Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2021-2022

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, Biological Big Data, Machine Learning, Generative Models, Personalized Medicine

Summary:

Biological data is accumulating at an unprecedented rate, escalating the role of data-driven methods in computational drug discovery. The urge to couple biological data to cutting-edge machine learning has spurred developments in data integration and knowledge representation, especially in the form of heterogeneous, multiplex and semantically-rich biological networks. Today, thanks to the propitious rise in knowledge embedding techniques, these large and complex biological networks can be converted to a vector format that suits the majority of machine learning implementations. Indeed, we have generated biological embeddings (i.e. bioactivity signatures) that capture complex relationships between small molecules and other biological entities such as targets or diseases (Duran-Frigola et al. 2020 Nat Biotechnol; Bertoni et al. bioRxiv 2020). However, only a tiny fraction of the possible chemical space has been so far explored, meaning that most compounds able to modulate biological activities (i.e. drugs) are yet to be discovered. Accordingly, the main objective of this project is to couple our bioactivity signatures to inverse design algorithms to generate new chemical entities with a desired functionality. In particular, we aim at generating new chemical entities (NCEs) to modulate the activity of a specific set of targets, selected from a combination of perturbagen profiles, to revert the pathological state induced by Alzheimer´s disease (AD) and other complex disorders. All in all, the incorporation of machine learning methods to the drug discovery process will trigger the development of thousands of novel compounds, finally enabling precision medicine. The MSc student shall be responsible, with the help and guidance of senior members of the lab, for the implementation of ML-based Generative Models (i.e. cVAEs or GANs) to create new small molecules that fulfill the required polypharmacological properties to revert AD pathological signatures.

References:

Duran-Frigola M et al. Extending the small-molecule similarity principle to all levels of biology with the Chemical Checker, Nature Biotechnology, 2020, 38(9):1087-1096, DOI: https://doi.org/10.1038/s41587-020-0502-7 Bertoni M et al. Bioactivity descriptors for uncharacterized compounds, BioRxiv, 2020, DOI: https://doi.org/10.1101/2020.07.21.214197 Duran−Frigola M et al. Formatting biological big data for modern machine learning in drug discovery, WIREs Computational Molecular Science, 2018, DOI: 10.1002/wcms.1408

Expected skills::

Strong programming and scripting skills, with deep knowledge of Python. Knowledge of machine learning techniques and libraries will be an asset.

Possibility of funding::

Yes

Possible continuity with PhD: :

Yes

Comments:

In the frame of this project, the MSc student will gain deep technical knowledge of Python programming, the use of HPC queue systems, virtual machines (Open Nebula), Grid Containers (Dockers, Singularity). Additionally, he/she will acquire strong on bio- and chemo-informatics skills (small molecules bioactivity descriptors) and knowledge on Data Sciences technologies (machine (deep) learning, ML-based generative models).