

R à Québec
26 mai 2017



PathQuant : un outil bio-informatique pour l'annotation de paires gène-métabolite en utilisant les réseaux métaboliques

Sandra Therrien-Laperrière, Étudiante à la maîtrise, ICM et Université de Montréal

Directrice : Christine Des Rosiers, PhD

Fonction cellulaire - Plusieurs plans de régulation, du génotype au phénotype ...



Phénotype

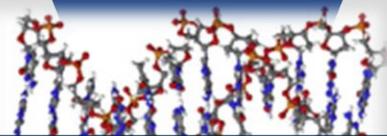
Métabolome
(>65 000)

Protéome

Transcriptome

Génome
(~25 000)

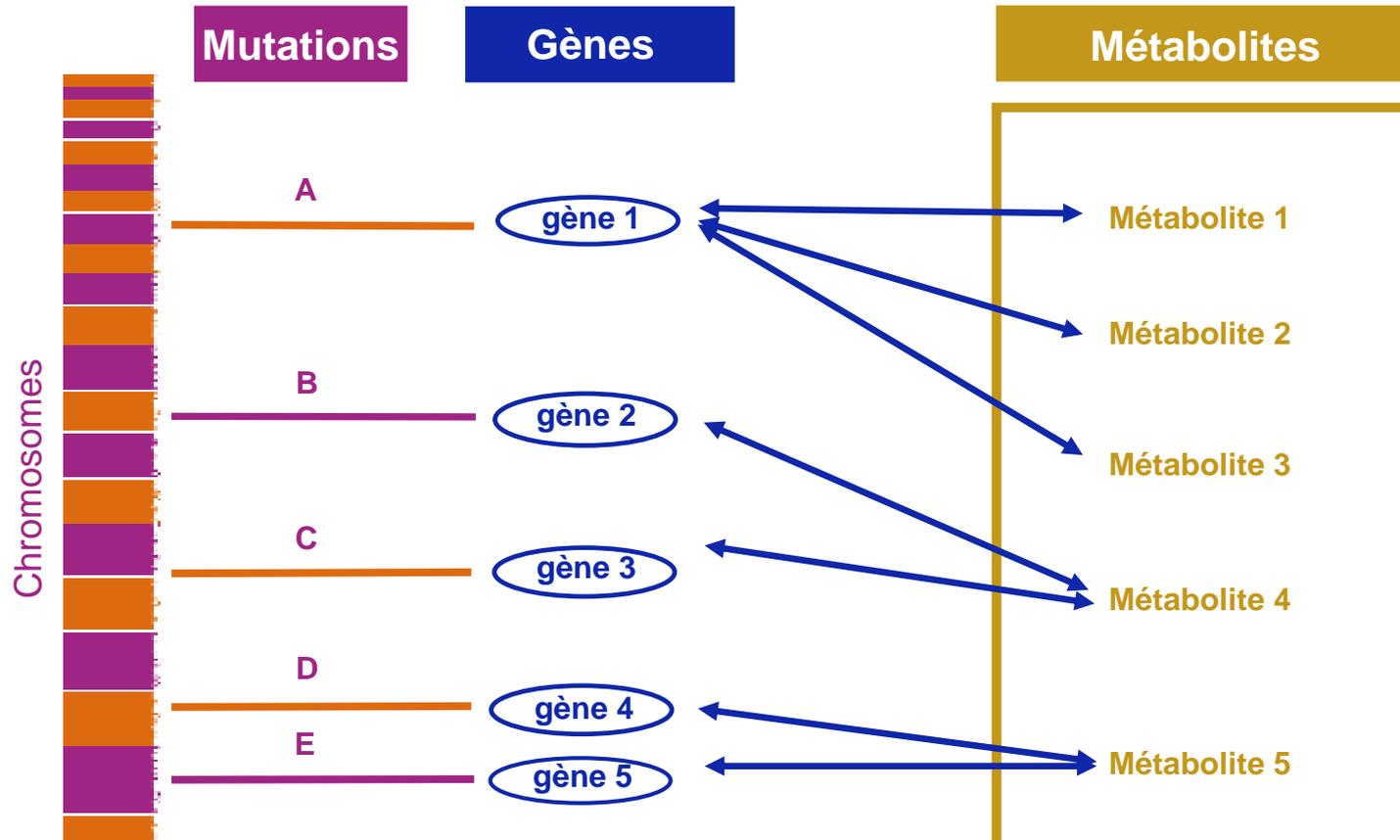
Génotype



mGWAS - Données d'associations entre gène et métabolite

Étude du génome

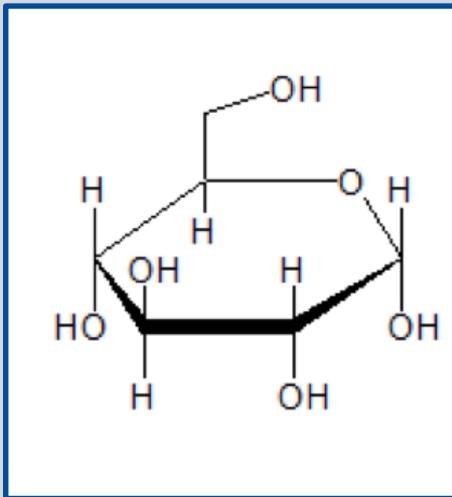
Étude du métabolome



Quelle est la valeur biologique des associations rapportées par les études mGWAS ?

Les réactions enzymatiques

Métabolite - Substrat:
Glucose



Gène

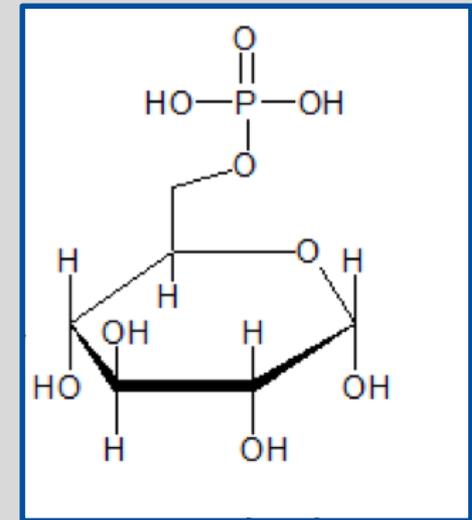


Enzyme

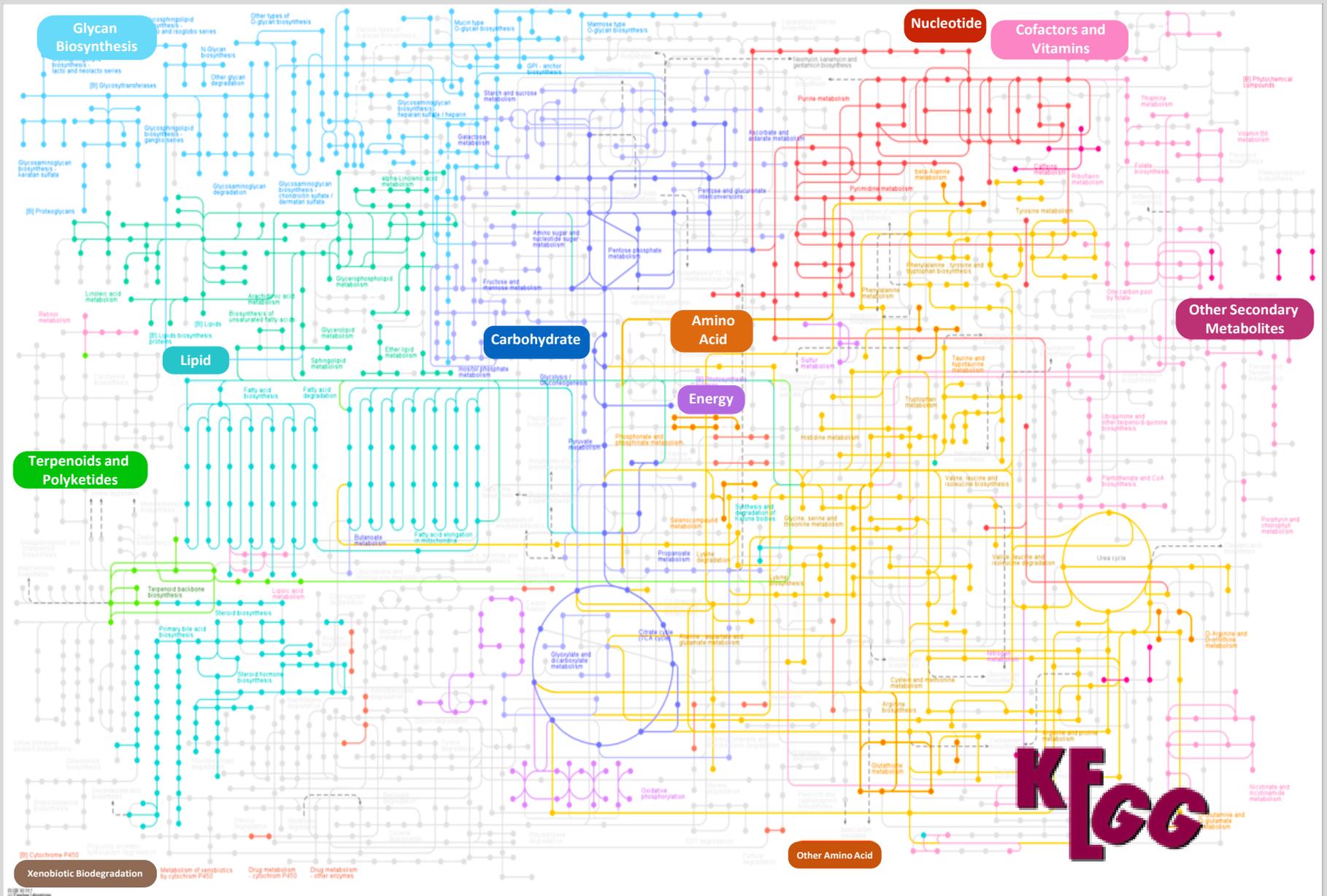
Hexokinase



Métabolite - Produit:
Glucose 6-phosphate



Reconstruction du métabolisme KEGG – "Overview"



Hypothèse - Objectif

Rationnel

- Actuellement, il y a un manque d'outils bio-informatiques pour annoter les paires gène-métabolite dans le contexte des connaissances actuelles des bases de données de voies métaboliques, ce qui inclut les données venant des mGWAS.

Hypothèse

srd : plus court chemin réactionnel

- On suppose que le **srd** entre la réaction catalysée par une enzyme qui est encodée par un gène et son métabolite associé dans les réseaux métaboliques révélera une **information biologique pertinente**.

Objectif

- Développer et tester une méthode automatique pour **annoter les paires entre gène et métabolite** en les **cartographiant** sur les **réseaux métaboliques**.

Fonctionnement de PathQuant

1. Entrée : paires gène-métabolite



Testé avec Shin et al., 2014

2. Classification par produit génétique



- Enzymes
- Transporteurs
- Autres

3. Cartographie des paires sur les réseaux métaboliques

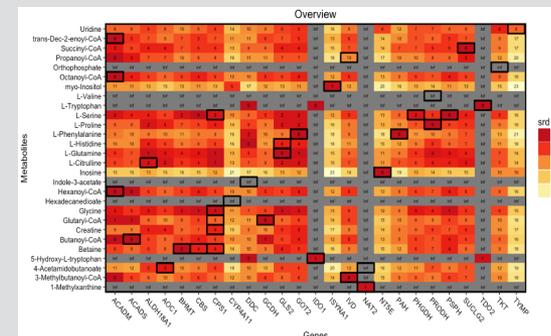


KEGG

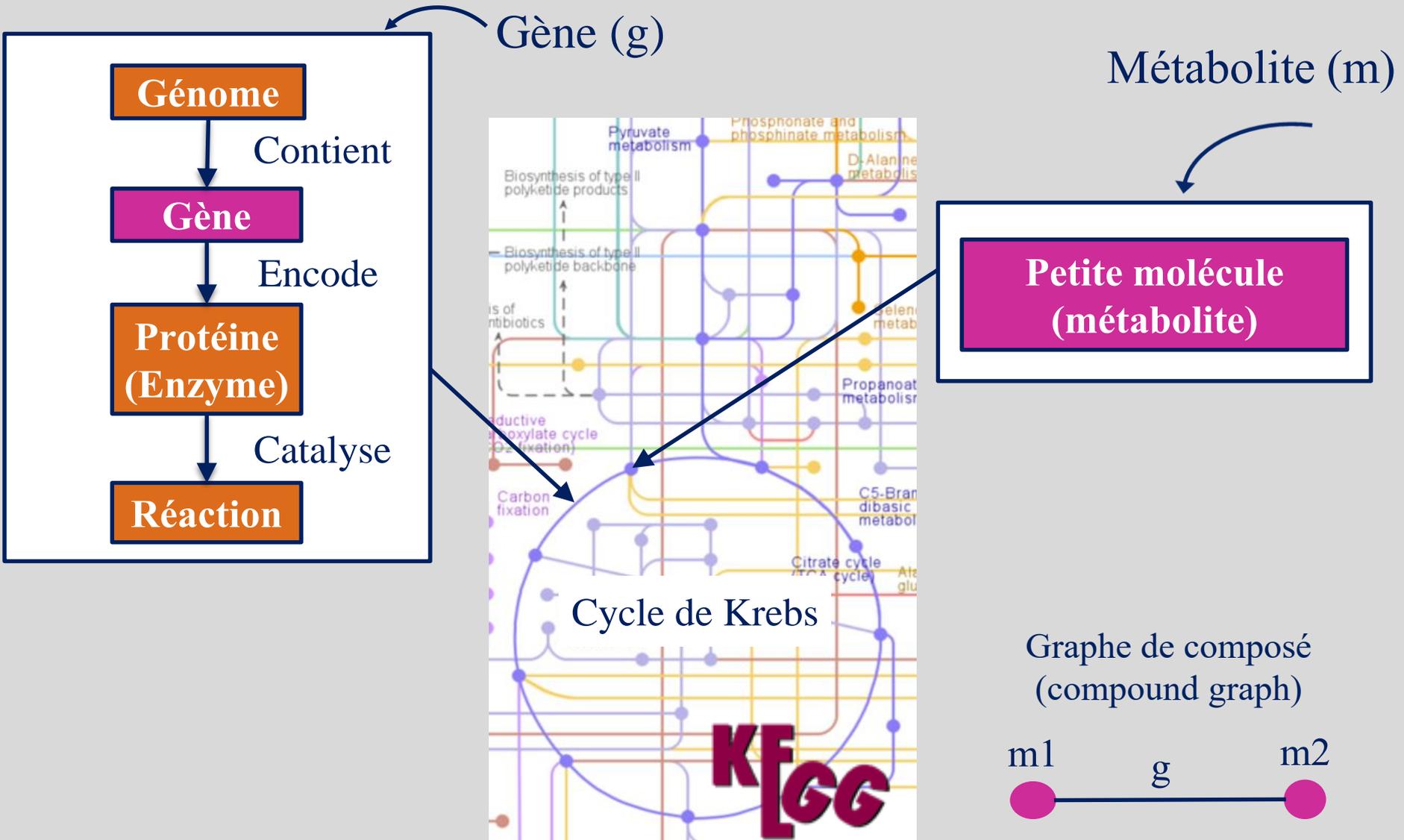
Kyoto Encyclopedia of
Genes and Genomes

4. Calcul du srd

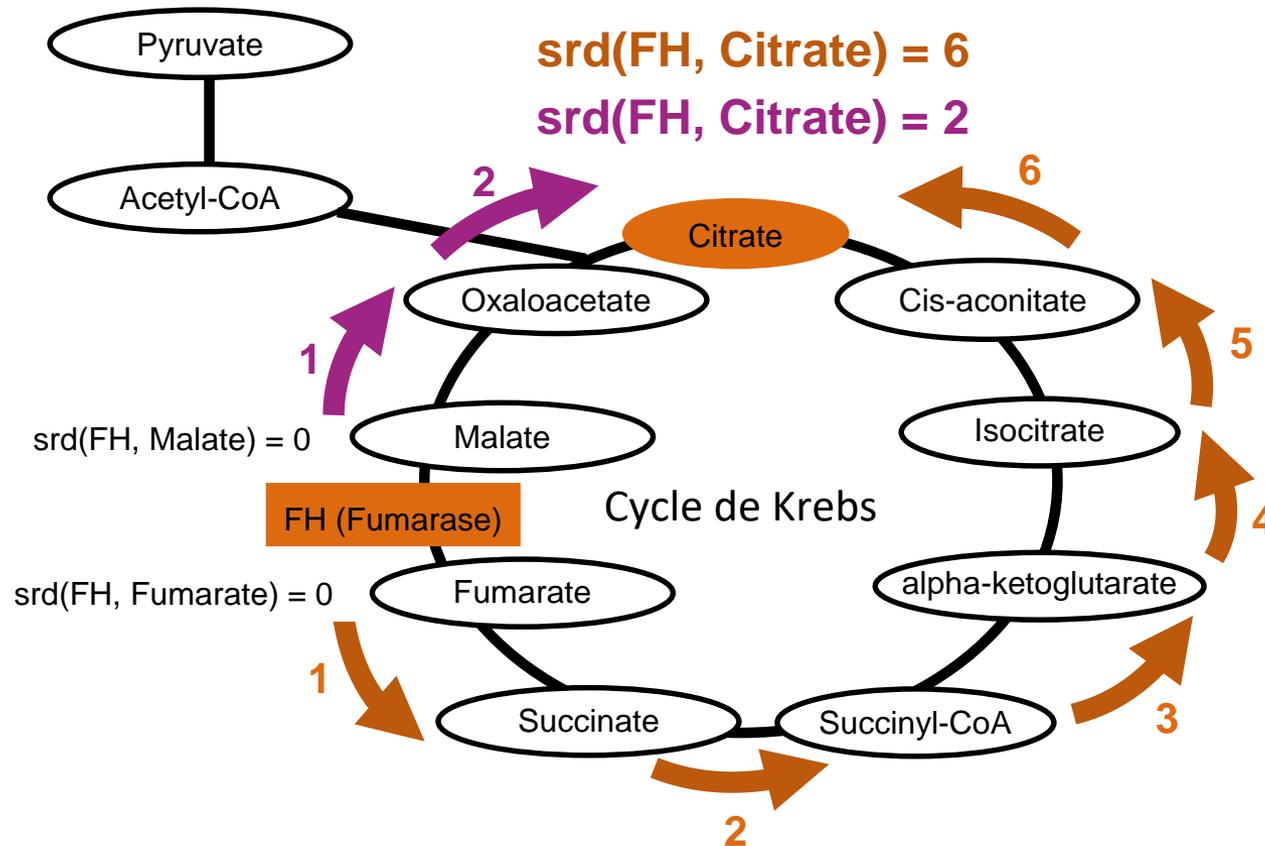
5. Sorties – textuelle et graphique



Reconstruction du métabolisme KEGG – « Overview »



Calcul du srd - Algorithme Dijkstra



Développement de PathQuant - Cadriciels utilisés

Suivant les recommandations de « bioconductor »

RStudio : environnement de développement intégré

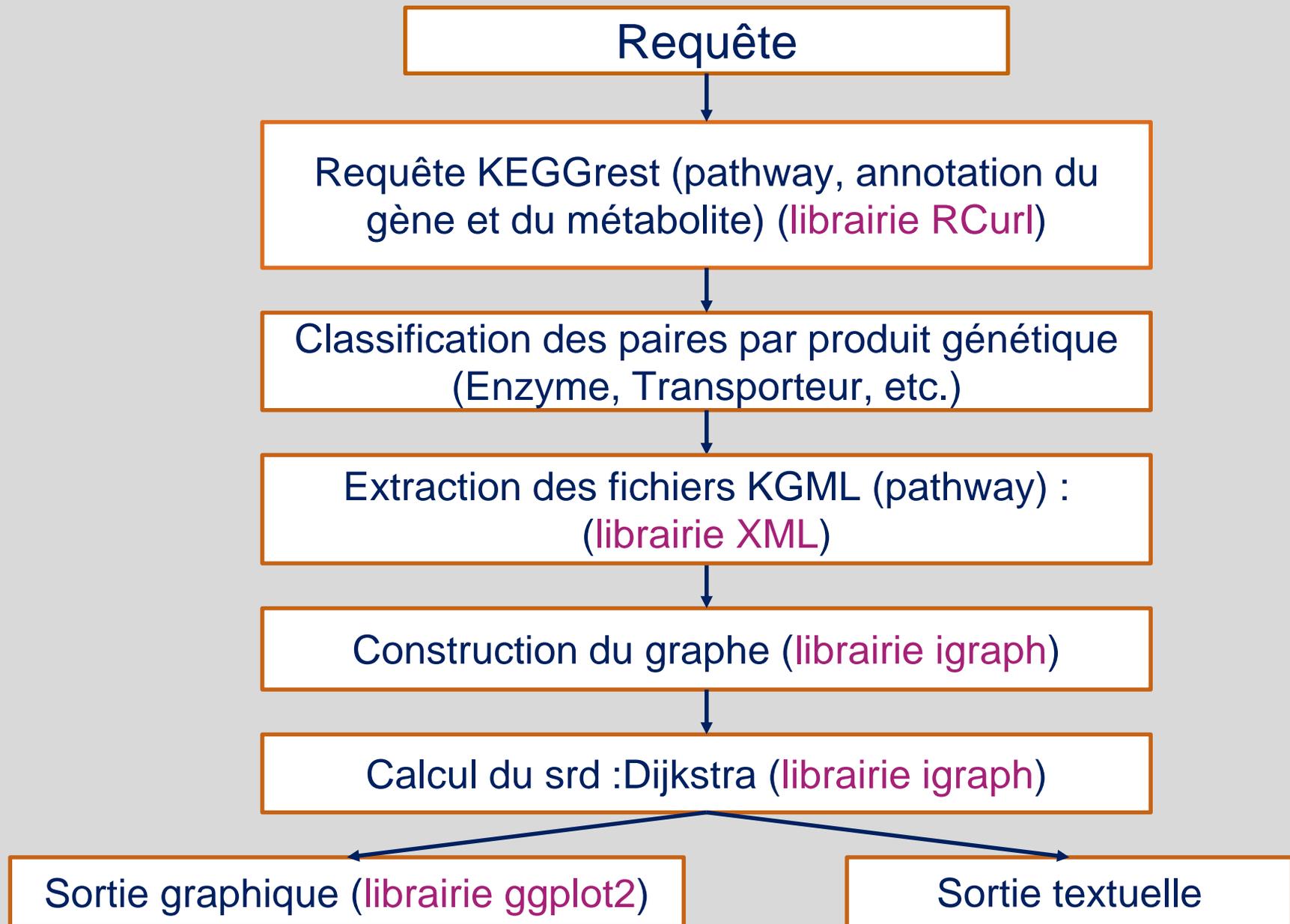
Github : gestionnaire de version de code

Librairie devtools :

- 1) permet la simulation de l'installation du package
- 2) permet l'utilisation de test unitaire (librairie testthat)
- 3) permet l'écriture de la documentation (librairie roxygen2)

Implémentation d'objet S3 pour la flexibilité du code.

Développement de PathQuant - Bibliothèques utilisées



Exemple d'application de PathQuant

Shin et al. : 299 paires gène-métabolite associées

132 gènes uniques

79
Enzymes

20
Transporteurs

33
Autres



183 Associations



86
Métabolites dans KEGG

54
Métabolites pas dans
KEGG

43
Métabolites inconnus

Sortie textuelle - Annotation

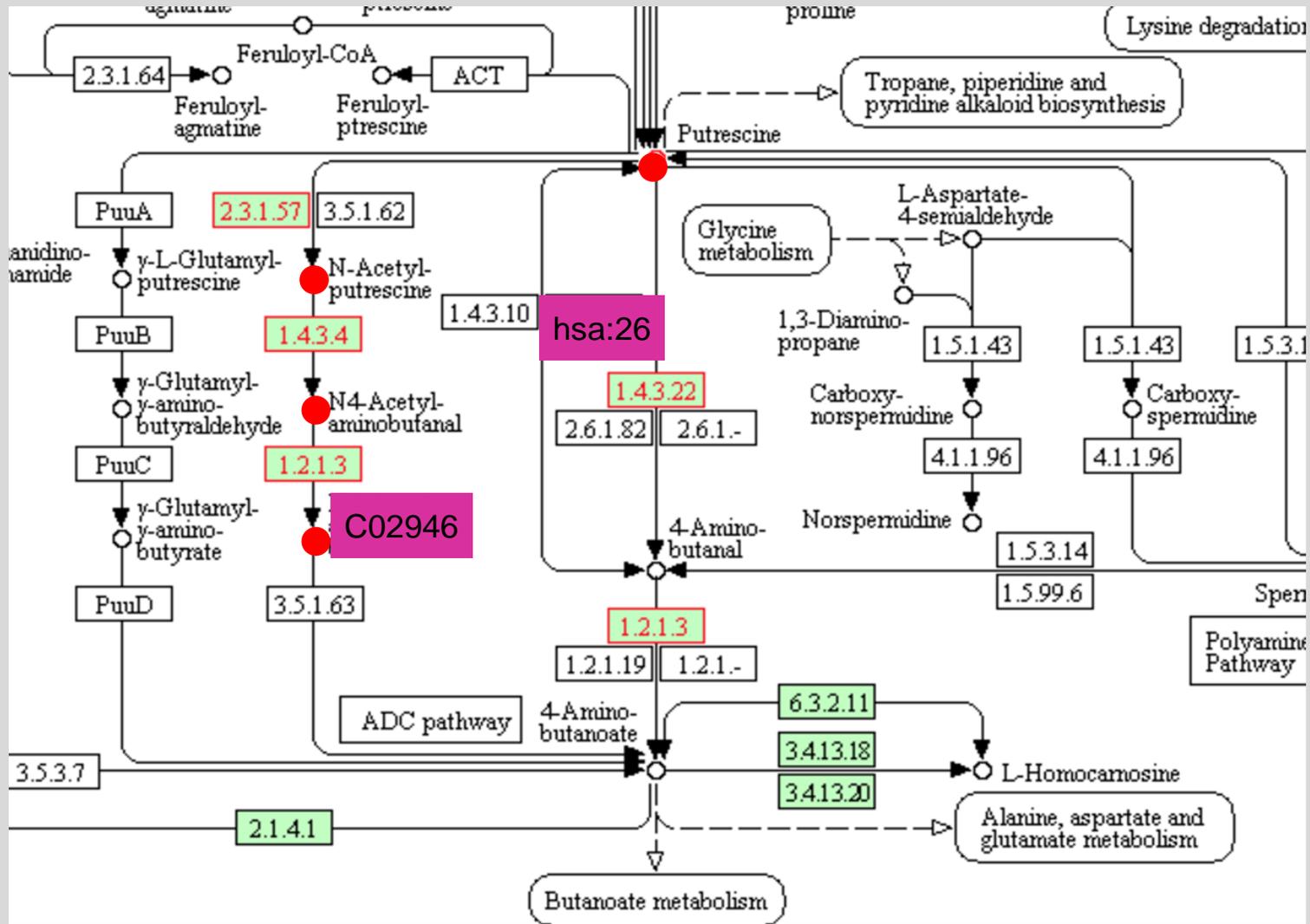
Entrée : liste des paires gène-métabolite (KEGG id)

Sortie : data.frame

| | | |
|---------------------------------|--|-----------|
| gene | hsa:26 | } KEGG id |
| metabolite | C02946 | |
| geneType | Enzyme | |
| geneEnzyme | EC1.4.3.22 | |
| geneBrite | br:hsa00001 - br:hsa01000 | |
| metaboliteClassification | Lipids | |
| srd (distance) | 3 | |
| pathway | Hsa:00330 | |
| geneCommonName (optionnel) | AOC1 | |
| metaboliteCommonName(optionnel) | 4-acetamidobutanoate | |
| path (optionnel) | C00134 -> [hsa:112483 hsa:6303] -> C02714 -> [hsa:4128 hsa:4129] -> C05936 -> [hsa:217 hsa:219 hsa:223 hsa:224 hsa:501] -> C02946 | |

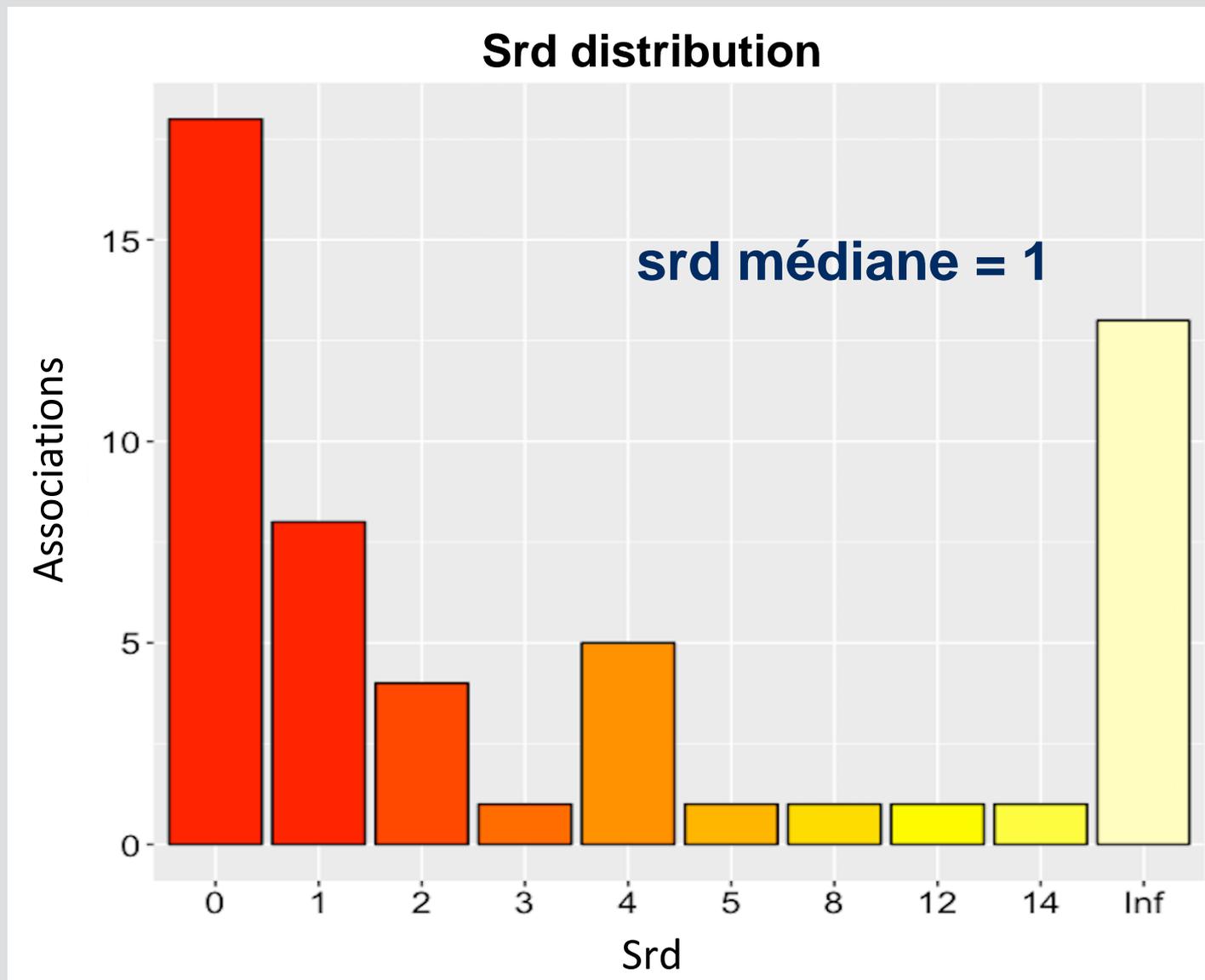
KEGGmapper pour cartographier le plus court chemin

Path : C00134 -> [hsa:112483 hsa:6303] -> C02714 -> [hsa:4128 hsa:4129] -> C05936 -> [hsa:217 hsa:219 hsa:223 hsa:224 hsa:501] -> C02946



Barplot - Distribution des srd calculées

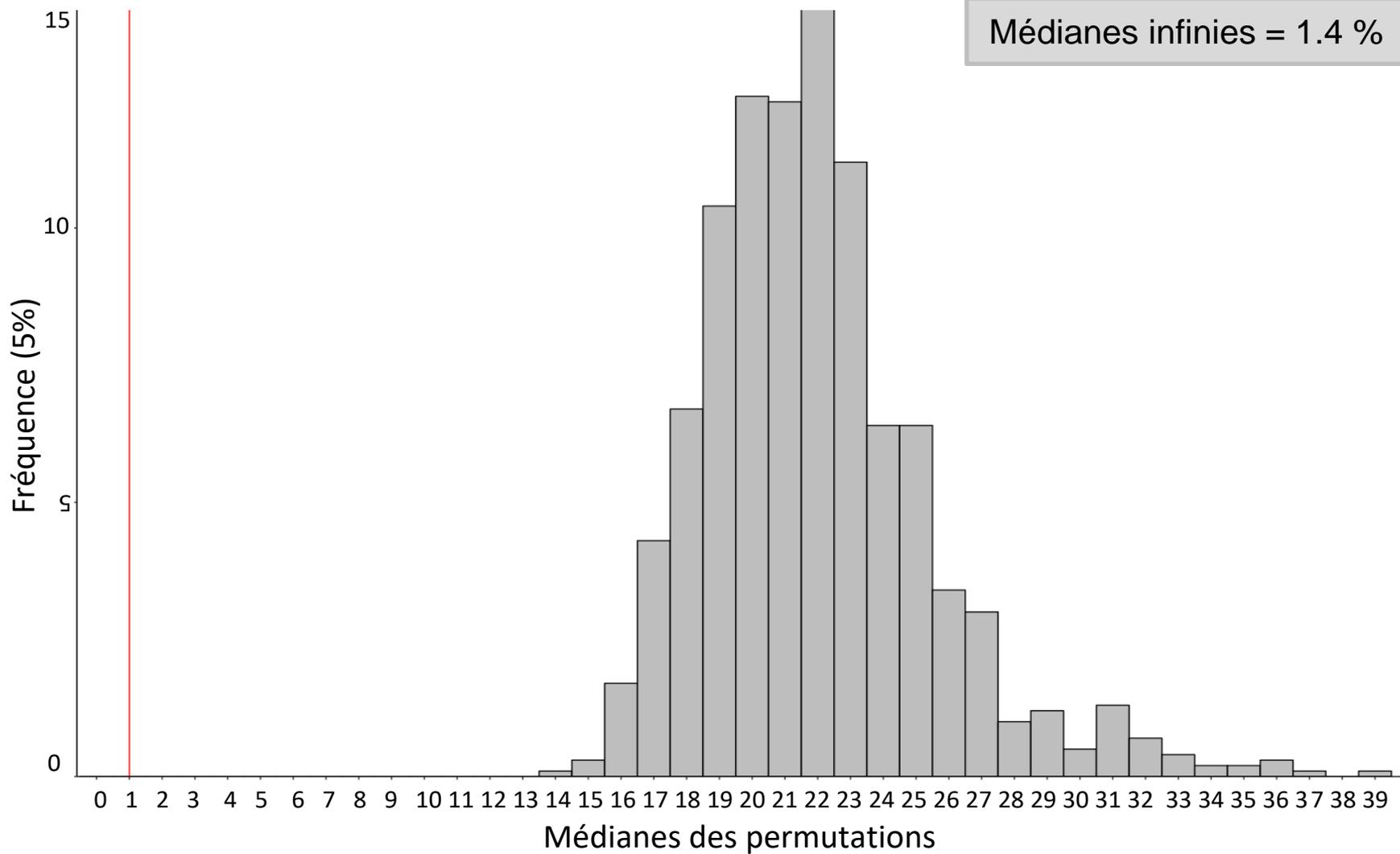
49/86 paires cartographiées sur une même carte



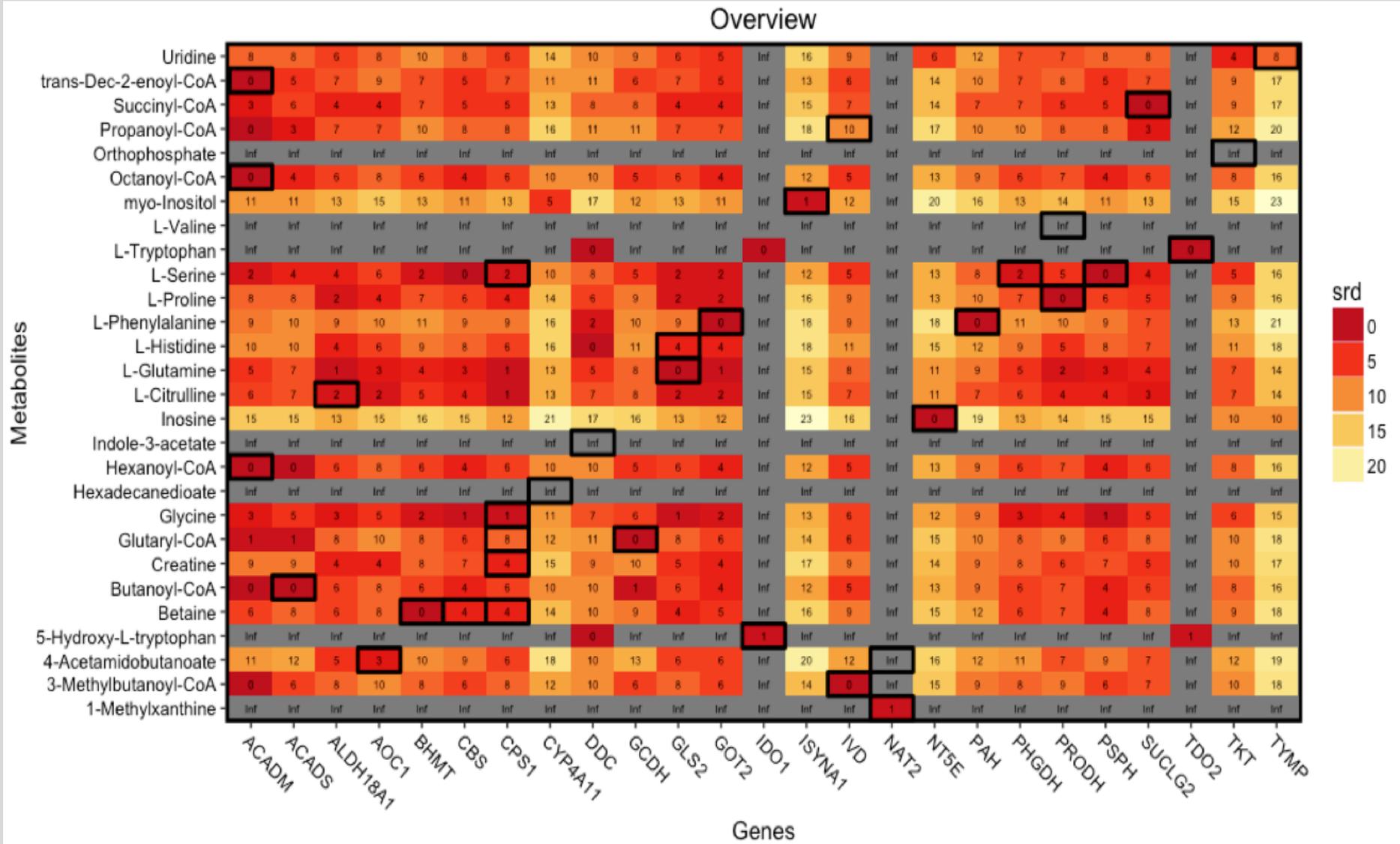
Analyse du réseau de l'« overview » – Test de 1000 permutations (35 paires)

Permutations : 1000

Médianes infinies = 1.4 %



Heatmap des paires cartographiées dans l'« overview »



Valeur biologique du srd

| distance | 49 Associations |
|-----------------------|-----------------|
| ≤ 5 | 36 |
| $6 < \text{srd} < 14$ | 2 |
| Inf | 11 |

Valeur biologique ?

1. Lien biologique direct

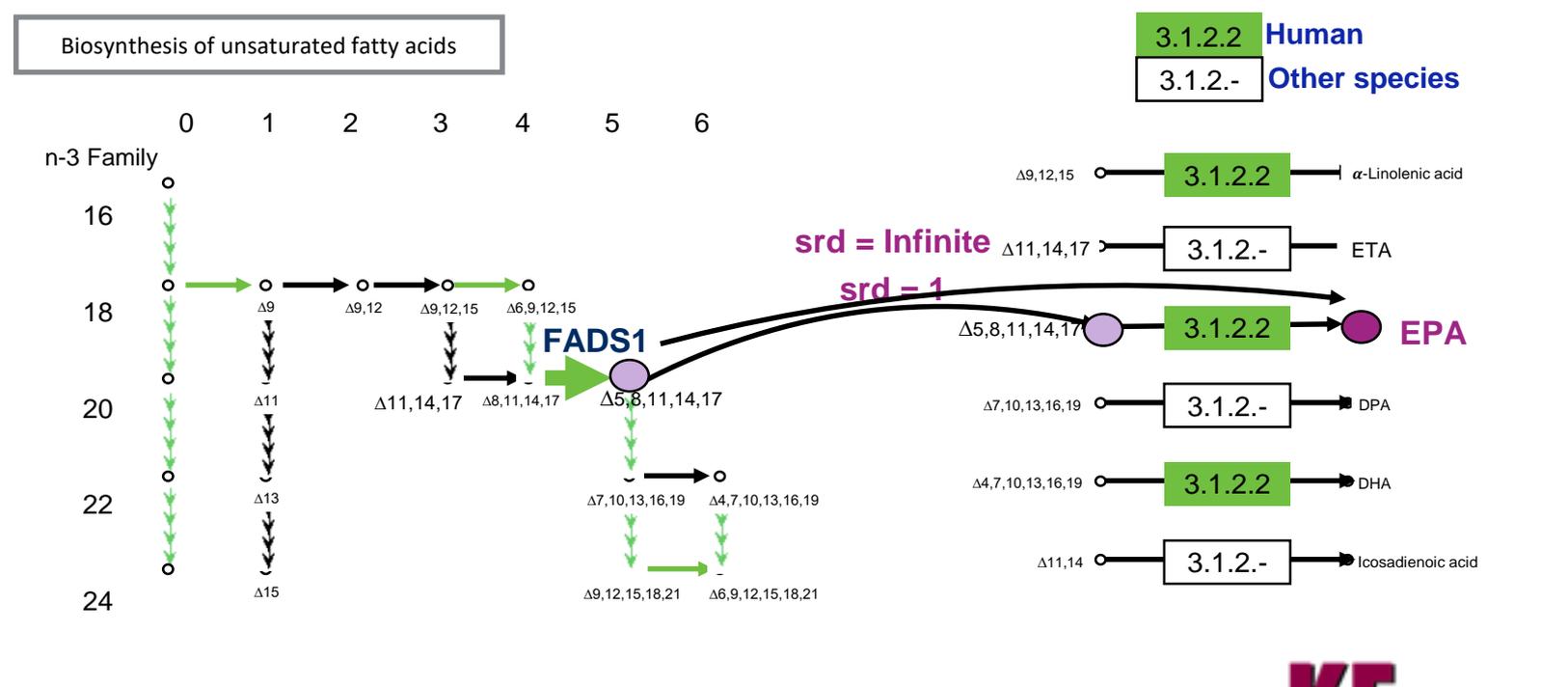
1. Lien biologique direct ou indirect

2. Il n'y a pas de lien biologique

Évaluation manuelle des résultats

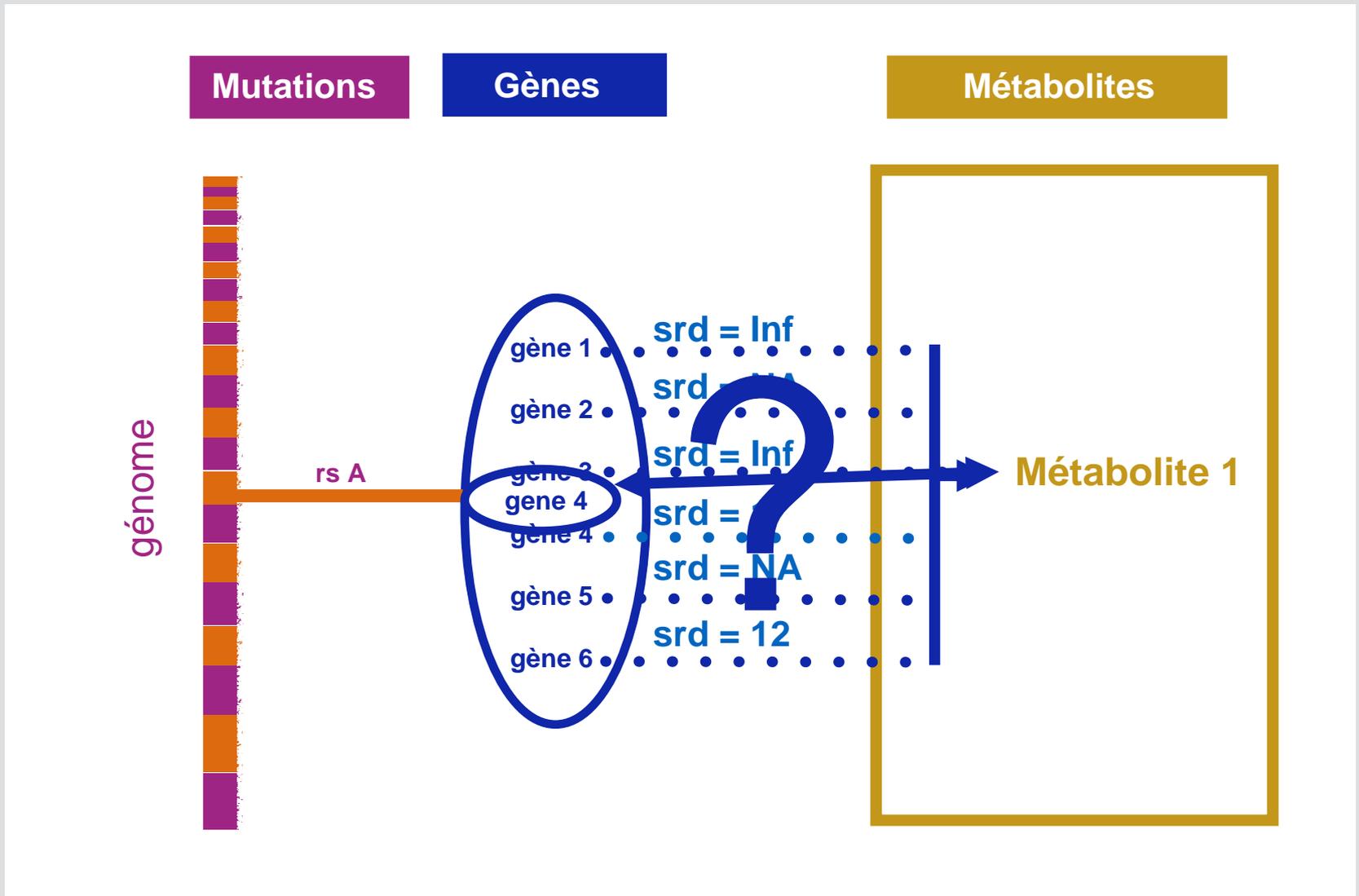
Évaluation manuelle – Topologie des réseaux

Associations : 1) FADS1 - EPA



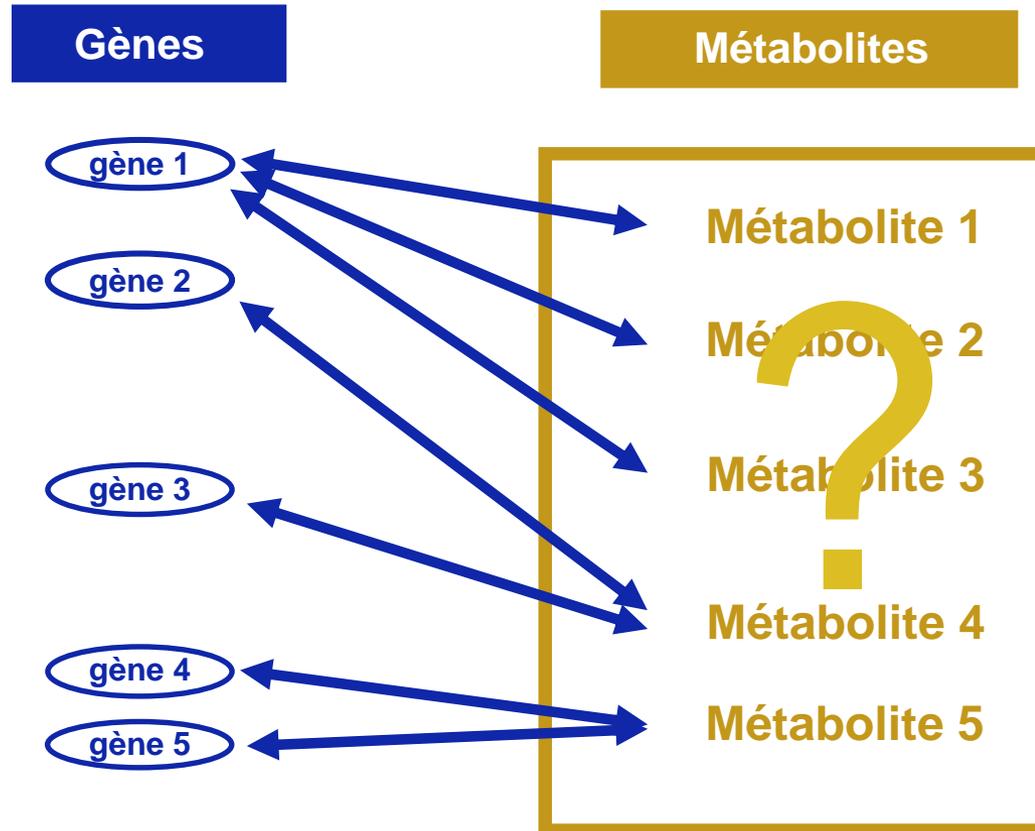
Autres applications

1) Priorisation du gène affecté par une mutation génétique

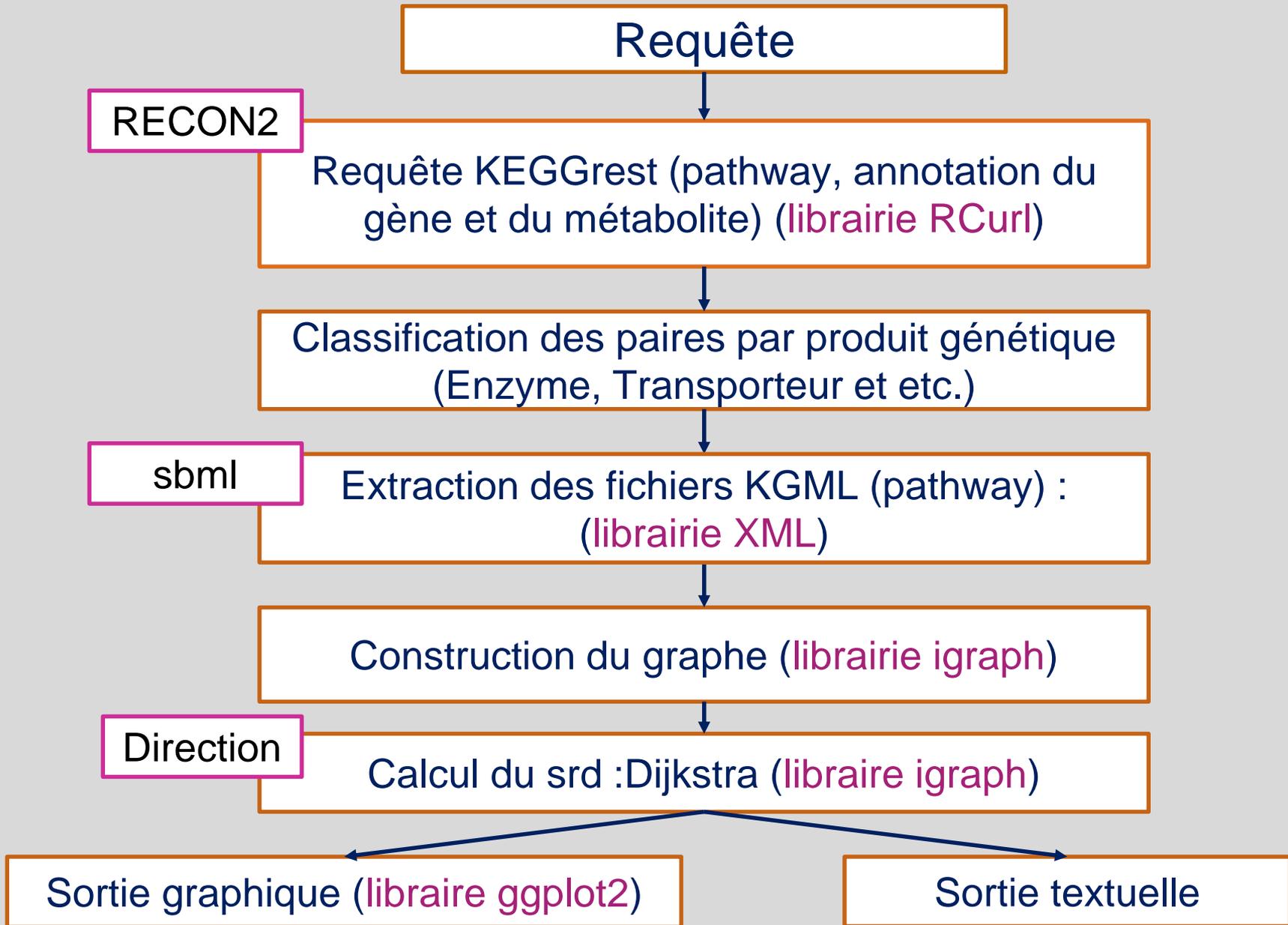


Autres utilisations

2) Priorisation d'une liste de métabolites à mesurer à partir d'une liste de gènes



Perspectives de développement



Conclusion

- ★ L'utilisation de R nous a permis de développer une librairie R robuste pour annoter des paires entre gène et métabolite ✓
- ★ Nous avons testé la librairie sur un jeu de données mGWAS ✓
- ★ Nous avons démontré que l'annotation des paires entre gène-métabolite produit une information pertinente à la compréhension biologique des associations rapportées par les mGWAS ✓
- ★ L'outil est disponible <https://github.com/sandraTL/pathquant> ✓

Remerciements

Université de Montréal

Directrice

- **Christine Des Rosiers, PhD**

Membres du laboratoire

- **Matthieu Ruiz, PhD**
- Lise Coderre, PhD
- Anik Forest, MSc
- Bertrand Bouchard, MSc
- Isabelle Robillard Frayne, MSc
- Caroline Daneault, MSc
- Yvette Mukaneza, candidate PhD
- Bruno Lefort, MD

Co-directeurs

- **John D. Rioux, PhD**
- **Guillaume Lettre, PhD**

Membres du laboratoire

- Gabrielle Boucher, MSc
- Julie Thompson-Legault, MSc

ETH, Zurich

- **Sarah Cherkaoui, candidate PhD**

INRA, Toulouse, France

- Fabian Jourdan, PhD
- Clément Frainay, candidat PhD



INSTITUT DE
CARDIOLOGIE
DE MONTRÉAL

Université 
de Montréal



Agilent Technologies



GenomeCanada



GenomeQuébec



GenomeBritish Columbia

Enseignement
supérieur,
Recherche et Science
Québec 



Crohn's and
Colitis Canada
Crohn et
Colite Canada



CIHR IRSC
Canadian Institutes of Health Research
Institut de recherche
en santé de Montréal
en santé du Québec



Government
of Canada

Faculté des études supérieures et postdoctorales