Master project 2021-2022

Personal Information

Supervisor	Julio Alonso Padilla
Email	julio.a.padilla@isglobal.org
Institution	Instituto de Salud Global de Barcelona
Website	https://www.isglobal.org/our-team/-/profiles/11902
Group	Iniciativa Chagas

Project

Computational systems biology

Project Title:

Image analysis for setting up an anti-Trypanosoma cruzi biological assay

Keywords:

Chagas disease, Trypanosoma cruzi, drug-discovery, image-based assay, image analysis.

Summary:

Chagas disease is a neglected infectious disease caused by the protozoan Trypanosoma cruzi (T. cruzi). It exerts its highets burden in Latin America where there are over 6 million people infected. Drug discovery for Chagas disease is an urgent medical need given the variable efficacy and frequent side effects associated with current treatments. With the aim to further qualify a battery of selected compounds with specific anti-T. cruzi activity we are developing an image-based high content assay. With it, we want to envisage the potential targets of the compounds, providing them with a very valuable information towards their further development as drugs.

References:

https://pubmed.ncbi.nlm.nih.gov/33445756/ https://pubmed.ncbi.nlm.nih.gov/32635780/ https://pubmed.ncbi.nlm.nih.gov/25615687/

Expected skills::

Database analysis, image parameters analysis, team working spirit, willingness to learn.

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Comments:

Participation in other research lines related to computational biology currently ongoing in the lab could be discussed, like the design of epitope-based biomedical

interventions or the datamining of RNAseq databases for the identification of potential trypanosomaviruses.

Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Marta Mele
Email	marta.mele@bsc.es
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:

Unraveling the role of long non-coding RNAs in response to Ebola infection

Keywords:

Transcriptomics, long non-coding RNA, viral infections, differential gene expression, de novo transcriptome assembly, Ebola, alternative splicing, linear models.

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 the information encoded in our genome determines gene expression variation and splicing patterns across all cell types in our body. To address this question, we use large-scale transcriptomic analysis and the latest single-cell sequencing technologies combined with methods development to study gene expression, splicing and cell type composition variation across human tissues and phenotypes. We ultimately want to understand how these expression changes in both coding and non-coding genes are associated with specific disease states1,2. The candidate will focus on developing an integrative analysis of gene expression variation in non-human primate tissues that have been infected with Ebola virus. Specifically, he/she will focus on the analysis of long non-coding RNAs that have been shown to participate in immune system activation but have not yet been studied in depth in the context of viral infections. This is a unique unpublished dataset recently generated in collaboration with the Sabeti lab (Broad Institute of Harvard and MIT). Viral hemorrhagic fevers, such as Ebola, are infamous for their astonishingly high fatality rates, yet the precise mechanisms and pathways altered during infection still lack a comprehensive characterization. The few transcriptional studies carried out to date have been limited to a few cell types and to protein-coding genes3,4. A recent study has explored the transcriptional changes that occur upon Ebola infection at the single-cell level4, shedding light on Ebola tropism, its replication dynamics, and the elicited immune response. Although long non-coding RNAs (lncRNAs) are

emerging as a key class of immune response regulators5, the role of the non-coding transcriptome in response to Ebola infection remains elusive. In our lab, we have studied the landscape of lncRNA expression in Chinese rhesus macaque (Macaca mulatta) over the course of Ebola infection6. Through a full de novo annotation, we have discovered over 10,000 novel lncRNAs which, along with previously annotated lncRNAs, serve as an atlas to study the non-coding transcriptome in macaque. Using PBMC single-cell data, we have found that lncRNAs are dysregulated upon Ebola infection across multiple immune cell types, in a cell type and disease stage specific manner, and that dysregulated lncRNAs are found close to protein-coding genes involved in immune response. In this project, the candidate will computationally analyze hundreds of bulk transcriptomes from 12 different tissues (adrenal gland, brain, kidney, liver, lung, lymph node, ovary, skin, spinal cord, spleen, testis, and whole blood) of macaques infected with Ebola at several time points. First we aim to detect lncRNA up-regulated or down-regulated in response to the viral infection, and identify early, mid and late onset lncRNA that could be the key players in the immune response to Ebola virus. Second, we want to discern between systemic and tissue-specific responses, and build gene expression networks that connect protein-coding genes with their lncRNA regulators. Ultimately, the question that we want to tackle in this project is what is the role of lncRNA in mediating the immune response elicited upon Ebola virus infection. What you will learn: Development of computational pipelines to analyze and interpret large datasets, especially from bulk RNA-sequencing and single-cell RNA-seq as well. Working in a High Performance Computing environment. Interpretation of multi-omics data. Scientific collaboration, effective communication of research findings into internal and external meetings, scientific writing and critical thinking.

References:

1. Melé, M. et al. The human transcriptome across tissues and individuals. Science (80-.). 348, 660–665 (2015). 2. Melé, M. et al. Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. Genome Res. 27, 27–37 (2017). 3. Caballero, I. S. et al. In vivo Ebola virus infection leads to a strong innate response in circulating immune cells. BMC Genomics 17, 707 (2016). 4. Kotliar, D. et al. Single-Cell Profiling of Ebola Virus Disease In Vivo Reveals Viral and Host Dynamics. Cell, 83, 1383–1401, (2020). 5. Chen, Y. G., Satpathy, A. T. & Chang, H. Y. Gene regulation in the immune system by long noncoding RNAs. Nat. Immunol. 18, 962–972 (2017). 6. Santus, L. et al. Single-cell profiling of long non-coding RNAs during Ebola virus infection in non-human primates reveals dysregulation of the non-coding transcriptome in key immune-related pathways. In preparation.

Expected skills::

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 Availability to start in July 2021 is preferred

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor Josep Vilardell

Email josep.vilardell@ibmb.csic.es

Institution ICREA / IBMB

Website https://www.ibmb.csic.es/en/department-of-molecular-genomics-dmg/molecular-mechanisms-of-pre-mrna-splicing/

Group RNA Splicing

Project

Computational systems biology

Project Title:

A Genomics Perspective of pre-mRNA Splicing

Keywords:

spliceosome, ribosome, pre-mRNA, ribosomal protein, genome

Summary:

The information content of genomes can be greatly expanded by pre-mRNA splicing. Virtually all human pre-mRNAs need to be spliced to become mRNAs. Furthermore, most pre-mRNAs can be spliced into different mRNAs by alternative splicing. Therefore, it is hardly surprising that perturbations in splicing are linked to disease. However, we know little on how the splicing of particular sets of RNAs may be affected, and even less on how a number of splicing changes are coordinated during development or disease. To start addressing this question, we are analyzing WGS and RNASeq data from a number of cancer datasets. Although we are interested in all events of regulated splic-ing, we pay special attention to those related to the biosynthesis and function of the ri-bosome. A cycling cell depends on a suitable set of ribosomes to provide the necessary amount of structural and functional proteins before mitosis; paradoxically, making this machinery requires most of the cell's energy (as an illustrative example, a growing HeLa cell is making 1.6 x105 ribosomal proteins (RP) per minute). Thus, we expect that fast-growing cell, subjected to a strong selection (such as a tumor cell), will tweak this process trying to get any advantage. However, the analysis of the RP transcriptome presents specific challenges because (a), it includes the mRNAs that are amongst the cell's most abundant, and while the ribosome has one copy of each RP, the corre-sponding mRNAs are in variable amounts; (b), the RP pre-mRNAs undergo little alterna-tive splicing; and (c), the majority of human pseudogenes originate from them, which introduces ambiguity when mapping reads to the genome. Our initial results suggest that processing of this set of transcripts is altered in cancer in unexpected ways, and we plan on strengthening our conclusions by expanding our analyses. In this context there are many opportunities for those with a strong motivation to document genomic strategies that control the transcriptome of specific gene families, like those related to the ribosome or the spliceosome. The tasks involve quality analysis of raw RNASeq data, mapping using standard tools (for example, Hisat, STAR, and those related to direct sequencing of RNA), statistical analysis (Ballgown, Salmon, Vast-tools, DexSeq, or others), and modeling. Subject to progress, we would explore the use of transcriptomics data as a disease prognosis tool; namely, is a distinct distribution of transcripts indicative of a particular disease out-come?

References:

* Hussain, S. (2018) "Native RNA- Sequencing Throws its Hat into the Transcriptomics Ring" TiBS 1434. https:// /doi.org/10.1016/j.tibs.2018.02.007 * Guimaraes, JC. and Zavolan, M. (2016) "Patterns of ribosomal protein expression specify normal and malignant human cells" Genome Biol. 17:236-248 * Gupta, V. and J. R. Warner (2014). "Ribosome-omics of the human ribosome." RNA 20: 1004-1013. * Bitton, D. A., et al. (2014). "LaSSO, a strategy for genome-wide mapping of intronic lariats and branch points using RNA-seq." Genome Res 24(7): 1169-1179. * Acuna, L. I. and A. R. Kornblihtt (2014). "Long range chromatin organization: a new layer in splic-ing regulation?" Transcription 5. * Kawashima T et al (2014) Widespread use of non-productive alternative splice sites in Saccharo-myces cerevisiae. PLoS Genet. 2014 Apr 10;10(4):e1004249. * Zhang, J. and J. L. Manley (2013). "Misregulation of pre-mRNA alternative splicing in cancer." Cancer Discov 3(11): 1228-1237. * Fu, R. H., et al. (2013). "Aberrant alternative splicing events in Parkinson's disease." Cell Trans-plant 22(4): 653-661. * Plass, M., et al. (2012). "RNA secondary structure mediates alternative 3'ss selection in Saccha-romyces cerevisiae." RNA 18(6): 1103-1115.

Expected skills::

In addition to a good Bioinformatics background, an ability to work independently and to integrate diverse data (mutational analyses, epigenetics, transcriptomics, and others) is desirable. Knowledge of bash and R will be helpful.

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Comments:

Our group provides an field with significant questions to be addressed by a bioinformatician developing a career. We can steer the student in the right direction to use her/his skills, helping to make sure that the effort is relevant. We cannot, however, provide refined Bioinformatics background. These questions relate to the study of the spliceosome from a Systems Biology point-of-view: is there a limit in the number of exons that can be processed into the same transcript? What determines the splicing order of the introns? What may be the role of the pre-mRNA structure? Has the number of RNA Pol II molecules transcribing a gene any relation with its splicing? Is splicing connected to translation in more ways than we presently know?

Master project 2021-2022

Personal Information

Supervisor	BEJANIN Alexandre
Email	ABejanin@santpau.cat
Institution	Research Institute of the Hospital de la Santa Creu i Sant Pau
Website	https://santpaumemoryunit.com/es/
Group	Sant Pau Memory Unit

Project

Computational systems biology

Project Title:

Neuroimaging of vascular disease in sporadic and genetic Alzheimer's disease

Keywords:

Alzheimer's disease, Vascular lesions, Brain, Neuroimaging, Segmentation

Summary:

Cerebral small vessel disease (SVD) is now recognized to be the most important vascular contributor to dementia. Substantial evidence shows that SVD and Alzheimer's disease share risk factors and have additive, if not synergetic, effects on cognitive impairment and neurodegeneration. Yet, the intersection between SVD and AD pathophysiological processes remains unclear. Consequently, a proper understanding of how SVD exerts its action on the aging brain, interacts with Alzheimer's disease, and leads to clinical symptoms is urgently needed. In the present project, we will focus on one specific radiological markers of SVD: the white matter hyperintensities (WMH). The term of WMH refers to abnormal clusters of hyperintense signal in white matter tissue on T2-weighted or fluid attenuated inversion recovery (FLAIR) MRI. WMH are frequently found in patients with Alzheimer's disease but their specific relationship with the disease processes remains debated. The aim of the current project is therefore to better characterize the emergence of WMH in the context of Alzheimer's disease, and delineate their effect on the neuronal loss. For this, the candidate will use and manipulate the MRI data (T1, T2, FLAIR) of >100 patients with sporadic Alzheimer's disease or Down Syndrome (i.e., genetic Alzheimer's disease). Specifically, the master student will: • Familiarize with Alzheimer's disease pathophysiology in a highly interdisciplinary environment. • Learn about brain image processing tools • Implement different methods to segment WMH and compare their application in the context of Down syndrome • Familiarize with statistical approaches to assess the relationships between WMH and different biomarkers of Alzheimer's disease

Expected skills::

Interest for neuroimaging and/or neurodegenerative disease, curiosity, ability to work independently but also in group, programming skills (R or matlab) are not mandatory but will be valorized, basic statistical knowledge, ability to interact in english

Possibility of funding::

No

Possible continuity with PhD: :

Yes

Comments:

We have several other neuroimaging projects. Don't hesitate to contact me if you are interested in learning about neuroimaging in neurodegenerative diseases.

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Andrés Ozaita
Email	andres.ozaita@upf.edu
Institution	UPF
Website	https://www.upf.edu/neurophar
Group	Laboratory of Neuropharmacology

Project

Computational systems biology

Project Title:

Intellectual disability can be caused by genetic mutations

Keywords:

intellectual disability, treatment, synaptic proteome, synaptic transcriptome, splicing

Summary:

Synaptic Intellectual disability may derive from specific genetic alterations, as found in neurodevelopmental disorders such as fragile X syndrome (FXS) and Down syndrome (DS), both disorders associated to relevant alterations in synaptic plasticity. Mouse models of these disorders mimicking the genetic alterations found in humans have demonstrated relevant tools to understand the physiopathology of the disorders and to test pharmacological approaches that may improve cognitive performance. In the lab we have described an approach to improve cognitive performance in models of FXS and DS, but the impact of these treatments in the biology of the synapse has not been addressed. We are now investigating, using high throughput proteomic and transcriptomic analysis of sorted synaptic contacts, the characteristics of pathological synapses, and the effects that pharmacological treatments have in improving synaptic plasticity in both models of intellectual disability.landmarks of intellectual disability

References:

Navarro-Romero A, Vázquez-Oliver A, Gomis-González M, Garzón-Montesinos C, Falcón-Moya R, Pastor A, Martín-García E, Pizarro N, Busquets-Garcia A, Revest JM, Piazza PV, Bosch F, Dierssen M, de la Torre R, Rodríguez-Moreno A, Maldonado R, Ozaita A. Cannabinoid type-1 receptor blockade restores neurological phenotypes in two models for Down syndrome. Neurobiol Dis. 2019 May;125:92-106. doi: 10.1016/j.nbd.2019.01.014. Epub 2019 Jan 25. PMID: 30685352. Busquets-Garcia A, Gomis-González M, Guegan T, Agustín-Pavón C, Pastor A, Mato S, Pérez-Samartín A, Matute C, de la Torre R, Dierssen M, Maldonado R, Ozaita A. Targeting the endocannabinoid system in the treatment of fragile X syndrome. Nat Med. 2013 May;19(5):603-7. doi: 10.1038/nm.3127. Epub 2013 Mar 31. PMID: 23542787.

Expected skills::

Bioinformatics

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 2021-2022

Personal Information

Supervisor	Ana I. Caño-Delgado
Email	ana.cano@cragenomica.es
Institution	CRAG (Centre for Research in Agricultural Genomics)
Website	https://www.cragenomica.es/research-groups/brassinosteroid-signaling-in-plant-development
Group	Brassinosteroid signaling in plant development

Computational systems biology

Project Title:

Development and implementation of bioinformatic tools for the functional analysis of plant adaptation to climate change

Keywords:

Drought, brassinosteroids, multi-omic data, bioinformatics

Summary:

Drought and elevated temperatures consequence of climate change cause major loses in agriculture and threaten foot security worldwide. The group of Ana I. Caño-Delgado at CRAG (Centre for Research in Agricultural Genomics, UAB Campus, Barcelona) has an international reputation in the study of hormone signalling pathways necessary for plant growth and adaptation to abiotic stresses. Her lab uses system biological approaches for investigating how BRI1-type of membrane receptors modulate plant growth under severe drought stress without affecting growth. By using bioinformatics, a wealth of multi-omic data has been generated in her lab that offers an excellent opportunity to design new functional genomics tools: 1) to unveil new in silico biotechnological approaches to improve the plant signalling under stress and 2) to shed light into the present understanding of the regulatory networks of brassinosteroid (BR) signalling with a bioinformatics point of view.

References:

1. Gupta A, Rico-Medina A, Caño-Delgado AI. The physiology of plant responses to drought. Science. 2020 Apr 17;368(6488):266-269. doi: 10.1126/science.aaz7614. PMID: 32299946. 2. Lozano-Elena F, Caño-Delgado AI. Emerging roles of vascular brassinosteroid receptors of the BRI1-like family. Curr Opin Plant Biol. 2019 Oct;51:105-113. doi: 10.1016/j.pbi.2019.06.006. PMID: 31349107. 3. Fàbregas N, Lozano-Elena F, Blasco-Escámez D, Tohge T, Martínez-Andújar C, Albacete A, Osorio S, Bustamante M, Riechmann JL, Nomura T, Yokota T, Conesa A, Alfocea FP, Fernie AR, Caño-Delgado AI. Overexpression of the vascular brassinosteroid receptor BRL3 confers drought resistance without penalizing plant growth. Nat Commun. 2018 Nov 8;9(1):4680. doi: 10.1038/s41467-018-06861-3. PMID: 30409967.

Expected skills::

Knowledge in bioimformatics analysis of omics data (RNAseq, proteomics, metabolomics); Coding experience in R or python; Knowledge in data analysis and statistics; Database managing and developing; Ability to report results in a clear and summarized manner working independently and within a collaborative research team; Knowledge in web tools is desirable

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Comments:

The candidate will work with Veredas Coleto Alcudia, an expert Bionformatician doing the PhD in the lab.

Master project 2021-2022

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:

PREDICTING AND MONITORING CANCER USING MACHINE LEARNING ON NANOPORE SEQUENCING DATA

Keywords:

nanopore sequencing, RNA modifications, small RNA, machine learning, deep learning, cancer, prognosis, sample classification

Summary:

Dysregulation of small RNA abundances and their RNA modifications is a well-known feature in cancer cells, which leads to enhanced expression of specific oncogenic transcripts and proteins [1,2]. Despite the well-established association between small RNA dysregulation and cancer progression and malignancy, small RNA abundances and modifications are still not being used as screening, diagnostic or prognostic markers for cancer detection or progression, mainly due to the lack of a simple, unbiased and cost-effective method to quantify small RNA abundances and their modifications. Our laboratory has pioneered the use of direct RNA sequencing for the detection and quantification of RNA abundances and their modifications, including both development of improved library preparation protocols as well as the development of novel algorithms to predict and quantify RNA modifications [3-5]. Here we propose to use native RNA nanopore sequencing technology to predict the malignancy of biological samples in a high-throughput, rapid, multiplexed and cost-effective manner. Specifically, the candidate MSc student will benefit from a recently developed method in our lab to sequence small RNAs using nanopore sequencing. The candidate will then develop and apply deep learning algorithms to classify small RNA profiles into "normal", "tumoral" and "metastasic". Once the classification model is benchmarked and validated using cell lines, the methodology will then be applied to patient-derived samples.

References:

1. Begik O, Lucas MC, Ramirez JM, Liu H, Mattick JS and Novoa EM#. Integrative analyses of the RNA modification machinery reveal tissue- and cancer-specific signatures. Genome Biology 2020, 21:97. doi: 10.1186/s13059-020-02009-z 2. Gingold et al., A Dual Program for Translation Regulation in cellular profileration and differentiation. Cell 2014, 158(6):1281-1292. 3. 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 4. Smith MA*, Ersavas T*, Ferguson JM*, Liu J, Lucas MC, Begik O, Bojarski L, Barton K and Novoa EM#. Molecular barcoding of native RNAs using nanopore sequencing and deep learning. Genome Research 2020 30(9): 1345-1353 5. Begik O*, Lucas MC*, Ramirez JM, Milenkovic I, Cruciani C, Vieira HGS, Medina R, Liu H, Sas-Chen A, Mattick JS, Schwartz S and Novoa EM#. Quantitative profiling of native RNA modifications and their dynamics using nanopore sequencing. bioRxiv 2021, 189969 (accepted in Nature Biotechnology)

Expected skills::

python (required), R (required), prior experience with machine learning is a plus but not required, familiarity with third-generation sequencing (e.g. nanopore) is a plus but not required

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Comments:

Option for funding, as well as option for PhD continuity.

Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Pia Cosma
Email	pia.cosma@crg.es
Institution	CRG
Website	http://www.crg.eu/en/maria_pia_cosma
Group	Reprogramming and Regeneration

Project

Computational systems biology

Project Title:

Identification of master regulators of reprogramming

Keywords:

retina, reprogramming, regeneration, master regulators, gene networks

Summary:

We use gene regulatory network to identify master regulators of reprogramming and pluripotency. We are now investigating master regulators that can be enhances to induce the regeneration of the retina in mammals.

Expected skills::

bioinformatics, math lab,

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Master project 2021-2022

Personal Information

Supervisor	Arnau Montagud, Miguel Ponce de León, Alfonso Valencia
Email	arnau.montagud@bsc.es
Institution	Barcelona Supercomputing Center (BSC)
Website	https://www.bsc.es/discover-bsc/organisation/scientific-structure/computational-biology
Group	Computational Biology

Project

Computational systems biology

Project Title:

Use of data-tailored multiscale models to study drug transport

Keywords:

multiscale modelling, drug simulations, transport mechanisms, phamacodynamics

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 at the simulation of how drugs interact with different cell signalling pathways. The focus of present thesis will be the study of different mechanisms of drug transport into cells and how these interfere with the normal functioning of the cells. 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 predict drug combination responses are still computationally intense, as models typically rely on extensive drug perturbation data[1, 2]. In addition and relying on literature and databases, patient-specific dynamical models were developed[3] previously tailoring a general cancer model[4] 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 stratifications[3]. Currently and in the frame of PerMedCoE project (https://permedcoe.eu/use-cases/), we are using this same framework to obtain cell-line-specific dynamical models and performing simulated drug perturbation studies[5, 6] identifying sets of concentrations where this synergy is maximal and considering population-level constraints and behaviours. Nevertheless, our modelling framework would benefit from a detailed description of drug transport mechanisms to tally it to the latest pharmacodynamic data available. In present project, the candidate will build different drug transport models that characterise a set of drugs in different cell lines model using a multiscale modelling framework, PhysiBoSS[7], that combines agent-based[8], Boolean[5, 6] and environmental dynamics[9] modelling. In our group, we have already built a model of the transport mechanism of the Tumour Necrosis Factor or TNF, a protein that mediates tumour cells' death and we are aware that literature usually lacks all of the information needed to perfectly adjust these models' parameters. Thus, the candidate will use a large-scale model exploration optimised for HPC[10] to fit and explore the model's parameter space and to potentially approximate pharmacodynamic data that is not currently available. The candidate will first gather from databases and literature on biophysical and pharmacodynamic information on mechanisms and parameters that allow for the tailoring of the multiscale simulation to each drug. Then, the candidate will build dynamical models that capture the transport mechanisms of several drugs (in C++), use them to explore different combinations of drugs (synergistic, antagonistic, etc.) and use scripts already in place (mostly in python, but also a few in bash, perl and R) and new ones developed by the student to analyse the results of the simulations. At the end of the thesis, the student will have built more realistic transport models that help identifying more realistic drug responses potentially relevant to the clinic. What you will learn: - Computational Biology: use of databases and literature search, use of models to train critical thinking on hypotheses' generation, integrate experimental data in models and validate them, collaborative software development using GIT, design and use of computational pipelines for high performance computing. - Boolean modelling: simulate models using probabilistic techniques, study the effects of perturbations on Boolean models for cell signalling, postulate hypotheses about the underlying mechanisms of drug transport, using experimental data to filter and validate proposed mechanisms, analyse the results. - Multiscale modelling: integrating Boolean with agent-based and environmental dynamics modelling, performing 2D and 3D simulations of cells, explore regular and irregular environment architectures and their effect in the drug delivery. - Scientific dissemination: presenting results to critical audiences, participating in the writing of the research article resulting from the work.

References:

References: 1. Flobak Å, Baudot A, Remy E, Thommesen L, Thieffry D, Kuiper M, et al. Discovery of Drug Synergies in Gastric Cancer Cells Predicted by Logical Modeling. PLOS Comput Biol. 2015;11:e1004426. 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. 2017;33:2410–2. 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. 2019;9:1965. 4. Fumia HF, Martins ML. Boolean Network Model for Cancer Pathways: Predicting Carcinogenesis and Targeted Therapy Outcomes. PLoS ONE. 2013;8:e69008. 5. Stoll G, Viara E, Barillot E, Calzone L. Continuous time Boolean modeling for biological signaling: application of Gillespie algorithm. BMC Syst Biol. 2012;6:116. 6. Stoll G, Caron B, Viara E, Dugourd A, Zinovyev A, Naldi A, et al. MaBoSS 2.0: an environment for stochastic Boolean modeling. Bioinformatics. 2017;33:2226–8. 7. Letort G, Montagud A, Stoll G, Heiland R, Barillot E, Macklin P, et al. PhysiBoSS: a multi-scale agent-based modelling framework integrating physical dimension and cell signalling. Bioinformatics. 2019;:bty766. 8. Ghaffarizadeh A, Heiland R, Friedman SH, Mumenthaler SM, Macklin P. PhysiCell: An open source physics-based cell simulator for 3-D multicellular systems. PLOS Comput Biol. 2018;14:e1005991. 9. Ghaffarizadeh A, Friedman SH, Macklin P. BioFVM: an efficient, parallelized diffusive transport solver for 3-D biological simulations. Bioinformatics. 2016;32:1256–8. 10. Ozik J, Collier N, Wozniak JM, Macal C, Cockrell C, Friedman SH, et al. High-throughput cancer hypothesis testing with an integrated PhysiCell-EMEWS workflow. BMC Bioinformatics. 2018;19:483.

Expected skills::

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

Possibility of funding::

Yes

Possible continuity with PhD: :

Yes

Master project 2021-2022

Personal Information

Supervisor	Mireia Valles-Colomer
Email	mireia.vallescolomer@unitn.it
Institution	CIBIO (University of Trento)
Website	http://segatalab.cibio.unitn.it/
Group	Segata Lab

Project

Computational systems biology

Project Title:

The neuroactive potential of the underexplored fraction of the human microbiome and its association with health and disease

Keywords:

Microbiome, gut-brain axis, metagenomics

Summary:

There is evidence of bidirectional communication between the gut microbiota and the nervous system (gut-brain axis), potentially playing a role in brain development, brain physiology, and behavior. One of the main mechanisms of gut-brain communication is the metabolism of molecules with neuroactive properties (neuroactive compounds) by members of the gut microbiota. Metagenomic sequencing provides insight into both species composition and their functional potential, but dedicated tools are crucial for gut-brain axis systematic analysis and interpretation. To this end, we assembled the first reference catalogue of neuroactivity of human gut microorganisms, which allowed us to perform the first population-level study on the link between the gut microbiome and quality of life and depression (Valles-Colomer et al, 2019). Metagenomics is rapidly evolving, and a wide diversity of the human microbiome remains unexplored. By gathering >9,000 samples from 47 different metagenomic datasets across the world, we assembled >150,000 genomes from metagenomes, and found >75% of them to belong to so far uncharacterized species (Pasolli et al, 2019). The neuroactive potential of this underexplored diversity, as well as its contribution to host health and disease remains unchartered territory. The aims of this project are 1) to expand the manually-curated neuroactivity framework (gut-brain modules) to recently-described microbiota-derived compounds influencing host health, 2) to assemble a reference catalogue of neuroactivity of so far uncharacterized members of the microbiome, and 3) to perform an association study between microbiome neuroactivity and host health and disease on a set of publicly-available and host lab datasets. The project will be carried out under supervision of Dr Mireia Valles-Colomer at the Segata Lab (CIBIO, University of Trento, Italy), an international and multidisciplinary work environment with access to high-performance computing resources and cutting-edge data. Remote working arrangements are possible.

References:

Valles-Colomer, M. et al. The neuroactive potential of the human gut microbiota in quality of life and depression. Nat. Microbiol. (2019). doi:10.1038/s41564-018-0337-x. Pasolli, E. et al. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. Cell 176, 649–662 (2019).

Expected skills::

Familiarity with Python/R programming, notions of microbiology and multivariate statistics..

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Master project 2021-2022

Personal Information

Supervisor	Alfonso Valencia, Davide Cirillo & Jon Sánchez
Email	alfonso.valencia@bsc.es; davide.cirillo@bsc.es; jon.sanchez@bsc.es
Institution	Barcelona Supercomputing Center
Website	http://life.bsc.es/compbio
Group	Computational Biology

Project

Computational systems biology

Project Title:

Graph-based approaches for the study of drug interactions in comorbidity relations

Keywords:

Comorbidity, omic data, drug-drug interactions

Summary:

The candidate will join Alfonso Valencia's lab (Computational Biology) within the Life Sciences Department at the Barcelona Supercomputing Center. The project will focus on the analysis of the role of drug treatments and drug-drug interactions in comorbidity relations. Comorbidity is defined as the altered risk of developing secondary diseases in patients suffering from a previous one. The co-occurrence of two or more conditions in the same patient increases with age and hinders the choice of the proper treatment, as drugs used to treat different disorders can interact and generate adverse effects. As an example, the drugs used to treat type 2 diabetes and atrial fibrillation (glyburide and warfarin), two well known comorbid conditions, interact, increasing both the risk of bleeding and blood levels of hypoglycemic agents. Disease comorbidity relations have been analyzed using insurance claims and medical records, connecting diseases based on their temporal association [1]. The molecular bases of such associations have been analyzed making use of disease-associated genes (alone or in combination with protein-protein interaction networks) [2-3]. In the last years, our group has analyzed comorbidity relations and their potential association to drug-treatments using microarray [4-6] and RNAseq data. Additionally, we have analyzed Electronic Health Records from Catalonia, Blumenau (Brazil) and Indianapolis (USA), identifying increased risk of co-administration of drugs known to interact with ageing. The mentioned study has identified gender-dependent patterns of co-administration of drug-drug interactions that correlate with diagnoses and comorbidities. The main goal of the project is to better understand the role and effect of drug treatments and drug-drug interactions on comorbidity relations, looking for alternative drug-treatments that will decrease the risk of harmful interactions. To this end, we will combine disease-specific protein-protein interaction networks extracted from the Integrated Interactions Database [7] with drug-gene, drug-disease and drug-drug associations extracted from DrugBank [8] (https://go.drugbank.com/). We will focus on cardiovascular disorders and their associated comorbid conditions (e.g. asthma, high blood pressure, diabetes...) and

treatments (digoxin, furosemide, glyburide, warfarin…). Disease-diagnose and drug-treatment information extracted from the Information System for Research in Primary Care (SIDIAP) will be used to evaluate the treatment associated risk of developing secondary diseases and the use of alternative drug-treatments to decrease comorbidity risk. The SIDIAP contains primary care Electronic Health Records from 2008 to 2018 covering 75% of the Catalan population.

References:

1. Jensen, A. B. et al. Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. Nat. Commun. 5, 4022 (2014) 2. Goh, K. et al. The human disease network. Proc Natl Acad Sci U S A. 104(21): 8685-90 (2007) 3. Menche, J. et al. Disease networks. Uncovering disease-disease relationships through the incomplete interactome. Science. 347(6224): 1257601 (2015) 4. Ibañez, K. et al. Molecular evidence for the inverse comorbidity between central nervous system disorders and cancers detected by transcriptomic meta-analyses. PLoS Genet. 10(2):e1004173 (2014) 5. Sánchez-Valle, J. et al. A molecular hypothesis to explain direct and inverse co-morbidities between Alzheimer's disease, glioblastoma and lung cancer. Sci. Rep. 7(1): 4474 (2017) 6. Sánchez-Valle, J. et al. Interpreting molecular similarity between patients as a determinant of disease comorbidity relationships. Nat. Commun. 11, 2854 (2020) 7. Kotlyar, K. et al. IID 2018 update: context-specific physical protein-protein interactions in human, model organisms and domesticated species. Nucleic Acids Res. 47(D1):D581-D589 (2019) 8. Wishart, D. et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 46(D1): D1074-D1082 (2018)

Expected skills::

Background in biomedicine and basic knowledge of molecular biology // Basic computational skills (Python, R and bash programming) // Ability to analyse and interpret biological data // Ability to access and evaluate scientific literature

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Comments:

Please include supervisors and Alba Jené in any future communication

---

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Josep Vilardell
Email	josep.vilardell@ibmb.csic.es
Institution	ICREA / IBMB
Website	https://www.ibmb.csic.es/en/department-of-molecular-genomics-dmg/molecular-mechanisms-of-pre-mrna-splicing/
Group	pre-mRNA splicing

Project

Computational systems biology

Project Title:

Splicing Systems Biology

Keywords:

spliceosome, genome, intron, exon, splicing

Summary:

The information content of genomes can be greatly expanded by pre-mRNA splicing. Virtually all human pre-mRNAs need to be spliced to become mRNAs. Furthermore, most pre-mRNAs can be spliced into different mRNAs by alternative splicing, which thus constitutes an additional layer of regulation in gene expression. Pre-mRNA splicing is the task of the spliceosome, a ribonucleoprotein complex larger than the ribo-some, and itself the target of a large number of regulatory factors that will shape its splicing output. In addition, as splicing takes place co-transcriptionally, the rate of pre-mRNA synthesis impacts its processing as well. We propose a combination of Bioinformatics and "wet lab" approaches to reveal novel splicing strategies that may contribute to the function of genomes. Using RNASeq data from multiple sources and evolution data we will address a number of questions, including (but not restricted to) (1) why some introns have specific properties like secondary structures in specific locations; (2) what may be the consequences for the splicing of a transcript from a sequence that includes a downstream start of transcription, or a transcription terminator in either strand; and (3) can we find common features in sets of related pre-mRNAs?

Expected skills::

In addition to a good Bioinformatics background, an ability to work independently and to integrate diverse data (mutational analyses, epigenetics, transcriptomics, and others) is desirable. Knowledge ofpython, bash, and R will be helpful.

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2021-2022

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 analysis across individuals

Keywords:

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

Summary:

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 and splicing profiles 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 the 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 explore human single-cell RNA-sequencing datasets to explore the impact of phenotypes such as aging, smoking and gender on gene expression and cell type composition in blood. Second, we will use cell type deconvolution methods to map single-cell signatures to bulk expression data from individuals affected by a wide array of different conditions from diabetes to cardiovascular diseases, to get a deeper understanding of the disease aetiologies and discriminate between gene expression changes due to differences in cell type proportions, differences in gene expression levels or combinations of both.. Overall, in this project we will explore in depth the role of gene expression and cell type composition in determining why human individuals differ from one another in the context of health and disease. What you will learn: Development of computational pipelines to analyse and interpret large datasets, especially from single-cell RNA-seq, and bulk RNA-sequencing. Working in a High Performance Computing environment. Interpretation of multi-omics data. Scientific collaboration in the context of international consortia, effective communication of research findings in internal and external meetings, scientific writing, and critical thinking. Also, the master student will join the Melé lab journal clubs, lab meetings and lab lunches to talk about science but also have fun and discuss non-science related topics with the group.

References:

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

Expected skills::

Strong programming skills in bash, python, R, perl, or similar. 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. Experience working in HPC clusters will be appreciated. Experience with Next Generation Sequencing data analysis will be appreciated. Availability to start in July 2020 is preferred.

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Anaïs Baudot
Email	anais.baudot@univ-amu.fr
Institution	Marseille Medical Genetics
Website	https://www.marseille-medical-genetics.org/a-baudot/
Group	Network and Systems Biology for Diseases

Project

Computational systems biology

Project Title:

Defining the molecular landscape of premature aging disease through biological networks

Keywords:

Networks, systems biology, graph theory, clustering, premature aging disease.

Summary:

Premature aging (PA) syndromes, also called Progeroid syndromes, are a group of rare genetic disorders that phenotypically recapitulate some of the aspects of physiological aging at an early age. These syndromes are clinically and genetically heterogeneous (Navarro et al., 2006). They are usually monogenic, i.e., caused by mutations in single genes, but can affect few or many tissues, different loci can lead to similar diseases, and, contrarily, the phenotypes and severity can vary considerably across individuals carrying the same mutations. Genes and proteins do not act isolated in cells but rather interact with each other to perform their functions in molecular complexes, pathways, and other biological processes. Mutations in genes and proteins will thereby affect their interactions and consequently the biological processes in which they are involved (Zhong et al., 2009). Diseases hence arise from network perturbations, and studying the complex biological networks in which genes and proteins participate is a first step towards better understanding the genotype to phenotype relationships in diseases (Schadt, 2009). Biological interaction data are accumulating since the development of experimental techniques allowing their identification on a large-scale. These interactions are usually represented as large networks in which the nodes correspond to the genes or proteins, and the edges represent their physical or functional interactions. Biological networks are usually organized in communities, i.e. structured around groups of nodes more densely connected with each other than with the rest of the network. These groups of tightly connected nodes, usually called modules, contain genes/proteins likely to be involved in the same cellular functions or processes in cells (Hartwell et al., 1999). The accurate extraction of these modules is promising in biomedicine because studying the modules in which the mutated genes/proteins are involved can reveal the cellular and molecular mechanisms underlying diseases (Furlong, 2013). We propose a project in which we aim at systematically identifying the modules associated with the different PA diseases and their causative genes. We hypothesized that those modules would i) reveal the biological processes perturbed in these diseases, but also ii) define a comprehensive landscape of biological processes perturbed in PA disorders. These results may provide a better understanding of the disease molecular mechanisms and reveal their links with physiological aging processes. During the internship, the student will get familiar with biological networks and graph theory. He/She will apply different methods from this field, such as random walks or community identification. He/She will also get his/her hands on the different types of biological networks, which can better describe the complexity of biological systems, like multiplex or heterogeneous networks. Moreover, the selected candidate will apply the acquired knowledge to a concrete biological question: obtaining new insights about premature aging diseases. Therefore, the student will learn about these syndromes and their potential links with physiological aging. Our group has extensive experience in the development of computational methods to extract the knowledge contained in biological networks (Didier et al., 2018; Novoa-del-Toro et al., 2020; Valdeolivas et al., 2019). We are based in the Faculty of Medicine, located next to the University Hospital of Marseille, La Timone. We work in close collaboration with medical doctors and this hospital which is a European reference in the treatment of rare disease and, in particular, of premature aging diseases. We therefore believe that the selected candidate will have a suitable environment to develop the proposed

project. The selected candidate will work in close collaboration with Ozan Ozisik (PostDoc in AB team) and Alberto Valdeolivas (Roche, Switzerland).

References:

Didier, G., Valdeolivas, A., & Baudot, A. (2018). Identifying communities from multiplex biological networks by randomized optimization of modularity. F1000Research, 7, 1042. Furlong, L. I. (2013). Human diseases through the lens of network biology. Trends in Genetics: TIG, 29(3), 150–159. Hartwell, L. H., Hopfield, J. J., Leibler, S., & Murray, A. W. (1999). From molecular to modular cell biology. Nature, 402(6761 Suppl), C47–C52. Navarro, C. L., Cau, P., & Lévy, N. (2006). Molecular bases of progeroid syndromes. Human Molecular Genetics, 15 Spec No 2, R151–R161. Novoa-del-Toro, E.-M., Mezura-Montes, E., Vignes, M., Magdinier, F., Tichit, L., & Baudot, A. (2020). A Multi-Objective Genetic Algorithm to Find Active Modules in Multiplex Biological Networks. In Cold Spring Harbor Laboratory (p. 2020.05.25.114215). https://doi.org/10.1101/2020.05.25.114215 Schadt, E. E. (2009). Molecular networks as sensors and drivers of common human diseases. Nature, 461(7261), 218–223. Valdeolivas, A., Tichit, L., Navarro, C., Perrin, S., Odelin, G., Levy, N., Cau, P., Remy, E., & Baudot, A. (2019). Random walk with restart on multiplex and heterogeneous biological networks. Bioinformatics , 35(3), 497–505. Zhong, Q., Simonis, N., Li, Q.-R., Charloteaux, B., Heuze, F., Klitgord, N., Tam, S., Yu, H., Venkatesan, K., Mou, D., Swearingen, V., Yildirim, M. A., Yan, H., Dricot, A., Szeto, D., Lin, C., Hao, T., Fan, C., Milstein, S., … Vidal, M. (2009). Edgetic perturbation models of human inherited disorders. Molecular Systems Biology, 5(1), 321.

Expected skills::

Programming skills in Python or R are necessary. Familiarity with network science and/or omics data analysis is a plus. Fluency in English is required. The candidate should have a background in one of the following domains: Biology, Medicine, Biochemistry, Biotechnology, Pharmacy, Veterinary studies, engineering studies, Chemistry, Physics or Mathematics or related degrees, complemented by basic knowledge in the other domains.

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Alfonso Valencia, Davide Cirillo & Jon Sánchez
Email	alfonso.valencia@bsc.es; davide.cirillo@bsc.es; jon.sanchez@bsc.es
Institution	Barcelona Supercomputing Center
Website	http://life.bsc.es/compbio
Group	Computational Biology

Project

Computational systems biology

Project Title:

Generation of age- and gender-dependent temporal disease trajectories from the Catalan population

Keywords:

Comorbidity, Electronic Health Records, personalized medicine

Summary:

The candidate will join Alfonso Valencia's lab (Computational Biology) within the Life Sciences Department at the Barcelona Supercomputing Center. The project will focus on the extraction of age- and gender-dependent temporal disease trajectories from primary care data from the Catalan population. Comorbidity is defined as a higher-than-expected risk of developing a secondary condition in patients already suffering from previous disorders. In particular, aging plays a major role in the emergence of comorbid chronic diseases. Indeed, as a consequence of ageing, chronic diseases tend to accumulate, affecting life quality, increasing disability and hindering the choice of the proper treatment [1]. Historically, comorbidity relations have been analysed by conducting case-control studies on an assembled cohort followed for a period of time (variable depending on the relationships to be analysed). The creation and accumulation of Electronic Health Records, EHRs (longitudinal collections of electronic health information including disease diagnoses and laboratory test results among others), enables the systematic analysis of comorbidity relations. Tens of studies have analysed comorbidity relations between diseases using EHRs. Hidalgo et al. generated a static view of disease comorbidity relationships analysing insurance claims from a population of elderly citizens in the USA [2]. They observed different disease associations depending on patients' age and ethnicity and identified that more connected diseases tend to be associated with an increased mortality. Jensen et al. went one-step further and, using the EHR registry of the population of Denmark, composed by more than 6.2 million patients, studied the temporal association between diseases [3]. They identified disease pairs that tend to co-occur more than the expected by chance presenting a significant temporal association (disease A precedes disease B more times than the other way around). By concatenating significantly associated disease diagnoses, they identified sequences of diseases followed by hundreds of patients, known as disease trajectories. By means of these analyses, the authors evaluated the risk of developing secondary diseases depending on the sequence of previous diagnoses, identifying key disorders that are central to disease progression, such as chronic obstructive pulmonary disease. Following Jensen's study, Westergaard et al. analysed the effect of patients' gender in the risk of developing secondary conditions, generating gender-specific temporal disease trajectories [4]. Since diseases' age of onset varies in each disease, we expect that comorbidity relations and associated trajectories could be dependent on patients' age. As an example, young type 2 diabetes patients tend to develop depression and schizophrenia as comorbid conditions, while older ones tend to suffer from cardiovascular diseases [5]. To date, no study has evaluated on a systemic way the effect of ageing on temporal disease trajectories. The Information System for Research in Primary Care (SIDIAP) contains disease diagnoses, coded using the International Code of Diseases, 10th version, conducted in primary care centres covering the 75% of the Catalan population (more than 5.5 million patients) from 2008 to 2018. This project aims at identifying gender- and age-dependent comorbidity relationships and disease trajectories in the Catalan population. Making use of the SIDIAP data, the student will first generate temporal disease trajectories, following a case-control approach, that will be compared to the ones generated by Søren Brunak's lab [3,4] in the Denmark population. Secondly, the role of gender and ageing on comorbidity relations will be evaluated stratifying patients based on their gender and age and generating gender- and age-dependent temporal disease trajectories.

References:

1. Barnett K. et al. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet. 380: 37-43. (2012) 2. Hidalgo, C. A. et al. A Dynamic Network Approach for the Study of Human Phenotypes. PLoS Comput. Biol. 5, e1000353 (2009) 3. Jensen, A. B. et al. Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. Nat. Commun. 5, 4022 (2014) 4. Westergaard D. et al. Population-wide analysis of differences in disease progression patterns in men and women. Nat. Commun. 10, 666 (2019) 5. Klimek P et al. Quantification of Diabetes Comorbidity Risks across Life Using Nation-Wide Big Claims Data. PLoS Comput. Biol. 11(4): e1004125 (2015).

Expected skills::

Background in biomedicine // Basic computational skills (R, python and bash programming) // Ability to analyse and interpret large datasets // Ability to access and evaluate scientific literature

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Comments:

Please send any future communications to supervisors and Alba Jené.