

DATABASE

Open Access



riboviz: analysis and visualization of ribosome profiling datasets

Oana Carja^{1*} , Tongji Xing², Edward W. J. Wallace³, Joshua B. Plotkin¹ and Premal Shah^{2,4*}

Abstract

Background: 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.

Description: We have developed a database and a browser-based visualization tool, **riboviz**, that enables exploration and analysis of riboseq datasets. In implementation, **riboviz** consists of a comprehensive and flexible computational pipeline that allows the user to analyze private, unpublished datasets, along with a web application for comparison with published yeast datasets. Source code and detailed documentation are freely available from <https://github.com/shahpr/RiboViz>. The web-application is live at www.riboviz.org.

Conclusions: **riboviz** provides a comprehensive database and analysis and visualization tool to enable comparative analyses of ribosome-profiling datasets. This toolkit will enable both the community of systems biologists who study genome-wide ribosome profiling data and also research groups focused on individual genes to identify patterns of transcriptional and translational regulation across different organisms and conditions.

Keywords: Ribosome profiling, Translation quantification, Database, Visualization and comparison tool-kit

Background

Quantification of gene expression using RNA-seq has provided insights into most areas of modern biology [1]. However, ultimately, it is protein synthesis from mRNAs that is responsible for executing most cellular functions. Although mRNA abundance has been used as a proxy for protein production, the correlation between mRNA and protein levels is typically weak and varies widely, likely due to post-transcriptional regulation [2–4]. In contrast, ribosome profiling (**riboseq**) provides a direct method to quantify translation [5, 6]. Ribosome profiling takes advantage of the fact that a ribosome translating an mRNA protects around 30 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. Since its introduction in 2009, ribosome profiling has played a key role in driving several biological discoveries [7–26].

Analyses of ribosome profiling datasets can be challenging. In mammalian cells, there can be over 10 million unique footprints. The quantification and processing of these footprints requires computational and domain-specific knowledge.

Despite the similarity between ribosome footprinting and RNA-seq datasets, traditional bioinformatics tools developed for analyzing RNA-seq datasets are limited in their utility when applied to footprinting datasets. For instance, in RNA-seq datasets, variation in distribution of mapped reads along the length of a gene is typically attributed to random sampling. In contrast, several coding

*Correspondence: ocarja@sas.upenn.edu; premal.shah@rutgers.edu

¹Department of Biology, University of Pennsylvania, 204K Lynch Labs, 433 S University Ave, Philadelphia, PA 19104, USA

²Department of Genetics, Rutgers University, Piscataway, NJ, USA
Full list of author information is available at the end of the article

sequence features such as biased codon usage, presence of poly-basic amino-acids, and protein-domain architecture affect the distribution of footprinting reads along a transcript [27]. Recently, several tools such as GWIPS-viz [28], RiboGalaxy [29], and RPFdb [30] have been developed for both analysis and visualization of ribosome-profiling datasets. While GWIPS-viz and RPFdb use unified pipelines for processing and mapping footprinting datasets, source code for these tools and the underlying pipelines themselves are not publicly available. As a result, it is difficult to compare the effects of various mapping-related parameters on the overall analyses and visualization. Lack of open source code also limits the use of these tools for analyzing ribosome-profiling datasets in non-model organisms. In addition, tools such as RiboGalaxy and RPFdb are limited by computational resources available on the host servers and can lead to long lag times.

To address these limitations, we have developed an open-source bioinformatics toolkit, **riboviz**, for analyzing and visualizing ribosome profiling data. In implementation, **riboviz** consists of a comprehensive and flexible computational analyses pipeline along with a web application for visualization. The computational pipeline processes raw reads in FASTQ files, trims sequencing adapters, removes rRNA contaminants, aligns reads to ORFs, and generates summary statistics, and metagene and gene-specific QC plots for both RPF and mRNA datasets. Most of the individual steps of the pipeline are parallelized, thereby enabling iterative testing and faster data processing. The visualization tools are based on D3 javascript and R/Shiny and can be set up on any PC.

Construction and content

Mapping and parsing riboseq datasets

A major challenge in analyses of ribosome profiling datasets is mapping 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 footprinting reads on a gene across datasets remains a challenge. Using a combination of existing tools used for trimming and mapping reads such as *cutadapt* [31], *bowtie* [32], and *hisat2* [33], and custom perl scripts, we have developed a simple set of instructions for mapping reads. We have used this pipeline to remap both RNA-seq and footprinting 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 yeast 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>).

Utility and discussion

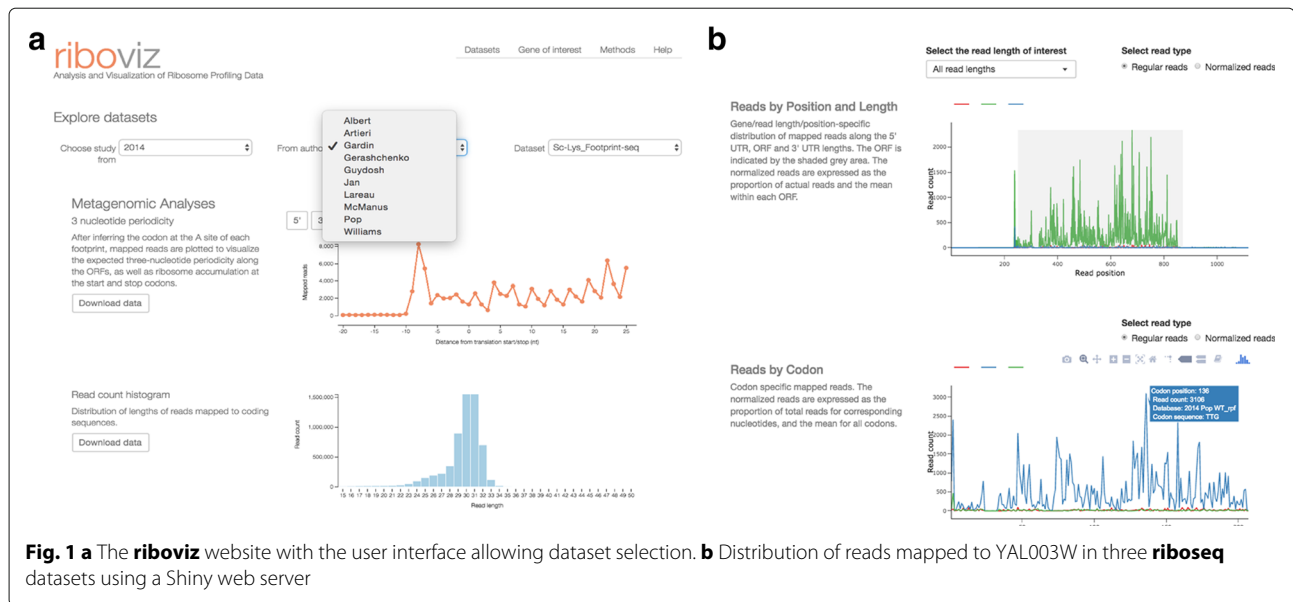
The web application is available at <https://riboviz.org/>. Through this web framework, a user can interactively explore publicly available yeast ribosome profiling datasets using JavaScript/D3 [34], JQuery (<http://jquery.com>) and Bootstrap (<http://getbootstrap.com>) for metagenomic analyses and R/Shiny 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 (Fig. 1).

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 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 [15, 35–37]. **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 up to nine 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.

Conclusions

Ribosome profiling provides a detailed snapshot of translation dynamics within a cell, and has been used to address fundamental questions related to regulation of gene expression in viruses, bacteria, as well as unicellular and multicellular eukaryotes. We have developed a comprehensive analyses and visualization tool – **riboviz** – to enable comparative analyses of ribosome-profiling datasets. This toolkit will enable both the community of



systems biologists who study genome-wide ribosome profiling data and also research groups focused on individual genes of interest to identify patterns of transcriptional and translational regulation across different organisms and conditions.

Acknowledgments

None

Funding

This work has been supported by a Penn Institute for Biomedical Informatics grant to OC, the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 661 179 to EW, and funding from the David & Lucille Packard foundation and the Army Research Office (W911NF-12-1-0552) to JBP, and NIH grant R35 GM124976, and start-up funds from Human Genetics Institute of New Jersey and Rutgers University awarded to PS.

Availability of data and materials

The web-application is live at www.riboviz.org. All the datasets, JavaScript and R source code and extra documentation are freely available from <https://github.com/shahpr/RiboViz>. All the datasets can also be directly downloaded from the website.

Authors' contributions

OC, JBP and PS conceived the study. OC, TX, EW, PS performed the analyses and wrote the code. OC, JBP and PS wrote the manuscript, with input from TX and EW. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Biology, University of Pennsylvania, 204K Lynch Labs, 433 S University Ave, Philadelphia, PA 19104, USA. ²Department of Genetics, Rutgers University, Piscataway, NJ, USA. ³School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, UK. ⁴Human Genetics Institute of New Jersey, Piscataway, NJ, USA.

Received: 20 March 2017 Accepted: 17 October 2017

Published online: 25 October 2017

References

- Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Rev Genet.* 2009;10:57–63.
- Greenbaum D, Colangelo C, Williams K, Gerstein M. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol.* 2003;4:117.
- Csárdi G, Franks A, Choi DS, Airoldi EM, Drummond DA. Accounting for experimental noise reveals that mRNA levels, amplified by post-transcriptional processes, largely determine steady-state protein levels in yeast. *PLoS Genet.* 2015;11(5):e1005206.
- Franks A, Airoldi E, Slavov N. Post-transcriptional regulation across human tissues. *PLoS Comput Biol.* 2017;13(5):e1005535.
- Ingolia NT, Ghaemmaghami S, Newman JR, et al. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science.* 2009;324:218–23.
- Ingolia NT. Ribosome profiling: new views of translation, from single codons to genome scale. *Nat Rev Genet.* 2014;15:205–13.
- Dunn JG, Foo CK, Belletier NG, et al. Ribosome profiling reveals pervasive and regulated stop codon read-through in *Drosophila melanogaster*. *eLife.* 2013;2:e01179.
- Williams CC, Jan CH, Weissman JS. Targeting and plasticity of mitochondrial proteins revealed by proximity-specific ribosome profiling. *Science.* 2014;346(6210):748–51.
- Guydosh NR, Green R. Dom34 rescues ribosomes in 30 untranslated regions. *Cell.* 2014;156:950–62.
- Zinshteyn B, Gilbert WV. Loss of a conserved tRNA anticodon modification perturbs cellular signaling. *PLoS Genet.* 2013;9:e1003675:9.
- Pelechano V, Wei W, Steinmetz LM. Widespread Cotranslational RNA Decay Reveals Ribosome Dynamics. *Cell.* 2015;161:1400–12.
- Gerashchenko MV, Lobanov AV, Gladyshev VN. Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. *Proc Natl Acad Sci USA.* 2012;109:17394–9.

13. Stern-Ginossar N, Weisburd B, Michalski A, et al. Decoding human cytomegalovirus. *Science*. 2012;338:1088–93.
14. Bazzini AA, Johnstone TG, Christiano R, et al. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J*. 2014;33:981–93.
15. Artieri CG, Fraser HB. Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. *Genome Res*. 2014a;24(12):2011–21.
16. Artieri CG, Fraser HB. Evolution at two levels of gene expression in yeast. *Genome Res*. 2014b;24(3):411–21.
17. Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*. 2011;147:789–802.
18. Li GW, Burkhardt D, Gross C, et al. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*. 2014;157:624–35.
19. McManus CJ, May GE, Spealman P, et al. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. *Genome Res*. 2014;24:422–30.
20. Pop C, Rouskin S, Ingolia NT, et al. Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol Syst Biol*. 2014;10:770.
21. Guttman M, Russell P, Ingolia NT, et al. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell*. 2013;154:240–51.
22. Shalgi R, Hurt JA, Krykbaeva I, et al. Widespread regulation of translation by elongation pausing in heat shock. *Mol Cell*. 2013;49:439–52.
23. Michel AM, Choudhury KR, Firth AE, et al. Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res*. 2012;22:2219–29.
24. Brar GA, Yassour M, Friedman N, et al. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science*. 2012;335:552–7.
25. Shah P, Ding Y, Niemczyk M, et al. Rate-limiting steps in yeast protein translation. *Cell*. 2013;153:1589–1601.
26. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Rev Genet*. 2012;13:227–32.
27. Weinberg DE, Shah P, Eichhorn SW, et al. Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. *Cell Rep*. 2016;14:1–13.
28. Michel AM, Fox G, Kiran AM, et al. GWIPS-viz: development of a ribo-seq genome browser. *Nucleic Acids Res*. 2013;42(D1):D859–D864.
29. Michel AM, Mullan JP, Velayudhan V, et al. RiboGalaxy: a browser based platform for the alignment, analysis and visualization of ribosome profiling data. *RNA Biol*. 2016;13(3):316–319.
30. Xie SQ, Nie P, Wang Y, et al. RPFdb: a database for genome wide information of translated mRNA generated from ribosome profiling. *Nucleic Acids Res*. 2015;44(D1):D254–D258.
31. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;17(1):pp–10.
32. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357–9.
33. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357–60.
34. Bostock M, Ogievetsky V, Heer J. D3: Data-Driven Documents. *IEEE Trans Vis Comput Graph (Proc. InfoVis)*. 2011;17(12):2301–2309.
35. Gerashchenko MV, Gladyshev VN. Translation inhibitors cause abnormalities in ribosome profiling experiments. *Nucleic Acids Res*. 2014;e134:42.
36. Zheng W, Chung LM, Zhao M. Bias detection and correction in RNA-Sequencing data. *BMC Bioinformatics*. 2011;12:290.
37. Ingolia NT. Genome-wide translational profiling by ribosome footprinting. *Methods Enzymol*. 2010;470:119–42.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

