riboviz: analysis and visualization of ribosome profiling datasets

Oana Carja *[1], Tongji Xing[2], Joshua B. Plotkin[1], and Premal Shah †[2,3]

[1]Department of Biology, University of Pennsylvania, Philadelphia, 19104

[2]Department of Genetics, Rutgers University, Piscataway, NJ, USA

[3]Human Genetics Institute of New Jersey, Piscataway, NJ, USA

Abstract

Using high-throughput sequencing to monitor translation in vivo, ribosome profiling can provide critical insights into the dynamics and regulation of protein synthesis in a cell. Since its introduction in 2009, this technique has played a key role in driving biological discovery, and yet it requires a rigorous computational toolkit for widespread adoption. We developed a processing pipeline and browser-based visualization, **riboviz**, that allows convenient exploration and analysis of riboseq datasets. In implementation, **riboviz** consists of a comprehensive and flexible backend analysis pipeline that allows the user to analyze their private unpublished dataset, along with a web application for comparison with previously published public datasets.

Availability and implementation: JavaScript and R source code and extra documentation are freely available from https://github.com/shahpr/RiboViz, while the web-application is live at www.riboviz.org.

Introduction

Analyses of mRNA abundances have been used to gain insight into almost every area of modern biology (1). But ultimately, it is protein synthesis that is the central purpose of mRNAs in a cell. Although mRNA abundance has been used as a proxy for protein production, the correlation between mRNA and protein levels is weak, likely due to post-transcriptional regulation (2; 3; 4). Ribosome profiling (**riboseq**) now provides a direct method to quantify translation, the next obvious step following quantification of transcript abundances (5; 6). This technique rests on the fact that a ribosome translating a fragment of mRNA protects around 30

---

*Corresponding author: ocarja@sas.upenn.edu

†Corresponding author: premal.shah@rutgers.edu

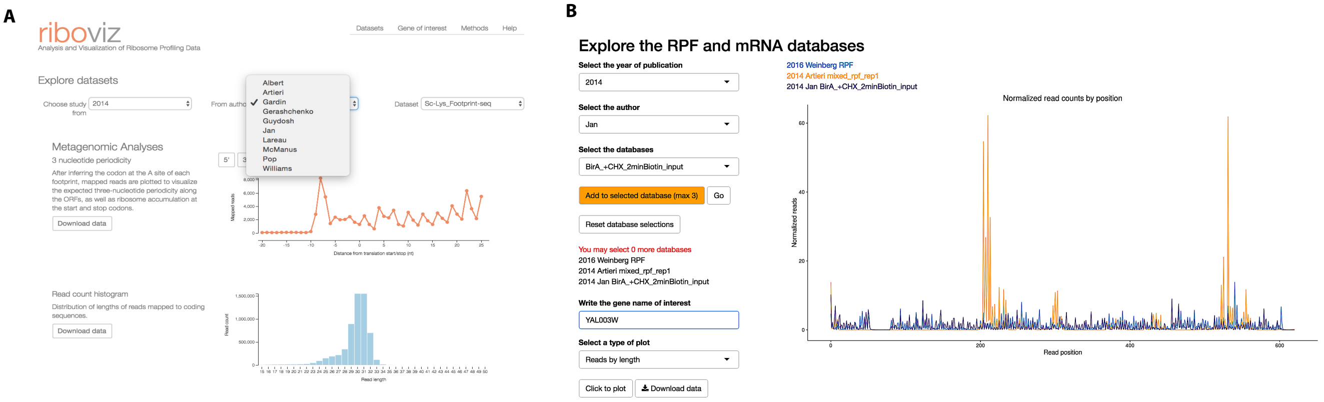
nucleotides of the mRNA from nuclease activity. High-throughput sequencing of these ribosome protected fragments (called ribosome footprints) offers a precise record of the number and location of the ribosomes at the time at which translation is stopped. Mapping the position of the ribosome-protected fragments indicates the translated regions within the transcriptome. Ribosomes spend different periods of time at different positions, leading to variation in the footprint density along mRNA transcripts. These data provide an estimate of how much protein is being produced from each mRNA (5; 6). Importantly, ribosome profiling is as precise and detailed as RNA sequencing. Even in a short time since its introduction in 2009, ribosome profiling has played a key role in driving biological discovery (13; 14; 15; 16; 17; 18; 19; 20; 21; 12; 22).

However, ribosome profiling is not without its limitations. In mammalian cells, there can be over 10 million unique footprints. The quantification and processing of these footprints remains a challenge and requires computational and domain-specific knowledge. Despite the similarity of ribosome profiling to RNA-seq, for which there is a well established set of bioinformatic tools, ribosome profiling data differs in how it is distributed across the genome. The map and density of ribosomal footprints across the genome contain much more additional information that RNA-seq alone, and traditional bioinformatics pipelines are not designed to handle such data.

We developed a bioinformatic toolkit, **riboviz**, for analyzing and visualizing ribosome profiling data, an important step for this technology to reach a broad audience of practicing biologists. In implementation, **riboviz** consists of a comprehensive and flexible backend analysis pipeline along with a web application for visualization.

Methods

Mapping and parsing riboseq datasets. A major challenge in analyses of ribosome-profiling datasets is mapping of individual footprints to ribosomal A, P and E site codons. While several ad hoc rules have been developed to assign reads to particular codons based on the read lengths, these rules are not implemented consistently across studies and as a result, comparing foot-printing reads on a gene across datasets remains a challenge. Using a combination of existing tools used for trimming and mapping reads such as *cutadapt* and *bowtie*, and in-house perl scripts, we have developed a simple set of instructions on mapping reads. We have used this pipeline to remap both RNA-seq and foot-printing datasets from published yeast studies to allow comparison of reads mapped to individual genes across different conditions and labs. In addition, researchers can download individual datasets in a flexible hierarchical data format (HDF5) and gene-specific estimates in flat *.tsv* files. The code and documentation for this pipeline are hosted on Github, with a public bug tracker and community contribution (https://github.com/shahpr/RiboViz).



Choose study from: 2014

From author: Gardin

Dataset: Sc-Lys_Footprint-seq

Metagenomic Analyses

3 nucleotide periodicity

After inferring the codon at the A site of each footprint, mapped reads are plotted to visualize the expected three-nucleotide periodicity along the ORFs, as well as ribosome accumulation at the start and stop codons.

Read count histogram

Distribution of lengths of reads mapped to coding sequences.

Explore the RPF and mRNA databases

Select the year of publication: 2014

Select the author: Jan

Select the databases: BirA_+CHX_2minBiotin_input

You may select 0 more databases

2016 Weinberg RPF
2014 Artieri mixed_rpf_rep1
2014 Jan BirA_+CHX_2minBiotin_input

Write the gene name of interest: YAL003W

Select a type of plot: Reads by length

Figure 1: **A.** The **riboviz** website with the user interface allowing dataset selection. **B.** Distribution of reads mapped to YAL003W in three **riboseq** datasets using a Shiny web server.

Web application and visualization toolkit

The web application is available at www.riboviz.org. Through this web framework, the user can interactively explore publicly available ribosome profiling datasets using JavaScript/D3 (31), JQuery (http://jquery.com) and Boostrap (http://getbootstrap.com) for metagenomic analyses and R/Shiny for gene-specific analyses. The visualization framework of **riboviz** allows the user to select from available riboseq datasets and visualize different aspects of the data. Researchers can also download a local version of the Shiny application to analyze their private unpublished dataset alongside other published datasets available through the **riboviz** website.

riboviz allows visualization of metagenomic analyses of (i) the expected three-nucleotide periodicity in footprinting data (but not RNA-seq data) along the ORFs as well as ribosome accumulation of ribosomal footprints at the start and stop codons, (ii) the distribution of mapped read lengths to identify changes in frequencies of ribosomal conformations with treatments, (iii) position-specific distribution of mapped reads along the ORF lengths, and (iv) the position-specific nucleotide frequencies of mapped reads to identify potential biases during library preparation and sequencing. **riboviz** also shows the correlation between normalized reads mapped to genes (in reads per kilobase per million RPKM) and their sequence-based features such as their ORF lengths, mRNA folding energies, number of upstream ATG codons, lengths of 5' UTRs, GC content of UTRs and lengths of poly-A tails. Researchers can explore the data interactively and download both the whole-genome and summary datasets used to generate each figure.

In addition to the metagenomic analyses, the R/Shiny integration allows researchers to analyze both footprinting and RNA-seq reads mapped to specific genes of interest, across different datasets and conditions. The Shiny application allows researchers to visualize reads mapped to a given gene across three datasets to compare (i) the distribution of reads of specific lengths along the ORF, (ii) the distribution of lengths of reads mapped to that gene as well as (iii) the overall abundance of that gene relative to its abundance in a curated set of wild-type datasets.

Summary

Ribosome profiling has been used in viruses, bacteria, yeast, mice, plants, and human cells. This new technique requires a rigorous computational toolkit to reach full promise. **riboviz** will increase the accessibility of this new technology, increase research reproducibility, and offer an broadly useful toolkit for both the community of systems biologists who study genome-wide ribosome profiling data and also to research groups focused on individual genes of interest. By developing and distributing this sets of tools we hope to remove the need for custom script generation by independent researchers, and to increase the pace of genomics research.

Funding

This work has been supported by start-up funds from Human Genetics Institute of New Jersey and Rutgers University awarded to PS and a Penn Institute for Biomedical Informatics grant to OC. JBP acknowledges funding from the David & Lucille Packard foundation, Army Research Office, and DARPA.

References

[1] Wang Z., Gerstein M., and Snyder M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nature Rev. Genet, 10, 57-63.

[2] McCann K.L., and Baserga S.J. (2013) Mysterious ribosomopathies. Science 341, 849-850.

[3] Cleary J.D. and Ranum L.P.W. (2013) Repeat-associated non-ATG (RAN) translation in neurological disease. Hum.Mol.Genet. 22, R45-R51.

[4] Bolze A. et al. (2013) Ribosomal protein SA haploinsufficiency in humans with isolated congenital asplenia. Science 340, 976-978.

[5] Ingolia N.T., Ghaemmaghami S., Newman J.R., and Weissman J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324, 218-223.

[6] Ingolia N.T. (2014). Ribosome profiling: new views of translation, from single codons to genome scale. Nat. Rev. Genet. 15, 205-213.

[7] Dunn J.G., Foo C.K., Belletier N.G., Gavis E.R., and Weissman J.S. (2013). Ribosome profiling reveals pervasive and regulated stop codon read-through in Drosophila melanogaster. eLife 2, e01179.

[8] Williams C. C., Jan C. H., and Weissman J. S. (2014). Targeting and plasticity of mitochondrial proteins revealed by proximity-specific ribosome profiling. Science, 346(6210), 748-751.

[9] Guydosh N.R., and Green R. (2014). Dom34 rescues ribosomes in 30 untranslated regions. Cell 156, 950-962.

[10] Zinshteyn B., and Gilbert W.V. (2013). Loss of a conserved tRNA anticodon modification perturbs cellular signaling. PLoS Genet. 9, e1003675.

[11] Pelechano V., Wei W., and Steinmetz L.M. (2015). Widespread Cotranslational RNA Decay Reveals Ribosome Dynamics. Cell 161, 1400-1412.

[12] Gerashchenko M.V., Lobanov A.V. and Gladyshev V.N. (2012). Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. Proc. Natl Acad. Sci. USA 109, 17394 -17399.

[13] Stern-Ginossar N. et al. (2012) Decoding human cytomegalovirus. Science 338, 1088 -1093.

[14] Bazzini A.A. et al. (2014) Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. EMBO J. 33, 981 -993.

[15] Artieri C. G. and Fraser H. B. (2014a). Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. Genome Res, 24(12), 2011-2021.

[16] Artieri C. G. and Fraser H. B. (2014b). Evolution at two levels of gene expression in yeast. Genome Res, 24(3), 411-421.

[17] Ingolia N.T., Lareau L.F., and Weissman J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147, 789-802.

[18] Li G.W., Burkhardt D., Gross C., and Weissman J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell 157, 624-635.

[19] McManus C.J., May G.E., Spealman P., and Shteyman A. (2014). Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. Genome Res. 24, 422-430.

[20] Pop C., Rouskin S., Ingolia N.T., Han L., Phizicky E.M., Weissman J.S., and Koller D. (2014). Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. Mol. Syst. Biol. 10, 770.

[21] Guttman M., Russell P., Ingolia N.T., Weissman J.S. and Landeră E.S. (2013) Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. Cell 154, 240-251.

[22] Shalgi R. et al. (2013). Widespread regulation of translation by elongation pausing in heat shock. Mol. Cell 49, 439-452.

[23] Michel A.M. et al. (2012). Observation of dually decoded regions of the human genome using ribosome profiling data. Genome Res. 22, 2219-2229.

[24] Brar G.A. et al. (2012) High-resolution view of the yeast meiotic program revealed by ribosome profiling. Science 335, 552-557.

[25] Shah P., Ding Y., Niemczyk M., Kudla G. and Plotkin J.B. (2013) Rate-limiting steps in yeast protein translation. Cell 153, 1589-1601.

[26] Vogel C. and Marcotte E.M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nature Rev. Genet. 13, 227-232.

[27] Weinberg D.E., Shah P., Eichhorn S.W., Hussmann J.A., Plotkin J.B., Bartel D.P. (2016) Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. Cell Reports 14: 1-13.

[28] Gerashchenko M.V., and Gladyshev V.N. (2014). Translation inhibitors cause abnormalities in ribosome profiling experiments. Nucleic Acids Res. 42, e134.

[29] Zheng W., Chung L.M. and Zhao M. (2011) Bias detection and correction in RNA-Sequencing data. BMC Bioinformatics 12, 290.

[30] Ingolia N.T. (2010) Genome-wide translational profiling by ribosome footprinting. Methods Enzymol. 470, 119-42.

[31] Bostock M., Ogievetsky V., and Heer J. (2011). D3: Data-Driven Documents. IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis).