



## Programme des activités pour étudiant.es

9h45-11h15	Public : Licences et masters	<p><b>Conférence pour étudiant.es*</b></p> <p>« Statistique appliquée : Proposer de nouveaux modèles statistiques et découvrir les mécanismes de régulation des gènes grâce à une recherche pluridisciplinaire. »</p> <p>Sophie Lèbre, Enseignante-chercheuse, Université Paul Valéry</p> <p><i>Amphithéâtre Dumontet</i></p>
11h30-12h30	Public : Master et participant.es du Forum	<p><b>Table ronde « Maths et entreprises »*</b></p> <p>Animée par Véronique Maume-Deschamps, Enseignante-chercheuse (Université Lyon1)</p> <p><i>Amphi 36.01</i></p>
12h30	Réservé aux Master et aux participant.es du Forum	<p>Buffet</p> <p><i>Hall bâtiment 36</i></p>

\* programme et résumés pages suivantes



## « Statistique appliquée : Proposer de nouveaux modèles statistiques et découvrir les mécanismes de régulation des gènes grâce à une recherche pluridisciplinaire. »

Sophie Lèbre, Enseignante-chercheuse, Université Paul Valéry

L'expression des gènes, c'est-à-dire la façon dont ils fonctionnent, est orchestrée par des régions spécifiques de la séquence ADN. Cette séquence est constituée à partir de quatre types de "bases nucléotidiques" notées par les lettres A, C, G et T. L'ensemble des combinaisons possibles garantit une grande variété de fonctions et de types cellulaires. Les mécanismes qui contrôlent l'expression des gènes sont donc complexes et sont encore largement méconnus aujourd'hui. Les identifier permettrait de faire avancer la recherche en génétique et d'améliorer des traitements de médecine personnalisée.

Grâce aux avancées technologiques, on peut maintenant observer les cellules en action et mesurer précisément la quantité de certaines molécules présentes dans la cellule à un instant donné. En répétant ces mesures plusieurs fois, on obtient de gros volumes de données qui représentent une information précieuse. Cependant, ces données sont complexes et leur analyse requiert le développement de nouveaux outils mathématiques et informatiques.

Dans le cadre d'une collaboration entre des chercheurs en mathématique, informatique et biologie, nous développons des modèles statistiques permettant d'expliquer les mécanismes de régulation des gènes à partir de la composition de la séquence ADN. Notre objectif est d'identifier des caractéristiques de la séquence ADN qui sont associées à un processus biologique donné, comme la réponse à un traitement médical chez l'homme, ou l'adaptation au changement climatique chez la plante.

Cela requiert notamment :

- l'identification et la construction de variables explicatives (feature engineering) à partir de données publiquement disponibles (base de données telles que ENCODE, TCGA, FANTOM, GTEx) ou développées par des collaborateurs biologistes (Menichelli et al. 2021, Bejjani et al., NAR2021, Cassan et al. 2023).
- la modélisation statistique d'un processus biologique
- l'inférence statistique des modèles proposés et l'identification des variables clés (méthodes de régularisation telles que le lasso, group-lasso, elasticnet, ... et calcul d'importance des variables).

Cassan, O., Lecellier, C.-H. , Bréhélin. , L., Martin, A. , Lèbre, S. (2024) Optimizing data integration improves Gene Regulatory Network inference in Arabidopsis thaliana. *Bioinformatics*, 40 (7).  
<https://academic.oup.com/bioinformatics/article/40/7/btae415/7698029>

Roméro, R. , Menichelli, C., Marin, J.-M. , Lèbre, S., Lecellier, C.-H. , Bréhélin., L. (2024) TFscope: systematic analysis of the sequence features involved in the binding preferences of transcription factors. *Genome Biology*, 25 (187).  
<https://genomebiology.biomedcentral.com/articles/10.1186/s13059-024-03321-8>

Cassan, O., Lèbre, S., Martin., A (2021) Inferring and analyzing gene regulatory networks from multi-factorial expression data : a complete and interactive suite. *BMC genomics*, 22:387.  
<https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-021-07659-2>

Bessière, C., Taha, M., Petitprez, F., Vandel, J., Marin, J.-M., Bréhélin, L., Lèbre, S., Lecellier, C.-H. (2018) Probing instructions for expression regulation in gene nucleotide compositions. *PLoS Computational Biology*, 14 (1).  
<https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005921>

**Table ronde « Maths et entreprises », avec les participantes suivantes :**

- **Véronique Maume-Deschamps** en animatrice (ancienne directrice d'[AMIES](#) - Agence pour les Mathématiques en Interaction avec l'Entreprise et la Société).
- **Lucie Quibel**, ingénieure modélisatrice au sein de ITK. ITK provides farmers and their advisors with analyses to measure and improve profitability, production and supply chain performance, to support sustainable agriculture. They use the results of their agricultural research activity to create powerful algorithmic models, which facilitate strategic and operational decision-making for stakeholders in the world of agriculture and food.
- **Annaëlle Dambreville**, Full-stack Developer | Data Scientist | PhD in applied statistical modeling. Elle travaille chez PUR: PUR is a certified B Corp and global leader in nature-based solutions. We are experts at empowering local communities to develop long-term socio-environmental projects — including agroforestry, land preservation, and sustainable agricultural practices.
- **Margaux Brégère**, Ingénieure chercheure en apprentissage statistique, et qui a eu le prix de thèse maths entreprise société en 2021. Elle travaille chez EDF R&D.
- **Adeline Fermanian**, head of research at Califrais. Califrais est l'opérateur digital et logistique officiel du Marché de Rungis. Califrais décarbone la supply chain alimentaire. Leur mission est d'inventer les solutions technologiques soutenues par l'IA pour optimiser les flux alimentaires à grande échelle. Leurs recherches se situent à l'intersection de l'apprentissage automatique, de l'optimisation logistique et de l'écologie, avec des sujets tels que la prévision de la demande, l'optimisation des stocks, les problèmes de routage.