

Proposition de stage de Recherche en Laboratoire - Master 1

Romain Yvinec - INRA Nouzilly

1^{er} décembre 2016

Analyse statistique de données de transcriptome, traductome, et protéome. Application au réseau FSH.

L'unité Physiologie de la Reproduction et des Comportements (PRC) est une unité mixte de recherches du centre INRA Val de Loire, basée à Nouzilly (20 km de Tours). L'unité PRC comporte 11 équipes de recherche dont les thématiques sont organisées autour de l'axe de contrôle de la reproduction et des comportements, allant de l'analyse comportementale et neurobiologique à l'analyse des mécanismes moléculaires et génétiques liés à la fonction de reproduction et sa régulation. Au sein de l'unité, l'équipe Biologie et Bioinformatique des Systèmes de Signalisation (BIOS) développe une stratégie pluridisciplinaire de biologie systémique combinant des approches expérimentales (aux échelles moléculaire, cellulaire, physiologique), et de modélisation mathématique et informatique, afin d'acquérir un savoir exhaustif et quantitatif des phénomènes biologiques médiés par les récepteurs couplés aux protéines G, en particulier ceux liés à la reproduction.

Les objectifs de l'équipe BIOS sont de comprendre la dynamique des réseaux de signalisation cellulaire induits par la stimulation de ces récepteurs par un ligand donné, en définissant précisément les effets intracellulaires induits et leurs conséquences biologiques. Cette compréhension permet à l'équipe BIOS, en parallèle, de développer de nouvelles approches pharmacologiques afin de contrôler finement l'activité de récepteurs liés à la reproduction et pour réduire les effets néfastes associés à l'utilisation de leur ligand (naturel ou de synthèse) chez les animaux d'élevage et chez l'humain.

Analyse différentielle de données "omics" Le sujet de stage de recherche proposé porte sur l'analyse différentielle de données haut-débit, de type transcriptomique (mesure de quantité d'ARNm dans la cellule), de type traductomique (mesure de la quantité d'ARNm traduite dans la cellule) et protéomique (mesure de quantité de protéines dans la cellule).

Les approches expérimentales de type séquençage à haut-débit se heurtent à des questions d'analyses statistiques inédites, si l'on veut déterminer des différentiels dans une condition stimulée par rapport à une condition contrôle. La situation désormais classique consiste à séquencer les ARNm ou protéines produits (à l'échelle d'une cellule ou d'une population de cellules) dans deux conditions expérimentales différentes. On cherche alors à identifier les gènes pour lesquels l'activité (quantité de produits géniques) a évolué, soit augmenté soit diminué. Bien que des méthodes d'analyse différentielle de transcriptome à partir de données RNA-Seq existent déjà [1-6], elles sont encore perfectibles, et aucune méthode à ce jour ne fait complètement consensus au sein de la communauté. La difficulté résulte de plusieurs facteurs : le faible nombre de répétitions des expériences (dû principalement à leur coût), le très grand nombre de tests statistiques effectués en parallèle (au moins un par gène), la possible inter-dépendance entre les gènes, les différentes normalisations dues aux cycles d'amplifications et d'échantillonnage des ARNm ou encore le faible niveau d'expression d'une grande partie d'ARNm (seulement quelques copies par échantillon). Ce problème méthodologique est encore plus criant pour l'analyse des données de traductome (recrutement des ribosomes aux ARNm) car il a encore été peu abordé. On cherchera ainsi à développer des méthodes rigoureuses permettant d'attester si l'analyse de l'expression différentielle pour chaque gène est possible au regard des données et de quantifier celle-ci le cas échéant, en limitant le nombre de faux positifs.

Enfin, nous nous attacherons à mettre en regard plusieurs jeux de données de natures différentes (RNA-seq et protéome obtenu par spectrométrie de masse). En particulier, la confrontation de données de transcriptome, de traductome et de protéome doit permettre d'obtenir un niveau de corrélation statistique de l'activité des molécules satisfaisant, que les seules données de transcriptome et protéome ne permettent pas d'obtenir. De plus, la prise en compte simultanée de ces trois types de données permettra de proposer différents mécanismes

de régulation (transcriptionnel, traductionnel, post-traductionnel et dégradation des molécules) présents entre deux conditions différentes.

Les méthodes développées seront appliquées à des données de transcriptome, traductome, et protéome disponibles au sein de notre équipe, dans des cellules de Sertoli de rat en culture primaire stimulées in vitro par l'hormone FSH [7-9]. Si le temps le permet, nous chercherons à identifier de manière robuste le réseau de gènes et de protéines activés et inhibés dans les cellules de Sertoli suite à la stimulation par FSH.

- 1- Li et al., Biometrics 70(4) :872-80, 2014.
- 2- Canale and Dunson, J Am Stat Assoc 106(496) :1528-1539, 2011.
- 3- Blei, Commun ACM 55(4) : 77-84, 2012.
- 4- Benjamini and Hochberg, J.R.Stat. S B 57(1), 289-300, 1995.
- 5- McCarthy et al. Nucl. Acids. R. 40(10) 4288-4297, 2012.
- 6- Bullard et al. BMC Bioinfo. 11(94), 2010.
- 7- Musnier, León et al. Mol. Endocrinol., 26 (669-680) 2012.
- 8- Leon et al. J. Mol. Endocrinol. 52(373-382) 2014.
- 9- Le Borgne et al. Development, 141(2096-107) 2014.

Détails pratiques Contact : Romain Yvinec

romain.yvinec@tours.inra.fr

02 47 42 75 05

Equipe BIOS : <http://bios.tours.inra.fr/>

Page personnelle : <http://yvinec.perso.math.cnrs.fr/>