Instituto de Medicina Molecular João Lobo Antunes

Master Project Proposal

Title: Sequence determinants of transcriptional pausing

Synopsis:
Transcription is the process by which an enzyme called RNA Polymerase II (Pol II) makes an RNA copy of a gene. This copy can later be used to produce a protein. As Pol II slides along the gene, synthesizing the copy, it does not progress smoothly. Instead, it may behave more like a car in rush hour traffic: speeding up, slowing down, and even coming to a stop from time to time. Most famously, Pol II often pauses 20-120 nucleotides downstream of the transcription start site ("promoter-proximal pausing") (Chen et al. 2018). However, frequent pausing is also observed at various locations within the gene itself. These pauses are much less understood, although it is thought that they are linked to overcoming nucleosome barriers and/or to splicing. **Could the locations of pauses be partially encoded in the sequence itself?**

Such an encoding could be direct, with Pol II itself reacting to the nucleotide composition of the sequence. It could also happen indirectly: the sequence could contain short motifs that are recognized and bound by proteins, which then affect the elongation rate (the rate at which Pol II progresses along the gene). Encouragingly, a study in yeast (Cohen et al. 2018) found some evidence for the existence of "a code for transcription elongation speed", whereby the elongation rate of Pol II is (partially) determined by the frequencies of short ~5-nucleotide sequence motifs. It remains unclear, however, whether these results can be generalized to other organisms. The functional importance of any such elongation code is also unexplored.

We have used the Native Elongating Transcript sequencing (NET-seq) method in *Drosophila melanogaster* to determine the locations of elongating polymerases (Nojima et al. 2015). We then used a peak-calling algorithm on this data to determine regions where Pol II is likely pausing. **In this project, the student will test whether there is evidence for specific sequence signals at or near pause sites.** Two complementary approaches are proposed. Firstly, the student will compare the frequencies of short nucleotide motifs within and outside of pause sites. This is to check for motifs that are over-represented at pause sites and could be signals that promote pausing. The second approach is to compare levels of sequence conservation inside and outside of pause regions. If the pausing happens as a result of sequence signals and is functionally important, then the need to preserve these signals should lead to increased sequence conservation.

Supervisor: *Rosina Savisaar, Maria Carmo-Fonseca Lab, rsavisaar@medicina.ulisboa.pt*
Webpage of the group

Bibliography:

Chen FX, Smith ER, Shilatifard A. 2018. Born to run: Control of transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. 19:464–478.
Cohen E, Zafrir Z, Tuller T. 2018. A code for transcription elongation speed. RNA Biol. 15:81–94.
Nojima T, Gomes T, Grosso ARF, Kimura H, Dye MJ, Dhir S, Carmo-Fonseca M, Proudfoot NJ. 2015. Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. Cell 161:526–540.

Remunerated or volunteer training: volunteer