

Master project 2020-2021

Personal Information

Supervisor	José Manuel Mas
Email	raquel.valls@anaxomics.com
Institution	Anaxomics Biotech
Website	www.anaxomics.com
Group	Data Science Department

Project

Computational systems biology

Project Title:

Generation of a Real World Data repository from RNASeq analysis

Keywords:

artificial intelligence, high-throughput, real world data, GEO database

Summary:

GEO database contains the description of millions of experiments including high-throughput data. Some of these experiments are based on primary cells and represent a source of Real World Human Data (RWD), being this type of data of special interest for FDA and EMA during drug development process. GEO database includes around 150.000 patients containing RNASeq data. The protein expression pattern from these data is an interesting information. The student enrolled in this project will be responsible to extract the protein expression pattern from the RNASeq datafiles grouping them by their labelled phenotypes. The student will determine the existing relationship in the protein expression pattern between these labelled patients and other patients with the same labels but with different high-throughput data.

Expected skills::

programming python, c++ or Matlab

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

