Significant and pervasive effects of RNA degradation on Nanopore direct RNA sequencing

Yair D.J. Prawer[1], Josie Gleeson[1], Ricardo De Paoli-Iseppi[1] and Michael B. Clark[1]

1.Centre for Stem Cell Systems, Department of Anatomy and Physiology, The University of Melbourne, Parkville, VIC, Australia

INTRODUCTION

- Oxford Nanopore's direct RNA sequencing (DRS) is a powerful tool capable of sequencing native RNA molecules.
- DRS has many advantages:
  - Designed to sequence complete or **full-length molecules**.
  - Accurately measures both **gene and isoform expression levels**.
  - DRS sequences native RNA molecules meaning **no need for RT-PCR**
- The precision and reliability of **DRS** can be heavily **dependent upon RNA integrity** (RIN is a measure of RNA integrity and ranges from 1-10).
- Often RNA from in-vivo or clinical samples have undergone some degradation.

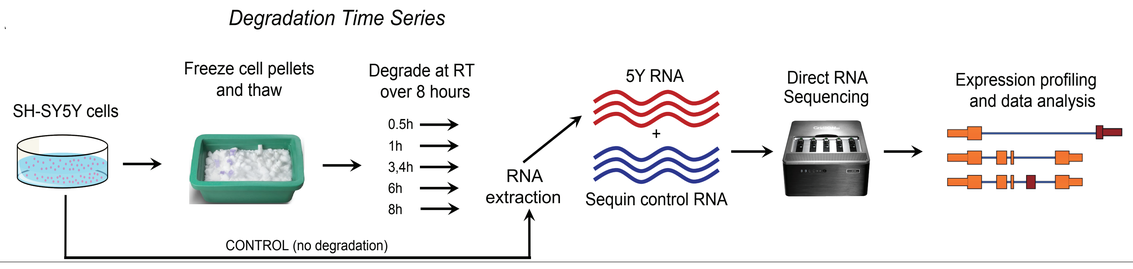
KNOWLEDGE GAP

- We currently don't have a deep understand of how degradation impacts DRS.
- It is unclear how best to apply DRS to samples that have undergone degradation and if the impact of degradation can be corrected for.

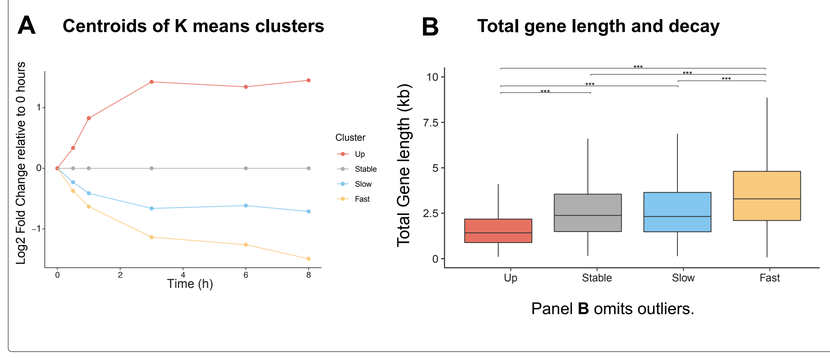
PROJECT AIM

- It is essential to determine how best to apply DRS in a degradation context so the technology can be applied appropriately across a range of RNA qualities.

Visual methodology



Degradation rates are associated with gene and isoform architecture



- Length impacts degradation rate[2].
- Genes and isoforms with:
  - Higher %GC are more stable
  - Shorter length/3'UTR are more stable
- Short and high GC content genes are overrepresented at later time points

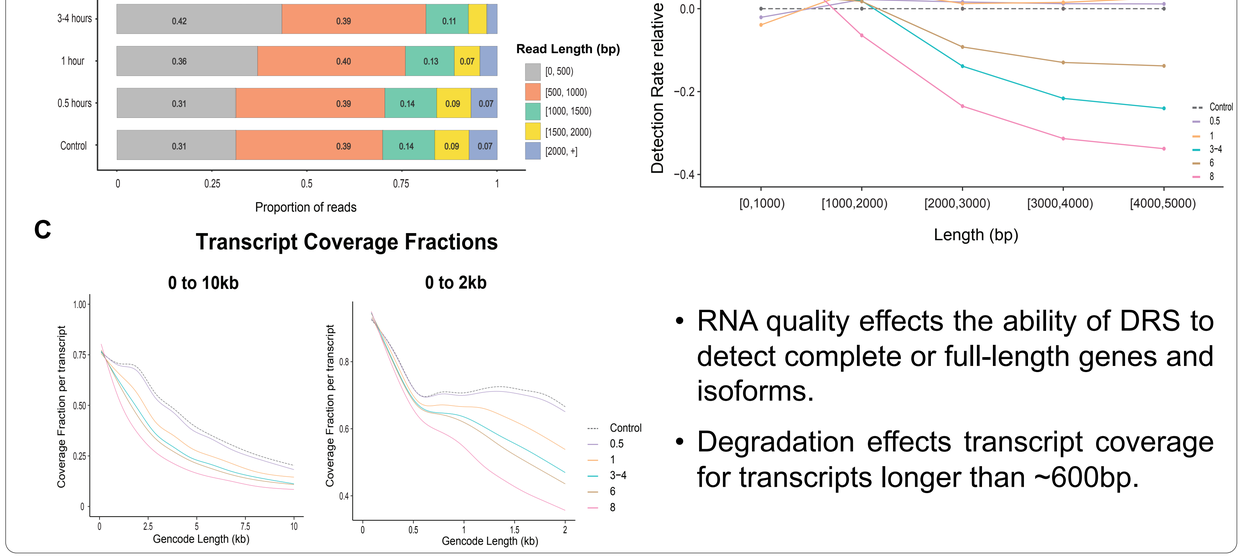
Degradation introduces short isoform bias and impedes full-length isoform detection

- Degradation negatively impacts library QC metrics including N50 and percentage of full-length reads.

Table 1. Direct RNA sequencing metrics

Time post degradation treatment (h)	Average RIN	N50 (bp)	Median Aligned length (bp)	Full length transcripts (%)	Median coverage fraction per read	Median coverage fraction per transcript	Average Number of Genes detected
0 (n=3)	9.8	1,378	773	25	0.68	0.53	14,393
0.5 (n=2)	9.6	1,460	774	24	0.67	0.52	15,342
1 (n=2)	9.3	1,203	697	20	0.57	0.42	15,840
3-4 (n=3)	8.7	1,009	628	19	0.53	0.38	16,028
6 (n=3)	8.2	925.6	607	20	0.53	0.35	14,956
8 (n=2)	7.25	777	518	16	0.47	0.29	15,189

- Samples that have undergone degradation are biased toward shorter read lengths.
- Lower quality samples detect more shorter genes and isoforms.



- RNA quality effects the ability of DRS to detect complete or full-length genes and isoforms.
- Degradation effects transcript coverage for transcripts longer than ~600bp.

Effects of degradation can impede differential expression analysis but can be corrected for using sample RIN

- Samples cluster based on sample quality (PC1) and replicate effects (PC2).
- Sample degradation biases Differential Expression (DE) analysis. Many genes and isoforms detected as DE even without a biological change between time points.

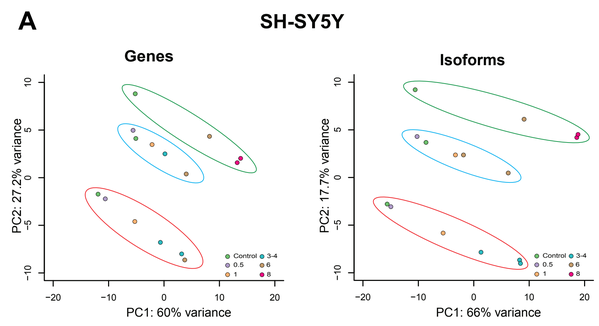
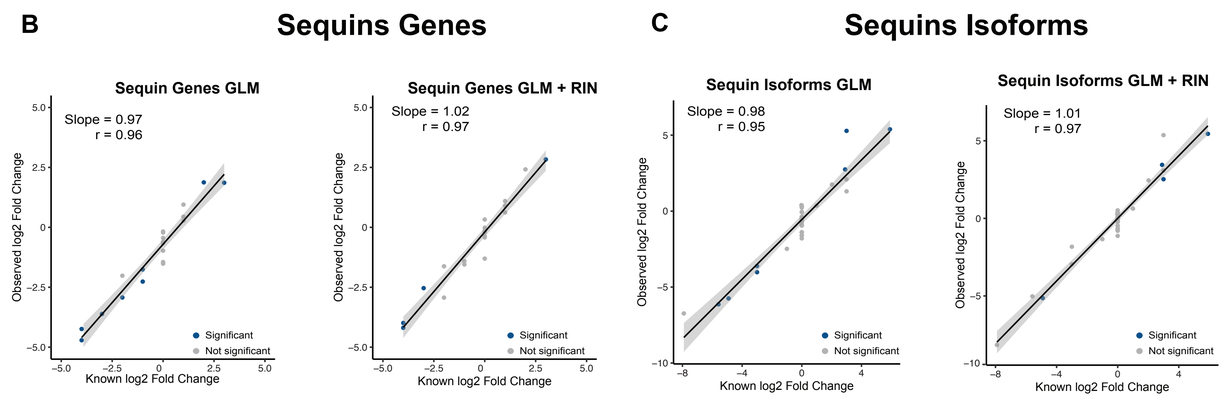


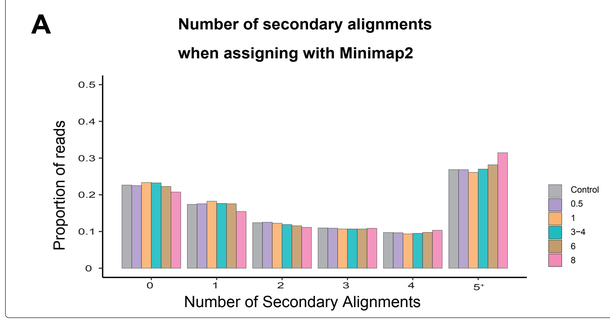
Table 2. Diff Genes and Isoforms

Time Point	GLM DE genes (ft counts)	GLM DE Isoforms (NanoCount)	GLM +RIN DE genes (ft counts)	GLM +RIN DE Isoforms (NanoCount)
0 vs 0.5h	0	0	0	0
0 vs 1h	36	28	4	0
0 vs 3-4h	164	201	12	10
0 vs 6h	169	261	0	0
0 vs 8h	381	573	0	0

- Incorporating RIN into the GLM does not bias the DE framework but does reduce sensitivity[2,3].
- True DE can still be detected while correcting for degradation.



Read assignment becomes more challenging as samples degrade



- The number of secondary alignments can be used as a proxy for read assignment[1].
- As samples degrade, the proportion of secondary alignments increases.
- Degradation makes read assignment more ambiguous.

SUMMARY AND CONCLUSION

- Degradation effects have **pervasive** and **significant outcomes** for Direct RNA sequencing.
- Degradation biases read length, effects which genes and isoforms are detected, biases differential expression analysis and introduces ambiguity when assigning reads.
- Degradation effects can be corrected for by incorporating sample quality into the analysis.
- DRS can be applied to a range of RNA qualities. **RIN > 9 is recommended** but can be applied appropriately with RIN > 7 if degradation impacts are considered.

1. Gleeson, J., et al. (2021). "Accurate expression quantification from nanopore direct RNA sequencing with NanoCount." Nucleic Acids Research.
2. Gallego Romero, I., et al. (2014). "RNA-seq: impact of RNA degradation on transcript quantification." BMC Biology **12**(1): 42.
3. Jaffe, A. E., et al. (2017). "qSVA framework for RNA quality correction in differential expression analysis." Proceedings of the National Academy of Sciences **114**(27): 7130.

yairp@student.unimelb.edu.au

Australian Government National Health and Medical Research Council NHMRC
BRAIN & BEHAVIOR RESEARCH FOUNDATION Awarding NARSAD Grants