

Master project 2020-2021

Personal Information

Supervisor	Davide Piscia
Email	davide.piscia@cnag.crg.eu
Institution	CNAG-CRG
Website	www.cnag.cat
Group	Data Analysis Team, Bioinformatics Unit

Project

Computational genomics

Project Title:

Deep learning models applied to rare diseases: where do we stand?

Keywords:

deep learning, rare diseases, variant effect prediction, non-coding regions

Summary:

Deep learning models have been used extensively in image recognition and natural language processing, but in the last years they have been also applied to genomics. In the last years some interesting initiatives were started such as kipoi (<https://kipoi.org>) and selene (<https://selene.flatironinstitute.org>) whose aim is to facilitate the use of deep learning in biological contexts. One of the objectives of this project is to evaluate and summarize the state of the art of genomics deep learning, especially for variant effects prediction in non-coding regions . Once the most promising models have been selected, the candidate will have to apply it to some of the rare disease whole genome datasets hosted at CNAG-CRG. In this task she/he will have to be able to run the models in the CNAG-CRG HPC cluster (GPUs enabled) and do a first assessment of the model predictions. The long-term goal of this work is to integrate deep learning as a functionality in the RD-Connect GPAP platform (<https://platform.rd-connect.eu>).

Expected skills::

The candidate is expected to have good computational skills, especially in python.

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

