

IC-3i International PhD Program
PhD thesis project
2018 Call for application



**Cause-effect analyses of breast cancer clinical records from
Institut Curie**

General information

Call	2018
Reference	2017-05-ISAMBERT&REYAL
Keyword(s)	Data analysis, Causal inference analysis, Clinical records, Breast cancer, Heterogeneous data types

Director(s) and team

Thesis director(s)	Hervé Isambert & Fabien Reyal
Research team	Evolution of Biomolecular Networks, RNA Dynamics
Research department	UMR168 – Physico Chimie Curie Lab

Description of the PhD thesis project

Despite continuing medical advances, cancer has become the first cause of death in western Europe and affects many patients with comorbid medical conditions such as diabetes, hypertension and cardiovascular diseases, who are under-represented in clinical trials.

This highlights the need to directly analyze real life biological and medical data. While numerous methods have been developed to identify correlations in such heterogeneous datasets, a central challenge remains to uncover unsuspected cause-effect relationships from medical records. It is now considered a priority to guide clinical understanding and treatments by novel and innovative data analysis and computational methods. This project aims, more specifically, at identifying possible causal relationships between breast cancer treatments and other seemingly unrelated medications to circumvent detrimental interference or, on the contrary, leverage therapeutic synergy between drug treatments and ultimately improve the clinical outcomes of breast cancer patients.

The Isambert lab recently developed statistical and computational methods based on the analysis of multivariate information which unify causal and non-causal network learning frameworks while including the effects of unobserved latent variables. In this project, in collaboration with the Reyal lab, Institut Curie, we will extend our causal inference approaches to the framework of deep learning and generative adversarial network methods to obtain robust cause-effect inference.

To this end, generative adversarial network methods will be adapted to the context of heterogeneous clinical data. In particular, the integration of categorical, continuous and textual data requires novel methodological developments. As a proof-of-concept, we have performed a preliminary causal analysis on a hand-curated dataset of 1,200 breast cancer patients, which will be extended to about 2,700 patients and, eventually, to nearly 70,000 patients in the course of the project.

International, interdisciplinary & intersectoral aspects of the project

This interdisciplinary project involves i) theoretical developments and implementation of causal inference analyses (based on information theory, computer science / machine learning and text analysis approaches) applied to ii) the analysis of biological and clinical records from breast cancer patients from Institut Curie Hospital.

Its international dimension concerns data visualization which will be developed in collaboration with Federal Universities in Porto Alegre and Natal, Brazil, which have groups with strong expertise in innovative data visualization.

The intersectoral component of the project will involve the Paris-based Quattrocento company, which develops innovative text analysis approaches relevant for the extraction of meaningful information from medical records.

Recent publications

1. 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*, under revision (2017).
2. 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).
3. Hamy-Petit A-S, Belin L, Bonsang-Kitzis H, Paquet C, Pierga J-Y, Lerebours F, Cottu P, Rouzier R, Savignoni A, Lae M, **Reyal F**: Pathological complete response and prognosis after neoadjuvant chemotherapy for HER2-positive breast cancers before and after trastuzumab era: results from a real-life cohort *Br J Cancer*. 114(1): 44-52 (2016).
4. Affeldt S, Verny L, **Isambert H**: 3off2: A network reconstruction algorithm based on 2-point and 3-point information statistics. *BMC Bioinformatics*, 17 Suppl 2:12 (2016).
5. Affeldt S, **Isambert H**: Robust reconstruction of causal graphical models based on conditional 2-point and 3-point information. *Proceedings of the 31th conference on Uncertainty in Artificial Intelligence (UAI)*, Amsterdam. Morgan Kaufmann (2015).

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.