

# Projets 2018-2019

Bioinformatique Appliquée 2

# Déroulement de l'UE

12 groupes : 10 trinômes et 2 binômes

## **Séances de TP**

24 octobre => 21 novembre

## **Présentation orale de la stratégie (13 novembre 2018) - en anglais**

5 min de présentation / 10 min de questions

## **Rapport écrit en PDF (16 décembre 2018) - en français ou en anglais**

10 pages environ

## **Présentation orale du projet (21 décembre 2018) - en français ou en anglais**

10 min de présentation / 10 min de questions

# Sujet n°1 : Modélisation des protéines OcKAI2 sur les récepteurs aux strigolactones

Les plantes parasites des genres *Phelipanche*, *Orobanche* et *Striga* sont responsables de dégâts considérables dans de nombreux agrosystèmes. Une des adaptations de ces organismes au mode de vie parasitaire est une germination des graines qui s'effectue en réponse à la perception de composés présents dans la rhizosphère de la future plante hôte. Les composés les plus largement perçus par ces plantes parasites sont associés aux strigolactones. Les strigolactones sont des phytohormones impliquées dans la régulation de nombreux processus au sein des plantes mais également dans la mise en place de la symbiose avec les champignons mycorhiziens. Les plantes parasites détourneraient donc ce signal allélopathique exsudé par leur future plante hôte pour optimiser la possibilité de se fixer aux racines de cette dernière. Des récepteurs, de type  $\alpha/\beta$ -hydrolase, homologues à ceux impliqués dans la perception des strigolactones (D14/KAI/HTL) chez les plantes non parasite seraient responsables de cette perception. Disposant du génôme d'*O. cumana*, les séquences des paralogues *KAI2* ont été obtenues. L'objectif de ce projet sera de modéliser les structures des différentes protéines OcKAI2 puis d'identifier celles qui peuvent se lier à des strigolactones ou à un autre stimulant de germination le dehydrocostus lactone. Ce dernier stimulant est spécifique d'*O. cumana*.

## Sujet n°2 : Analyse comparative des domaines CHASE des récepteurs aux cytokinines

- Les orobanches (*Orobanche* et *Phelipanche* ssp) sont des plantes parasites obligatoires, dépourvues de chlorophylle et entièrement dépendantes de leur hôte pour leur développement et qui dans certains agrosystèmes sont devenues des adventices dévastatrices. Parmi ces espèces, quelques-unes dont *Phelipanche ramosa* et *Orobanche cumana* sont considérées comme de véritables fléaux agronomiques dans les cultures oléagineuses (colza, tournesol). Après l'étape de germination qui résulte d'une induction par des molécules stimulatrices produites par l'hôte (strigolactones), l'autre étape clé du mécanisme parasitaire est l'établissement de l'interaction entre la plante parasite et la plante hôte. La racine de l'orobanche entre en effet en contact avec la racine hôte et met en place une structure appelée haustorium qui permettra d'établir des connexions vasculaires nécessaires à l'orobanche afin de prélever chez l'hôte l'eau et les nutriments nécessaires à sa propre croissance. Ce processus « d'haustoriogenèse » qui permet donc au parasitisme de se mettre en place est lui aussi sous le contrôle de molécules produites par l'hôte et notamment des cytokinines.

## Sujet n°2 : Analyse comparative des domaines CHASE des récepteurs aux cytokinines (suite du projet)

- Les cytokinines (CKs) sont des hormones végétales contrôlant un grand nombre de processus du développement de la plante tels que la division cellulaire et l'activité méristématique, la croissance, la ramification, l'activité stomatique, la nutrition ou la sénescence. Toutefois, l'exsudation de ces molécules dans la rhizosphère et leur perception par d'autres plantes n'avaient pour l'heure pas été démontrées. Une étude récente au sein de notre laboratoire a montré que ce nouveau signal allélopathique existait et que le parasite serait donc en mesure de différencier les CKs émises par l'hôte de ses propres CKs. Ces hormones sont normalement perçues par des récepteurs de type Histidine Kinase (HK2, HK3 et HK4) dont la structure a été clairement décrite chez *Arabidopsis thaliana*. Plus précisément, l'interaction entre les CKs et le récepteur se fait au niveau d'un domaine transmembranaire (CHASE) qui est responsable de la spécificité de la protéine réceptrice pour un type de CK. Une étude de « Structure Activity Relationship » (SAR) a déjà été menée au sein de notre laboratoire lors de l'haustoriogenèse de *P. ramosa* et les séquences des gènes codant ces récepteurs ont été obtenues chez *P. ramosa* et *O. cumana*. L'objectif de ce projet sera d'identifier les domaines CHASE des différentes protéines PrHKs et OcHKs puis de comparer ces derniers avec ceux d'autres plantes disponibles dans les banques de données afin de déterminer les CKs pouvant s'y fixer préférentiellement voire d'aller jusqu'à une étude phylogénétique si cela est possible.

# Projet 3 : ATIS

LES ATIS POUR « MRNAS PREDICTS ALTERNATIVE TRANSLATION INITIATION SITES ».

**Etat des lieux** - Sur la base des publications récentes portant sur la découverte de ces séquences ATIS, pourriez vous mettre en place une recherche systématique de ce type de motif d'initiation ?

**Objectif** - Les canaux sodiques et potassiques cardiaques responsables de la genèse du potentiel d'action myocardique peuvent ils être modulés par ces phénomènes? Il peut s'agir également de toute molécule impliquée dans le structure cardiaque responsable de pathologie. Une liste non exhaustive vous est proposée ci joint.

Proposez un projet d'investigation à plus grande échelle, sur ce thème, permettant d'apporter potentiellement des éléments de réponses pour faire progresser cette problématique en gardant une visée cardiovasculaire.

**Références introductives et non exhaustives :**

Wegrzyn JL, Drudge TM, Valafar F, Hook V. - BMC Bioinformatics. 2008 May 8;9:232.

Bioinformatic analyses of mammalian 5'-UTR sequence properties of mRNAs predicts alternative translation initiation sites.

# Projet 4 : miRNA

## PRÉDICTION DE LA MISE EN PLACE D'UN NETWORK miRNA EN CARDIOVASVULAIRE

**Etat des lieux** - Sur la base des implications de plus en plus abondantes décrites pour ces ARN non codant, pourriez vous mettre en place une recherche systématique de ce type de régulation sur la base des moteurs de recherche disponible?

**Objectif** - Les canaux sodiques et potassiques cardiaques responsables de la genèse du potentiel d'action myocardique peuvent ils être modulés par ces phénomènes?

Proposez un projet d'investigation à plus grande échelle, sur ce thème, permettant d'apporter potentiellement des éléments de réponses pour faire progresser cette problématique en effectuant une étude étendue la plupart de ces canaux.

### **Références introductives et non exhaustives :**

Papoutsidakis N, Deftereos S, Kaoukis A, Bouras G, Giannopoulos G, Theodorakis A, Angelidis C, Hatzis G, Stefanadis C. MicroRNAs and the heart: small things do matter. Curr Top Med Chem. 2013;13(2):216-30.

Kim GH. MicroRNA regulation of cardiac conduction and arrhythmias. Transl Res. 2013 May;161(5):381-92.

# Projets 3 et 4 : listing de gène à éventuellement retenir

Des gènes ont été découverts et reliés à l'expression de divers phénotypes associés à des troubles du rythme, tels que, le syndrome du QT long de type 3, le syndrome de Brugada, la maladie de Lenègre ou la mort subite du nourrisson.

BrS	CACNB2
BrS	GPD1L
BrS	HCN4
BrS	KCNE3
BrS	RANGRF
BrS	SCN1B
BrS	SCN3B
BrS / ERS	KCNJ8
BrS / IVF	KCND3
BrS / IVF	KCNE1L
LQT	AKAP9
LQT	ANK2
LQT	CACNA1C
LQT	KCNE1
LQT	KCNE2
LQT	KCNH2
LQT	KCNJ2
LQT	KCNJ5
LQT	KCNQ1
LQT	SCN4B
LQT	SNTA1
LQT ?	NOS1AP
LQT / BrS / DCM / IVF...	SCN5A
LQT / HCM	CAV3



# Projet 5 : Analyse bibliographique, miRNA et myopathie

## RÉALISATION D'UNE MAP DES miRNA DANS LA MYOPATHIE

**Etat des lieux** - Sur la base des publications récentes portant sur la caractérisation des miRNA dans le système musculaire et ou cardiaque, pouvez vous essayer de mettre en place un réseau d'interaction miRNA/target pour ces organes?

**Objectif** – Mettre en place un outil d'investigation bibliographique qui générera les données qui seront secondairement nécessaires afin d'établir une « carte » voir un « réseau d'interaction », cette approche pouvant alors être étendue in fine à toute autre problématique.

Proposez un processus semi-automatisé, voir automatisé sur ce thème, permettant d'apporter potentiellement des éléments de réflexion pour faire progresser cette problématique en gardant une visée cardiologique dans un premier temps.

# Projet 6 : Analyse bioinformatique et microsatellite

## **Systemic Delivery of MicroRNA Using Recombinant Adeno-associated Virus Serotype 9 to Treat Neuromuscular Diseases in Rodents...**

Proc Natl Acad Sci U S A. 2018 Apr 17;115(16):4234-4239. doi: 10.1073/pnas.1716617115. Epub 2018 Apr 2.

### **Intron retention induced by microsatellite expansions as a disease biomarker.**

Sznaider ŁJ<sup>1</sup>, Thomas JD<sup>2</sup>, Carrell EM<sup>3</sup>, Reid T<sup>2</sup>, McFarland KN<sup>4</sup>, Cleary JD<sup>2</sup>, Oliveira R<sup>2</sup>, Nutter CA<sup>2</sup>, Bhatt K<sup>3</sup>, Sobczak K<sup>5</sup>, Ashizawa T<sup>6</sup>, Thornton CA<sup>3</sup>, Ranum LPW<sup>2</sup>, Swanson MS<sup>1</sup>.

**Etat des lieux** - Expansions of simple sequence repeats, or microsatellites, have been linked to ~30 neurological-neuromuscular diseases. While these expansions occur in coding and noncoding regions, microsatellite sequence and repeat length diversity is more prominent in introns with eight different trinucleotide to hexanucleotide repeats, causing hereditary diseases such as myotonic dystrophy type 2 (DM2), Fuchs endothelial corneal dystrophy (FECD), and C9orf72 amyotrophic lateral sclerosis and frontotemporal dementia (C9-ALS/FTD).

**Objectif** – Mettre en place un outil d'investigation de recherche systématique de ces séquences afin d'établir un « réseau de régulation » cette approche pouvant alors être étendue in fine à toute autre problématique. => This screening will constitute a rapid and inexpensive biomarker for intronic repeat expansion disease.

Proposez un processus semi-automatisé, voir automatisé sur ce thème, permettant d'apporter potentiellement des éléments de réflexion pour faire progresser cette problématique.

## **Sujet 7 : Analyses génomique dans l'insuffisance veineuse primitive : Analyses des miRNA**

L'insuffisance veineuse primitive est une pathologie fréquente touchant 24 à 57% des femmes et 12 à 26% des hommes. C'est une pathologie caractérisée par une incontinence valvulaire, une désorganisation de la paroi des veines saphènes des membres inférieurs associées à une hypo-réactivité du vaisseau. Il n'y a pas de véritable traitement des veines variqueuses. L'étiologie de la pathologie est à ce jour inconnu.

Une analyse d'expression des ARNm grâce à une puce d'expression Affymetrix (HuGene 2.1 ST 24 array-plate) a été entreprise à partir de veines saines et veines variqueuses. Après une normalisation des données, un fichier de données regroupant les miRNAs détectés et les valeurs d'expression a été généré (en LOG base 2).

**L'objectif de ce projet est d'analyser l'expression de ces miRNAs et de sélectionner les meilleurs miRNA candidats potentiellement impliqués dans le développement de la pathologie.**

Conditions :                   veines variqueuses avec ou sans endothélium (VE et -E)  
                                      veines saines avec ou sans endothélium (VE et -E)

Fichier : NormDataAnnot miRNA.txt

## **Sujet 8 : Analyses génomique dans l'insuffisance veineuse primitive : Analyses des miRNA**

L'insuffisance veineuse primitive est une pathologie fréquente touchant 24 à 57% des femmes et 12 à 26% des hommes. C'est une pathologie caractérisée par une incontinence valvulaire, une désorganisation de la paroi des veines saphènes des membres inférieurs associées à une hypo-réactivité du vaisseau. Il n'y a pas de véritable traitement des veines variqueuses. L'étiologie de la pathologie est à ce jour inconnu.

Une analyse d'expression des ARNm grâce à une puce d'expression Affymetrix (HuGene 2.1 ST 24 array-plate) a été entreprise à partir de veines saines et veines variqueuses. Après une normalisation des données, un fichier de données regroupant l'ensemble des ARNs détectés et les valeurs d'expression a été généré (en LOG base 2).

**L'objectif de ce projet est d'analyser l'expression des non-coding RNAs et de sélectionner les meilleurs candidats potentiellement impliqués dans le développement de la pathologie.**

Conditions :                   veines variqueuses avec ou sans endothélium (VE et -E)  
                                      veines saines avec ou sans endothélium (VE et -E)

Fichier : NormDataAnnot RNA NA

## Sujet 9 : Analyses génomique dans l'insuffisance veineuse primitive : Analyses des LncRNA

L'insuffisance veineuse primitive est une pathologie fréquente touchant 24 à 57% des femmes et 12 à 26% des hommes. C'est une pathologie caractérisée par une incontinence valvulaire, une désorganisation de la paroi des veines saphènes des membres inférieurs associées à une hypo-réactivité du vaisseau. Il n'y a pas de véritable traitement des veines variqueuses. L'étiologie de la pathologie est à ce jour inconnu.

Une analyse d'expression des ARNm et des ARN non-codants par micro-arrays a été entreprise par Li et ses collaborateurs, en 2014. Ils se sont focalisés sur les variations de LncRNA. A partir de mots clés (apoptose, prolifération, matrice extracellulaire etc...) ils ont sélectionné quelques candidats LncRNA puis se sont focalisé sur l'étude du LncRNA-GAS5.

**En reprenant les données de bases des expressions des LncRNA et en retravaillant les critères de sélection des LncRNA candidats (mots clé, niveau de différentiel d'expression, gènes cibles de la pathologie etc...), pourriez-vous proposer une sélection d'un ou plusieurs LncRNA pouvant faire l'objet d'une étude plus approfondie.**

[Aberrantly expressed lncRNAs in primary varicose great saphenous veins.](#) Li X, Jiang XY, Ge J, Wang J, Chen GJ, Xu L, Xie DY, Yuan TY, Zhang DS, Zhang H, Chen YH.

PLoS One. 2014 Jan 31;9(1):e86156. [Low expression of lncRNA-GAS5 is implicated in human primary varicose great saphenous veins.](#) Li L, Li X, The E, Wang LJ, Yuan TY, Wang SY, Feng J, Wang J, Liu Y, Wu YH, Ma XE, Ge J, Cui YY, Jiang XY. PLoS One. 2015 Mar 25;10(3):e0120550.

# Sujet 10(a,b): Inférence d'interactions microbiennes

Les études de métagénomiques mettent en évidence la présence et l'absence de différentes souches dans un microbiome. Pour résumer les abondances, on peut établir des graphes de co-occurrence via :

- Rapid inference of direct interactions in large-scale ecological networks from heterogeneous microbial sequencing data, Janko Tackmann, Joao Frederico Matias Rodrigues, Christian von Mering. bioRxiv 390195; doi: <https://doi.org/10.1101/390195>

Cependant, ces techniques reposent sur des corrélations et non sur des causalités. Elles ne permettent pas de comprendre les interactions entre organismes. Pour cela, il est important d'intégrer une autre dimension comme les informations contenues dans les réseaux métaboliques des bactéries. Pour cela, certaines bases de données existent :

- Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Dräger A, Mih N, et al. Recon3D enables a three-dimensional view of gene variation in human metabolism. Nat Biotechnol. Nature Publishing Group; 2018 Mar;36(3):272–81.

- <https://github.com/cdanielmachado/carveme/tree/master/carveme>

Ce projet vise à proposer une stratégie pour détecter les interactions via les échanges de métabolites via Flux Variability Analysis. La stratégie sera à tester sur des données de microbiotes intestinales fournies par le laboratoire TENS.

# Projet 11 : Modélisation de la réponse épithéliale

L'étude de la colonisation de l'intestin est de plus en plus importante:

1.Yassour M, Jason E, Hogstrom LJ, Arthur TD, Tripathi S, Siljander H, et al. Strain-Level Analysis of Mother-to-Child Bacterial Transmission during the First Few Months of Life. *Cell Host & Microbe*. 2018 Jul 11;24(1):146–154.e4.

2.Ferretti P, Pasolli E, Tett A, Asnicar F, Gorfer V, Fedi S, et al. Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. *Cell Host & Microbe*. 2018 Jul 11;24(1):133–5.

Pour aller au delà, il est possible de mettre en place des organoïdes pour étudier les interactions entre les bactéries et l'épithélium. Pour cela il faut proposer un protocole in silico qui sera ensuite testé in vivo

1. Sélectionner les souches pour proposer un écosystème minimal (3 bactéries)
2. Construire un modèle métabolique pour les 3 bactéries à partir de la base de données VMH
3. Immerger un métabolisme épithéliale dans l'écosystème

Le but est de caractériser les effets d'une petite communauté bactérienne sur le métabolisme épithéliale.

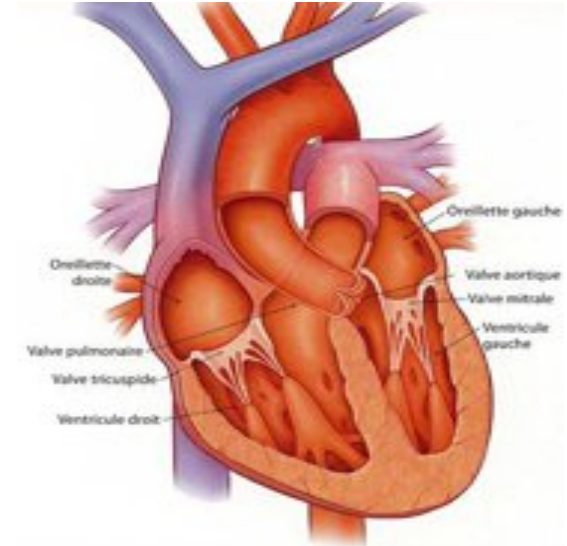
## Projet 12 : Gènes candidats pour le rétrécissement aortique calcifié

Le rétrécissement aortique calcifié est une pathologie fréquente pour laquelle le seul traitement est le remplacement chirurgical de la valve. Ces dernières années, plusieurs gènes de susceptibilité ont été décrits (*NOTCH1*, *PALMD*, *LPA*...). Plusieurs voies biologiques semblent impliquées : inflammation, fibrose, calcification.

L'étude d'une famille a identifié un intervalle de liaison de 28 Mb sur le chromosome 5.

**Parmi les gènes de cet intervalle, identifier celui ou ceux qui pourraient être impliqués dans le phénotype observé.**

Genomics: the next step to elucidate the etiology of calcific aortic valve stenosis. Bossé Y, Mathieu P, Pibarot P. J Am Coll Cardiol. 2008 Apr 8;51(14):1327-36. PMID: 18387432.

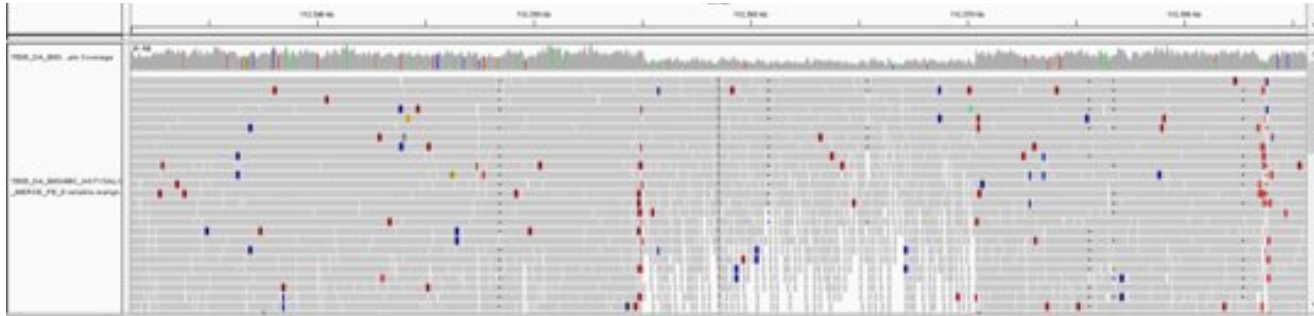




## Projet 13 : Effet d'une délétion intergénique ?

En 2008, un intervalle de liaison a été identifié sur le chromosome 4, dans une famille présentant un syndrome du QT long congénital de type 4, lié aux mutations du gène *ANK2*, codant pour l'Ankyrine-B ("family 2", [PMID 8832177](#)). Cependant aucune mutation n'a été détectée dans les exons du gène *ANK2*, situé dans l'intervalle de liaison, pour cette famille.

Tout récemment, par séquençage du génome entier de patients de cette même famille, une délétion de 15 kb a été détectée dans une région intergénique de l'intervalle de liaison. **Déterminez les effets possibles de cette délétion : Contient-elle des régions régulatrices ? Quels gènes pourraient être affectés ?**



# Projet 14 : Annotation de sites de fixation de FT

Non-coding regions have been associated with a cardiac disorder at risk of sudden cardiac death. An ATAC-seq experiment identified peaks of interest in these regions. Those peaks are open chromatin regions, potential regulatory regions and are susceptible to be bound by transcription factors.

**Given the peak regions, you will :**

- (1) find potential transcription factors binding sites**
- (2) compare the results between tools**
- (3) present advantages / disadvantages**
- (4) identify SNPs that could be disrupting those TFs binding sites**

- Peak file
- Human genome build : **GRCh37**
- Contact: [adrien.foucal@univ-nantes.fr](mailto:adrien.foucal@univ-nantes.fr)
- Non-exhaustive list of binding site research tools : MEME-suite, HOMER...

