

(Cryptic) splicing project

Assessing patterns associated with cryptic splicing using Nanopore direct RNA datasets

Background:

Splicing of genes is a highly precise process that cleaves out introns marked at the borders by GT and AG dinucleotides. However, recently we have shown that for transgenes, there is some degree of cryptic, non-canonical splicing patterns [1]. Previously, many of these patterns have been rejected as technical noise; however, validation of this 'noise' using direct RNA sequencing in addition to classic TA-cloning has established this noise to be a biological phenomenon.

Project aims:

In this project, the student will explore the underlying patterns of cryptic splicing in endogenous genes using computational approaches.

- Generation of novel dataset of cryptic splicing events by exploring publicly available Nanopore direct RNA datasets
- Identification of the features defining these events, including sequence determinants of cryptic vs frequent splicing
- Assessing potential relationships between cryptic splicing and exon size, RNA expression and/or GC content
- Assessing sequence determinants of cryptic VS frequent splicing.

Additional opportunities:

The student could additionally compare patterns in endogenous exons to splicing patterns in transgenes, by using publicly available datasets such as the SARS-CoV2 S protein datasets from our previous study.

While sequence patterns underlying splicing have previously been explored [2], this will be the first systematic study of the patterns underlying cryptic splicing using Nanopore direct RNA datasets.

We are looking for a student with some experience in python programming and a basic understanding of genetics & molecular biology. Experience working with sequencing data is a plus (alignment).

Responsible supervisors: Ahmed Mahfouz (a.mahfouz@tudelft.nl) and Kärt Tomberg (k.tomberg@lumc.nl)

References:

[1] Tomberg, K., Antunes, L., Pan, Y., Hepkema, J., Garyfallos, D. A., Mahfouz, A., & Bradley, A. (2021). Intronization enhances expression of S-protein and other transgenes challenged by cryptic splicing. *bioRxiv*.

[2] Jaganathan, K., Panagiotopoulou, S. K., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., ... & Farh, K. K. H. (2019). Predicting splicing from primary sequence with deep learning. *Cell*, 176(3), 535-548.