

Classification adaptative de séries temporelles : application à l'identification des gènes exprimés au cours du cycle cellulaire

Alpha Diallo^{*,**}, Ahlame Douzal-Chouakria^{*} et Françoise Giroud^{**}

^{*}TIMC-IMAG TIMB(CNRS-UMR 5525), Université Joseph Fourier Grenoble 1
F-38706 La Tronche Cedex, France
(alpha.diallo, ahlame.douzal)[@]imag.fr
<http://www-timc.imag.fr/Ahlame.Douzal/>

^{**}TIMC-IMAG RFMQ (CNRS-UMR 5525), Université Joseph Fourier Grenoble 1
F-38706 La Tronche Cedex, France
francoise.giroud[@]imag.fr
<http://www-timc.imag.fr/Francoise.Giroud/index.html>

Résumé. Ce travail s'inscrit dans le cadre de l'étude de la division cellulaire assurant la prolifération des cellules. Une meilleure compréhension de ce phénomène biologique nécessite l'identification des gènes caractérisant chaque phase du cycle cellulaire. Le procédé d'identification est généralement basé sur un ensemble de gènes dits gènes de référence, sélectionnés expérimentalement et considérés comme caractérisant les phases du cycle cellulaire. Les niveaux d'expression des gènes étudiés sont mesurés durant le cycle de la division cellulaire et permettent de construire des profils d'expression. Chaque gène étudié est affecté à la phase du cycle cellulaire correspondant au groupe de gènes de référence le plus similaire. Cette approche classique souffre de deux limites. D'une part, les mesures de proximité les plus couramment utilisées entre profils d'expression de gènes sont basées sur les écarts en valeurs sans tenir compte de la forme des profils. D'autre part, dans la littérature il n'y a pas consensus quant à l'ensemble des gènes de référence à considérer. Dans cet article, notre but est de proposer une classification adaptative, basée sur un indice de dissimilarité incluant les proximités en valeurs et en forme des profils d'expression de gènes, permettant d'identifier les phases d'expression des gènes étudiés, et de présenter un nouvel ensemble de gènes de référence validé par une connaissance biologique.

1 Introduction

Les puces à ADN permettent de mesurer simultanément le niveau d'expression de plusieurs milliers de gènes dans un type cellulaire et un contexte physiologique et/ou pathologique particulier. Ces mesures du transcriptome permettent notamment d'étudier la cinétique des phénomènes cellulaires comme le cycle cellulaire (Spellman et al., 1998) pour ainsi défi-