2020-11-30

EuReCa International PhD Program

# PhD thesis project

## 2022 Call for application

## Causal inference for live cell imaging and single-cell multi-omics data

### General information

| | |
|---|---|
| **Call** | 2022 |
| **Reference** | 2022-07-ISAMBERT_HERSEN |
| **Keyword(s)** | Machine and Deep learning; Causal inference; Information theory; Live cell imaging; Single cell multi-omics. |

### Director(s) and team

| | |
|---|---|
| **Thesis director(s)** | Hervé Isambert & Pascal Hersen |
| **Research team** | Reconstruction, Analysis and Evolution of Biological Networks |
| **Research department** | UMR168 - Physico-Chimie Curie Lab |

### Description of the PhD thesis project

Live cell imaging microscopy and next generation sequencing technologies, now routinely used in cell biology labs, produce massive amounts of time-lapse images and gene expression data at single cell resolution. However, this wealth of state-of-the-art biological data remain largely under-explored due to the lack of unsupervised methods and tools to analyze them without preconceived hypothesis. This highlights the need to develop new Machine Learning and Artificial Intelligence strategies to better exploit the richness and complexity of the information contained in time-resolved cell biology data.

The Isambert lab recently developed novel causal inference methods and tools (https://miic.curie.fr) to learn cause-effect relationships in a variety of biological or clinical datasets, from single-cell transcriptomic and genomic alteration data (Verny et al 2017, Sella et al 2018) to medical records of patients (Cabeli et al 2020). These machine learning methods combine multivariate information analysis with interpretable graphical models (Li et al 2019) and outperform other methods on a broad range of benchmarks, achieving better results with only ten to hundred times fewer samples.

The objective of the present PhD project is to extend these causal inference methods to analyze time-resolved cell biology data, for which the information about cellular dynamics can facilitate the discovery of novel cause-effect functional processes. These novel causal inference methods for time series data will then be applied to analyze two types of high-through put time-resolved cell biology data: 1- time-lapse images of cellular systems (Marinkovic et al 2019) from the Hersen lab (Institut Curie) to analyze cell cycle progression and apoptosis in tumor-on-chip devices, in collaboration with the Martinelli lab (University of Rome Tor Vergata, Italy), and 2- single-cell multi-omics data on a cellular therapy against multiple sclerosis in collaboration with the Fillatreau lab (Institut Necker).

## International, interdisciplinary & intersectoral aspects of the project

This interdisciplinary PhD project will consist in:

i) theoretical developments of causal inference analyses for time-resolved cell biology data (based on machine learning, AI approaches and Information theory) applied to

ii) the analysis of time-lapse images of cellular systems (Hersen lab) and time-resolved single-cell multi-omics data (Fillatreau lab, Institut Necker).

The extraction of cellular morpho dynamic properties from raw images will be done through advanced segmentation and vision-based AI techniques in collaboration with the Martinelli lab, University of Rome Tor Vergata, Italy. The project will also benefit from the expertise of Coservit, an IT software company based in Grenoble, developing a new generation of tools based on inferring causal relations between connected objects.

## Recent publications

1. Cabeli V, Verny L, Sella N, Uguzzoni U, Verny M, **Isambert H**: Learning clinical networks from medical records based on information estimates in mixed-type data. PLoS Comput Biol 16(5):e1007866 (2020).

2. Marinkovic ZS, Vulin C, Acman M, Song X, Di Meglio JM, Lindner AB, **Hersen P**: A microfluidic device for inferring metabolic landscapes in yeast monolayer colonies. Elife 8:e47951 (2019).

3. Li H, Cabeli V, Sella N, **Isambert H**: Constraint-based causal structure learning with consistent separating sets. Advances in Neural Information ProcessingSystems (NeurIPS) 32, 14257-14266 (2019).

4. Sella N, Verny L, Uguzzoni G, Affeldt S, **Isambert H**: MIIC online: a web server to reconstruct causal or non-causal networks from non-perturbative data. Bioinformatics 34(13):2311-2313 (2018).

5. Verny L, Sella N, Affeldt S, Singh PP, **Isambert H**: Learning causal networks with latent variables from multivariate information in genomic data. PLoS Comput Biol 13(10):e1005662 (2017).

## Expected profile of the candidate

Applicants should have a strong background in machine learning or computer science and a keen interest to analyze complex heterogeneous data of biological and medical interests. Applicants should be proficient in programming and willing to interact with scientists from different disciplines, from data scientists to medical doctors. Applicants are expected to show a clear capacity for independent and creative thinking. Experience on causal inference analysis is a plus but not required as long as the applicant has a strong motivation to learn.