Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Eva Maria Novoa
Email	eva.novoa@crg.eu
Institution	Center for Genomic Regulation (CRG)
Website	https://www.crg.eu/en/programmes-groups/novoa-lab
Group	Epitranscriptomics and RNA Dynamics

Project

Computational systems biology

Project Title:

PREDICTING AND MONITORING CANCER USING MACHINE LEARNING ON NANOPORE SEQUENCING DATA

Keywords:

nanopore sequencing, RNA modifications, small RNA, machine learning, deep learning, cancer, prognosis, sample classification

Summary:

Dysregulation of small RNA abundances and their RNA modifications is a well-known feature in cancer cells, which leads to enhanced expression of specific oncogenic transcripts and proteins [1,2]. Despite the well-established association between small RNA dysregulation and cancer progression and malignancy, small RNA abundances and modifications are still not being used as screening, diagnostic or prognostic markers for cancer detection or progression, mainly due to the lack of a simple, unbiased and cost-effective method to quantify small RNA abundances and their modifications. Our laboratory has pioneered the use of direct RNA sequencing for the detection and quantification of RNA abundances and their modifications, including both development of improved library preparation protocols as well as the development of novel algorithms to predict and quantify RNA modifications [3-5]. Here we propose to use native RNA nanopore sequencing technology to predict the malignancy of biological samples in a high-throughput, rapid, multiplexed and cost-effective manner. Specifically, the candidate MSc student will benefit from a recently developed method in our lab to sequence small RNAs using nanopore sequencing. The candidate will then develop and apply deep learning algorithms to classify small RNA profiles into "normal", "tumoral" and "metastasic". Once the classification model is benchmarked and validated using cell lines, the methodology will then be applied to patient-derived samples.

References:

1. Begik O, Lucas MC, Ramirez JM, Liu H, Mattick JS and Novoa EM#. Integrative analyses of the RNA modification machinery reveal tissue- and cancer-specific signatures. Genome Biology 2020, 21:97. doi: 10.1186/s13059-020-02009-z 2. Gingold et al., A Dual Program for Translation Regulation in cellular profileration and differentiation. Cell 2014, 158(6):1281-1292. 3. Liu H*, Begik O*, Lucas MC, Ramirez JM, Mason CE, Wiener D, Schwartz S, Mattick JS, Smith MA and Novoa EM#. Accurate detection of m6A RNA modifications in native RNA sequences. Nature Comm 2019, 10:4079. doi:10.1038/s41467-019-11713-9 4. Smith MA*, Ersavas T*, Ferguson JM*, Liu J, Lucas MC, Begik O, Bojarski L, Barton K and Novoa EM#. Molecular barcoding of native RNAs using nanopore sequencing and deep learning. Genome Research 2020 30(9): 1345-1353 5. Begik O*, Lucas MC*, Ramirez JM, Milenkovic I, Cruciani C, Vieira HGS, Medina R, Liu H, Sas-Chen A, Mattick JS, Schwartz S and Novoa EM#. Quantitative profiling of native RNA modifications and their dynamics using nanopore sequencing. bioRxiv 2021, 189969 (accepted in Nature Biotechnology)

Expected skills::

python (required), R (required), prior experience with machine learning is a plus but not required, familiarity with third-generation sequencing (e.g. nanopore) is a plus but not required

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Comments:

Option for funding, as well as option for PhD continuity.
