

Characterizing Poorly Structured 3' UTRs in Structure-Mediated RNA Decay

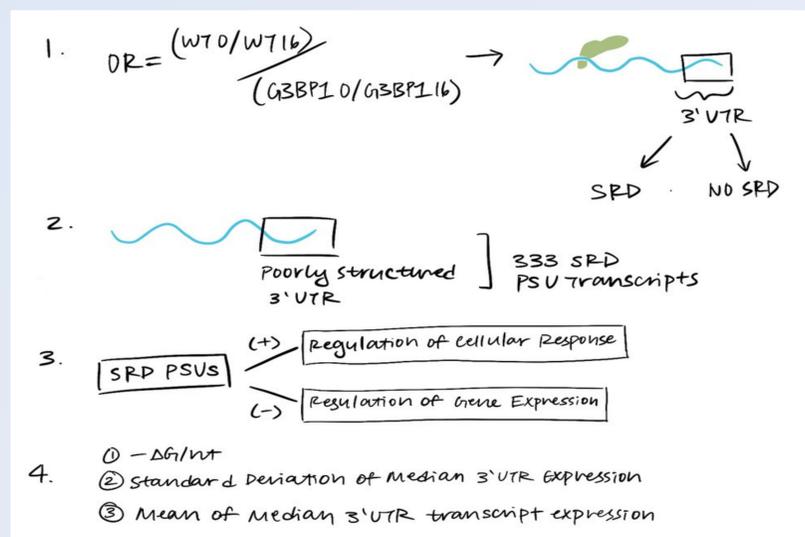
Caroline Wang, Department of Public Health Studies

Mentor: Dr. Anthony Leung, Veronica Busa, Johns Hopkins Bloomberg School of Public Health Department of Biochemistry and Molecular Biology

Background and Abstract

The 3' untranslated regions (3'UTRs) of mRNA direct post-transcriptional outcomes, which affect mRNA metabolism by regulated stability and influencing the expression of RNAs [1]. Structure Mediated RNA Decay (SRD) is a structure-based post-transcriptional mechanism that depends on overall base-pairing density and relies on UPF1 and G3BP1 proteins [2]. SRD is predicted to regulate over one-third of mRNA transcripts and is linked to differences in breast cancer outcomes, emphasizing its importance. In in-vivo experiments, SRD demonstrated that highly structured 3'UTRs (HSUs, $-\Delta G/\text{nt} > 0.25$) generally underwent SRD while poorly structured 3'UTRs (PSUs, $-\Delta G/\text{nt} < 0.25$) generally did not [2]. Despite this general pattern, some HSUs did not undergo SRD while some PSUs did in vivo experiments – defying our general expectations and suggesting that there may be better ways of quantifying SRD than our current computational models provide. In this project, I characterize PSUs subject to SRD and discuss better prediction models for SRD.

Results



Interpretation & Conclusions

While HSUs have been shown to be more subject to SRD than PSUs, there is evidence that PSU transcripts may be subject to SRD too – which suggests that there may be more to understand about the process and mechanism of SRD and identification of protein involved in SRD. After identifying an Odds Ratio cutoff value of 1.2 as a metric for determining whether a transcript would be subject to SRD, we validated our results by examining well-known positive and negative controls for PSUs and HSUs and examining whether the odds ratio accurately identified the 14 controls. These results not only validate our Odds Ratio metric, but also suggest that we could continue to use our Odds Ratio metric in the future to identify transcripts subject to SRD.

With an identified Odds Ratio cutoff of 1.2, application of our new metric to PSU transcripts suggested that roughly 11% of total PSU transcripts are subject to SRD, or 333 transcripts in total. This suggests that dG/nt , which is our current indicator of transcripts being subject to SRD, may not be the most accurate metric – specifically, dG/nt may be imperfect marker for PSUs and has room for improvement.

To better understand the functions that SRD PSUs may play a role in, I inputted a dG/nt based ranked list of the SRD PSUs into a gene ontology database named “GO-rilla”, and then used the Revigo platform to visually represent the linked function – which included cell population proliferation, G protein signaling, and pattern specification processes. Most importantly, we also found that SRD PSUs are also implicated in negative regulation of gene expression, further validating their role in SRD. With SRD already implicated in roles in racial differences in breast cancer outcomes and predicted to influence lncRNAs, circRNAs, and mRNAs, these results further point towards SRD as an important mechanism with global roles in our understanding of disease and gene regulation.

Methodology

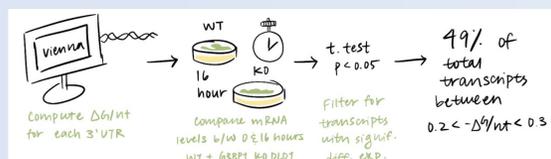
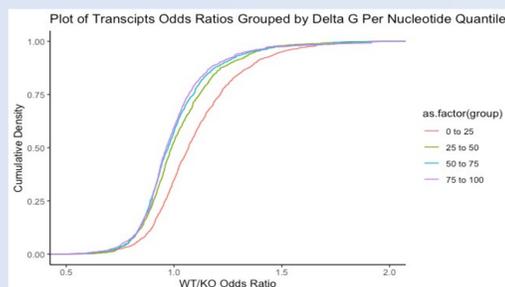
While $\Delta G/\text{nt}$ serves as a valuable predictor for SRD targets, its predictions are imperfect, since some PSUs seem to undergo SRD. $\Delta G/\text{nt}$ is dependent on numerous factors, including length, GC content, and nucleotide order – however, when each of these factors, when tested individually, could not fully explain the influence in determining whether a 3' UTR would be regulated by SRD. To define whether transcripts undergo SRD, I used timed data from an *in vivo* Wildtype (WT) and Knockout (KO) G3BP1 DLD1 colon cancer cell assay. Two sets of WT and KO DLD1 cells samples had 16 hours to modify transcript expression, with additional actinomycin D (actD) serving to ensure to further creation of mRNA confounded any transcript expressions.

To determine whether a 3' UTR undergoes SRD, I calculated the difference between expression at 0 and 16 hours for every transcript in the WT and KO G3BP1 DLD1 cells. To avoid non-significant differences in transcript levels, I conducted t-test for each transcript in the WT/KO data and filtered for transcripts that had statistically significantly decreased expression at 16 hours post-actD treatment. Since transcripts subject to SRD will decay faster in WT cells compared to G3BP1 KO cells (since G3BP1 serves as the rate-limiting function), I created an odds metric.

$$\text{Odds Ratio} = (\text{DLD1 WT 0 hours} / \text{DLD1 WT 16 hours}) / (\text{DLD1 KO 0 hours} / \text{DLD1 KO 16 hours})$$

To determine an Odds Ratio cutoff value that would indicate whether a transcript was subject to SRD, we conducted two major tests.

- 1) Group Delta G per nucleotide values of each transcript into different bins and created a density plot comparing the Odds Ratio densities by dG/nt quantile cutoffs.
- 2) Utilize the Kolmogorov-Smirnov test to establish the point of greatest vertical deviation between the lowest and highest quartile of the odds ratio spread. Calculate the Odds Ratio at which maximum vertical deviation occurred.



Sources and Acknowledgements

I would like to thank the Woodrow Wilson Fellowship and Hopkins Discovery Award for their generous support of my project. I would also like to thank my wonderful mentors, Veronica Busa and Dr. Anthony Leung from the Leung lab.

1. Beaudoin J-D, Novoa EM, Vejnar CE, et al. Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. *Nature Structural & Molecular Biology*. 2018;25(8):677-686. doi:10.1038/s41594-018-0091-z
2. Felekis, K., Touvana, E., Stefanou, C., and Deltas, C. (2010). microRNAs: a newly described class of encoded molecules that play a role in health and disease. *Hippokratia* 14, 236–240
3. Fischer JW, Busa VF, Shao Y, Leung AKL. Structure-Mediated RNA Decay by UPF1 and G3BP1. *Molecular Cell*. 2020;78(1):70-84.e6. doi:10.1016/j.molcel.2020.01.021

