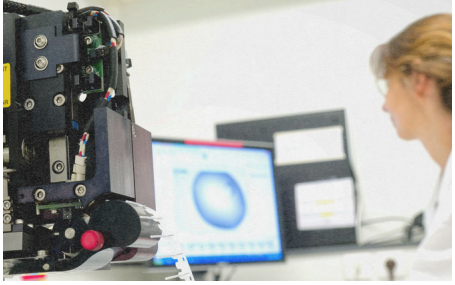
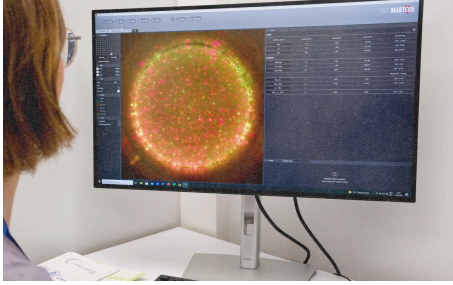
SFRBIOSCIENCES Lyon





SFR Scientific Day 2024

Advanced data analysis for modern biology

7 June 2024

Amphitéatre Mérieux, ENS site Monod, Lyon

8:30-9:00: Welcome

9:00: Introduction to the SFR - Yann Leverrier, SFR Director

9:15: Keynote lecture: The use (and abuse) of artificial intelligence in biomedical image analysis- Damian Dalle Nogare (Human Technopole, National Facility for Data Handling and Analysis)

10:00: How to train interdisciplinary students - Anna-Sophie Fiston-Lavier (Université de Montpellier)

10:30: Coffe break

11:00: CAN (Conseil Analayses Numériques) presentation - L.Modolo (ENS, LBMC)

11:05: Parallel and AI-driven Image Analysis - David Cluet (LBMC)

11:25: Using Bayesian modeling approaches to disentangle transcriptional signals in RNAseq data - Mathilde Paris (IGFL)

11:45: Analysis of metagenomic background in public RNASeq data in sarcoidosis patients - Thomas El Jammal (LBTI)

12:05: Company Flash Talks

12:15: Lunch Break, coffee, poster and exhibition tables

14:00: 4 Flash Talks Technology Development SFR-funded projects

14:30: Keynote lecture: AI as a diagnostic tool in hematology - Pierre Sujobert (CIRI, HCL)

15:15: Coffee break

15:45: How to analyse polymorphic transposable elements in 1000 genomes?- Marie Verneret (IVPC)

16:05: Molecular Dynamics Simulations as Computational Microscopy -Jackson Crowley (MMSB)

16:25: Structure of the replication complex of vesicular stomatitis virus by cryogenic electron microscopy- Louis-Marie Bloyet (CIRI)

16:45: Numerical reconstruction of plant seeds, from multiangle acquisition to FEM-ready meshes. - Elsa Gascon (RDP)

17:05: Concluding remarks

TECNIPLAST Clini Sciences Miltenyi Biotec BMG LABTECH The Microplate Reader Company

cnrs Université de Lyon Université Claude Bernard Lyon 1 Inserm La science pour la santé From science to health Université de Lyon ENS DE LYON

https://www.sfr-biosciences.fr/

Flash talks Technology Development SFR funded project 2023

- 3D-Imaging of embryonic tooth germs with light-sheet technology at Platim - Sophie Pantalacci
- Virometry Prospects. Counting VLPs by Cytometry - Philippe Mangeot (CIRI)
- Analyzing pro-inflammatory cytokines secreted from human primary keratinocytes using the JESS Simple Western - Emma Fraillon (LBTI)
- Development of reliable guidelines for preparation, imaging, and deep analysis of organoid structures - Morgane Fouché

Company Flash talk

- Miltenyi Biotech

TECNIPLAST CliniSciences Miltenyi Biotec BMG LABTECH The Microplate Reader Company

cnrs Université Claude Bernard Lyon 1 Inserm La science pour la santé From science to health ENS DE LYON

Abstracts

The use (and abuse) of artificial intelligence in biomedical image analysis-

Damian Dalle Nogare (Human Technopole, National Facility for Data Handling and Analysis)

The application of artificial intelligence (AI) approaches in bioimage analysis applications has exploded in recent years. From the early days of supervised image restoration and enhancement to today's plethora of new tools leveraging new deep learning architectures and models, AI is becoming a ubiquitous part of the analysis toolbox. These approaches have not only improved on the performance in common image processing tasks such as image denoising and segmentation, but also enabled tasks such as automated classification, image decomposition, and image generation, with high fidelity. Due in large part to these new approaches, the way in which bioimage analysis projects are undertaken has undergone a seismic shift, creating enormous opportunity for scientific advancement, but also enormous risk, as the application of deep learning models on out-of-distribution datasets can (but do not always) produce misleading or incorrect results. This presents a challenge to the widespread adoption of these important technologies, both for developers, who must communicate these risks and build robust, transparent and FAIR systems, as well as for users, who must ensure that they choose the right models and approaches for their particular application.

Bioinformatics learning laboratory, an innovative teaching project for transdisciplinary student training

Fiston-Lavier Anna-Sophie, University of Montpellier – ISEM (institut des Sciences de l'Evolution) - IUF

In education, several innovative teaching methods have gained prominence in recent years.
Both aim to improve student engagement and understanding, but they differ considerably in their structure and implementation. Active learning is a concept that encompasses various techniques designed to actively involve students in the learning process. Rather than passively receiving information, students participate in discussions, group activities and practical exercises during lessons. Active learning fosters critical thinking, problem solving and collaboration, with an emphasis on applying knowledge in real-life scenarios such as a research laboratory. In the ever-expanding field of bioinformatics, this active learning is relevant in order to prepare students for transdisciplinarity exchanges that is now essential in bioinformatics research laboratories and companies. To this end, since 2018, together with Jean-Christophe Avarre, an IRD researcher, Anne-Sophie Gosselin-Grenet, head of the microbiology master's programme, and myself, a member of the teaching staff of the Bioinformatics master's programme of the University of Montpellier, we have been working to set up and run a teaching project called BILL for Bioinformatics Learning Lab. Every year, we invite Master students from both master program to work in mixed groups on a research project: Evolution of the Cyprinid herpesvirus 3 (CyHV-3) genome that is responsible for the KHV disease that has spread to many countries worldwide and made the Common carp (Cyprinus carpio L.) the most produced fish in the world, highly susceptible. We ask them to combine their bioinformatic and molecular biology skills to sequence different cultures of viruses that have evolved under different conditions, in order to later on search for variants that may be associated with the different conditions. With the support of the University of Montpellier's innovation programme and the Faculty of Science, in 2019, we were able to set up a dedicated space consisting of a practical work room next to a modular project room with computers. Thanks to its success, this project has been permanently integrated into the curriculum of several Masters courses and this year we were invited to join the Oxford Nanopore (ONT) training program to share our experience with other teachers around the world who have also incorporated ONT sequencing into their courses.

Abstracts

Parallel and AI-driven Image Analysis
David Cluet (LBMC)

The increase of microscopes resolution, combined with the usage of Ais, generates new image analysis challenges. The datasets are more and more complex and permit to decipher more precisely spatially and/or temporally biological processes. These technical breakthroughs allow the experimental scientists to request advanced analytic solution with finely tuned analytical tools. The wide spread of AI usage (Cellpose, Stardist, ...) permits now to optimize the efficiency of specific analytical steps, but require GPU based computer architectures. Thus, an image analysis pipeline should now integrate some tasks performed by AIs on GPU within finely tuned CPU based code, and ensure the reproducibility of the results. A project from LBMC permits the development of such a pipeline. Its architecture, benchmark, and user's experience will be presented.

Using Bayesian modeling approaches to disentangle transcriptional signals in RNAseq data
Mathilde Paris (IGFL)

RNA-seq is widely used to study gene expression because it gives access to measures of mRNA levels for the whole transcriptome. However the transcriptome is affected by many factors, some of which are known and of interest (for instance, the effect of a treatment), or are less interesting but well-defined and easily controllable (for instance the sampling hour). Some other factors are well-defined but poorly controlled at the experimental level (for instance the age of the cohort members), and finally, others are not even known. The transcriptome is the resultant of this complex network of influences so computational methods are required to disentangle signals a posteriori. We have faced this problem in our study of leg regeneration in the Crustacean Parhyale hawaiensis. We have used Bayesian modeling to dissect the contributions of regeneration and the animal's individual physiological state. During this presentation, I will present our strategy that combines a specific experimental design and the downstream corresponding Bayesian model. I will then quantify the effect of this modeling on disentangling the regeneration signal from the physiological state of the animal.

Analysis of metagenomic background in public RNASeq data in sarcoidosis patients
Thomas El Jammal (LBTI)

Abstracts

AI as a diagnostic tool in hematology

Pierre Sujobert (CIRI, HCL)

Medicine is widely regarded as one of the most promising fields of application for artificial intelligence technologies. The diagnostic process, which involves identifying patterns of association among clinical signs, biological markers, and medical images, appears particularly suited to AI approaches. Within the hematology laboratory of the Hospices Civils de Lyon, we have developed several AI projects for the diagnosis of malignant hematologic disorders (such as leukemias and lymphomas), which I will briefly present to highlight the interest of these approaches as well as their limitations.

Evaluation of a machine-learning model based on laboratory parameters for the prediction of acute leukaemia subtypes: a multicentre model development and validation study in France.
Alcazer V, Le Meur G, Roccon M, Barriere S, Le Calvez B, Badaoui B, Spaeth A, Kosmider O, Freynet N, Eveillard M, Croizier C, Chevalier S, Sujobert P. Lancet Digit Health. 2024 May;6(5):e323-e333. doi: 10.1016/S2589-7500(24)00044-X. PMID: 38670741

Diagnosis of acute promyelocytic leukemia based on routine biological parameters using machine learning.
Cheli E, Chevalier S, Kosmider O, Eveillard M, Chapuis N, Plesa A, Heiblig M, Andre L, Pouget J, Mossuz P, Theisen O, Alcazer V, Gugenheim D, Autexier N, Sujobert P. Haematologica. 2022 Jun 1;107(6):1466-1469. doi: 10.3324/haematol.2022.280406. PMID: 35199507

How to analyse polymorphic transposable elements in 1000 genomes?

Marie Verneret (IVPC)

Transposable elements (TEs) are repetitive sequences constituting a significant fraction of eukaryotic genomes. Their ability to move and replicate makes them crucial drivers of genomic diversity and yet challenging for bioinformatic detection. TEs are playing important roles in genomic evolution as they may, for example, regulate gene expression, thereby generating transcriptomic variability. To mitigate deleterious effects, TEs are strongly regulated. However, some of them remain transpositionally active, leading to variations in insertion patterns between individuals named transposon insertion polymorphism (TIPs). In this study, our goal is to detect polymorphic insertions of class I TEs, more precisely endogenous retroviruses (ERV), in ruminant species using whole genome sequencing data from 1000 individuals. Although numerous bioinformatic tools have been developed to detect TIPs, their efficiency relies on TE and genomic sequence characteristics, as well as the target species underscoring the importance of benchmarking before undertaking large-scale genome analyses. In order to test the different programs, we first characterized consensus sequences of 23 ERV families in ruminants and annotated more than 20,000 ERVs in the reference genomes. We then performed a comparative analysis to evaluate the most effective tool and parameters to detect these TIPs on simulated data and real short-read data from 10 cattle. Analysis of the simulated ones revealed better results, highlighting a common tendency of simulated data-based benchmarks to overestimate the tool performance and overtake the true complexity and diversity of TEs. The overall sensitivity on real data reached 80% but it was associated to false positive rates ranging from 25% to 40%. Increasing sequencing coverage improved the sensitivity but at the expense of precision. Fixing a minimum copy frequency in populations enhanced precision but makes the detection of low-frequency insertions challenging. Our study emphasizes the importance of selecting appropriate tools and thresholds based on research questions as well as the necessity for robust TE annotations in the reference genomes.

Abstracts

Molecular Dynamics Simulations as Computational Microscopy

Jackson Crowley (MMSB)

Molecular dynamics (MD) simulations have emerged as a powerful tool for studying the structure and dynamics of biomolecular systems at sub-nanometer resolution, on time scales from picoseconds to milliseconds. Notably, MD facilitates the study of specific properties of membranes and proteins that are difficult to study otherwise with experimental methods. For this reason we can consider MD simulations as a "computational microscope", wherein we can study the structure, dynamics, and interaction of biomolecular system by predicting and viewing the movement of individual molecules. This allows for specific forms of analysis providing insight into complex properties such as membrane asymmetry2. The insight gained from MD simulations of biomolecular systems is becoming more critical as computational power is increasing and becoming more accessible, making MD a powerful tool for studies of biomolecular systems.
In my presentation I will illustrate the power of MD simulations as a computational microscope using three examples related to biological membranes. First, I will show how we used simulations to predict and understand the function of Ist2, a membrane protein involved in lipid transport. Second, I will show how we can use MD simulations to provide quantitative prediction on membrane physical properties, notably the elastic moduli. Third, I will show results (from my PhD thesis work) on MD simulations of lipid droplets (LDs), the organelles responsible for lipid metabolism1. Our MD simulations revealed how the budding of LDs requires the buildup of phospholipids on one side of the membrane. This buildup leads to membrane asymmetry, that is notoriously difficult to study experimentally. Detailed analysis of MD trajectories reveals important differences in phospholipid composition between the LD surface and the leaflets of the endoplasmic reticulum from which LDs bud.

1: Nieto, V., Crowley, J., Santos, D.E.S., Monticelli, L. (2023). Birth of an organelle: molecular mechanism of lipid droplet biogenesis. BioRxiv.
2: Crowley, J., Hilpert, C., & Monticelli, L. (2024). Predicting lipid sorting in curved membranes. Methods in Enzymology.

Structure of the replication complex of vesicular stomatitis virus by cryogenic electron microscopy

Louis-Marie Bloyet (CIRI)

Several human pathogens such as measles, Ebola, and mumps viruses have a non-segmented negative-strand RNA genome and share sophisticated molecular mechanisms to transcribe and replicate their genome. This RNA genome is tightly enwrapped by a homopolymer of viral nucleoproteins (N), thus forming a large ribonucleoprotein complex termed nucleocapsid. The polymerase complex comprises the large protein (L), which performs all the catalytic activities, and its cofactor the phosphoprotein (P). The polymerase complex transcribes the genome into capped and polyadenylated mRNA and replicates the genome into a full-length, non-modified, and encapsidated antigenome. How the virus ensures the specificity and efficiency of the encapsidation of the nascent RNA during replication is still unknown. Using vesicular stomatitis virus (VSV) as a prototype, we investigated the structure of the replicase complex made of L, N, and the N-terminal region of P by cryo-electron microscopy. The structure of the complex reassembled in vitro is very heterogeneous and N can be found in a continuum of positions next to the C-terminal domain of L. The structure of a preferred conformation was solved and shows that the interactions stabilizing the orientation of N mainly involve residues in two regions of P.

Abstracts

Numerical reconstruction of plants seeds for morphogenesis analysis

Elsa Gascon, Guillaume Cerutti, Benoit Landrein, Olivier Ali.
Laboratoire Reproduction et Développement des Plantes, Université de Lyon, ENS de Lyon, UCB Lyon, CNRS, INRAE, INRIA, F-69342, Lyon, 69364 Cedex 07, France.

Arabidopsis thaliana seeds are organs of choice in the study of morphogenesis and its regulation. During development, seed growth is promoted by the pressure generated in the endosperm under compression and restricted by the mechanosensitive stiffening of the walls in the surrounding seed coat (Creff, A. & Ali, O. et al. 2023). While this mechanosensitive regulation, formalized by an incoherent feedforward loop, accounts for final seed size and shape emergence; experimental data suggest that it could be controlled by the inner layers of the seed coat. To assess the role of inner tissue organization on seed development, we built relevant data structures that allow fine analysis at the cellular level of the entire seed coat. We developed a numerical study driven by an image analysis pipeline combining high-resolution microscopy and computational methods to extract meshes compatible with mechanical simulations. Combining multiangle image acquisition, neural network cell wall prediction, watershed segmentation, and tissue surface meshing, we reconstructed numerically the full 3D multilayered structure of Arabidopsis seeds encompassing tissue topology and unraveling geometrical properties at the cellular scale.

Creff A, Ali O, Bayle V, Ingram G, Landrein B. Evidence that endosperm turgor pressure both promotes and restricts seed growth and size. Nat Commun. 2023.