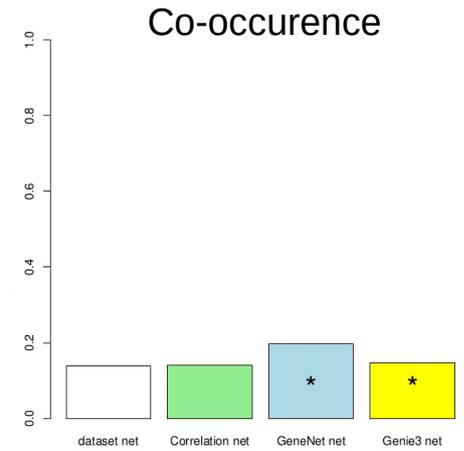
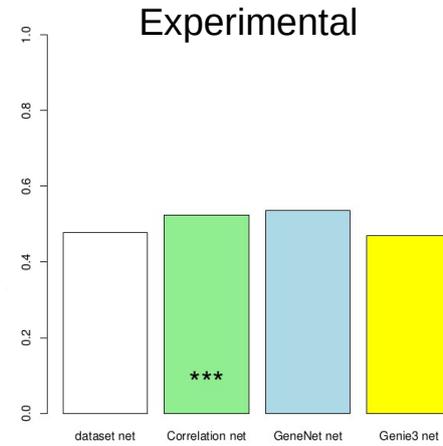
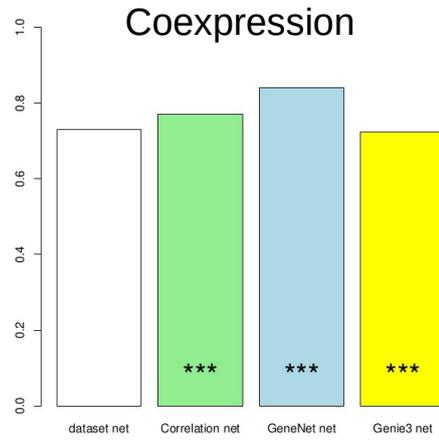
16.6 %

19.3 %

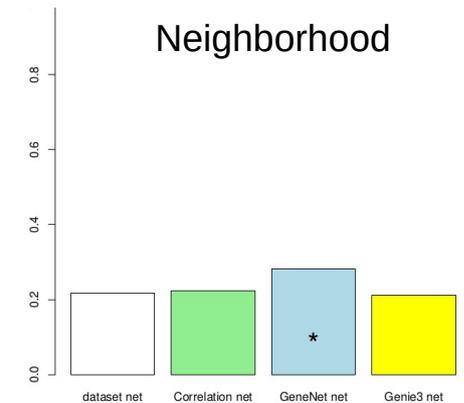
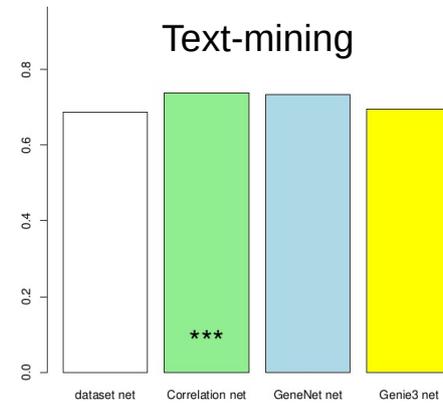
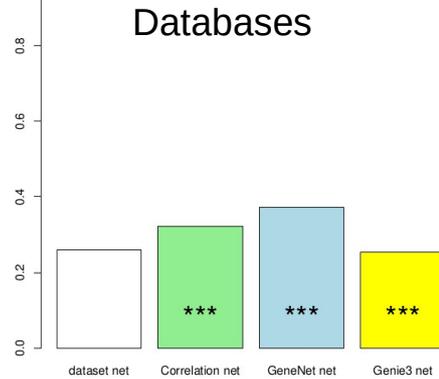
Genie3 identifie davantage d'interactions et la part les relations du réseau de référence est plus importante

# Quels types d'interactions mises en évidence par chacune des méthodes ?

Proportion des scores non nuls



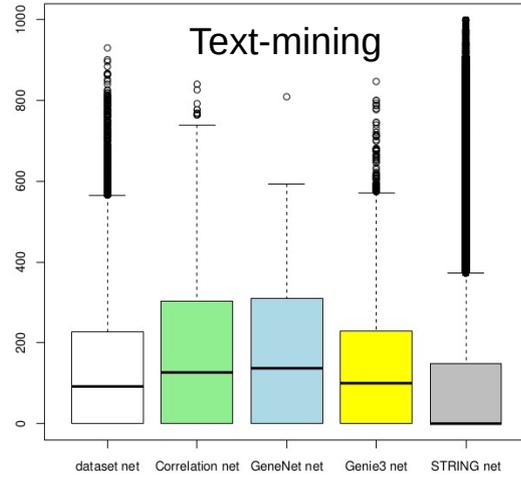
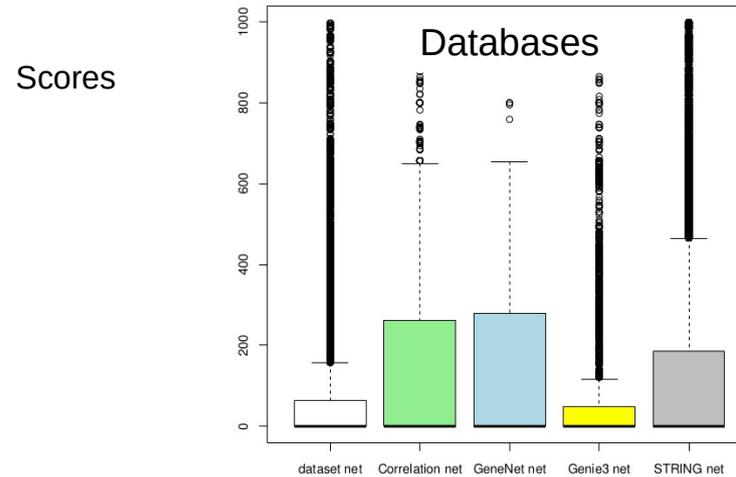
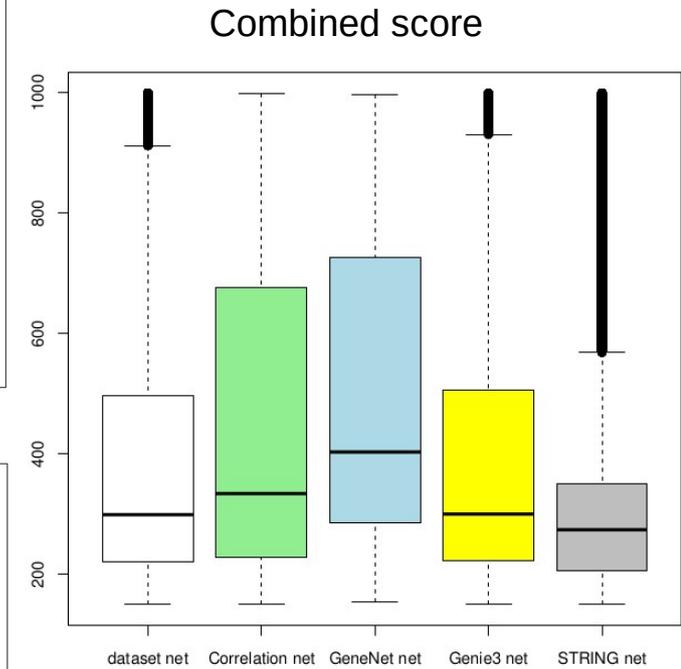
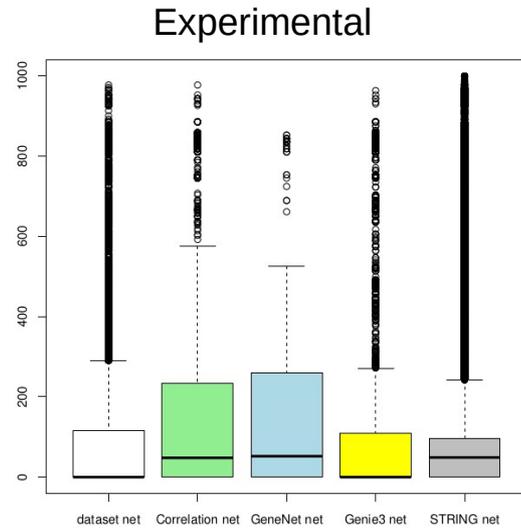
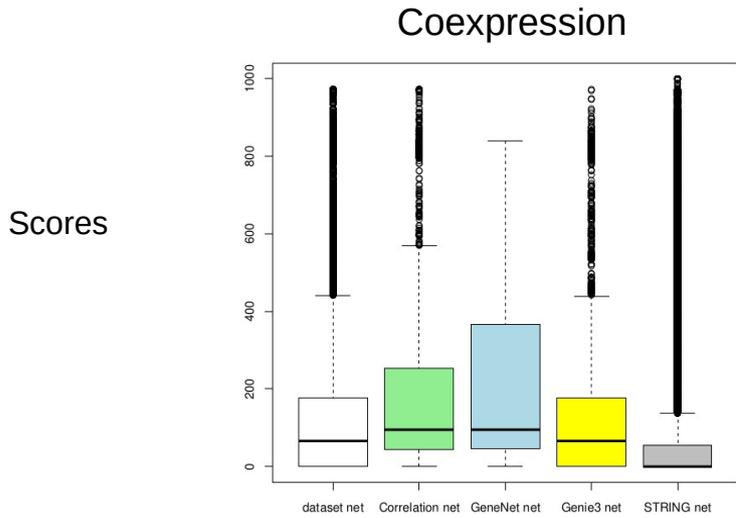
Proportion des scores non nuls



Liens fonctionnels  
mis en évidence avec  
correlation

Prédictions basées sur la  
génomique mis en évidence  
avec GeneNet

# Quels types d'interactions mises en évidence par chacune des méthodes ?



# Conclusion

|                      | Pearson                        | GeneNet               | Genie3 |
|----------------------|--------------------------------|-----------------------|--------|
| Nb interactions      | +++                            | +                     | +++    |
| Sens biologique      | +                              | ++                    | +++    |
| Types d'interactions | Fonctionnelles et coexpression | Phylo-genomic related | Tous   |

Pearson : identification de nombreuses relations (fonctionnelles et coexpression mais potentiellement des « indirectes »)

Genie 3 : beaucoup de relations, probablement moins de relations « indirectes »

GeneNet : identification de relations de types différents

# Perspectives

- Comparer les méthodes prenant en compte le temps
  - jeu de données adapté car effet jour + zone peut-être intégré par une estimation de l'âge

# Méthodes de construction de réseaux

V.A. Huynh-Thu and G. Sanguinetti.  
*Gene regulatory network inference: an  
introductory survey*  
Methods Mol Biol. 2019

## Méthodes basées sur les données

- corrélation (Pearson, WGCNA)
- information mutuelle (ARACNE, CLR, MRNET)
- régression (directed networks, time series, ...)

## Méthodes probabilistes

- gaussian graphical models (GeneNet)
- réseaux bayésiens (CatNet, Banjo)

## Méthodes basées sur le machine learning

- Random forest (Genie3)

## Modèles dynamiques (time series data)

- réseaux dynamiques bayésiens (G1DBN, GRENITS)
- méthodes basées sur les équations différentielles (TSNI, Inferelator)

# Perspectives

- Comparer les méthodes prenant en compte le temps
  - jeu de données adapté car effet jour + zone peut-être intégré par une estimation de l'âge
- Etudier l'effet du traitement sur les réseaux
  - prise en compte du traitement dès la construction du réseau (regressions avec interaction traitement)
  - construction de réseaux indépendants et comparaison ?
- Application de la méthode à des réseaux de types 'clusters' comme WGCNA

# Remerciements

Michel Zivy  
Olivier Turc

Mélina Gallopin  
Mélisande Blein-Nicolas

Malak Fakh (M2 Gen2E)  
Ilhem Riahi (M1 bioinfo)



université  
PARIS-SACLAY