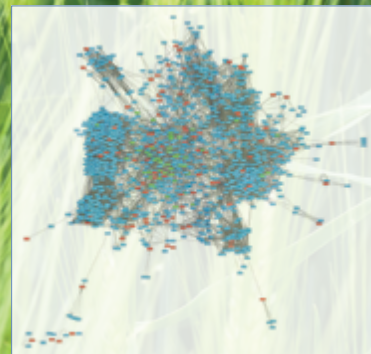
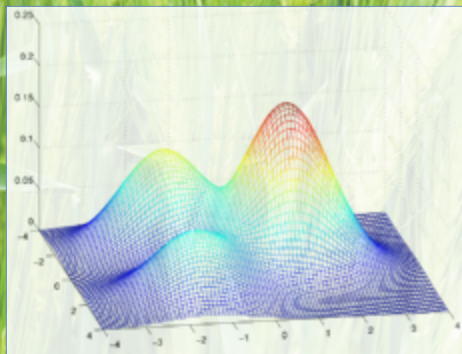
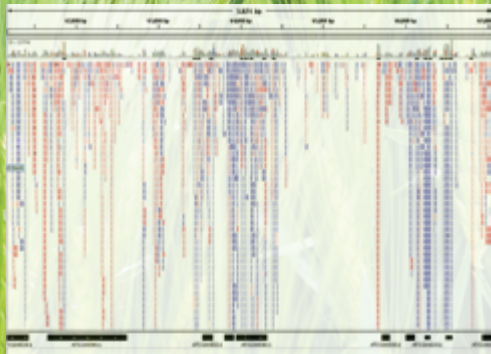


Ecole-Chercheur 2017

De l'expression des gènes aux réseaux

Modules



25/04/2017

SACLAY PLANT SCIENCES

Table des matières

Informations générales	3
Module 1 : Les technologies de séquençage actuelles et futures et leurs protocoles associés	4
Module 2 : La bioinformatique du RNAseq	5
Module 3 : Normalisation et analyse différentielle des données RNAseq	6
Module 4 : Analyse de la coexpression de données RNAseq	7
Module 5 : Les réseaux de régulation	8
Module 6 : Les intégrations de données omiques	9

Organisateurs

Marie-Laure Martin-Magniette et Etienne Delannoy

Lieu

Institut de Sciences des Plantes de Paris-Saclay (IPS2)
> Amphithéâtre (matin) et Salle rouge (après-midi)

Langue

Français (diapositives des présentations en anglais)

Format

Matin > Cours en amphithéâtre, 2 fois 90 minutes avec une pause de 15 minutes
Après-midi > 2 TP de 90 minutes pour un groupe de 20 personnes

Les modules (cours et TP) sont organisés selon les différentes étapes de l'analyse mais sont indépendants.

L'inscription est obligatoire.

Vous pouvez vous inscrire pour un seul module ou pour plusieurs.

De même, vous pouvez vous inscrire seulement pour les présentations du matin en amphithéâtre si vous le souhaitez. Par contre, si vous vous inscrivez pour une session de l'après-midi (en petit groupe), nous vous demandons d'assister obligatoirement aux présentations du matin du module correspondant.

Si vous vous êtes inscrits à une ou plusieurs sessions de l'après-midi, nous vous demandons de rédiger une lettre de motivation expliquant vos raisons de vouloir y participer. Si le nombre d'inscrits est supérieur au nombre de places disponibles, nous utiliserons ces lettres pour sélectionner les participants. Dans cette lettre de motivation, n'oubliez pas de donner des informations sur le(s) projet(s) sur le(s)quel(s) vous souhaitez appliquer les méthodes présentées.

Public visé

Pour les cours du matin : Tout membre de l'Université Paris-Saclay dans la limite de la capacité d'accueil de l'amphithéâtre, soit 190 personnes.

Pour les TP de l'après-midi : Ouverts en priorité aux membre du LabEx Sciences des Plantes de Saclay dans la limite de 20 personnes (s'il reste des places, celles-ci seront ouvertes à tout membre de l'Université Paris-Saclay).

Prérequis

Pour tous : Connaître les notions de moyennes et variance. Pas de lecture recommandée.

Pour les TP de l'après-midi : venir avec un ordinateur avec un OS Windows, Mac ou Linux sur lequel sera installé R avec les librairies requises pour le TP. Savoir utiliser R en lignes de commande ou Rstudio.

Modalités d'inscription

Inscription obligatoire en ligne avant le 17 avril 2017 au soir (avec chargement de la lettre de motivation le cas échéant).

Modules

Module 1 : Les technologies de séquençage actuelles et futures et leurs protocoles associés

Dates : 30 mai 2017

Intervenants : Mathilde Clément et Jérémie Bazin

Contenu de la formation : Ce premier module a pour objectif de présenter les technologies de séquençage et de leurs capacités en terme technique et biologique.

> **Matin (cours en amphithéâtre) :**

9h-10h30 : **Les technologies de séquençage existantes et à venir.** Cet exposé permettra de faire un état de lieu des technologies actuellement disponibles. Il sera proposé une description de leurs caractéristiques en nombre de reads, leur longueur et leur qualité. La facilité de mise en œuvre et l'infrastructure informatique seront discutées. Enfin les avantages et inconvénients de chacune des technologies pour le RNA-seq seront présentées. La dernière partie de l'exposé sera une prospective sur les technologies à venir.

10h45-12h15 : **Les protocoles.** Cet exposé permettra de comprendre le principe général des banques de séquençage. La première partie sera consacrée aux étapes des banques Truseq stranded et leur susceptibilité à introduire de la variabilité technique. La seconde partie de l'exposé présentera d'autres types de banques (small RNA, ultra low, single cell...). La troisième partie abordera les possibilités d'étudier d'autres phénomènes moléculaires par exemple les empreintes des RNA binding proteins, les sites de fixation de protéines ou autres marques, les empreintes ribosomales, la stabilité des ARNs ou encore la détermination des extrémités des ARNs. La dernière partie de l'exposé abordera les possibilités de faire des banques sans utilisation de kits commerciaux.

> **Après-midi :**

Des visites de la plateforme transcriptomique de l'IPS2 seront proposées et des discussions seront organisées avec les intervenants et les membres de la plateforme.

Module 2 : La bioinformatique du RNAseq

Dates : 29 juin 2017

Intervenants : Véronique Brunaud et Claire Toffano-Nioche

Contenu de la formation : Ce second module a pour objectif de présenter les techniques bioinformatiques associées au RNA-seq.

> **Matin (cours en amphithéâtre) :**

9h-10h30 : **Bioinformatique pour l'analyse d'expression.** Cet exposé permettra de faire un état de lieu des étapes bioinformatiques pour quantifier l'expression des gènes à partir de données RNAseq. A partir de l'expérience des intervenants et de la littérature, il sera proposé une description de chacune de ces étapes qui sont la qualité, le mapping/assemblage et le comptage. Les outils disponibles et l'infrastructure informatique nécessaires seront discutées.

10h45-12h15 : **Autres exploitations bioinformatiques du RNAseq.** Cet exposé permettra de voir comment la technologie du RNAseq peut être utilisée pour l'identification et la quantification des isoformes et l'analyse des petits RNAs.

> **Après-midi (atelier de 20 personnes) :**

Discussion à partir d'exemples. Les participants seront sollicités avant le module pour préciser les points qu'ils souhaitent voir aborder ou approfondir.

Module 3 : Normalisation et analyse différentielle des données RNAseq

Dates : 28 septembre 2017

Intervenants : Julie Aubert, Etienne Delannoy et Marie-Laure Martin-Magniette

Contenu de la formation : Ce troisième module a pour objectif de présenter la correction des biais techniques et la détection des gènes impactés dans leur transcription quand plusieurs conditions sont étudiées.

> **Matin (cours en amphithéâtre) :**

9h-10h30 : **La normalisation.** L'objectif de la normalisation est de corriger les biais techniques. Le biais le plus important à corriger dans un projet RNAseq est la différence du nombre de reads mappés pour les différents échantillons considérés. Cet exposé s'appuiera sur la publication de Dillies et al. (2012) dans Briefings in Bioinformatics qui a mené une étude d'évaluation des différentes méthodes de normalisation proposées. Après la présentation de cette publication et de ses conclusions, nous discuterons de la mise en œuvre.

10h45-12h15 : **L'analyse différentielle.** L'objectif de l'analyse différentielle est d'identifier les gènes dont l'expression est dépendante des conditions étudiées. Cet exposé s'appuiera sur la publication de Rigai et al. (2016) dans Briefings in Bioinformatics qui a mené une large étude de simulation pour identifier les étapes importantes de l'analyse différentielle. Après la présentation de cette publication et de ses conclusions, nous verrons la mise en œuvre sur des exemples.

> **Après-midi (TP pour un groupe de 20 personnes) :**

Le TP sera consacré à la normalisation et l'analyse différentielle d'un jeu de données RNAseq. Nous identifierons ensemble les étapes de l'analyse et les implémenterons en R avec le package edgeR.

Module 4 : Analyse de la coexpression de données RNAseq

Dates : 12 octobre 2017

Intervenants : Andréa Rau (à confirmer), Etienne Delannoy et Marie-Laure Martin-Magniette

Contenu de la formation : L'analyse de la coexpression a pour objectif d'identifier des groupes de gènes ayant un même profil d'expression car ils sont de potentiels partenaires fonctionnels. Ce quatrième module a pour objectif de montrer à l'aide d'exemples l'information biologique apportée par ce type d'analyse puis de présenter les méthodes.

> **Matin (cours en amphithéâtre) :**

9h-10h30 : **Exploitation de l'analyse de la coexpression dans des projets biologiques.** L'analyse de la coexpression est souvent l'étape après l'analyse différentielle. Cela permet d'identifier des groupes de gènes qui ont le même profil d'expression sur l'ensemble des échantillons du projet. A l'aide d'analyses de coexpression publiées, nous discuterons des approches utilisées et des résultats obtenus.

10h45-12h15 : **La coexpression.** Après une introduction générale des méthodes de classification non supervisée, l'exposé focalisera sur les modèles de mélange qui sont des modèles probabilistes qui permettent d'identifier des groupes de gènes ayant un même profil d'expression. Le principe du modèle de mélange et son utilisation pour modéliser des données microarray et RNAseq seront présentés. Nous présenterons également les sorties du modèle et leur interprétation d'un point de vue de la qualité de l'analyse.

> **Après-midi (2 x TP pour un groupe de 20 personnes) :**

Deux TP sont proposés :

1) Un TP consacré à la mise en œuvre de modèles de mélange pour l'analyse de données RNAseq. L'analyse sera faite à l'aide du package coseq.

2) Un TP consacré à la prise en main du module GEM2net de CATdb (<http://tools.ips2.u-psud.fr//GEM2NET/>) qui donne accès aux résultats d'une étude de coexpression de données microarrays d'Arabidopsis soumis à un stress (Zaag et al, 2015, NAR). Après une présentation du projet, les participants pourront interroger la base de données avec leur liste de gènes d'intérêt et comprendre comment l'utiliser. Ce sera aussi l'occasion pour les développeurs de la base d'avoir des retours pour de futurs développements. Ce TP aura lieu dans la salle de TP de l'IPS2.

Module 5 : Les réseaux de régulation

Dates : 30 novembre 2017

Intervenants : Guillem Rigail, Marie-Laure Martin-Magniette et Etienne Delannoy

Contenu de la formation : L'inférence d'un réseau de régulation a pour objectif d'identifier des interactions entre gènes à partir de données transcriptomiques. Ce cinquième module a pour objectif de présenter l'information biologique apportée par ce type d'analyse puis de présenter les principes de ces méthodes.

> **Matin (cours en amphithéâtre) :**

9h-10h30 : **Exemples de réseaux de régulation.** A l'aide de réseaux de régulation publiés, nous présenterons les approches, le type de données nécessaires et les résultats obtenus.

10h45-12h15 : **Les méthodes d'inférence de réseaux.** Nous présenterons le concept des méthodes utilisées pour inférer un réseau de régulation en précisant le type de données utilisées, les possibles difficultés techniques de mise en œuvre et la manière dont il faut interpréter les résultats. Nous discuterons de l'apport de la modélisation pour évaluer la robustesse des interactions identifiées.

> **Après-midi (TP pour un groupe de 20 personnes) :**

Le TP sera consacré à la notion de la corrélation partielle utilisée dans les modèles graphiques Gaussiens au travers de rééchantillonnage de jeux de données réels et de données simulées avec le package R huge.

Module 6 : Les intégrations de données omiques

Dates : 14 décembre 2017

Intervenants : Sébastien Déjean

Contenu de la formation : L'intégration de données signifie mettre ensemble des données de natures différentes sur des entités biologiques différentes pour en extraire de la connaissance. Ce sixième module a pour objectif de présenter plusieurs principes d'intégration sur la base d'exemples utilisant des données et des méthodes différentes.

> **Matin (cours en amphithéâtre) :**

9h-10h30 : **Exemples d'intégration de données omiques.** A l'aide d'exemples tirés de la littérature, nous présentons des approches, le type de données utilisées et les résultats obtenus.

10h45-12h15 : **Intégration de données omiques avec MixOmics.** MixOmics est un package R qui rassemble un grand nombre de méthodes capables d'intégrer des données omiques de types différents obtenues sur les mêmes échantillons. L'objectif de cet exposé est de présenter les concepts de ces méthodes.

> **Après-midi (TP pour un groupe de 20 personnes) :**

Le TP sera consacré à la découverte du package R MixOmics.



SACLAY PLANT SCIENCES



LabEx Sciences des Plantes de Saclay (SPS) - Institut Jean-Pierre Bourgin - Bâtiment 2 - INRA Centre de Versailles-Grignon
Route de St-Cyr (RD 10) - 78026 Versailles Cedex - France - Tél: + 33(0)1 30 83 30 82 - sps@versailles.inra.fr

www.saclayplantsciences.fr