RiboTools: A Galaxy toolbox for qualitative Ribosome Profiling analysis

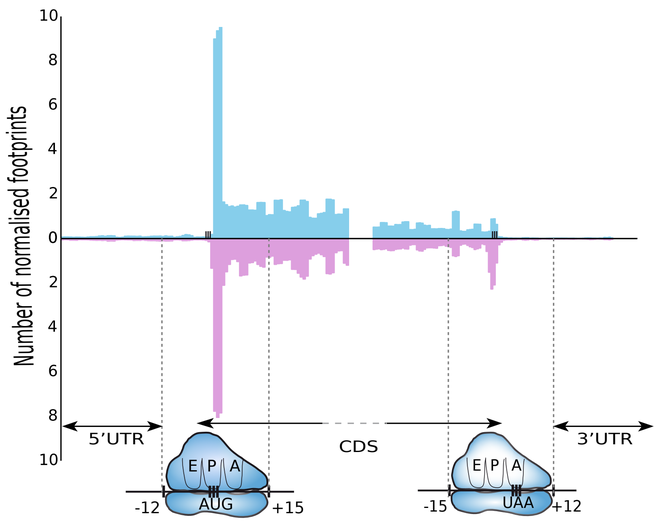
Rachel Legendre, Agnès Baudin-Baillieu, Isabelle Hatin and Olivier Namy

Institute for Integrative Biology of The Cell, UMR CNRS 9198, University Paris-Sud, 91405 Orsay, Cedex, France
Contact: olivier.namy@i2bc.paris-saclay.fr and rachel.legendre@i2bc.paris-saclay.fr

Introduction

Ribosome Profiling (Ribo-seq) provides genome-wide information about translational regulation. It can identify many features affecting **translational quality**, such as ribosomal pauses, dual coding regions, translation start sites at AUG or non-AUG codons, translation of non-coding RNA and translational ambiguities. Here we describe **RiboTools**, an efficient **Galaxy package** to helps manage primary analyses of Ribo-seq data and addressing common issues, such as the identification of translational ambiguities, or of stop codon readthrough events, and the codon occupancy of ribosomal A, P, and E-sites.

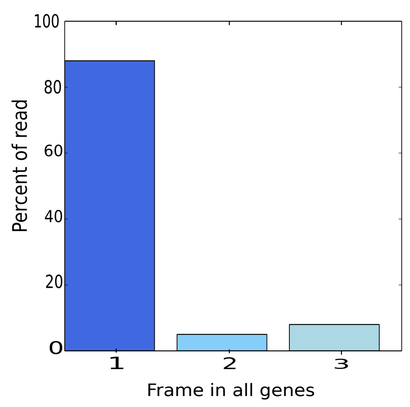
RiboTools requires that **ribosome-protected fragments** (footprints) have already been mapped to a reference genome resulting a sorted BAM file. Feature annotation (GFF3) and genome sequence (FASTA) are required to focus the analysis to the translated parts of genome. Footprints are 28-30 nucleotides long and their 5' end shows a **triplet periodicity**.



Density of coverage of the A-site position by footprints for the whole yeast genome, 50 bases on either side of the start (AUG) and stop (UAA) codons

Kmer tool: choose the best footprints

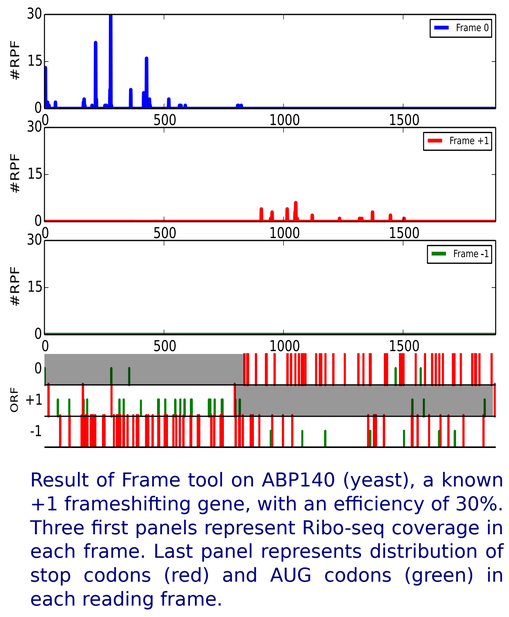
Depending on the organism studied and sample preparation, the footprints may be in any frame. *Kmer* tool calculates the number of footprints in each frame, for each transcript in the GFF3 file.
The *kmer* tool helps to determine:
i) the distribution of footprint lengths
ii) the percentage of footprints in each reading frames for a given length. This is based on the position of the 5' end footprint in the BAM file.



Distribution of 28-mer footprints within the three reading frames in all genes on a yeast [PSI+] strain.

Frame tool: translational ambiguities

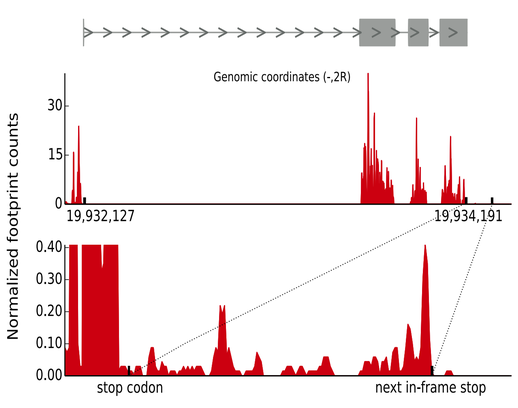
Frame is used to detect genes with translational ambiguities (i.e footprints in different phases). *Frame* uses the footprint size defined by the user with *Kmer*, corresponding to the footprint population with the best triplet periodicity. *Frame* uses these footprints to identify transcripts for which the proportion of out-of-frame footprints is higher than expected (cut-off parameter).



Result of Frame tool on ABP140 (yeast), a known +1 frameshifting gene, with an efficiency of 30%. Three first panels represent Ribo-seq coverage in each frame. Last panel represents distribution of stop codons (red) and AUG codons (green) in each reading frame.

Stop_supp tool: get readthrough events

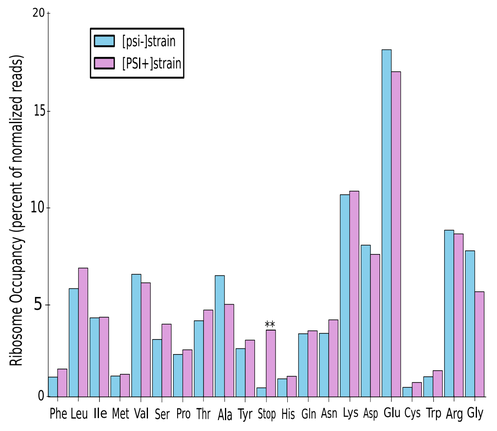
Stop_supp is designed for the detection of genes potentially displaying stop codon readthrough events. This tool searches for footprints in 3' UTR regions. For all transcript, *Stop_supp* looks for footprints after the annotated stop codon and checks whether all the criteria (footprints after the annotated stop and in the next in-frame stop, no methionine in the next five codons and homogeneous coverage of the extension) are respected.



Stop readthrough in twinstar gene on Drosophila melanogaster. Two first panels are global gene organisation and footprint coverage respectively. Last panel displays a zoom to the stop readthrough extension coverage.

Codon tool: get codon occupancy

The *codon_density* tool determines whether codon occupancy differs between two conditions (with a pair of bam file, or more in case of replicas). To identify A, P or E ribosomal sites, the user selects the length of footprints to be used and the position of the A-site (the P and E sites are deduced from the position of the A site). The A-site is generally estimated to be 15 nucleotides away from the 5’ end of the footprints.



Ribosomal A-site occupancy in two isogenic yeast strains. Codons are grouped by amino acid. We can see that stop codons are more frequently found at the A-site in [PSI+] than in [psi-]. This demonstrates that termination process is less efficient in presence of [PSI+].

Conclusion

The methods provided by RiboTools[1] are designed for the accurate analysis of k-mer length distribution, translation ambiguities and translation readthrough events. They evaluate codon occupancy at a specific ribosome site using Ribo-seq data.
All RiboTools require on BAM files from classic mappers and GFF3 annotation (GTF soon accepted). RiboTools could be used on any organism, if a reference genome and its annotation are available. We recently used RiboTools to analyse the translational impact of the [PSI+] prion in yeast[2].
In order to help the complete analysis of Ribo-seq data, we provide a public history for pre-processing, with one million footprints for yeast on the Galaxy main server to use as an example: https://usegalaxy.org/u/rlegendre/p/riboseq.
RiboTools is available from the public Galaxy test ToolShed repository for share it across all Galaxies.
RiboTools is also available on RiboGalaxy, a dedicated Galaxy server to Ribo-seq data.

[1] Legendre, R., Baudin-Baillieu, A., Hatin, I., & Namy, O. (2015). RiboTools: A Galaxy toolbox for qualitative ribosome profiling analysis. **Bioinformatics**, (March), 1–3. doi:10.1093/bioinformatics/btv174
[2] Baudin-Baillieu, A., Legendre, R., Kuchly, C., Hatin, I., Demais, S., Mestdagh, C., Gautheret D., and Namy, O. (2014). Genome-wide Translational Changes Induced by the Prion [PSI(+)]. **Cell Reports**, 8(2)

Acknowledgments

Université Paris Sud · cnrs dépasser les frontières · Agence Nationale de la Recherche ANR · eBIO · Ribo Galaxy · ucc Coláiste na hOllscoile Corcaigh, Éire University College Cork, Ireland