**Research Project Proposal**
Academic year 2020-2021

***Máster en Métodos Computacionales en Ciencias***

**Project Nº 46**

**Title:** Deep learning to elucidate a key mechanism to overcome cellular stress.

**Department/ Laboratory** *Department of Gene Therapy and Regulation of Gene Expression. Laboratory of Tomás Aragón: The response to protein misfolding in neurodegenerative diseases.*

**Director 1** *Tomás Aragón Amonarriz*
**Contact:** *taragon@unav.es*
**Codirector:** Mikel Hernáez Arrazola
**Contact:** *mhernaez@unav.es*

**Summary**

The construction and training of deep neural networks has been enabled by the explosion of data, algorithmic advances and substantial increases in computational capacity that now can be applied to solve essential biological processes. Here, we use deep learning technologies to elucidate the key regulatory step of a mechanism by which our cells overcome stress.

The capacity of our cells to respond to endoplasmic reticulum (ER) stress relies on a unique, non-canonical splicing reaction, where a stress-regulated protein, IRE1, cleaves the intron of a unique mRNA, that encodes the transcription factor XBP1; splicing of XBP1 mRNA and the expression of XBP1s, a potent transcription factor needed to overcome stress and plays a key role in a wide variety of diseases. The high specificity, efficiency and biomedical relevance of this reaction is based on a key-lock interaction between IRE1 and XBP1 RNA. But, how is this high specificity achieved?

To answer this question, we will use the sequence of hundreds of IRE1 and XBP1 homologs from a wide set of animal, plants and fungal species. Through evolution, the aminoacid sequence of IRE1 has changed, as well as the nucleotide sequence of the cleavage sites in XBP1 intron. By making a matrix of these variations, we will build degenerative models into a Deep learning algorithm to predict which string of aminoacids encodes the capacity of IRE1 to recognize and cleave XBP1 RNA. In a broader perspective, we hope that Deep Learning will help us to understand the “Rosetta Stone” code that makes RNAses specific.