Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

# Master project 2021-2022

## Personal Information


| | |
|---|---|
| **Supervisor** | Olivia Belbin |
| **Email** | obelbin@santpau.cat |
| **Institution** | IIB-SantPau |
| **Website** | www.santpaumemoryunit.com |
| **Group** | Neurobiology of Dementias |


## Project

# Computational genomics

**Project Title:**

A multimodal study to determine the polygenic risk for cognitive, structural and functional brain changes in Alzheimer's disease patients.


**Keywords:**

Alzheimer's disease, synapse, polygenic risk, neuroimaging, proteomics


**Summary:**


Despite significant efforts over the last decade, there is still substantial 'missing heritability' for late-onset Alzheimer's disease (LOAD). As synapse loss is an early event in LOAD, we hypothesized that synapse-encoding genes will be enriched for LOAD risk-modifying loci that could account for this missing heritability. To test this hypothesis, we first characterized the synaptic proteome by combining data from proteomic studies of synaptic fractions isolated from mouse, rat and human brain tissue with gene ontologies from public databases. The resulting synaptic proteome comprised 537 proteins with a known synaptic function and that are expressed at the synapse (Lleó et al 2019). To construct the "synaptic PRS" model, we extracted summary statistics from the International Genetics of Alzheimer's Project genome-wide association meta-analysis of 74,046 patients for the 2,993 single nucleotide polymorphisms (SNPs) that reside within the 537 synapse-encoding genes. Synaptic PRS using these SNPs were calculated using PRSice-2 software (Choi and O'Reilly, 2019, Euesden, et al., 2015) as previously described (Chaudhury, et al., 2019, Lawingco et al., 2020). An unbiased threshold for p-values in the genome-wide meta-analysis was used to prioritize SNPs that gave the best fitting PRS model (highest Nagelkerke r2 value) and the β-statistic was used to generate weighting estimates for each SNP. The optimal model ("synaptic PRS") was tested in 2 independent data sets of controls and pathologically confirmed LOAD. The mean Synaptic PRS was 2.3-fold higher in LOAD compared to controls ($p<0.0001$) with a predictive accuracy of 72% in the target dataset (n=439) and 73% in the validation dataset (n=136), a 5-6% improvement compared to the current best known LOAD risk factor, APOE ($p<0.00001$) and a 3% improvement compared to an unrestricted model. The synaptic model comprises 8 variants from 4 previously identified (BIN1, PTK2B, PICALM, APOE) and 2 novel (DLG2, MINK1) LOAD loci involved in glutamate signaling ($p = 0.01$) or APP catabolism or tau binding ($p = 0.005$). As the simplest PRS model with good predictive accuracy to predict LOAD, the synaptic PRS could be used to identify individuals at risk of LOAD before symptom onset. These data are published in Lawingco et al 2020. In the proposed project, the candidate will generate the synaptic PRS for 360 clinically diagnosed AD patients from the SPIN cohort (Alcolea et al 2019 Alzheimers Dement (N Y)) using PRSice-2 software, implemented in R, and will perform a multimodal study to determine the association of the synaptic PRS with baseline and longitudinal change in cognitive performance, brain atrophy (structural magnetic resonance imaging), brain glucose metabolism (Positron Emission Tomography) as well as cerebrospinal fluid markers of LOAD pathology (Aβ1-42, AB42:40, t-tau, p-tau) and axonal (NF-L) and synapse (VAMP-2) degeneration in the same individuals. Using R, the candidate will assess whether the predictive capacity of the synaptic PRS can be improved by inclusion of SNP*SNP statistical interaction terms into the model. The candidate will use the same methodology to generate and compare synaptic PRS models for other neurodegenerative diseases such as frontotemporal lobar degeneration related syndromes and Lewy body dementias. Finally, there is scope for the candidate to integrate the PRS with proteomic data from cerebrospinal fluid and synaptic fractions to further elucidate the functional basis of the PRS variants.


**References:**


Lleó, A., Núñez-Llaves, R., Alcolea, D., Chiva, C., Balateu-Paños, D., Colom-Cadena, M., Gomez-Giro, G., Muñoz, L., Querol-Vilaseca, M., Pegueroles, J., Rami, L., Lladó, A., Molinuevo, J., Tainta, M., Clarimón, J., Spires-Jones, T., Blesa, R., Fortea, J., Martínez-Lage, P., Sánchez-Valle, R., Sabidó, E., Bayés, À., Belbin, O. 2019.

Changes in synaptic proteins precede neurodegeneration markers in preclinical Alzheimer's disease cerebrospinal fluid. Mol Cell Proteomics. 2019 Mar;18(3):546-560. Choi, S.W., O'Reilly, P.F. 2019. PRSice-2: Polygenic Risk Score software for biobank-scale data. Gigascience 8(7). Euesden, J., Lewis, C.M., O'Reilly, P.F. 2015. PRSice: Polygenic Risk Score software. Bioinformatics 31(9), 1466-8. Chaudhury, S., Brookes, K.J., Patel, T., Fallows, A., Guetta-Baranes, T., Turton, J.C., Guerreiro, R., Bras, J., Hardy, J., Francis, P.T., Croucher, R., Holmes, C., Morgan, K., Thomas, A.J. 2019. Alzheimer's disease polygenic risk score as a predictor of conversion from mild-cognitive impairment. Transl Psychiatry 9(1), 154. T Lawingco, S Chaudhury, KJ Brookes, T Guetta-Baranes, R Guerreiro, J Bras, J Hardy, P Francis, A Thomas, O Belbin, K Morgan. Genetic variants in glutamate-, Aβ-, and tau-related pathways determine polygenic risk for Alzheimer's disease. Neurobiol Aging. 2020 Nov 12:S0197-4580(20)30391-2. D Alcolea, J Clarimón, M Carmona-Iragui, I Illán-Gala, E Morenas-Rodríguez, I Barroeta, R ribosa-Nogué, I Sala, M.B Sánchez-Saudinós, L Videla, A Subirana, B Benejam, S Valldeneu, S Fernández, T Estellés, M Altuna, M Santos-Santos L García-Losada, A Bejanin, J Pegueroles, V Montal, E Vilaplana, O Belbin, O Dols-Icardo, S Sirisi, M Querol-Vilaseca, L Cervera-Carles, L Muñoz, R Núñez, S Torres, M. Valle Camacho, I Carrió, S Giménez, C Delaby, R Rojas-Garcia, J Turon-Sans, J Pagonabarraga, A Jiménez, R Blesa, J Fortea, A Lleó. The Sant Pau Initiative on Neurodegeneration (SPIN) cohort: A data set for biomarker discovery and validation in neurodegenerative disorders. Alzheimer Dement (NY) 5 (2019) 597-609.

**Expected skills::**

Basic R programming skills, conceptual understanding of genome-wide association studies and polygenic risk scores, regression modelling (glm, nlme, rlr), correlative analyses and basic statistics. Prior knowledge of Alzheimer's disease genetics/biology would be highly valued.

**Possibility of funding::**

No

**Possible continuity with PhD: :**

To be discussed