

## Thème 4 : Application à la bio-informatique

### Chapitre 9 : Une méthode implicative pour l'analyse de données d'expression de gènes

Gérard Ramstein

LINA, Polytech'Nantes  
Rue Christian Pauc BP 50609 44306 Nantes cedex 3, France  
gerard.ramstein@univ-nantes.fr

**Résumé.** Nous présentons une méthode d'extraction d'associations basée sur l'analyse statistique implicative et la notion de rang. Nous avons adapté le concept d'intensité d'implication à des classements pour découvrir des relations partielles robustes vis à vis du bruit et des variations d'amplitude. Appliquée aux données de puces à ADN, cette méthode met en évidence des relations entre des formes d'expressions particulières de gènes. Ces associations peuvent être révélatrices de mécanismes de corégulation génique et donc contribuer à l'analyse biomédicale. Nous montrons que cette définition de l'intensité d'implication apporte une connaissance plus fine des relations entre les gènes que les méthodes usuelles de corrélation et qu'elle permet notamment de discriminer entre différents phénotypes avec une précision comparable aux techniques de classification les plus abouties dans ce domaine.

## 1 Introduction

La technologie des puces à ADN permet d'analyser l'expression simultanée de milliers de gènes. L'étude du transcriptome représente un enjeu considérable, tant du point de vue de la compréhension des mécanismes du vivant que des applications cliniques et pharmacologiques. Malheureusement, les données d'expression sont entachées de multiples bruits. D'une part, la complexité du protocole expérimental conduit à une réduction de la précision des mesures. D'autre part, la variabilité naturelle de l'activité cellulaire induit des différences notables d'amplitude d'expression entre les gènes, phénomène également perceptible en considérant plusieurs patients présentant le même phénotype. Cette étude propose une méthode d'analyse implicative des règles d'association sur les données du transcriptome. Elle utilise l'approche de Régis Gras (1996) en considérant non pas les mesures elles-mêmes, mais le rang des observations. Cette optique permet de s'affranchir des valeurs numériques en considérant des zones de classement dans les mesures d'expression. S'intéresser au classement a l'avantage d'améliorer la robustesse des algorithmes en les rendant insensibles à des transformations monotones des données. On peut émettre une analogie avec un tableau notes scolaires. Chaque enseignant possède son propre système de notation et aura une sévérité différente vis à vis des réponses