Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Patrick Aloy
Email	patrick.aloy@irbbarcelona.org
Institution	Institute for Research in Biomedicine (IRB Barcelona)
Website	https://sbnb.irbbarcelona.org
Group	Structural Bioinformatics & Network Biology

Project

Pharmacoinformatics & systems pharmacology

Project Title:

Formatting Biological Big Data to Enable Systems Pharmacology

Keywords:

Systems pharmacology, complex diseases, biological and disease signatures

Summary:

The amount and complexity of the biological data generated in the last years, due to the popularization of high-throughput pipelines, is virtually flooding biomedical research. Indeed, the growth of biological databases is steeper than ever before, and the repertoire of possible read-outs spans all levels of biology. However, the nature of biological data is remarkably complex, and dealing with diversity, inconsistency and incompleteness, among other issues, demands heavy specialist processing, and prevents a widespread predictive approach to disease biology. Indeed, this deluge of data has not spurred the development of truly precision therapies, and the inherent limitations of the prevailing reductionist approaches have highlighted the need of moving away from the 'one disease, one target, one drug' paradigm and consider the complexity of human pathologies and physiological responses. The current project builds on the hypothesis that the disease-causing perturbations leave detectable traces at different - and variable - levels of biological complexity (i.e. activation/inhibition of signaling pathways, transcriptional changes, etc) that capture both the direct effect of the perturbation and a global reaction of the system. Accordingly, the main aim of the project is to collect genuinely heterogeneous datasets, and offer a generic and intuitive means to bridge the gap between biological big data repositories and state-of-the-art machine-learning tools. Besides, we shall develop a generalized connectivity mapping, as a form of virtual phenotypic screening, to discover novel chemical or genetic modulators able to revert the specific signatures of disease and 'cancel out' the phenotypic traits of the disorder. The successful candidate shall be responsible for the implementation of a pipeline to collect and process biological big data, and to encapsulate it in the form of heterogeneous biological embeddings. Overall, we shall develop a novel strategy to integrate the deluge of biological data in a format that is readily suitable for modern machine learning. Additionally, he/she will develop a General Connectivity Mapping (GCMap) strategy to link biological and chemical signatures from the Chemical Checker (htts://chemicalchecker.org), so that the biological context of each small molecule can be incorporated as a descriptor. We shall then explore the added value of these biological descriptors to identify therapeutic opportunities to treat complex diseases.

References:

- Duran-Frigola M, et al. Formatting biological big data for modern machine learning in drug discovery. WIREs Comp Mol Sci

(2018), e1408. - Duran-Frigola M, et al. Extending the small molecule similarity principle to all levels of biology. Nat Biotechnol (2020) In press. Available at bioRxiv.

Expected skills::

Highly motivated. Fluency in English. Good programming and scripting skills, with knowledge of Python and databases management (e.g. postgresSQL).

Possibility of funding::

Yes

Possible continuity with PhD: :

Yes