To make a short story long: simultaneous short and long RNA profiling on Nanopore devices

Morgan MacKenzie[1], Susan Tigert[2], Debbie Lovato[2], Hamza Mir[1], Kamyar Zahedi[1,3], Sharon L. Barone[1,3], Marybeth Brooks[1,3], Manoocher Soleimani[1,3], Christos Argyropoulos[1,2]

[1]Department of Internal Medicine, Division of Nephrology, School of Medicine, University of New Mexico, Albuquerque, NM, 87131, [2]University of New Mexico Clinical & Translational Science Center, [3]Research Services, Veterans Healthcare Medical Center, Albuquerque, NM 87108, NM 88061

Background

- Long coding and short, non-coding (microRNAs) RNAs yield valuable information about the abundance and novelty of the transcriptome and the epigenetic regulation of it, respectively
- Noncoding RNAs are of interest for clinical research applications, as their relative stability and tissue-specific nature make them viable candidates for disease-state biomarkers
- Consideration of epigenetic regulation requires examination of the quantitative relationships between noncoding and long coding RNAs from a few input samples
- Current methodology for detecting all types of RNAs are limited to biotype specific sequencing experiments, which are incompatible with this analysis, time intensive and costly
- Therefore, a workflow that can simultaneously profile short and long RNA species would meet a currently unmet need
- **Overarching Aim:** to deliver a sequencing workflow and its bioinformatics ecosystem that allows for the simultaneous analysis of coding and non-coding RNAs ("universal RNAseq") for epigenetic research applications
- Here, we describe our proposed SMART-seq *(1)* modifications within the context of Nanopore sequencing using the PCR-cDNA (SQK-PCS109) library preparation kit

Introduction

- We propose the universal, single tube polyadenylation of RNAs in a sample as a means for making all species present compatible for use in SMART-Seq protocols
- We evaluate this approach through a series of 2x2 factorial-design experiments of various RNA inputs, consisting of mixes of 10 microRNAs of interest *(2)* and 92 spike-in long RNAs (ERCCs)
- This approach was further validated against two biological samples derived from total RNA isolated from the jejunum of ten C57Bl6/J mice involved in carbohydrate metabolism studies
- Biological samples were analyzed through three platforms for comparative performance analysis:
  - •The proposed workflow, here on notated as 'PALS-NS'
  - •ONT's unmodified PCR-cDNA Sequencing Protocol ('ONT PAP(-)', or simply 'PAP(-)')
  - •A HiSeqTM 2500 Illumina system ('Illumina')

METHODS

- Polyadenylation (PAP), reverse transcription (RT) and strand switching reactions were done in a single tube, **Fig. 2A**
- All synthetic experiments were run on R9.4.1 flow cell Chemistry except two flow cells used to sequence biological samples without a polyadenylation step that were of R10.4 chemistry.
- A custom bioinformatics pipeline was developed **Figure 2B** to extract the inserts from the Nanopore reads **(Fig. 2C)**
- Over-dispersed Poisson and negative binomial regression models were used to compute **bias factors,** deviations of observed from expected counts based on inputs

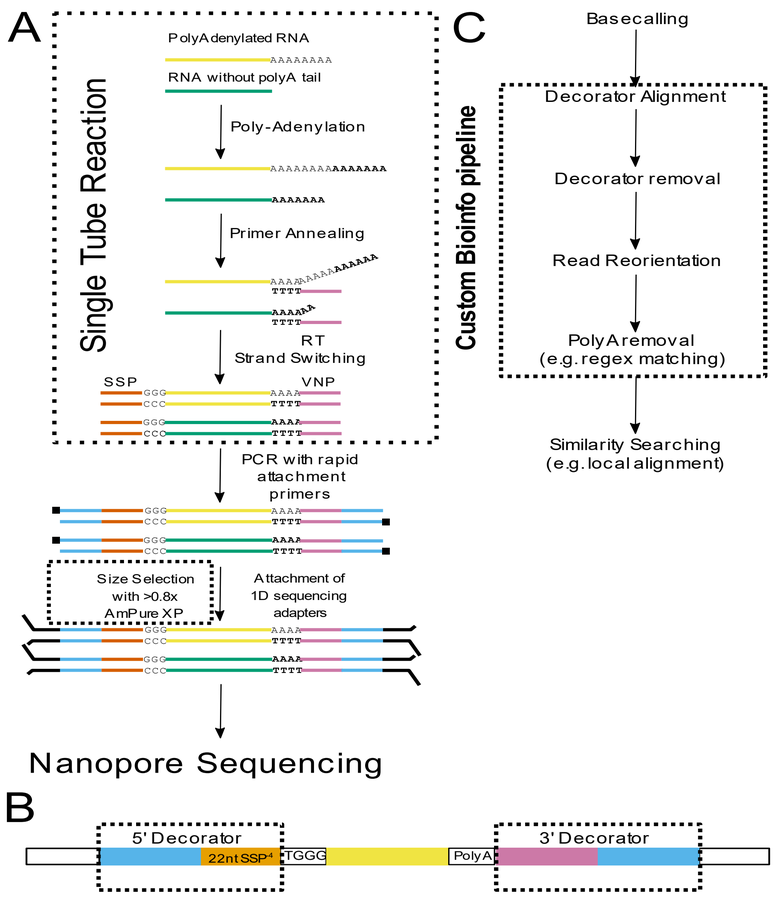

Figure 2 PALS-NS experimental workflow **(A)**, text based model of a well-formed read **(B)** and custom bioinformatics pipeline **(C)**.

Dashed boxes indicate modifications to biochemical protocols, read models and bioinformatics, statistical methods for the analyses of counts that enable PALS-NS to be applied to any RNA sources of interest

RESULTS

- PALS-NS is able to detect RNA species in proportion to their input **(Fig. 3)** with the expected number of false positives
- Truncation of longer transcripts (>700bp) increases in highly concentrated and complex samples (i.e., with short and long species present), implicating competition during the RT step **(Fig. 4, Fig. 5)**
- While all libraries detected the same protein coding RNAs (**Fig. 6A)**, PALS-NS improved the detection of long noncoding **(Fig. 6B)** and microRNAs **(Fig. 6C)**.
- PALS-NS captures non-coding RNAs **(Fig. 7, component A)**, that Illumina or the unmodified Nanopore cDNA sequencing protocol, PAP (-) cannot quantify well
- miRNA bias factors in PALS-NS were comparable to those obtained from the gold standard, degenerate 5' end ligation protocol '4 N' *(3)* **(Table 1)**

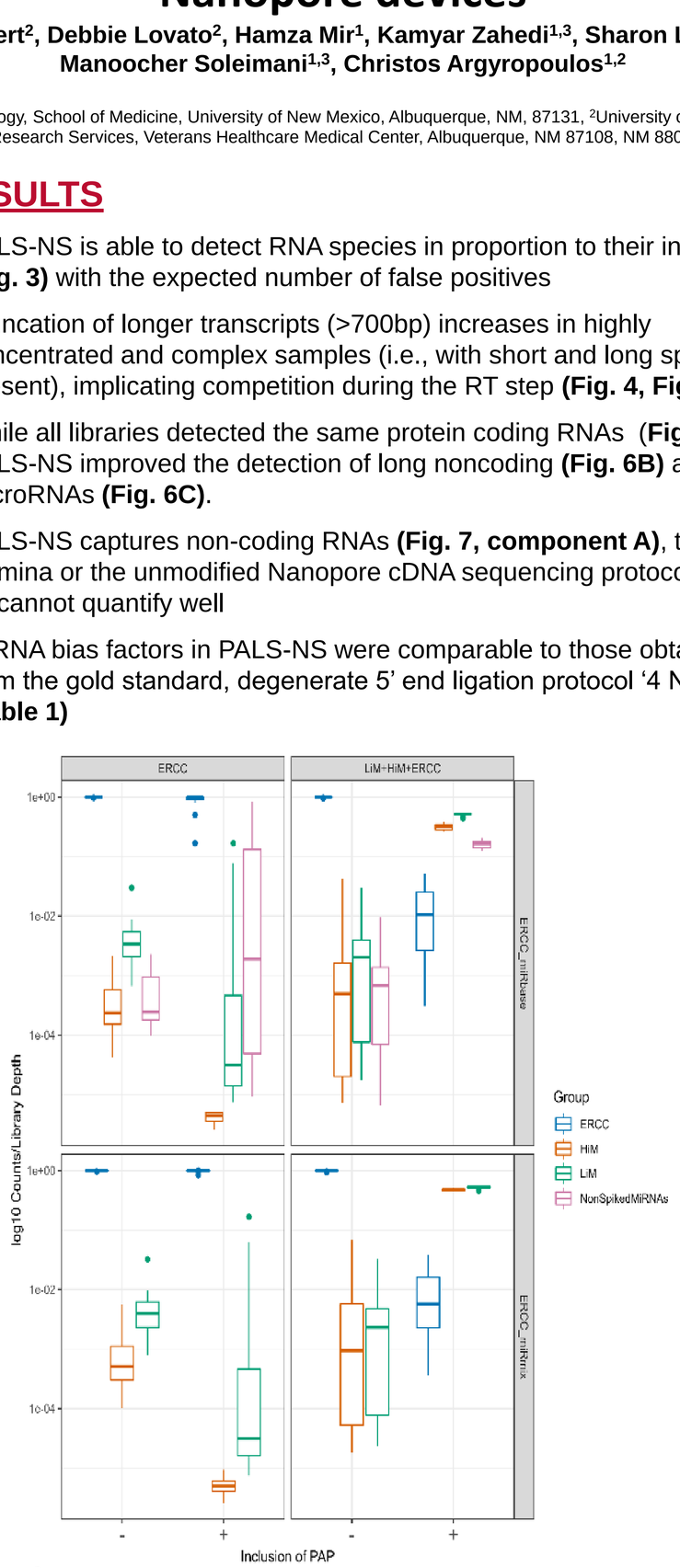

Figure 3

Representation of groups of RNA as a proportion of library depth in the 2x2 experiments

A sensitivity analysis was also performed by mapping against the entire miRbase and counts of RNAs not present in the mix ("NonSpikedMiRNAs") were tabulated as a separate category

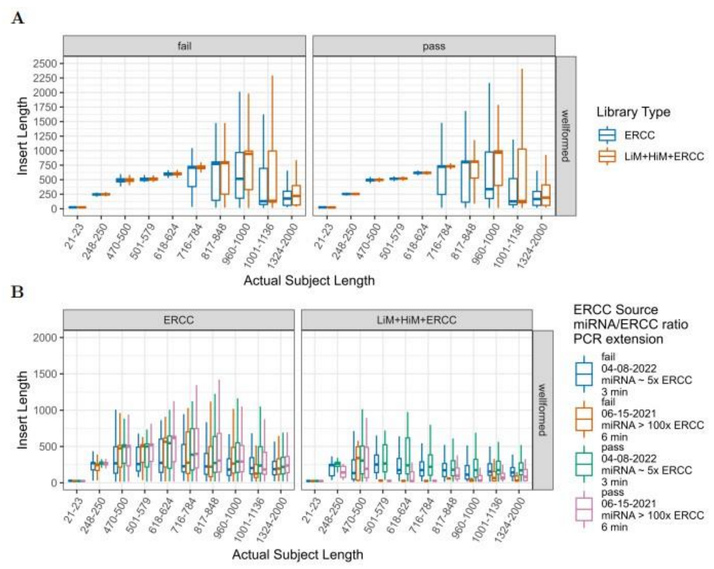

Figure 4 Comparison of ERCC insert length to subject length for (A) sham polyadenylation and (B) 2x2 experiments.

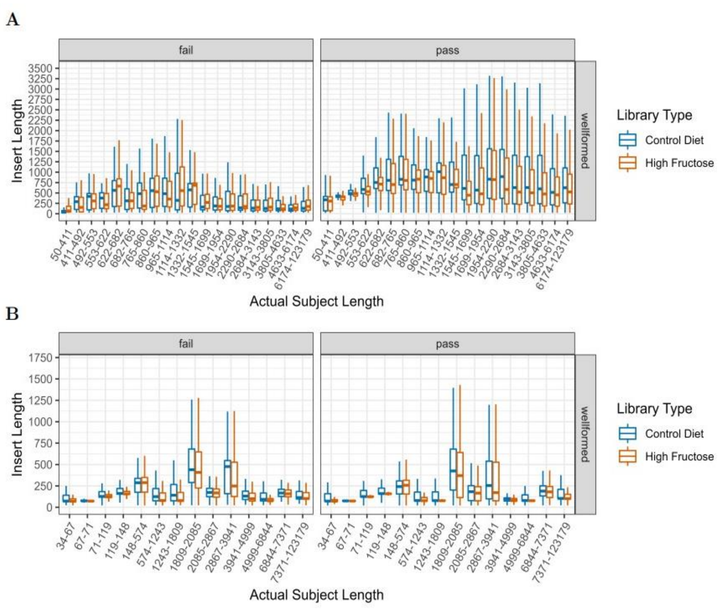

Figure 5 Comparison of insert length to subject length for biological samples run through (A) Base ONT protocol (No PAP) and (B) PALS-NS protocol.

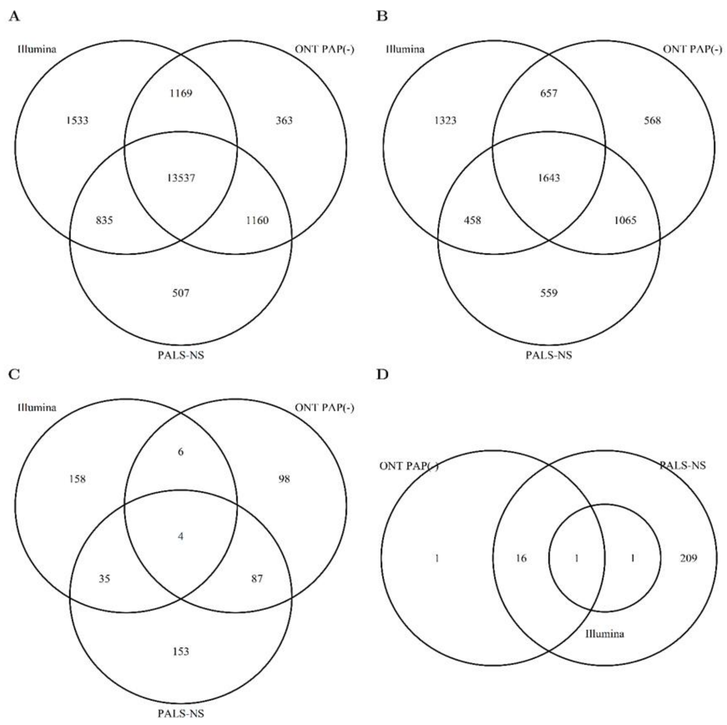

Figure 6 Venn diagram showing the overlap of individual RNAs detected in libraries constructed from a polyA enriched samples (Illumina), long RNA sequencing on a Nanopore device without polyadenylation i.e., ONT PAP(-), and PALS-NS for protein coding RNAs (**A**), long non -coding RNAs (**B**), microRNAs (**C**) and ribosomal RNAs (**D**).

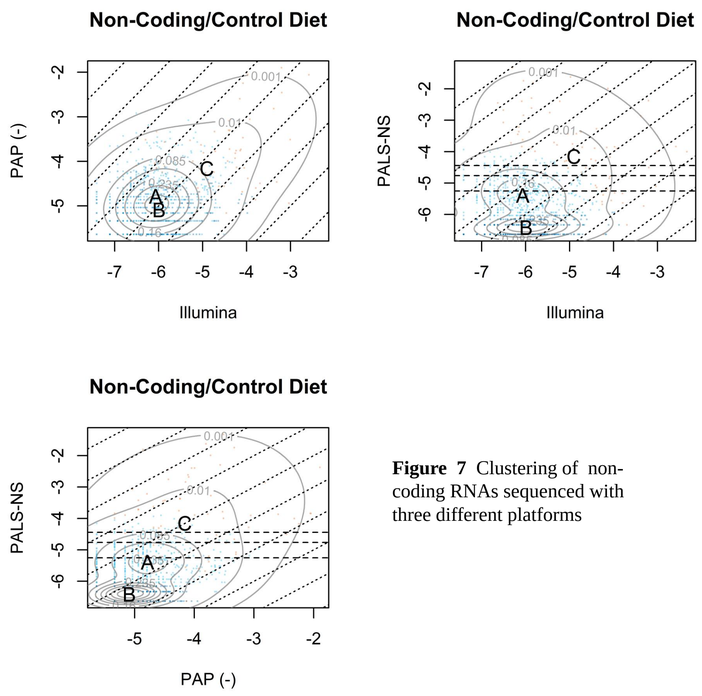

Figure 7 Clustering of non-coding RNAs sequenced with three different platforms

miRNA	PALS-NS	4N
hsa-miR-132-5p	-0.092	-0.037
hsa-miR-145-5p	0.200	-0.301
hsa-miR-192-5p	-0.098	-0.447
hsa-miR-21-5p	0.091	-0.655
hsa-miR-324-3p	-1.261	0.786
hsa-miR-744-5p	0.098	-0.029

Table 1 Comparison of Bias Factors for short RNAs in the PALS-NS and a 4N randomized adapter ligation protocol from previous experimentation *(3)*

CONCLUSIONS

- PALS-NS is capable of simultaneously profiling short and long RNAs from a single tube reaction using simple modifications of existing SMART-seq protocols
- PALS-NS extends the dynamic range of reads detection to non-coding RNAs with limited length and sequence-dependent bias (data not shown)
- Bias for short RNAs comparable to the gold standard Illumina protocol (4N) developed by NIH's exRNA consortium
- PALS-NS provides unique insights into the biochemical mechanisms of library preparation protocols

Acknowledgments

- Financial support by DCI Inc. #C-3765 "A community based study of the Epidemiology of CKD in rural New Mexico"
- Research performed at the University of New Mexico Clinical and Translational Science Center (CTSC): **Cite the Grant:** This project is supported by an award from the National Center for Advancing Translational Sciences, National Institutes of Health under grant number UL1TR001449.

References

1. Giraldez, M., Spengler, R., Etheridge, A. *et al.* Comprehensive multi-center assessment of small RNA-seq methods for quantitative miRNA profiling. *Nat Biotechnol* **36**, 746–757 (2018). https://doi.org/10.1038/nbt.4183
2. Shaffi, S.K., Galas, D., Etheridge, A. and Argyropoulos, C. (2018) Role of MicroRNAs in Renal Parenchymal Diseases—A New Dimension. *International Journal of Molecular Sciences*, **19**, 1797.
3. Argyropoulos, C., Etheridge, A., Sakhanenko, N. and Galas, D. (2017) Modeling bias and variation in the stochastic processes of small RNA sequencing. *Nucleic Acids Research*, **45**, e104.