Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

# Master project 2021-2022

## Personal Information


| | |
|---|---|
| **Supervisor** | Mar Albà |
| **Email** | mar.alba@upf.edu |
| **Institution** | Hospital del Mar Medical Research Institute (IMIM) |
| **Website** | evolutionarygenomics.imim.es |
| **Group** | Evolutionary Genomics |


## Project

# Computational genomics

**Project Title:**

Using Nanopore RNA sequencing to explore the transcriptome


**Keywords:**

Transcriptomics; transcript discovery; long reads; Nanopore; gene expression.



**Summary:**

The field of transcriptomics is rapidly advancing thanks to the development of high throughput RNA sequencing techniques. However, most studies are based on Illumina short reads, which limits our understanding of the transcriptome. Short reads can often be mapped to different gene isoforms, which generates ambiguous results, and the use of a fixe set of gene annotations prevents detecting transcripts that are specific of a given individual or disease. Long read sequencing tecnologies, such as Nanopore, can generate reads that correspond to the whole transcript, and thus can solve many of the problems currently associated with the use of short reads. For example, Nanopore reads have recently been employ to unravel the complexity of the transcriptome of SARS-CoV-2 (Kim et al., 2020) or to identify alternative splicing events in cancer (Tang et al., 2020). Whereas several programs are available for transcriptome reconstruction using long reads, we have recently developed the first method that does not need a reference genome and which can identify transcripts in genomes that have undergone structural rearrangements (de la Rubia et al., 2020). The aim of the project will be to investigate the complexity of the transcriptome using Nanopore RNA sequencing reads and compare the results to those obtained Illumina reads. We have already generated the sequencing data for 4 different yeast species, including S. pombe, which contains about 1,000 genes with multiple introns. The project provides an unique opportunity to explore the power of long read tecnologies for novel transcript discovery and quantification.


**References:**


de la Rubia, I., Indi, J.A., Carbonell, S., Lagarde, J., Albà, M.M., Eyras, E. (2020). Reference-free reconstruction and quantification of transcriptomes from long-read sequencing. bioRxiv, https://doi.org/10.1101/2020.02.08.939942 Kim, D., Lee, J-Y., Yang, J-S., Kim, J.W., Kim, V.N., Chang H. (2020). The architecture of SARS-CoV-2 transcriptome. Cell 181, 914–921. Tang , A.D., Soulette, C.M., van Baren, MJ., Hart, K.,.. , Brooks, A.M. (2020). Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. Nature Communications 11(1):1438.


**Expected skills::**

Interest in transcriptomics and long read technologies; basic knowledge of a programming language; basic knowledge of R; good level of English.

**Possibility of funding::**

Yes

**Possible continuity with PhD: :**

To be discussed