Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Jana Selent
Email	jana.selent@upf.edu
Institution	IMIM-UPF
Website	www.jana-selent.org
Group	GPCR Drug Discovery Group

Project

Computational systems biology

Project Title:

Creating the framework for a multidimensional understanding of signal transduction

Keywords:

signal transduction, multidimensional, molecular, spatial, and temporal level

Summary:

Remarkably, signal transduction systems use a relatively limited repertoire of intracellular signaling components. This small number of effector proteins nevertheless enables a cellular signaling apparatus that is flexible and versatile. Versatility is achieved by modulation at the molecular, spatial, and temporal levels of the macro-molecular interactions at each node in the pathway. In effect, a limited number of nodes, each with several alternative downstream pathways, can give rise to a vast number of distinct signaling pathways. Although versatile and complex, the biological role of signal transduction demands specificity and precision in signaling. In this respect, many questions remain open about the interplay of the molecular, spatial, and temporal levels of the macro-molecular interactions. This knowledge gap is tackled by the European Research Network on Signal Transduction (ERNEST) which counts currently more than 400 researchers with different expertise in the field. We are looking for a motivated student who is interested in supporting the endeavor of ERNEST. The Master student will have the unique opportunity to interact with known researchers across Europe. The student will be in charge of developing a framework for collecting, organizing and visualizing diverse signaling data. He/she should have knowledge in HTLM/CSS, web page design, MySQL 5.5, database design, Python. C/C++ and JavaScript is a plus. An important benefit of this projects is that the master student will be introduced to a wide European network in signal transduction which provides with valuable contacts and diverse opportunities for future job openings.

References:

Sommer et al. The European Research Network on Signal Transduction (ERNEST): Toward a Multidimensional Holistic Understanding of G Protein-Coupled Receptor Signaling (2020) (https://pubs.acs.org/doi/pdf/10.1021/acsptsci.0c00024)

Expected skills::

He/she should have knowledge in HTLM/CSS, web page design, MySQL 5.5, database design, Python. C/C++ and JavaScript is a plus.

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Mireya Plass
Email	mplass@idibell.cat
Institution	Idibell
Website	
Group	Gene regulation of Cell Identity

Project

Computational systems biology

Project Title:

Functional characterization of RNA binding proteins in neural differentiation

Keywords:

single-cell trabscriptomics, gene regulation, RNA binding proteins, alternative polyadenylation

Summary:

Background The Plass lab, located at Bellvitge Biomedical Research Institute (IDIBELL), investigates the role of RNA binding

proteins (RBPs) in the regulation of cell identity using a combination of computational and high-throughput experimental approaches. In particular, the lab is interested in understanding how RNA processing mechanisms, including splicing and alternative polyadenylation (APA), impact neuronal cell differentiation and how this process is altered in the development of neurodegenerative diseases. We known that around 70% of human genes are regulated by APA. Despite the growing amount of evidence showing that APA changes according to the proliferation and differentiation status of a cell, it is still not clear if APA contributes to this process and which are RBPs involved. Goals The project proposed will focus on developing new computational approaches to identify RBPs involved in the differentiation of neural cell types in single-cell transcriptomics data (scRNA-seq) and characterize their target genes across different cell types. During this project, the student will learn how to analyze scRNA-seq datasets using existing tools and will develop new computational methods to identify regulatory interactions in single-cell datasets. Approach 1.- Characterize cell and tissue specific RBPs Using a collection of published scRNA-seq datasets, the student will characterize the expression of RBPs across human and mouse cell types from different tissues and differentiation states using existing packages for single-cell transcriptomic analyses. She/he will use these data to make an expression atlas of RBPs and identify cell type and tissue specific RBPs, by comparing the expression of the RBPs across datasets. 2.- Identify RBPs target genes across cells Using a combination of published and newly-developed methods, the student will identify RBP-RBP and RBP-target interactions across cell types that could suggest a functional relation. Assuming that we find interactions in which the expression of an RBP affects the expression of a gene, the student will develop a computational method to recover them from single-cell datasets. We will benchmark the method developed by assessing the ability to identify known interactions between RBPs and genes. Next, we will use a computational pipeline developed in the lab (Plass et al. unpublished), that allows quantifying the expression of individual APA isoforms. Using a similar approach as described before, we will now identify robust associations between specific RBP and individual APA isoforms across cells. 3. – Identify RBP – gene/isoform interactions relevant for neural differentiation The student will use a computational lineage-reconstruction to understand the relationships between different neural cell types in a time-dependent manner, i.e. understand which cell types give rise to other cell types. These methods can also be used to order cells according to their differentiation status. Once she/he has obtained a ordering of cells, she/he will develop a new method to identify interactions between an RBP and a target gene or isoform in time, i.e., identify cases in which the expression of an RBP in time n affects the expression of a target gene/isoform in time n + t. In this way, we will be able to identify new correlations that may be related to the cellular differentiation process.

References:

Derti, A. et al. A quantitative atlas of polyadenylation in five mammals. Genome Res 22, 1173–1183 (2012). Ji, Z., Lee, J. Y., Pan, Z., Jiang, B. & Tian, B. Progressive lengthening of 3’ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. Proc Natl Acad Sci U S A 106, 7028–7033 (2009). Miura, P., Shenker, S., Andreu-Agullo, C., Westholm, J. O. & Lai, E. C. Widespread and extensive lengthening of 3’ UTRs in the mammalian brain. Genome Res 23, 812–825 (2013). Plass, M. et al. Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. Science (80-) 360, eaaq1723 (2018). Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol 20, 59 (2019).

Expected skills::

We are looking for a master student with experience in high-throughput sequencing data analyses. Candidates are expected to have experience in R and a scripting language such as Python or Perl. Prior knowledge on post-transcriptional regulation or method development will be a plus. Interest in gene regulation and working in a multidisciplinary team will be valued, as interaction with experimental researchers will be required for the success of the project.

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Master project 2020-2021

Personal Information

Supervisor	Arnau Montagud
Email	arnau.montagud@bsc.es
Institution	BSC - Barcelona Supercomputing Center
Website	https://www.bsc.es/
Group	Computational Biology

Project

Computational systems biology

Project Title:

Simulation of drug interactions in multiscale model tailored to prostate cell-lines

Keywords:

multiscale modelling, drug simulations, gastric cancer, parameter databases

Summary:

General context The candidate will join the area of Precision Medicine in Alfonso Valencia's Computational Biology group within the Life Sciences Department at the Barcelona Supercomputing Center. This research line encompasses the development of different strategies and approaches to improved personalized diagnosis of disease, as well as treatment selection for particular patients, based on their individual characteristics. Computational systems biomedicine relies on the development of in silico models to integrate different sources of experimental information and produce patient-specific mechanistic explanations of cellular behaviour used to design new targeted therapies. In the context of cancer, cell signalling as well as metabolic models have been reconstructed for different cancer types and healthy tissues. Simulation of these models using different computational approaches (e.g. Boolean formalism, Constraint-Based Modelling) have supported the development of targeted therapies that attack specific biological pathways in the cell. The candidate will focus on using and further developing a set of tools aimed for the simulation drug inhibitions of different cell lines. These simulations will explore varying concentrations of single drug inhibition and combinations of them. Scientific context Discovery of efficient anti-cancer drug combinations is a major challenge, since experimental testing of all possible combinations is clearly impossible. Recent efforts to computationally predict drug combination responses retain this experimental search space, as model definitions typically rely on extensive drug perturbation data1,2. Relying on background knowledge extracted from literature and databases, patient-specific dynamical models were developed3 previously tailoring a general cancer model4 to breast-cancer patients. In this work, the study of solutions of the Boolean model led to identifications of particularities among patients and their clinical stratifications3. Currently, we have used this same framework to obtain prostate-cell-line-specific dynamical models and are starting to perform drug perturbation studies. Nevertheless, due to the limitations of the simulation tools used5,6, this study neither identifies sets of concentrations where this synergy is maximal nor it considers population-level constraints and behaviours. In present project, the candidate will simulate varying concentration of inhibitors in the different cell lines model using a multiscale modelling framework, PhysiBoSS7, that that combines agent-based8, Boolean5,6 and environmental dynamics9 modelling. The candidate will first gather from databases and literature biophysical information on parameters that allows for the tailoring of the multiscale simulation to each cell-line such as uptake rates, growth rates, etc. Then, the use of scripts already in place (in python, bash, perl, R) and new ones developed by the student will allow exploring different concentrations of drugs to find maximal synergies specific for each cell-line that would help identifying drug responses potentially relevant in the clinic.

References:

1. Flobak, Å. et al. Discovery of Drug Synergies in Gastric Cancer Cells Predicted by Logical Modeling. PLOS Comput. Biol. 11, e1004426 (2015). 2. Flobak, Å., Vazquez, M., Lægreid, A. & Valencia, A. CImbinator: a web-based tool for drug synergy analysis in small- and large-scale datasets. Bioinformatics 33, 2410–2412 (2017). 3. Béal, J., Montagud, A., Traynard, P., Barillot, E. & Calzone, L. Personalization of logical models with multi-omics data allows clinical stratification of patients. Front. Physiol. 9, 1965 (2019). 4. Fumia, H. F. & Martins, M. L. Boolean Network Model for Cancer Pathways: Predicting Carcinogenesis and Targeted Therapy

Outcomes. PLoS ONE 8, e69008 (2013). 5. Stoll, G., Viara, E., Barillot, E. & Calzone, L. Continuous time Boolean modeling for biological signaling: application of Gillespie algorithm. BMC Syst. Biol. 6, 116 (2012). 6. Stoll, G. et al. MaBoSS 2.0: an environment for stochastic Boolean modeling. Bioinformatics 33, 2226–2228 (2017). 7. Letort, G. et al. PhysiBoSS: a multi-scale agent-based modelling framework integrating physical dimension and cell signalling. Bioinformatics bty766 (2018) doi:10.1093/bioinformatics/bty766. 8. Ghaffarizadeh, A., Heiland, R., Friedman, S. H., Mumenthaler, S. M. & Macklin, P. PhysiCell: An open source physics-based cell simulator for 3-D multicellular systems. PLOS Comput. Biol. 14, e1005991 (2018). 9. Ghaffarizadeh, A., Friedman, S. H. & Macklin, P. BioFVM: an efficient, parallelized diffusive transport solver for 3-D biological simulations. Bioinformatics 32, 1256–1258 (2016).

Expected skills::

Knowledge of molecular and cell biology // Strong interest in the information gathering, analysis, modelling and simulation of biological systems. // Programming skills (python, R, bash and perl for the scripts and software tools are written in C++). // Ability to access and evaluate scientific literature.

Possibility of funding::

Yes

Possible continuity with PhD: :

Yes

Comments:

The project will be supervised by Arnau Montagud, co-supervised by Miguel Ponce de León and Alfonso Valencia.

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

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:

Molecular pattern recognition from high-throughput data of patients in a Real World Database

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. After the isolation and preparation of GEO database, it is necessary carrying out tasks associated with the validation of our RWD repository. This validation process is based on pathway enrichment analyses and the study of over/under expression of proteins, and they will be done by using artificial intelligence (AI) techniques. The student enrolled in this project will be the responsible to validate a subset of patients associated with certain specific pathologies (pending to decide). The student will select the patients from our RWD repository and will use AI techniques to compare patients' data in front of data from healthy people.

Expected skills::

programming python, c++ or Matlab

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Katsuyuki Shiroguchi
Email	katsuyuki.shiroguchi@riken.jp
Institution	RIKEN

Website	https://www.bdr.riken.jp/en/research/labs/shiroguchi-k/index.html
Group	Lab for Prediction of Cell Systems Dynamics

Project

Computational systems biology

Project Title:

Studying cell dynamics by combining live imaging and single-cell RNA-seq

Keywords:

Single cell, RNA sequencing, Optical microscope, Technology development, Challenge,

Summary:

As part of the internship, the student will use our developed single-cell picking system, which combines live imaging and single-cell whole gene expression analysis, to study molecular mechanisms of cell dynamics in a cell population, e.g., cell activation, differentiation, or cell-cell interaction, related to the immune system, cancer, or organoids.

Expected skills::

High motivation and good social manners

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Comments:

Good opportunity for those who want to try both experiments and computational analyses

---

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

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

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Jordi Garcia Ojalvo
Email	jordi.g.ojalvo@upf.edu
Institution	Universitat Pompeu Fabra
Website	https://www.upf.edu/web/dsb
Group	Dynamical Systems Biology

Project

Computational systems biology

Project Title:

Self-organization and decision making in cells and tissues

Keywords:

Developmental processes, microbiology, autoimmune diseases, single-cell behavior, dynamics

Summary:

We offer research projects devoted to understanding, using mathematical modeling, how cells make decisions and how cellular populations self-organize in time and space. The specific system and biological process to be studied will depend on the interest of the student and the availability of data/questions from our own lab and the labs of our international collaborators, at the time of the project design.

Expected skills::

Programming experience in Python, C or Julia would be useful, but it's not essential.

Possibility of funding::

No

Possible continuity with PhD: :

Yes

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Carlos Barata Martí
Email	cbmqam@cid.csic.es
Institution	IDAEA-CSIC
Website	www.idaea.csic
Group	Toxicology Group

Project

Computational systems biology

Project Title:

Funcional transcriptomic analyssis in the crustacean Daphnia magna

Keywords:

functional gene annotation, genome, Daphnia, curate

Summary:

Developing a friendly use pipeline for processing RNA seq data in Daphnia magna which involves ensambling reads, maping, quantifying counts and functional annotation and interpretation using Blast. KEGG and other bioinformatic databases. In addition the student will work on the annotation of the probes of a 8 x 60 K Agilent eArray containing the full set of the 41317gene models representing the full transcriptome of Daphnia magna. Three years ago we developed this array and we were able to annotate 50% of its probes. However, the genome of Daphnia is changing continuously and hence it has to be re-annotated using the existing gene Bancs. The idea is to use the Daphnia magna genome (wfleabase) for a primary annotation of probes and then using NCBI Blast tools using translated proteins or genes across taxa (mainly arthropods). The end product is to provide gene names associated to each probe and its homologous in Drosophila, humans and other species. We also intent to annotate the gene codes to perform GERONTOLOGY, KEGG and other functional analyses

References:

Campos B, Fletcher D, Piña B, Tauler R, Barata C. Differential gene transcription across the life cycle in Daphnia magna using a new all genome custom-made microarray. BMC Genomics 2018; 19. Campos B, Garcia-Reyero N, Rivetti C, Escalon L, Habib T, Tauler R, et al. Identification of metabolic pathways in daphnia magna explaining hormetic effects of selective serotonin reuptake inhibitors and 4-nonylphenol using transcriptomic and phenotypic responses. Environmental Science and Technology 2013; 47:

9434-9443. Campos B, Rivetti C, Tauler R, Piña B, Barata C. Tryptophan hydroxylase (TRH) loss of function mutations in Daphnia deregulated growth, energetic, serotoninergic and arachidonic acid metabolic signalling pathways. Scientific Reports 2019; 9. Fuertes I, Campos B, Rivetti C, Pinã B, Barata C. Effects of Single and Combined Low Concentrations of Neuroactive Drugs on Daphnia magna Reproduction and Transcriptomic Responses. Environmental Science and Technology 2019a; 53: 11979-11987. Fuertes I, Jordão R, Piña B, Barata C. Time-dependent transcriptomic responses of Daphnia magna exposed to metabolic disruptors that enhanced storage lipid accumulation. Environmental Pollution 2019b; 249: 99-108. Piña B, Barata C. A genomic and ecotoxicological perspective of DNA array studies in aquatic environmental risk assessment. Aquatic Toxicology 2011; 105: 40-49.

Expected skills::

knowled in r, pyton, automatic functional annotation, gerontaology, KEGG

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Eva Maria Novoa
Email	eva.novoa@crg.eu
Institution	Center for Genomic Regulation (CRG)
Website	https://www.crg.eu/en/programmes-groups/novoa-lab
Group	Epitranscriptomics and RNA Dynamics

Project

Computational systems biology

Project Title:

Understanding the role of RNA folding in neurodegenerative diseases using third-generation sequencing technologies (oxford nanopore)

Keywords:

Oxford Nanopore sequencing; RNA modiications; RNA structure; machine learning;

Summary:

RNAs are not simple intermediary molecules between DNA and protein, but are in fact functional molecules capable of regulating central cellular processes. Because RNA is a single-stranded molecule, it tends to fold back on itself, forming stable secondary and tertiary structures by internal base pairing and other interactions. The function of RNAs can vary depending on the specific folding that the molecule, and therefore accurate RNA structural maps are needed to understand the complexity, function, and regulation of these molecules. Unfortunately, current methods generating RNA structure maps employ second-generation sequencing technologies (e.g. Illumina), which are unable to produce information on highly repetitive regions of the genome. Here we will use Oxford Nanopore Technologies (ONT), capable of producing full-length RNA molecule reads, to produce RNA structure maps for highly repetitive regions, such as those involved in neurodegenerative diseases such as Amyotrophic Lateral Sclerosis (ALS) or Fronto-temporal dementia (FTD).

References:

1. Liu H*, Begik O, Lucas MC, Ramirez JM, Mason CE, Wiener D, Schwartz S, Mattick JS, Smith MA and Novoa EM#. Accurate detection of m6A RNA modifications in native RNA sequences. Nature Comm 2019, 10:4079. doi:10.1038/s41467-019-11713-9 2. Beaudoin JD*, Novoa EM*, Vejnar CE, Yartseva V, Takacs CM, Kellis M and Giraldez AJ. Analyses of mRNA structure dynamics identify the embryonic RNA regulome. Nat Struct Mol Biol 2018, 25, 677-686 3. Smith MA*, Ersavas T*, Ferguson JM*, Liu J, Lucas MC, Begik O, Bojarski L, Barton K and Novoa EM#. Barcoding and demultiplexing Oxford Nanopore native RNA sequencing reads with deep residual learning. bioRxiv 2019, 864322 (under review in Genome Research) 4. Cozzuto L, Liu H, Pryszcz LP, Hermoso Pulido T, Ponomarenko J and Novoa EM#. MasterOfPores: a workflow for the analysis of Oxford Nanopore direct RNA sequencing datasets bioRxiv 2019, 828336 (accepted in Front in Genet)

Expected skills::

Mandatory: Python, R. Desirable: machine learning, handling of third-generation sequencing data

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor Marta Melé

Email marta.mele.messeguer@gmail.com

Institution Barcelona Supercomputing Center

Website https://www.bsc.es/discover-bsc/organisation/scientific-structure/transcriptomics-and-functional-genomics-lab-tfgl

Group Transcriptomics and Functional Genomics Lab

Project

Computational systems biology

Project Title:

Single-cell transcriptomic meta-analysis of human disease across tissues

Keywords:

single-cell transcriptomics, Cell type deconvolution, disease, aging, smoking, meta-analysis,

Summary:

The candidate will join Marta Melé's Transcriptomics and Functional Genomics lab in the Life Sciences Department at the Barcelona Supercomputing Center. The lab is interested in understanding how individual variation in gene expression can explain phenotypic differences between individuals both in the context of health and disease. To address this question, we use large-scale transcriptomic analysis and latest single-cell sequencing technologies combined with methods development to study gene expression, splicing and cell type composition variation across human tissues and phenotypes. In this project, we will perform a large-scale analysis of single-cell RNA-sequencing datasets across tissues to address how individual variation in gene expression can explain phenotypic differences between individuals. First, we will analyze hundreds of single-cell RNA-sequencing datasets to explore the impact of aging, smoking, gender and certain disease conditions to changes in gene expression and cell type composition in blood. Second, we will use cell type deconvolution methods to map single-cell signatures in expression data across many tissues from individuals with different conditions including diabetes and cardiovascular diseases. Overall, this project will explore in depth what is the role of gene expression and cell type composition in defining why human individuals are different from one another and how this impacts disease progression. What you will learn: Development of computational pipelines to analyze and interpret large datasets specially from single-cell RNA-seq, and bulk RNA-sequencing. Working in a high performance computing (HPC) environment. Interpret multi-omics data, working through scientific collaboration, effective communicating research, writing scientific articles and critical thinking.

References:

Melé, M. et al. The human transcriptome across tissues and individuals. Science (80-.). 348, 660–665 (2015).

Expected skills::

Availability to start in July 2020 is encouraged Strong programming skills in bash, python, R, perl, or similar, some experience working in HPC clusters Some experience with Next Generation Sequencing data analysis Excellent communication skills in spoken and written English Capacity to contribute to research projects with novel research ideas and analysis Capacity to work as a team in a highly collaborative and diverse environment

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Antonio Julià
Email	toni.julia@vhir.org
Institution	Vall Hebron Research Institute
Website	www.urr.cat
Group	Grup de Recerca de Reumatologia

Project

Computational systems biology

Project Title:

Identification of synergistic drug combinations in autoimmune diseases through single-cell analysis

Keywords:

Autoimmune Disease; Combinatorial drug therapy; Single Cell RNA-seq; Network analysis; Personalized medicine

Summary:

Autoimmune diseases (ADs) are chronic inflammatory diseases that are present at a high frequency in our population. They include diseases like rheumatoid arthritis, psoriasis, lupus and inflammatory bowel disease. They cause a significant reduction in the quality of life of many patients and a significant increase in comorbidities. In the last decade there has been a big increase in number of therapies available to treat ADs. However, these therapies only work for a subset of patients and, in many cases, after a period of time their efficacy diminishes. In our group we are convinced that one solution to this major health problem would be the use of drug combinations. By identifying pairs of drugs that synergize their effect we could provide a more powerful therapy. The present master's thesis project is focused in this interesting research problem. To do so, the student will use single cell RNA-seq data on tissue and immune system cell samples, and different statistical and data mining tools to identify the most likely drug combination for a specific autoimmune disease. During this project, the student will learn a very novel type of data, will acquire advanced data analysis skills and interact with several other bioinformatics specialists in the group as well as clinical researchers.

References:

We have been recently granted a 5-year EU project on combinatorial therapies. We are the coordinators of this translational project, which includes single-cell data analysis. http://doctis.eu/

Expected skills::

Programming skills in R and Python Statistical analysis of data Biological knowledge

Possibility of funding::

No

Possible continuity with PhD: :

Yes

Comments:

While we don't provide funding during the Master's thesis, our aim is to integrate the candidate into our team and provide him/her with funding to be able to pursue his/her PhD.

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Alberto Santos Delgado
Email	alb.santosdel@gmail.com
Institution	Oxford Big Data Institute
Website	https://www.bdi.ox.ac.uk/
Group	Multi Omics Analytics

Project

Computational systems biology

Project Title:

Computational Prediction of Host-pathogen Protein-protein Interactions in Human

Keywords:

Microbiology, Protein-protein interactions, Machine learning, Biomedical databases

Summary:

Protein-protein interactions (PPIs) define the complexity of biological processes in healthy and disease states. Intra-species PPIs describe partially this complexity but symbiotic and pathogenic interactions contribute as well by either benefiting or harming the host. The study of inter-species PPIs are of special interest in the case of pathogens since knowledge extracted from these interactions can lead to new therapeutic targets to avoid their negative effects on the host. There exists an extensive collection of intra-species PPIs enabled by the rapid development of high-throughput experimental technologies. However, experimental identification of inter-species interactions is not simple and computational prediction becomes then necessary (Cuesta-Astroz et al 2018). For instance, we proposed a homology-based prediction method to obtain human-parasite PPIs in 15 parasitic species (Cuesta-Astroz and Santos et al 2019). Additionally, this method incorporated biological context relevant in the parasites' life cycles to obtain accurate spatially-resolved interactions. Here, we want to implement a method that benefits from features derived from intra-species interactions to predict inter-species interactions focusing on human-pathogen interactions. The vast amount of intra-species interactions compiled in publicly available databases can be used to train sequence-based classification algorithms that can then be tested in both intra- and inter-species interactions also available in several resources. Furthermore, this method can be improved by annotating the predicted interactions with known biological context such as infection, survival and pathogenic mechanisms. This project will be carried out in close collaboration with Dr. Yesid Cuesta-Astroz from the University of Antioquia and the Colombian Institute of Tropical Medicine (Medellin, Colombia).

References:

Computational and Experimental Approaches to Predict Host-Parasite Protein-Protein Interactions. Cuesta-Astroz Y, Oliveira G Analysis of Predicted Host–Parasite Interactomes Reveals Commonalities and Specificities Related to Parasitic Lifestyle and Tissues Tropism Cuesta-Astroz Y, Santos A, Oliveira G, and Jensen LJ Comparing two deep learning sequence-based models for protein-protein interaction prediction Richoux F, Servantie C, Borès C, and Téletchéa S

Expected skills::

Python, Machine learning, Fast.ai, PyTorch

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed