Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Carles Hernandez-Ferrer and Leslie Matalonga
Email	carles.hernandez@cnag.crg.eu, leslie.matalonga@cnag.crg.eu
Institution	CNAG-CRG
Website	www.cnag.cat
Group	Data Analysis Team - Bioinformatics Unit

Project

Web development & bioinformatic tools

Project Title:

Implementation of an automated genomic (re)analysis support system for rare disease diagnostic.

Keywords:

rare diseases, genome-phenome analysis

Summary:

It is estimated that 350 million people worldwide suffer from one of the approximately 7000 existing rare diseases (RDs). As 80% of RDs are thought to have a genetic origin, particular emphasis has been placed on the rapidly expanding development of genomic technologies. However, the interpretation of the genome is still a real challenge for molecular geneticists, and innovative bioinformatics solutions combining genomic and clinical data are crucial for reaching accurate patient diagnosis. At the CNAG-CRG and in the context of several EU projects (RD-Connect, Solve-RD, EJP-RD) we developed a data sharing and analysis tool, the RD-Connect Genome- Phenome Analysis platform (GPAP: https://platform.rd-connect.eu/), to provide methods and standardised analyses of phenotypic and (gen)omic data in order to facilitate mutation detection within the context of rare diseases. As part of the Solve-RD project: “solving the unsolved rare diseases” (www.solve-rd.eu), rare disease patient data from 19.000 genomic datasets will be reanalysed using the RD-Connect GPAP. To this purpose, high throughput SNV-indel data (re)analysis solutions are being implemented in the system to automatically allow real-time queries to a high number of samples. A first prototype has been built enabling the filtering of thousands of genomic datasets by specific filters including variant pathogenicity and population databases. The master student joining this project will be working on the development of this innovative approach by improving the tool that queries the RD-Connect API to enable more complex queries (involve family members’ genotypes, specific regions of the genome, other individuals with phenotypic similarity, etc.). This work will be done in close collaboration with clinicians and researchers of four ERNs (European Reference Networks) involved in the Solve-RD project enabling continuous analysis feedback and molecular diagnosis confirmation.

Expected skills::

Python programming Data base architecture Genetics background Team working skills

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed