PhD position in Deep Learning methods for long non-coding RNA prediction in cancer

Keywords : Bioinformatics deep-learning data integration long non-coding
RNA genomics cancer precise medicine

Context

Recently, many long ncRNAs (lncRNAs), larger than 200 nucleotides, have been identified as potential regulators. But unlike small ncRNAs, their characterization by structure and function is far from established. The determination of the structure, 2D or 3D, of an lncRNA by experimental methods (crystallography, NMR) or bioinformatics methods is a major challenge, since it helps to elucidate its function. RNAs from the same family indeed share the same structure, giving them the same function, the structure guiding in particular the interactions of this RNA with proteins or other RNAs.

RNAs, and more precisely non-coding RNAs (ncRNAs, RNA untranslated into proteins), have aroused growing interest in the international scientific community in recent years, due to their proven involvement in many biological processes and the important role they can play in pathological processes such as cancer. They are thus increasingly considered as potential therapeutic targets or biomarkers (diagnostic and prognostic markers).

Objectives

The final objective of the project will be to implement generic methods and tools for the prediction of lncRNAs. The tools developed will be made available to the scientific community via our EvryRNA platform: http://EvryRNA.ibisc.univ-evry.

The methods developed will be applied to cancer and will provide a better understanding of the involvement of RNAs in this pathology. Cancer in a given tissue is a heterogeneous disease; several cancer subtypes can be identified. Treatments and diagnosis should be tailored to each subtype. In this project, we will be interested in lncRNAs in a frequent cancer, bladder cancer (4th cancer in terms of incidence in men) as well as in pediatric cancer, retinoblastoma. A small number of lncRNAs predicted to be potentially involved will be functionally validated by the team of biologists. We hope ultimately to be able to offer clinicians new diagnostic or prognostic markers and enable them to better understand the biological causes of the disease in order to optimize treatments.

In this project, we propose to develop computational methods based on Deep Learning to predict and characterize lncRNAs by integrating different data: sequence, 2D and 3D structure, interaction with coding or non-coding genes. and genetic and epigenetic alterations. The development of methods to predict the 3D structure of RNAs, such as those developed by DeepMind (the AI subsidiary of Google), could also be considered.

The project will be in collaboration with ADLIN entreprise and Institut Curie.

Candidate profile

Master 2 in Data Sciences, Bioinformatics/Computational Biology or Computer Sciences (or equivalent).

Background in computer science and data sciences, in particular in machine learning and deep-learning. Knowledge in bioinformatics and biology will be highly appreciated.

Application

Procedure : The application (CV+motivation letter) should be sent to fariza.tahi@univ-evry.fr

Deadline : December, 26th 2021

Contact

Fariza TAHI, Pr. AROB@S Team, IBISC Laboratory, fariza.tahi@univ-evry.fr