

## **Prise en compte du caractère compositionnel des données d'expression différentielle des gènes : une approche par recherche de communautés.**

Emmanuel CURIS<sup>1,2,3</sup>, Cindie Courtin<sup>2</sup>, Pierre A. GEOFFROY<sup>2</sup>, Jeverson MOREIRA<sup>2</sup>,  
Cynthia MARIE-CLAIRE<sup>2</sup>, Bruno SAUBAMÉA<sup>2</sup>

1. Laboratoire de biomathématiques, faculté de pharmacie de Paris, université Paris Descartes
2. INSERM-Paris Descartes-Paris Diderot, UMR 1144 « Variabilité de réponse aux psychotropes »
3. Département de biostatistiques et d'informatique médicale, hôpital Saint-Louis, Assistance publique – hôpitaux de Paris

De nombreuses expériences, en biologie, s'intéressent à la quantification des acides ribonucléiques (A. R. N.) dans un échantillon biologique. Plusieurs méthodes sont utilisées : puces (*microarrays*), réaction de polymérisation en chaîne quantitative après rétrotranscription (RT-qPCR), séquençage massif après amplification (RNAseq)... Très souvent, le but de ces expériences est de comparer les niveaux d'expression de ces A. R. N. entre différentes conditions : types cellulaires, absence ou présence d'un traitement ou d'une pathologie... Plusieurs approches statistiques classiques ont été proposées, pour chacune des méthodes expérimentales («  $\Delta\Delta Ct$  » en RT-qPCR, par exemple).

Le plus souvent, les valeurs obtenues après ces expériences sont interprétées comme des concentrations ou des quantités (en unités arbitraires éventuellement). Cependant, la préparation des échantillons demande dans la majorité des cas une étape d'extraction d'une quantité prédéterminée d'A. R. N. de l'échantillon avant de commencer la quantification. De ce fait, les valeurs obtenues sont en réalité des *fractions molaires* — en effet, si l'on suppose que l'échantillon contient K sortes d'A. R. N. et si l'on note  $q_i$  la quantité de l'A. R. N. de type  $i$ , avec  $1 \leq i \leq K$ , alors dans l'aliquot qui est effectivement

quantifié ce même A. R. N. est présent à la concentration  $x_i = \frac{q_i}{\sum_{j=1}^K q_j}$  (sa fraction molaire dans le

mélange). De ce fait, toute modification de la concentration d'un seul A. R. N., quel qu'il soit, dans l'échantillon d'intérêt se traduira par la modification de *toutes* les concentrations dans l'aliquot quantifié. Conséquemment, les modifications de quantité d'un A. R. N. individuel ne sont pas interprétables.

Techniquement, l'ensemble des fractions molaires,  $(x_i)_{1 \leq i \leq K}$ , est un échantillon compositionnel. Il est cependant incomplet, car en général les méthodes de quantification ne détectent pas l'intégralité des K différents types d'A. R. N. Cela rend délicat l'utilisation des techniques usuelles d'analyse des données compositionnelles (travaux d'Aitchison, [1] par exemple).

Afin de traiter ces données, nous proposons une approche fondée sur la théorie des graphes. Cette approche, qui généralise la méthode du gène de référence utilisée en RT-qPCR, est une approche en deux temps. La première étape est de calculer tous les rapports  $x_i/x_j$  ( $1 \leq i < j \leq K$ ) pour chacun des échantillons et d'analyser les variations de ces rapports avec la technique statistique appropriée au plan expérimental et à la question posée. À partir des résultats de cette analyse, typiquement des degrés de signification obtenus, un graphe est construit : les nœuds du graphe sont les K types d'A. R. N. ; deux nœuds sont reliés si la méthode statistique n'a pas permis de mettre en évidence une modification du rapport correspondant. La recherche de groupes de nœuds associés dans le graphe (par exemple, de communautés ou de cliques) permet de définir des groupes d'A. R. N. qui se comportent de façon similaire ; des groupes distincts montrent au contraire un comportement différent des A. R. N. qui les constituent.

Nous illustrerons l'application de cette méthode en RT-qPCR et en RNAseq sur quelques exemples.

### **Références**

- [1] John AITCHISON, « The Statistical Analysis of Geochemical Data », *Mathematical Geology*, 1984, vol. 16 n° 6, p. 531-564.