

# Un outil pour la visualisation de relations entre gènes

Marie Agier\*, Jean-Marc Petit\*\*

\*LIMOS, UMR 6158 CNRS, Univ. Clermont-Ferrand II

\*\*LIRIS, UMR 5205 CNRS, INSA Lyon

agier@isima.fr, jean-marc.petit@insa-lyon.fr

**Résumé.** La reconstruction de réseaux de gènes est un des défis majeurs de la post-génomique. A partir de données d'expression issues de puces à ADN, différentes techniques existent pour inférer des réseaux de gènes. Nous proposons dans ce papier une approche pour la visualisation de réseaux d'interactions entre gènes à partir de données d'expression. L'originalité de notre approche est de superposer des règles avec des sémantiques différentes au sein d'un même support visuel et de ne générer que les règles qui impliquent des gènes dits centraux. Ceux-ci sont spécifiés en amont par les experts et permettent de limiter la génération des règles aux seuls gènes qui intéressent les spécialistes. Une implémentation a été réalisée dans le logiciel libre MeV de l'institut TIGR.

## 1 Introduction

Suite au succès rencontré par les techniques de puces à ADN pour mesurer l'expression des gènes à grande échelle, la **reconstruction de réseaux de gènes** à partir de ces données d'expression a suscité depuis quelques années un intérêt croissant. Dans des travaux antérieurs [1, 2], nous avons proposé une approche ayant pour but de découvrir **différents types de règles entre gènes**. Pour faciliter l'interprétation des règles par les experts, nous proposons dans ce papier une visualisation conviviale des règles générées. Nous montrons comment les règles peuvent être visualisées sous forme de graphe orienté présentant les diverses relations découvertes dans les données. L'originalité de notre proposition est de superposer différents types de règles dans un même support visuel. Nous proposons également aux utilisateurs de spécifier plusieurs gènes dits centraux, à partir desquels seront présentées uniquement les règles impliquant ces gènes centraux et limitant ainsi le coût de la génération des règles.

## 2 Approche proposée

Nous souhaitons avant tout réaliser un outil convivial et proposer ainsi une méthode de visualisation intuitive pour les experts. D'autre part, nous proposons d'appliquer un filtre sur les règles générées en fonction de cinq indices de qualité (support, confiance, lift, leverage et conviction). Ne seront donc visualisées que les règles les plus intéressantes pour les experts, il est donc suffisant de pouvoir visualiser les indices pour la règle ou l'attribut sélectionnés par un simple clic. L'interprétation des règles est une étape particulièrement délicate et très difficile, puisqu'une règle entre deux gènes impliquent également divers produits associés (protéines, facteurs de transcription...). C'est pourquoi les biologistes sont rarement intéressés par

Un outil pour la visualisation de relations entre gènes

les règles avec plus de 3 ou 4 attributs en partie gauche, l'interprétation devenant vite très difficile. Enfin, nous souhaitons offrir la possibilité aux utilisateurs de choisir avant la génération des règles, quelques gènes d'intérêts que nous appelons **gènes centraux**, qu'ils souhaitent voir apparaître en partie gauches ou droites des règles. Les biologistes pourront ainsi visualiser l'ensemble des règles associées à ces quelques gènes qui les intéressent plus particulièrement. Nous avons donc opté pour une visualisation par graphe qui nous semble être la méthode la plus facile à interpréter et la plus intuitive pour les experts, notre objectif étant aussi de nous positionner dans le cadre de la reconstruction de réseaux de gènes. L'originalité de notre approche réside dans le fait que nous proposons des graphes avec **plusieurs sémantiques**, que nous appelons **réseaux globaux**. Chaque sémantique est caractérisée par une couleur particulière. De plus, nous avons choisi de représenter les différents gènes composant les parties droites et gauches des règles, sous forme d'un cercle pour une meilleure lisibilité. Enfin, les gènes centraux sont colorés en rouge, permettant ainsi de les repérer plus facilement.

### 3 Implémentation

L'approche proposée a été développée spécifiquement pour les données d'expression de gènes. Un nouveau module appelé RG (Rule Generation) a été intégré à un logiciel gratuit et open-source consacré à l'analyse de données d'expression de gènes, le logiciel MeV. Les points-clés de ce module sont tout d'abord une interface conviviale permettant de choisir parmi **plusieurs sémantiques** et de spécifier les **gènes centraux**, le calcul de **5 indices de qualité** pour chaque règle générée, la visualisation de **réseaux globaux** incluant diverses sémantiques, enfin la possibilité de visualiser sous forme textuelle les règles impliquant un gène en particulier avec les différents indices de qualité associés. La version étendue du logiciel MeV avec le module RG, proposant la visualisation de différents types de règles entre gènes, est disponible sur le site : <http://www.isima.fr/agier/GeneRules>.

### Références

- [1] M. Agier, J.-M. Petit, V. Chabaud, C. Pradeyrol, Y.-J. Bignon, and V. Vidal. Vers différents types de règles pour les données d'expression de gènes-Application à des données de tumeurs mammaires. In *INFORSID*, pages 351–367, Biarritz, France, 2004.
- [2] M. Agier, J.-M. Petit, and E. Suzuki. Towards ad-hoc rule semantics for gene expression data. In *Proc. of the ISMIS*, pages 494–503, New-York, USA, 2005. Springer-Verlag.

### Summary

Reverse engineering of gene regulatory networks is one of the major challenges of the functional genomics. From gene expression data, various techniques exist to infer these networks. We propose in this paper an approach for the visualization of between-genes interaction networks, from gene expression data. The originality of our approach is to superimpose rules with different semantics within the same visual support and to generate only the rules which imply central genes. Those are specified by the experts and make it possible to limit the generation of the rules to the only genes which interest the experts. An implementation was carried out in the free software MeV of the TIGR institute.