



Physiologie de la Reproduction et des Comportements, UMR 7247, équipe "Biologie et Bioinformatique des Systèmes de Signalisation" <http://bios.tours.inra.fr>

### Sujet de stage de M2 : Elaboration d'un workflow dans Cytoscape pour l'analyse fonctionnelle des données biologiques à haut débit

Notre équipe met en œuvre une démarche de biologie des systèmes pour analyser les réseaux de signalisation induits par les récepteurs d'hormone gonadotropes, RLH et RFSH. Cette démarche comporte 2 piliers principaux : 1/ la modélisation mathématique dynamique qui permet de comprendre comment varient les concentrations relatives d'un petit nombre de molécules de signalisation dans le temps, et éventuellement dans l'espace ; 2/ l'analyse de données à haut débit de type transcriptomique ou (phospho)protéomique qui permet de retracer les relations fonctionnelles entre les effecteurs d'un même réseau de signalisation. A l'heure actuelle, ce second point souffre d'un manque de démarche structurée pour tirer le meilleur bénéfice de l'analyse fonctionnelle des réseaux de signalisation/ de gènes.

Nous disposons déjà de données à haut débit (protéomique, phosphoprotéomique, transcriptomique, traductomique) obtenues dans des modèles de cellules immortalisées ou de cellules natives exprimant le RFSH. Les données brutes ont été normalisées, cartographiées sur le génome de référence et analysées au plan statistique.

Le sujet de M2 consistera à établir un workflow standardisé pour l'analyse fonctionnelle des réseaux. Pour cela, l'étudiant devra être capable de discriminer les plug-ins les plus adaptés parmi les 140 auxquels le logiciel en accès libre Cytoscape est rattaché et que nous avons largement commencé à explorer. Par exemple pour des données de phosphoprotéomique, il s'agira en particulier :

- de reconstruire les relations entre protéines du réseau en respectant les sens de variation
- d'identifier les termes d'enrichissement fonctionnels en visualisant les relations entre pathways fonctionnels
- de prédire la présence de protéines constitutives du réseau en question mais qui n'ont pas pu être identifiées sur le plan expérimental, notamment les plus proche voisins
- de superposer des réseaux obtenus dans différentes conditions expérimentales (ex : avec ou sans FSH)
- de déterminer les paramètres du réseau : connectivité, modularité, identification des sous-réseaux, des hubs, etc)
- de rechercher des motifs, des cycles, etc

Pour l'analyse de données transcriptomiques, on cherchera à :

- identifier les motifs consensus sur une liste de gènes et les facteurs de transcriptions qui s'y rattachent
- identifier les plus proches « upstream regulators »
- clusteriser les gènes de niveau d'expression analogue
- créer des réseaux randomisés et les comparer au résultats des réseaux expérimentaux

La preuve de concept sera réalisée en reconstruisant le réseau protéines/ gènes-cibles et en le comparant avec un réseau que nous avons reconstruit à l'aide d'un logiciel commercial (IPA, Ingenuity Pathway Analysis). Ce travail sera d'une grande utilité car il devrait aboutir à la mise en commun d'une interface entre biologistes et modélisateurs de l'équipe. En effet, pour les biologistes, ce workflow offrira un ensemble d'analyses des données nettement approfondies par rapport aux logiciels commerciaux de type IPA ou Pathway Studio. Pour les modélisateurs, les plug-ins de

Cytoscape, de par leur sémantique, fournissent des paramètres quantitatifs sur les propriétés du réseau et constituent ainsi une fenêtre vers la modélisation dynamique.

**Contacts :**

Romain Yvinec, CR INRA : [Romain.Yvinec@inra.fr](mailto:Romain.Yvinec@inra.fr) ; tél : 33 2 47 42 75 05

Pascale Crépieux, DR CNRS, Co-responsable de l'équipe BIOS : [Pascale.Crepieux@inra.fr](mailto:Pascale.Crepieux@inra.fr)

tél : 33 2 47 42 75 14