



HAL
open science

Joint transcriptome and translome analysis: a reproducible pipeline

Julie Ripoll, Fati Chen, Céline Mandier, Eric Rivals

► To cite this version:

Julie Ripoll, Fati Chen, Céline Mandier, Eric Rivals. Joint transcriptome and translome analysis: a reproducible pipeline. JOBIM 2023 - 23es Journées Ouvertes en Biologie, Informatique et Mathématiques, Jun 2023, Nice, France. 2023. lirmm-04286339

HAL Id: lirmm-04286339

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-04286339v1>

Submitted on 15 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Joint transcriptome and translome analysis: a reproducible pipeline

Julie RIPOLL¹, Fati CHEN², Céline MANDIER¹ and Eric RIVALS^{1,3}

¹ MAB - LIRMM - Université Montpellier - CNRS, Montpellier, France

² ADVANSE - LIRMM - Université Montpellier - CNRS, Montpellier, France

³ IFB-CORE - Institut Français de Bioinformatique, Montpellier, France

Corresponding Author: julie.ripoll@lirmm.fr

RNA sequencing (RNA-seq) is often used to elucidate the regulation of gene expression, as it provides an in-depth view of transcribed RNAs and a relative quantification of each gene or transcript. However, several studies have found that variations in RNA transcript levels do not necessarily correlate well with protein levels [1], suggesting that translation also plays a key role. To study the role of translation in regulating gene expression, the translome can be studied by selecting ribosome-bound RNAs, releasing ribosomes, and then sequencing the selected population of RNAs as in RNA-seq. This approach, called polysome sequencing (POL-seq), is a census assay that provides relative expression levels of genes/transcripts during translation. Using the workflow manager Snakemake [2] and Conda [3] environments, we developed a bioinformatics pipeline to jointly analyze these transcriptome and translome fractions.

To promote reusability of individual steps, we propose a lightweight wrapper system to facilitate interoperability on different hardware without requiring an Internet connection. Moreover, it preserves the command inspection in Snakemake's dry-run mode.

This pipeline is divided into two parts for primary and secondary analysis. The primary analysis part consists of 4 steps: i) quality control, ii) cleaning of the sequencing reads, iii) mapping of the reads to the reference genome, and iv) counting of the mapped reads per gene. The secondary analysis part encapsulates all statistical analyses to estimate the differential expression of genes and then perform functional enrichment analyses (*i.e.* gene ontology and pathway databases).

For the statistical part, all samples are normalized together to avoid any comparison problems between fractions. We then combine the comparisons of each fraction, visualized by a log-log plot, and filter the data according to their transcriptional and translational expression levels. This allows us to classify differentially expressed mRNAs into eight categories. These categories represent mRNAs that are regulated by transcription only, by translation only, or by both transcription and translation. In the case of conjoint regulations, we determine whether these regulations have combined or opposing effects.

We will illustrate this pipeline with results from a publication with our collaborators [4].

Acknowledgements

This project was supported by the *Ligue Contre le Cancer* and by the Institut National du Cancer (INCa, FluoRib grant – PLBIO18-131). The development of pipeline for the community is supported by ANR-11-INBS-0013236.

References

- [1] Alexis Battle, *et al.* Impact of regulatory variation from rna to protein. *Science*, 347(6222):664–667, Feb 2015.
- [2] Johannes Köster, Sven Rahmann. Snakemake--a scalable bioinformatics workflow engine. *Bioinformatics*. 2012 Oct 1;28(19):2520-2. doi: 10.1093/bioinformatics/bts480. Epub 2012 Aug 20. Erratum in: *Bioinformatics*. 2018 Oct 15;34(20):3600. PMID: 22908215.
- [3] *Anaconda Software Distribution*. (2020). *Anaconda Documentation*. Anaconda Inc. Retrieved from <https://docs.anaconda.com/>
- [4] Gabriel Therizols, *et al.* Alteration of ribosome function upon 5-fluorouracil treatment favors cancer cell drug-tolerance. *Nature Communications*, 13(1):173, January 2022.