

# Recherche adaptative de structures de régulation génétique

Mohamed Elati<sup>\*,\*\*</sup>, Céline Rouveirol<sup>\*</sup>

<sup>\*</sup>LIPN — CNRS UMR 7030, Université Paris 13  
99, av. J-B Clément, F-93430 Villetaneuse  
prenom.nom@lipn.univ-paris13.fr

<sup>\*\*</sup>Institut Curie, CNRS UMR 144  
26 rue d’Ulm F-75248 Paris

**Résumé.** Nous avons proposé un algorithme original de Fouille de Données, LICORN, afin d’inférer des relations de régulation coopérative à partir de données d’expression. LICORN donne de bons résultats s’il est appliqué à des données de levure, mais le passage à l’échelle sur des données plus complexes (e.g., humaines) est difficile. Dans cet article, nous proposons une extension de LICORN afin qu’il puisse gérer une contrainte de co-régulation *adaptive*. Une évaluation préliminaire sur des données de transcriptome de tumeurs de vessie montre que les réseaux significatifs sont obtenus à l’aide d’une contrainte de corégulation adaptative de manière beaucoup plus efficace, et qu’ils ont des performances de prédiction équivalentes voire meilleures que celles obtenues par LICORN.

## 1 Introduction

Un des principaux objectifs de la biologie moléculaire consiste à comprendre la régulation des gènes d’un organisme vivant dans des contextes biologiques spécifiques. Les facteurs de transcription sont les régulateurs de la transcription qui vont réagir avec les promoteurs de la transcription des gènes cibles. Les techniques récentes d’analyse du transcriptome, telles que les puces à ADN permettent de mesurer simultanément les niveaux d’expression de plusieurs milliers de gènes. Nous avons déjà décrit le système LICORN (Elati et al., 2007a) qui se fonde sur un modèle de régulation locale coopérative : chaque gène peut être régulé par un ensemble des coactivateurs et/ou un ensemble de coinhibiteurs, ces corégulateurs agissent collectivement pour influencer leur(s) gène(s) cible(s). LICORN met en œuvre une approche originale de Fouille de Données afin d’inférer des relations de régulation coopérative à partir de données d’expression. Cet algorithme a été évalué avec succès sur des données publiques de transcriptome de levure. L’application de LICORN sur des données de transcriptome humaines est plus complexe, car le nombre de régulateurs connus est plus important, et nécessite un temps de calcul considérable. En effet, les gènes de faible support vont avoir un nombre très élevé de régulateurs candidats. Nous proposons dans ce travail d’étendre LICORN pour qu’il puisse traiter une contrainte de sélection de corégulateurs candidats adaptative pour chaque gène, prenant en compte le support du gène cible et bornant le nombre de corégulateurs candidats possibles. La suite de cet article est organisée comme suit. Dans la section 2, nous introduisons brièvement le principe de LICORN. Dans la section 3, nous détaillons l’extension de LICORN à la