# Inferring gene regulatory networks from single-cell data: a mechanistic approach

Ulysse Herbach*[1,2,3], Thibault Espinasse[1], and Olivier Gandrillon[2,3]

[1]Institut Camille Jordan (ICJ) – Institut National des Sciences Appliquées [INSA], Ecole Centrale de Lyon, Université Claude Bernard - Lyon I (UCBL), CNRS : UMR5208, Université Jean Monnet - Saint-Etienne – Bât. Jean Braconnier n 101 43 Bd du 11 novembre 1918 69622 VILLEURBANNE CEDEX, France

[2]Inria team DRACULA (Inria Grenoble Rhône-Alpes / Institut Camille Jordan) – Université Claude Bernard - Lyon I (UCBL), CNRS : UMR5534, CNRS : UMR5208, INRIA, Institut Camille Jordan – Institut Camille Jordan Université Claude Bernard Lyon 1 43 boulevard du 11 novembre 1918 69622 Villeurbanne cedex France, France

[3]Laboratoire de Biologie et Modélisation de la Cellule (LBMC) – CNRS : UMR5239, Institut national de la recherche agronomique (INRA) : UR5239, Université Claude Bernard - Lyon I (UCBL), École Normale Supérieure (ENS) - Lyon – ENS de Lyon 46 allée d'Italie 69364 LYON Cedex 07, France

**Résumé**

The behaviour of a cell population largely depends on decisions made at the individual level, resulting from a complex interaction network between genes within each cell. Gene network inference, i.e. reconstruction of both structure and underlying dynamics from data, has been a major challenge in systems biology for years. So far, most studies have been based on population-averaged data: now that new technologies enable to observe mRNA levels in individual cells, a revolution in terms of precision, this problem paradoxically remains more challenging than ever.

Besides, gene expression being fundamentally stochastic, any relevant modelling framework must have a probabilistic base to fully exploit such single-cell data, where biological stochasticity is preserved from averaging. We will focus on the construction of a dynamical model for a set of interacting genes, that should be relevant both from a mathematical and from a biological perspective. The starting point is a simple yet rich model of single gene expression, the well known "two-state model", for which one can compute analytically and infer the stationary distribution. The production of mRNA and proteins will be described with a two-state-like model, the paramaters of which will now depend on other genes. The result is a general network model where each link between two genes is directed and has an explicit biochemical interpretation, in terms of chemical reactions.

The network distribution then can be approximately computed and used to derive a promising statistical model for the data, where stochasticity is not just noise but also contains information. This work should eventually provide an efficient way to infer gene networks from single-cell transcriptomics data.

**Mots-Clés:** Stochastic gene expression, biochemical networks, Markov processes

*Intervenant