

Master project 2020-2021

Personal Information

Supervisor	Mar Albà
Email	mar.alba@upf.edu
Institution	IMIM-UPF
Website	evolutionarygenomics.imim.es
Group	Evolutionary Genomics, Research Programme on Biomedical Informatics

Project

Computational genomics

Project Title:

Building transcriptomes with Nanopore sequencing data

Keywords:

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

Summary:

Long read sequencing techniques, such as Oxford Nanopore Technologies (ONT), have great potential to sequence complex transcriptomes and to discover new transcripts beyond the annotated ones. While methods that work with Illumina short reads are quite mature, the development of software to work with Nanopore RNA sequencing reads is a very active area of research. We and co-workers have recently developed a method to build transcriptomes from RNA-derived Nanopore reads (cDNA and direct RNA) that does not require a reference genome and that could be used to investigate highly rearranged genomes (such as those in cancer cells) or species that currently lack a sequenced genome (de la Rubia et al., 2020). The aim of the project will be to compare methods based on Nanopore or Illumina reads for building eukaryotic transcriptomes in the absence of a reference genome and to identify novel, non-annotated, transcripts in species that already have a genome and reference annotations. We would like to determine when it is more convenient to use one sequencing technology over the other one, and if the combination of the two technologies - using Illumina reads to correct errors in Nanopore reads - is a real advantage. For this we will use already available datasets for yeast, human and mouse species, as well as datasets that are currently being generated in the group.

References:

de la Rubia, I., Indi, J.A., Carbonell, S., Lagarde, J., Albà, M.M., Eyra, E. (2020). Reference-free reconstruction and quantification of transcriptomes from long-read sequencing. bioRxiv, <https://doi.org/10.1101/2020.02.08.939942>

Expected skills:

Interest in computational genomics and transcriptomics; knowledge of a programming language; knowledge of R; good command of English.

Possibility of funding:

Yes

Possible continuity with PhD: :

To be discussed
