Nanopore native RNA sequencing of a human poly(A) transcriptome: RNA extraction, cDNA conversion and direct RNA and cDNA library preparation for Oxford Nanopore

CURRENT STATUS: POSTED

Rachael E. Workman
Department of Biomedical Engineering, Johns Hopkins University

Alison D. Tang
Department of Biomolecular Engineering, UCSC Genomics Institute, University of California, Santa Cruz

Paul S. Tang
Ontario Institute for Cancer Research

Miten Jain
UCSC Genomics Institute, University of California, Santa Cruz

ORCiD: https://orcid.org/0000-0002-4571-3982

John R. Tyson
Michael Smith Laboratories and Djavad Mowafaghian Centre for Brain Health, University of British Columbia

Roham Razaghi
Department of Biomedical Engineering, Johns Hopkins University

Philip C. Zuzarte
Ontario Institute for Cancer Research

Timothy Gilpatrick
Department of Biomedical Engineering, Johns Hopkins University

Alexander Payne
DeepSeq, School of Life Sciences, University of Nottingham

Joshua Quick
University of Birmingham
Norah Sadowski  
Department of Biomedical Engineering, Johns Hopkins University

Nadine Holmes  
DeepSeq, School of Life Sciences, University of Nottingham

Jaqueline Goes de Jesus  
University of Birmingham

Karen L. Jones  
Michael Smith Laboratories and Djavad Mowafaghian Centre for Brain Health, University of British Columbia

Terrance P. Snutch  
Michael Smith Laboratories and Djavad Mowafaghian Centre for Brain Health, University of British Columbia

Nicholas Loman  
University of Birmingham

Benedict Paten  
Department of Biomolecular Engineering, UCSC Genomics Institute, University of California, Santa Cruz

Matthew Loose  
DeepSeq, School of Life Sciences, University of Nottingham

Jared T. Simpson  
Ontario Institute for Cancer Research, Department of Computer Science, University of Toronto

Hugh E. Olsen  
Department of Biomolecular Engineering, UCSC Genomics Institute, University of California, Santa Cruz

Angela N. Brooks  
Department of Biomolecular Engineering, UCSC Genomics Institute, University of California, Santa Cruz

Mark Akeson  
UCSC Genomics Institute, University of California, Santa Cruz

✉️ makeson@soe.ucsc.edu

ORCID: https://orcid.org/0000-0001-9392-8075

Winston Timp
Department of Biomedical Engineering, Johns Hopkins University

Corresponding Author
ORCiD: https://orcid.org/0000-0003-2083-6027

DOI:
10.21203/rs.2.12872/v1

SUBJECT AREAS
Biotechnology  Molecular Biology

KEYWORDS
nanopore, direct RNA, cDNA, GM12878, third generation sequencing, sequencing
Abstract
High throughput cDNA sequencing technologies have dramatically advanced our understanding of transcriptome complexity and regulation. However, these methods lose information contained in biological RNA because the copied reads are often short and because modifications are not carried forward in cDNA. We address these limitations using a native poly(A) RNA sequencing strategy developed by Oxford Nanopore Technologies (ONT). Our study focused on poly(A) RNA from the human cell line GM12878, generating 9.9 million aligned sequence reads. These native RNA reads had an aligned N50 length of 1294 bases, and a maximum aligned length of over 21,000 bases. A total of 78,199 high-confidence isoforms were identified by combining long nanopore reads with short higher accuracy Illumina reads. We describe methods for extracting intact RNA, poly-A selection, cDNA conversion for a portion of sample, and library preparation for both direct RNA and cDNA libraries.

Introduction
The roles of RNA in cell function are numerous and complex. Beyond the fundamental importance of mRNA, tRNA, and ribosomal RNA in translation, several classes of non-coding RNA (ncRNA) regulate cellular processes including division, differentiation, and programmed cell death (Su et al. 2016). Nanopore RNA strand sequencing has emerged as an alternative single molecule strategy (Garalde et al. 2018; Smith et al. 2017; Jenjaroenpun et al. 2018). It differs from SBS-based platforms in that native RNA nucleotides, rather than copied DNA nucleotides, are identified as they thread through and touch a nanoscale sensor. Nanopore RNA strand sequencing shares the core features of nanopore DNA sequencing, i.e. a processive helicase motor regulates movement of a bound polynucleotide driven through a protein pore by an applied voltage. As the polynucleotide advances through the pore in single nucleotide steps, ionic current impedance reports on the segment of bases that occupy a narrow reading head as a function of time. This series of ionic current segments is then used to infer nucleotide sequence using an algorithm trained with known RNA molecules.

Here we describe RNA extraction, cDNA conversion and library preparation for both direct RNA and cDNA libraries for the human poly(A) transcriptome from the GM12878 cell line using the Oxford
Nanopore (ONT) platform. We demonstrate that long native RNA reads allow for discovery and characterization of RNA isoforms that are difficult to observe using short read cDNA methods (Steijger et al. 2013; Venturini et al. 2018). Because native RNA strands are directly read by nanopores, nucleotide modifications and 3’ poly(A) tail lengths can be determined directly from the ionic current signal absent additional processing steps. Data and resources are posted online at: (https://github.com/nanopore-wgs-consortium/NA12878/blob/master/RNA.md).

Reagents
RPMI media (Invitrogen cat# 21870076)
FBS (Lifetech cat# 12483020)
2mM L-Glutamax (Lifetech cat# 35050061)
TRI-Reagent (Invitrogen AM9738)
BCP (1-Bromo-3-chloro-propane)
CHCl3 (Chloroform)
NEXTflex Poly(A) Beads (BIOO Scientific Cat#NOVA-512980)
Ampure XP beads
LongAmp Taq Master Mix (NEB)
Superscript IV (Thermo Fisher)
SIRV set 3 control (Lexogen)
SQK-RNA001 kit (ONT)
SQK-PCS108 kit (ONT)
SQK-LSK308 kit (ONT)

Equipment
Refrigerated centrifuge
Hula mixer
Thermocycler
Vortex
Nanopore MinION sequencer

Procedure
**Cell Tissue Culture**

1. Culture GM12878 cells in RPMI media supplemented with 15% non heat-inactivated FBS and 2mM L-Glutamax.

2. Grow cells to a density of 1E6 / ml before subsequent dilution of ½ every ~3 days, and expand to 9 x T75 flasks (45 ml of media in each).

3. Centrifuge cells for 10 min at 100 x g (4C), wash in 1/10th volume of PBS (pH 7.4) and combine for homogeneity.

4. Evenly split cells between 8 x 15ml tubes and pellet at 100 x g for 10 mins at 4C. Unless proceeding directly with RNA isolation, snap freeze in liquid nitrogen and immediately store at -80C.

**RNA Isolation**

1. Add 4 ml of TRI-Reagent to a frozen pellet of 5E7 GM12878 cells and briefly vortex immediately. Incubate at room temperature for 5 minutes.

2. Add 400 μl BCP (1-Bromo-3-chloro-propane) or 200 μl CHCl3 (Chloroform) per ml of sample, vortex, incubate at room temperature for 5 minutes, briefly vortex again, and centrifuge for 10 minutes at 12,000 x g (4C).

3. Pool the aqueous phase in a LoBind Eppendorf tube and combine with an equal volume of isopropanol.

4. Mix sample, incubate at room temperature for 15 minutes, and centrifuge for 15 minutes at 12,000 x g (4C).

5. Remove supernatant, wash RNA pellet with 750 μl 80% ethanol and centrifuge for 5 minutes at 12,000 x g (4C).

6. Remove supernatant. Air dry pellet for 10 minutes, resuspend in nuclease free water (100 μl final volume), quantify with Qubit or Nanodrop, and either store at -80C or proceed to poly-A purification.

**Poly(A) RNA isolation**

1. Dilute 100 μg aliquots of total RNA in 100 μl of nuclease free water and poly-A select using NEXTflex Poly(A) Beads.

2. Elute resulting mRNA in Nuclease free water and store at -80oC.
**cDNA synthesis**

1. Perform first strand cDNA synthesis using Superscript IV and 100 ng of poly-A purified RNA combined with 0.5 ng of the SIRV set 3 control.

2. Perform reverse transcription and strand-switching with primers provided by ONT in the SQK-PCS108 kit.

3. After reverse transcription, amplify with LongAmp Taq Master Mix under the following conditions: 95C for 30 seconds, 11-15 cycles (95C for 15 seconds, 62C for 15 seconds, 65C for 15 minutes), 65C for 15 minutes, hold at 4C.

4. Perform 15 cycle PCR when using the SQK-PCS108 kit and 11 cycle PCR when using the SQK-LSK308 kit.

5. Purify PCR products using 0.8X AMPure XP beads, elute in nuclease-free water.

**Sequencing protocol**

**MinION native RNA sequencing of poly-A RNA**

1. Prepare biological poly-A RNA (500-775 ng) and a synthetic control (Lexogen SIRV Set 3, 5 ng) for nanopore direct RNA sequencing generally following the ONT SQK-RNA001 kit protocol, including the optional reverse transcription step recommended by ONT.

2. Replace Superscript III (ONT recommended) with Superscript IV for reverse transcription.

**MinION sequencing of cDNA**

1. Prepare cDNA sequencing libraries using 1 μg of cDNA following the standard ONT protocol for SQK-PCS108 (1D sequencing) or SQK-LSK308 (1D^2 sequencing).

2. Use 0.8X AMPure XP beads for cleanup, instead of recommended 0.4X.

**Troubleshooting**

**See Figure 3**

**Issue and Recommendations**

**All extraction steps**

Save the solutions at every step until the final quantifications are performed so that one could go back and troubleshoot if something goes awry.
RNA extraction yield low

If you saved the aqueous phase + isopropanol solution, try spinning it again for longer time.

Extracted RNA quality low

Extract with CHCl3 again and follow the protocol post that; or re-precipitate RNA and follow the protocol post that

rRNA peaks remaining after poly(A) selection

Perform poly(A) selection again

Time Taken

RNA extraction : 60-90 minutes

PolyA selection : 75-90 minutes

cDNA conversion : 3-5 hours

Direct RNA library preparation : 60-120 minutes (if performing RT)

cDNA library preparation : 60-75 minutes

Sequencing : 48 hours

Anticipated Results

*See Figures 1 and 2**

Step: Input, Yield

RNA extraction

5E6 cells (GM12878), 200-250 μg

PolyA selection

100 μg, 1-1.5 μg

cDNA conversion

1 μg, 5-10 μg

Library preparation

500 ng, 50K - 2M reads

References
Garalde, Daniel R., Elizabeth A. Snell, Daniel Jachimowicz, Botond Sipos, Joseph H. Lloyd, Mark Bruce, Nadia Pantic, et al. 2018. “Highly Parallel Direct RNA Sequencing on an Array of Nanopores.” *Nature Methods*, January. https://doi.org/10.1038/nmeth.4577.

Jenjaroenpun, Piroon, Thidathip Wongsurawat, Rui Pereira, Preecha Patumcharoenpol, David W. Ussery, Jens Nielsen, and Intawat Nookaew. 2018. “Complete Genomic and Transcriptional Landscape Analysis Using Third-Generation Sequencing: A Case Study of Saccharomyces Cerevisiae CEN.PK113-7D.” *Nucleic Acids Research*, January. https://doi.org/10.1093/nar/gky014.

Quick, Joshua, Aaron R. Quinlan, and Nicholas J. Loman. 2014. “A Reference Bacterial Genome Dataset Generated on the MinION trade mark Portable Single-Molecule Nanopore Sequencer.” *GigaScience* 3 (1): 1-6.

Smith, A. M., M. Jain, L. Mulroney, D. R. Garalde, and M. Akeson. 2017. “Reading Canonical and Modified Nucleotides in 16S Ribosomal RNA Using Nanopore Direct RNA Sequencing.” *bioRxiv*. http://biorxiv.org/content/early/2017/04/29/132274.abstract.

Steijger, Tamara, Josep F. Abril, Pär G. Engström, Felix Kokocinski, RGASP Consortium, Tim J. Hubbard, Roderic Guigó, Jennifer Harrow, and Paul Bertone. 2013. “Assessment of Transcript Reconstruction Methods for RNA-Seq.” *Nature Methods* 10 (12): 1177–84.

Su, Ye, Haijiang Wu, Alexander Pavlosky, Ling-Lin Zou, Xinna Deng, Zhu-Xu Zhang, and Anthony M. Jevnikar. 2016. “Regulatory Non-Coding RNA: New Instruments in the Orchestration of Cell Death.” *Cell Death & Disease* 7 (8): e2333.

Venturini, Luca, Shabhonam Caim, Gemy George Kaithakotttil, Daniel Lee Mapleson, and David Swarbreck. 2018. “Leveraging Multiple Transcriptome Assembly Methods for Improved Gene Structure Annotation.” *GigaScience* 7 (8). https://doi.org/10.1093/gigascience/giy093.

Acknowledgements
The authors are grateful for support from the following individuals. Libby Snell, Botond Sipos and Dan Turner (ONT) provided materials and advice relevant to the 3′ poly(A) standards used to test nanopolish-polya. Daniel Garalde (ONT) provided early advice on use of the MinION for RNA sequencing. Nicholas Conrad gave insight into the correlation of intron retention and poly(A) tail
length. Mark Diekhans reviewed the isoform analysis. Zofia M. Chrzanowska-Lightowlers, Tom Suzuki, and Shunpei Okada commented on early drafts of the manuscript. The authors thank Andrew Beggs, Louise Tee and Tom Nieto (University of Birmingham, UK) for providing cell cultures used in the Birmingham sequencing runs. The project was supported by the following grants: NIH HG010053 (ANB, BP, & MA), NIH 5T32HG008345 (ADT), NIH HG009190 (WT, JTS), NIH U54HG007990 (BP), U01 HL137183-02 (BP), Oxford Nanopore Research Grant SC20130149 (MA), National Institutes of Health Research Surgical Reconstruction and Microbiology Research Centre (JQ), Medical Research Council CLIMB Fellowship (NL), Wellcome Trust 204843/Z/16/Z (ML), BBSRC BB/N017099/1 and BB/M020061/1 (ML), the Canada Research Chair in Biotechnology and Genomics-Neurobiology (TPS), the Canadian Institutes of Health Research (#10677; TPS), the Canadian Epigenetics, Environment and Health Research Consortium (TPS), the Koerner Foundation (TPS), the Ontario Institute for Cancer Research through funds provided by the Government of Ontario (JTS).

Figures

**Anticipated results**

| Step               | Input                  | Yield              |
|--------------------|------------------------|--------------------|
| RNA extraction     | 5E6 cells (GM12878)    | 200-250 µg         |
| PolyA selection    | 100 µg                 | 1-1.5 µg           |
| cDNA conversion    | 1 µg                   | 5-10 µg            |
| Library preparation| 500 ng                 | 50K - 2M reads     |

Figure 1

RNA extraction and library preparation anticipated yields

| Native RNA Pass | 1D cDNA Pass |
|-----------------|--------------|

10
|                  |       |       |
|-----------------|-------|-------|
| Reads           | 10,302,647 | 15,152,101 |
| Bases (Gb)      | 10.61 | 14.13 |
| Mean Read Length| 1,030 | 933   |
| Median Read Length | 771  | 780   |
| Read Length N50 | 1,334 | 1,072 |
| Mean Aligned % Identity | 86.1 | 85.0 |
| Median Aligned % Identity | 86.6 | 85.5 |
| Mean Aligned Read Length | 987  | 791   |
| Median Aligned Read Length | 726  | 643   |
| Longest Aligned Read Length | 21,608 | 9,969 |
| Flowcells Used  | 30    | 12    |

Figure 2

Yield and read alignment statistics for native RNA and 1D cDNA
## Troubleshooting table

| Issue                              | Recommendations                                                                 |
|------------------------------------|-------------------------------------------------------------------------------|
| All extraction steps               | Save the solutions at every step until the final quantifications are performed so that one could go back and troubleshoot if something goes awry. |
| RNA extraction yield low           | If you saved the aqueous phase + isopropanol solution, try spinning it again for longer time. |
| Extracted RNA quality low          | Extract with CHCl3 again and follow the protocol post that ; or Re-precipitate RNA and follow the protocol post that |
| rRNA peaks remaining after poly(A) selection | Perform poly(A) selection again                                              |

---

**Figure 3**

**Troubleshooting guide**

Nanopore native RNA sequencing of a human poly(A) transcriptome

by Rachael E. Workman, Alison D. Tang, Paul S. Tang, +19