LAB PROTOCOL

*sel*Seq: A method for the enrichment of non-polyadenylated RNAs including enhancer and long non-coding RNAs for sequencing

Jason D. Limberis[1]*, **Alina Nalyvayko**[1], **Joel D. Ernst**[1], **John Z. Metcalfe**[2]

1 Division of Experimental Medicine, University of California, San Francisco, San Francisco, CA, United States of America, **2** Division of Pulmonary and Critical Care Medicine, Zuckerberg San Francisco General Hospital and Trauma Centre, University of California, San Francisco, San Francisco, CA, United States of America

* jason.limberis@ucsf.edu

Abstract

Non-polyadenylated RNA includes a large subset of crucial regulators of RNA expression and constitutes a substantial portion of the transcriptome, playing essential roles in gene regulation. For example, enhancer RNAs are long non-coding RNAs that perform enhancer-like functions, are bi-directionally transcribed, and usually lack polyA tails. This paper presents a novel method, *sel*Seq, that selectively removes mRNA and pre-mRNA from samples enabling the selective sequencing of crucial regulatory elements, including non-polyadenylated RNAs such as long non-coding RNA, enhancer RNA, and non-canonical mRNA.

Introduction

Noncoding RNAs (ncRNAs), a large subset of which are non-polyadenylated (polyA), are crucial regulators of RNA expression in a diverse range of life forms [1–3]. In mammals, ncRNAs constitute a substantial portion of the transcriptome [1], playing essential roles in gene regulation via chromatin modification transcriptional regulation, and post-transcriptional processing [4]. Long non-coding RNAs (lncRNAs) are the most prevalent and diverse class of ncRNAs and are typically over 200 nucleotides long and lack protein-coding potential [1]. A large subset of lncRNAs is not polyadenylated and regulate gene expression, with abnormal expression associated with several diseases [5–7] and are predictive of outcomes [8]. Enhancer RNAs (eRNAs) are lncRNAs that perform enhancer-like functions [5, 6], are bi-directionally transcribed, and usually lack polyA tails [9]. Due to their low copy numbers [10, 11], eRNAs are challenging to detect, with most being non-polyA. Nevertheless, measuring eRNAs using total RNA-seq has provided valuable insights into transcriptional regulation [12].

Current methods to explore ncRNAs and intermediate messenger RNA splicing reactions, such as total RNA sequencing or microarrays, have severe limitations. For example, microarrays require knowledge of the transcripts of interest a priori and have a small dynamic range. Total RNA sequencing quantifies all RNA in a sample, which, after ribosomal RNA depletion, is dominated by mRNA transcripts and includes a relatively low number of lncRNAs (a typical

human cell has 3×10^3-5×10^4 lncRNAs, while there are 3×10^5-1×10^6 mRNA transcripts). Thus, quantification of rare lncRNAs and isoforms is challenging. Here, we present a novel method that selectively removes polyadenylated RNA (e.g., mRNA and pre-mRNA; and optionally rRNA) from samples, allowing for the comprehensive characterization of low-abundance lncRNAs, eRNAs, and non-canonical mRNA splicing. Our method, *sel*Seq, offers a reliable and sensitive approach to exploring these crucial regulatory elements.
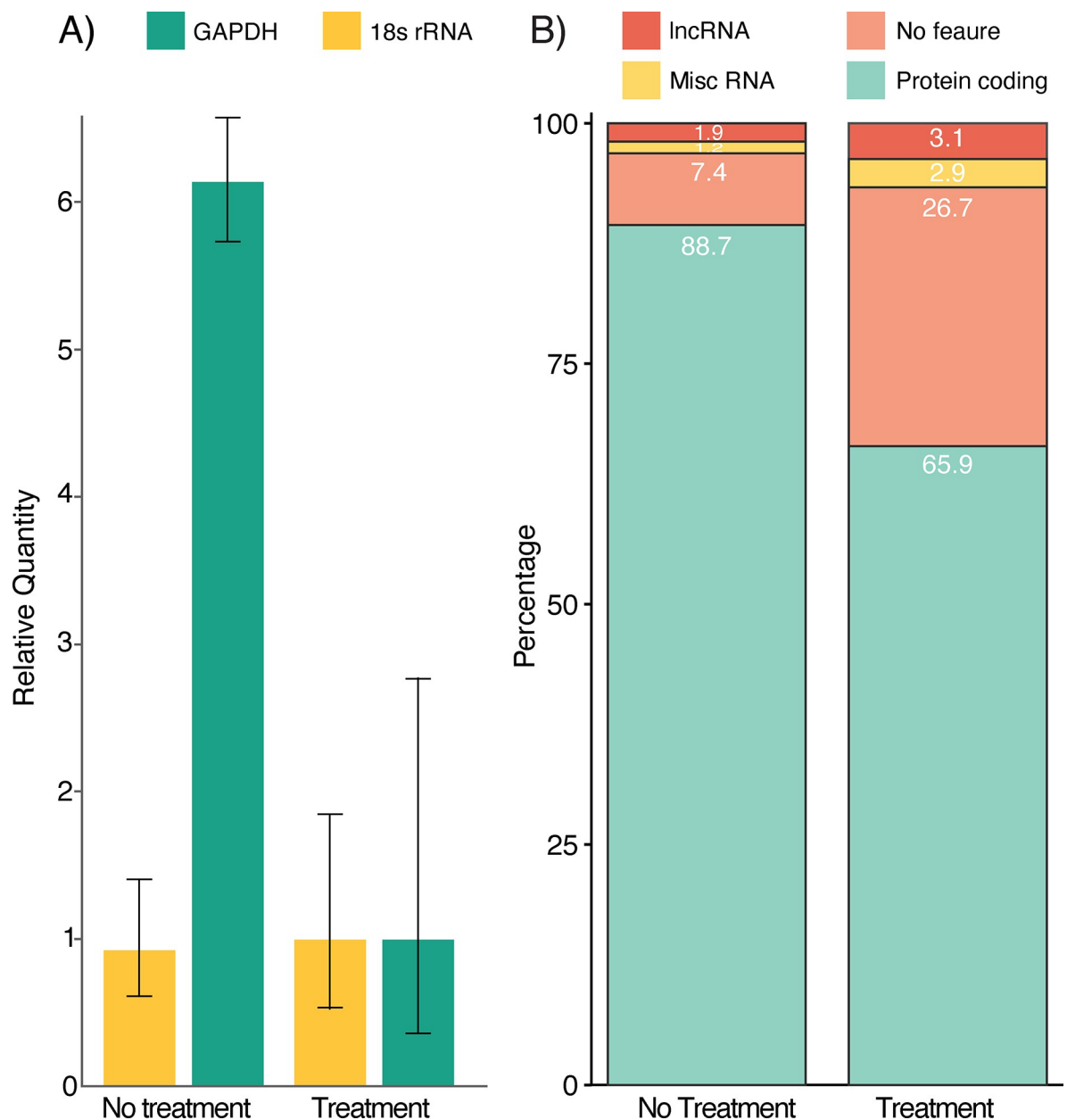
Fig 1. *sel*Seq results showing A) the decrease in polyA tailed *GAPDH* housekeeping gene and no decrease in non-polyA tailed 18s ribosomal RNA (n = 4; error bars show one standard deviation); and B) total RNA sequencing after rRNA depletion shows an increase in the proportion of long non-coding, unassigned (no feature), and miscellaneous RNA with a corresponding decrease in the protein-coding assignments, the remainder of which are likely non-polyadenylated transcripts.

https://doi.org/10.1371/journal.pone.0289442.g001

Materials and methods

The protocol described in this peer-reviewed article is published on protocols.io, DOI 10.17504/protocols.io.j8nlkwpk6l5r/v1, and is included for printing as S1 File with this article.

Expected results

There will be a decrease in the RNA quantity following the completion of the protocol. A reverse transcriptase quantitative PCR can be used to see the reduction of polyA housekeeping genes to non-polyA rRNA (**Fig 1A**) if the optional rRNA depletion step is not done as part of the protocol. Total RNA sequencing after rRNA depletion shows a large decrease in the proportion of polyA-tailed transcripts as illustrated in **Fig 1B**, where human liver total RNA (Ambion, USA) was used and aligned to the GENCODE nucleotide sequence of the Human GRCh38.p13 genome assembly version containing all regions, including reference chromosomes, scaffolds, assembly patches, and haplotypes. The proportion of lncRNA, no feature, and miscellaneous RNA have increased substantially, and the protein-coding assignments have decreased, with those remaining likely non-polyadenylated transcripts or degraded polyadenylated transcripts.

Supporting information

S1 File. The protocol in PDF format available from protocols.io is provided as supporting information File 1, with the caption: S1: Step-by-step protocol, also available on protocols.io. Sequence data is available on SRA under the accession number PRJNA949687.
(PDF)

Author Contributions

Conceptualization: Jason D. Limberis.

Formal analysis: Jason D. Limberis.

Funding acquisition: John Z. Metcalfe.

Investigation: Jason D. Limberis.

Methodology: Jason D. Limberis.

Project administration: Jason D. Limberis.

Software: Jason D. Limberis.

Supervision: Joel D. Ernst, John Z. Metcalfe.

Validation: Jason D. Limberis.

Visualization: Jason D. Limberis.

Writing – original draft: Jason D. Limberis, Alina Nalyvayko.

Writing – review & editing: Jason D. Limberis, Alina Nalyvayko, Joel D. Ernst, John Z. Metcalfe.

References

1. Hombach S, Kretz M. Non-coding RNAs: Classification, Biology and Functioning. Adv Exp Med Biol. 2016; 937: 3–17. https://doi.org/10.1007/978-3-319-42059-2_1 PMID: 27573892

2. Ponting CP, Oliver PL, Reik W. Evolution and Functions of Long Noncoding RNAs. Cell. 2009; 136: 629–641. https://doi.org/10.1016/j.cell.2009.02.006 PMID: 19239885

3. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009; 458: 223–227. https://doi.org/10.1038/nature07672 PMID: 19182780

4. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nature Reviews Genetics 2009 10:3. 2009; 10: 155–159. https://doi.org/10.1038/nrg2521 PMID: 19188922

5. Yan B, Tao ZF, Li XM, Zhang H, Yao J, Jiang Q. Aberrant expression of long noncoding RNAs in early diabetic retinopathy. Invest Ophthalmol Vis Sci. 2014;55. https://doi.org/10.1167/iovs.13-13221 PMID: 24436191

6. Lekka E, Hall J. Noncoding RNAs in disease. FEBS Letters. 2018. https://doi.org/10.1002/1873-3468.13182 PMID: 29972883

7. Gao J, Wang F, Wu P, Chen Y, Jia Y. Aberrant LncRNA expression in leukemia. Journal of Cancer. 2020. https://doi.org/10.7150/jca.42093 PMID: 32368311

8. Farrar JE, Smith JL, Othus M, Huang BJ, Wang Y-C, Ries R, et al. Long Noncoding RNA Expression Independently Predicts Outcome in Pediatric Acute Myeloid Leukemia. J Clin Oncol. 2023; JCO2201114. https://doi.org/10.1200/JCO.22.01114 PMID: 36795987

9. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014;507. https://doi.org/10.1038/nature12787 PMID: 24670763

10. Gorbovytska V, Kim SK, Kuybu F, Götze M, Um D, Kang K, et al. Enhancer RNAs stimulate Pol II pause release by harnessing multivalent interactions to NELF. Nat Commun. 2022;13. https://doi.org/10.1038/s41467-022-29934-w PMID: 35508485

11. Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at neuronal activity-regulated enhancers. Nature. 2010;465. https://doi.org/10.1038/nature09033 PMID: 20393465

12. Carullo NVN, Phillips RA, Simon RC, Roman Soto SA, Hinds JE, Salisbury AJ, et al. Enhancer RNAs predict enhancer–gene regulatory links and are critical for enhancer function in neuronal systems. Nucleic Acids Res. 2020;48. https://doi.org/10.1093/nar/gkaa671 PMID: 32810208