Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Alexandre Perera Lluna
Email	teresa.tarrago@exheus.com
Institution	Exheus SL
Website	https://exheus.com/
Group	Bioinformatics Department

Project

Computational genomics

Project Title:

Development and validation of predictive models for the analysis of transcriptomic data based on artificial intelligence.

Keywords:

transcriptomics, artificial intelligence, RNASeq, deep learning models, health ,

Summary:

The activity of Exheus S.L. focuses on analysing gene expression in blood samples by using Next Generation Sequencing (NGS) technologies together with artificial intelligence algorithms to produce reports on the health status of its users based on the expression levels of their genes. The information provided is actionable so that users can make changes in their eating habits, physical exercise, etc. and thus optimise their quality of life. In a first stage of the company's development, the methodology has been applied to the world of sport. At Exheus, we have a deep understanding of the effects of sport on gene expression, but now we intend to pursue new lines of research and identify more biomarker profiles with which to characterise other problems of the human organism such as the ageing process, cardiovascular diseases or metabolic diseases. The main objective of this project is to use information provided by the massive next-generation sequencing technology RNAseq to create a definition of normalised gene expression in a population, develop an artificial intelligence model with supervised learning and thus be able to characterise the differences that may pose a risk to health at an individual level, with the study of the deviation of that sample with respect to the rest of the population. To carry out this task, the starting point is to select the numerical method with which to quantify the differences at a statistical level, establish a normal population parameter, then build and train an artificial intelligence algorithm to differentiate between the different characteristic profiles of each case study and finally evaluate the effectiveness of the model to predict the particularities of a single sample. In other words, the big difference in the data processing proposed by Exheus, with respect to current solutions, is the generation of algorithms that group the data to analyse "what is normal" in an individual, and differentiate it from "what is not normal" to provide real-time information on the changes that occur in any organism.

References:

1- PLoS One. 2017 Oct 13;12(10):e0180322. doi: 10.1371/journal.pone.0180322. 2- Sergi Picart-Armada, Wesley K. Thompson, Alfonso Buil, Alexandre Perera-LLuna, The effect of statistical normalisation on network propagation scores (2020), Bioinformatics; btaa896,[Link] 3- Marín-Llaó J, Mubeen S, Perera-Lluna A, Hofmann-Apitius M, Picart-Armada S, Domingo-Fernández D. MultiPaths: a python framework for analyzing multi-layer biological networks using diffusion algorithms. Bioinformatics. 2020 Dec 26:btaa1069. doi: 10.1093/bioinformatics/btaa1069. Epub ahead of print. PMID: 33367476. 4- A Lopez-del Rio, M Martin, A Perera-Lluna, R Saidi, Effect of sequence padding on the performance of deep learning models in archaeal protein functional prediction (2020), Scientific Reports 10 (1), 1-14 5- Elizabeth Carolina Jiménez, Claudia Avella-Garcia, James Kustow, Sally Cubbin, Montse Corrales, Vanessa Richarte, Flavia Lorena Esposito, Imanol Morata, Alexandre Perera, Paloma Varela, Jose Cañete, Stephen V Faraone, Hans Supèr, Josep Antoni Ramos-Quiroga, Eye vergence responses during an attention task in adults with ADHD and clinical controls, Journal of attention disorders (2020), 1087054719897806 Sergio Picart-Armada, Steven J Barrett, David R Willé, Alexandre Perera-Lluna, Alex Gutteridge, Benoit H Dessailly, Benchmarking network propagation methods for disease gene identification (2019), PLoS computational biology 15 (9), e1007276 6- S Kanaan-Izquierdo, A Ziyatdinov, MA Burgueño, A Perera-Lluna, Multiview: a software package for multiview pattern recognition methods (2019), Bioinformatics 35 (16), 2877-2879 7- Josep Lupón, Giovana Gavidia-Bovadilla, Elena Ferrer, Marta de Antonio, Alexandre Perera-Lluna, Jorge

López-Ayerbe, Mar Domingo, Julio Núñez, Elisabet Zamora, Pedro Moliner, Evelyn Santiago-Vacas, Javier Santesmases, Antoni Bayés-Genis (2019) Heart Failure With Preserved Ejection Fraction Infrequently Evolves Toward a Reduced Phenotype in Long-Term Survivors, Circulation: Heart Failure 12 (3), e005652

Expected skills::

Good organizational skills. • Team player. • Drive and determination. • Desire to learn. • Enjoy solving problems; engaged and motivated. • Ability to communicate effective both verbally and in writing.Strong skills in data visualization tools. • Technical proficiency, scientific creativity, collaboration with others and independent thought. Good level of statistical programming (R, Python, SQL and others). managing, processing, performing quality control and analysis of large amounts of Next Generation Sequencing data. Fluent in English.

Possibility of funding::

Yes

Possible continuity with PhD: :

Yes

Comments:

We are looking for a talented Bioinformatics student to join our growing team of Data Scientists and to grow with us.