Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

# Master project 2021-2022

## Personal Information


| | |
|---|---|
| **Supervisor** | Luis A Pérez Jurado & Carlos Ruiz Arenas |
| **Email** | luis.perez@upf.edu / carlos.ruiza@upf.edu |
| **Institution** | UPF |
| **Website** | |
| **Group** | Genetics Unit |


## Project

# Computational genomics

**Project Title:**

Combination of common and rare genetic variants to improve the diagnosis of complex disease


**Keywords:**

Human genetics Polygenic risk score Pathogenic variants Complex diseases



**Summary:**

Complex diseases, such as obesity or autism, are caused in most cases by a combination of environmental and genetic factors. Genetic factors can be classified as pathogenic or disease susceptibility variants, depending on the strength of their association with the phenotype. On one hand, pathogenic variants are genetic variants enough to cause the disease with high penetrance. These variants are normally ultra rare (<0.1%), preventing the application of association studies and usually requiring family studies and/or other functional studies for their validation. On the other hand, disease susceptibility variants increase the risk to suffer the disease but they are not enough by themselves to cause it, requiring additive effects of other concurrent genetic or environmental factors. These susceptibility variants tend to be relatively common in the population (from 0.1% to 50%) and association studies, comparing their frequency in cases and controls, can be used to define and quantify their relation to disease. The results of these association studies can be collapsed in Polygenic Risk Scores (PRS). PRS are a measure of the risk alleles for a disease carried by an individual. Thanks to the availability of big public datasets, PRS have improved their performance and can identify individuals with high susceptibility to disease (Khera et al 2018). In recent years, both approaches have been independently applied to study complex diseases. For instance, the genetic heritability of autism has been estimated to be 3-10% due to de novo rare variants, 3-10% to inherited rare variants and around 50% due to common variants (Alonso-Gonzalez et al, 2018). Despite this success, a significant proportion of heritability is still missing. We hypothesize that missing heritability is mainly due to rare variants with low penetrance, i.e. variants that are only pathogenic in a specific genetic background (Fahed et al, 2020). In this project, we propose to combine the analysis of common and ultra rare variants to improve our understanding of some common diseases. We propose three main tasks: - Compare PRS between controls and cases with and without high penetrant variants - Prioritize variants considering PRS - Propose new candidate genes We will use autism as an example of a complex disease and we will apply our methods to data from public repositories (dbGAP, SFARI, UK Biobank) and internal data. The applicant who will work in this project will learn to perform variant calling, prioritize genetic variants, define and compute polygenic risk scores (PRS), work with public data, develop analysis pipelines and work with software containers.


**References:**


Alonso-Gonzalez A, Rodriguez-Fontenla C, Carracedo A. De novo Mutations (DNMs) in Autism Spectrum Disorder (ASD): Pathway and Network Analysis. Front Genet. 2018 Sep 21;9:406. doi: 10.3389/fgene.2018.00406. PMID: 30298087; PMCID: PMC6160549. Fahed AC, Wang M, Homburger JR, Patel AP, Bick AG, Neben CL, Lai C, Brockman D, Philippakis A, Ellinor PT, Cassa CA, Lebo M, Ng K, Lander ES, Zhou AY, Kathiresan S, Khera AV. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. Nat Commun. 2020 Aug 20;11(1):3635. doi: 10.1038/s41467-020-17374-3. PMID: 32820175; PMCID: PMC7441381. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, Kathiresan S. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018 Sep;50(9):1219-1224. doi: 10.1038/s41588-018-0183-z. Epub 2018 Aug 13. PMID: 30104762; PMCID: PMC6128408. Weiner DJ, Wigdor EM, Ripke S, Walters RK, Kosmicki JA, Grove J, Samocha KE,

Goldstein JI, Okbay A, Bybjerg-Grauholm J, Werge T, Hougaard DM, Taylor J; iPSYCH-Broad Autism Group; Psychiatric Genomics Consortium Autism Group, Skuse D, Devlin B, Anney R, Sanders SJ, Bishop S, Mortensen PB, Børglum AD, Smith GD, Daly MJ, Robinson EB. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. Nat Genet. 2017 Jul;49(7):978-985. doi: 10.1038/ng.3863. Epub 2017 May 15. PMID: 28504703; PMCID: PMC5552240.

**Expected skills::**

Good level of bash and R scripting and a good background in human genetics.

**Possibility of funding::**

No

**Possible continuity with PhD: :**

To be discussed

**Comments:**

Lab experiments to confirm the project results might be considered.