Post-doc Project – Aix Marseille University
Combined deep learning and synthetic-based approaches to unravel the genetic determinants of enhancer versus promoter activity of Epromoters

Lead supervisor: S. Spicuglia, Supervisor 2: B. Ghattas, Supervisor 3: A. González
Contacts: Salvatore.spicuglia@univ-amu.fr, aitor.gonzalez@univ-amu.fr, badih.ghattas@univ-amu.fr.
Application form: https://centuri-livingsystems.org/pdp2022-01/

Abstract

Regulation of gene transcription is accomplished by proximal (promoters) and distal (enhancers) regulatory elements. However, a strict dichotomy model is now challenged and a major question in the field is to define the genetic determinants of the different regulatory activities. The Spicuglia team has previously identified Epromoters as cis regulatory elements with both enhancer and promoter (E/P) activities and is currently using high-throughput approaches to evaluate both activities in thousands of wild-type and mutant DNA sequences. In this project, we will build a sequence-based deep learning model of Epromoters to unravel the genetic determinants of enhancer vs. promoter activities. The model will be challenged and refined in back-and-forth exchanges between model predictions, experimental validation and synthetic generation of Epromoters.

Keyworks

Cis-regulatory elements, genetic variants, machine learning, deep learning, synthetic biology

Objectives

1) To process NGS data from the dual reporter assay.
2) To create a deep learning model of DNA sequences to predict E/P activities.
3) To design de-novo synthetic regulatory sequences with defined E/P activities.
4) To evaluate experimentally E/P activities in synthetic sequences.
5) To analyse the model predictions to infer the logics of E/P activities at the DNA sequence level and assessment of natural genetic variants.

Expected profile

The postdoc candidate should have a PhD in bioinformatics or related fields, with a solid background in computer science, statistics and/or mathematics. The candidate should be interested in "omics" data analyses, genomics and gene regulation. Proven previous experience in manipulating NGS data and/or deep learning and in collaborations with experimental biologists is an advantage.

References

- Bernardo P. de Almeida, Franziska Reiter, Michaela Pagani, Alexander Stark. DeepSTARR predicts enhancer activity from DNA sequence and enables the de novo design of enhancers. doi:10.1101/2021.10.05.463203
- W Kopp, R Monti, A Tamburrini, U Ohler, A Akalin. Deep learning for genomics using Janggu. Nat Commun. 2020 Jul 13;11(1):3488. doi: 10.1038/s41467-020-17155-y.
- Andersson, R., Sandelin, A. Determinants of enhancer and promoter activities of regulatory elements. Nat Rev Genet 21, 71–87 (2020). https://doi.org/10.1038/s41576-019-0173-8
- Medina A, Santiago D, Puthier D, Spicuglia S. (2018) Wide-spread enhancer activity from core promoters. TiBS. 43(6):452-468.
- Core, L., Martins, A., Danko, C. et al. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet 46, 1311–1320 (2014). https://doi.org/10.1038/ng.3142