Nanopore long-read RNA-seq and absolute quantification delineate transcription dynamics in early embryo development of an insect pest

Anthony Bayega

McGill Genome Centre, Department of Human Genetics, McGill University, Montréal, Québec, Canada
<https://orcid.org/0000-0002-1247-4911>

Spyros Oikonomopoulos

McGill Genome Centre, Department of Human Genetics, McGill University, Montréal, Québec, Canada

Maria-Eleni Gregoriou

Laboratory of Molecular Biology and Genomics, Department of Biochemistry & Biotechnology, University of Thessaly, Larissa, Greece

Konstantina T. Tsoumani

Laboratory of Molecular Biology and Genomics, Department of Biochemistry & Biotechnology, University of Thessaly, Larissa, Greece

Yu Chang Wang

McGill Genome Centre, Department of Human Genetics, McGill University, Montreal, Canada

Antonis Giakountis

Laboratory of Molecular Biology and Genomics, Department of Biochemistry & Biotechnology, University of Thessaly, Larissa, Greece

Kostas D. Mathiopoulos (✉ kmathiop@bio.uth.gr)

Laboratory of Molecular Biology and Genomics, Department of Biochemistry & Biotechnology, University of Thessaly, Larissa, Greece

Jiannis Ragoussis (✉ ioannis.ragoussis@mcgill.ca)

McGill Genome Centre, Department of Human Genetics, McGill University, Montreal, Canada Department of Bioengineering, McGill University, Montréal, Québec, Canada

Method Article

Keywords: Nanopore, RNA sequencing, ERCC

Posted Date: May 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.pex-1409/v1>

Abstract

The Oxford Nanopore Technologies' long-read RNA sequencing (RNAseq) platform can yield full length transcripts that can be very instrumental in improving the characterization of non-model organisms. The resolution of RNAseq can be increased by addition of external RNA molecules of known concentration such as ERCC in order to obtain absolute gene expression quantification. This protocol details the procedure to use ONT long-read RNAseq with addition of ERCC external RNAs. This protocol can be used with total RNA extracted from any source for which transcripts with a poly(A) tail at the 3' end are the target. The protocol should be possible to complete in a single day (12 hours) followed by sequencing which takes another 48 hours.

Introduction

Reagents

Common reagents and materials

RNase Zap (Thermo Fischer Scientific, AM9780)

1M Tris-HCl pH 8.0 (Thermo Fischer Scientific, AM9855G)

Ethanol 100 % (reagent grade)

Agencourt AMPure XP beads (A63880, Beckman Coulter, 5 mL)

Eppendorf DNA LoBind Tubes 1.5 mL (022431021)

Filtered tips (10, 20, 100, 200, 1000 µL), and respective pipettes

VWR PCR 8-Tube Strip 0.2 mL (120 Strips) (VWR, 53509-304)

Qubit Assay Tubes (Thermo Fischer Scientific; Q32856)

Reagents to assess the quality of the extracted RNA

Reagents for RNA Quantification

Qubit RNA HS Assay Kit (Thermo Fischer Scientific, Q32852)

Reagents to examine the RNA profile

Agilent RNA ScreenTape Ladder (Agilent; 5067-5578)

Agilent RNA ScreenTape Sample Buffer (Agilent; 5067-5577)

Agilent RNA ScreenTape (Agilent; 5067-5576)

Reagents for cDNA synthesis

Oligos and reconstitution buffer for cDNA synthesis primers:

PolyT primer (V = A or C or G, N = A or C or G or T) (RNase-Free HPLC purification of the synthesized oligo is preferable. The oligo should be shipped lyophilized)

5'-AAGCAGTGGTATCAACGCAGAGTATGCAACGCAACT₍₃₀₎VN-3'

TSO oligo (TSO: Template-Switching oligonucleotide, the red marked bases are ribonucleotides. RNase-Free HPLC purification of the synthesized oligo is preferable. The oligo should be shipped lyophilized)

5'- AAGCAGTGGTATCAACGCAGAGTGGATTCTATCACGCrGrGrG -3'

THE RNA Storage Solution (Thermo Fischer Scientific, AM7000)

Reagents to remove DNA contamination in total RNA samples

DNA-free DNA Removal Kit (Thermo Fischer Scientific, M1906)

Enzymes and reagents for the cDNA synthesis reaction:

RNAse inhibitor 40 U/uL (Clontech, 2313A)

Advantage UltraPure PCR Deoxynucleotide Mix (10 mM each dNTP) (Clontech, 639125)

Water nuclease free PCR grade (eg. Affymetrix, 901578)

SuperScript IV (Thermo Fischer Scientific, 18090010)

Betaine (5M) (Sigma-Aldrich, B0300-1VL)

MgCl₂ (1M) (Thermo Fisher Scientific, AM9530G)

Reagents to spike RNA molecules of known abundance inside the sample RNA:

ERCC RNA Spike-In Mix (Thermo Fisher Scientific, 4456740)

Reagents for cDNA amplification:

Primer for cDNA amplification:

cDNA amplification primer (Standard Desalting of the synthesized oligo can be ordered. The oligo can be shipped lyophilized or reconstituted at standard 100µM concentration.)

5'- /5Phos/ TCGTCGGCAGCGTCCAAGCAGTGGTATCAACGCAGAGT-3'

Enzymes for cDNA amplification:

Advantage 2 PCR Kit (Clontech, 639207)

Reagents for cDNA Quality Control:

Reagents for cDNA quantification

Qubit HS DNA Assay Kit (Thermo, Q32851)

Reagents to examine the cDNA profile

Agilent D5000 ScreenTape (Agilent, 5067-5588)

Agilent D5000 Reagents (Agilent, 5067-5589)

Agilent D5000 Ladder (Agilent, 5067-5590)

Reagents for cDNA library preparation for the nanopore platform

End repair of the cDNA molecules

NEBNext End Repair Module (New England Biolabs, E6050S)

d(A) tailing of the cDNA molecules

NEBNext dA-Tailing Module (New England Biolabs, E6053S)

Ligate ONT adapters on the cDNA molecules

Ligation 1D Sequencing kit SQK-LSK108

NEB Blunt/TA Master Mix (New England Biolabs, M0367S)

Flow Cell Wash Kit (EXP-WSH002)

ONT MinION Sequencing

MinION SpotON FLO-MIN106 flow cells

MinION Mk1b

Equipment

Common equipment

Magnetic Stand for 1.5 mL tubes (e.g. Ambion P/N AM10026) and 0.2 mL tubes

Eppendorf Centrifuge 5424R or 5424 (or equivalent)

Micro-Centrifuge for 0.2 mL PCR tubes

BioRad Thermo Cyclor T100

Qubit Fluorometer (Life Technologies)

Mixer e.g. HulaMixer (Thermo Fischer Scientific), Vortex Mixer (VWR)

Agilent TapeStation 2200 and the corresponding assay tubes

Procedure

Preparation of reagents

a. The TSO oligo is reconstituted in “THE RNA Storage Solution” at a concentration of 1200 μM . *The information sheet from the manufacturer usually provides a dilution volume for a solution with a 100 μM oligo concentration. To create the solution with the 1200 μM oligo concentration, adjust the dilution volume accordingly by reducing 12X times the recommended volume presented on the information sheet.* Then 1 μl is diluted in 99 μl of RNA storage solution (100X dilution; final concentration: 12 μM) and stored in aliquots at 5.6 μl per tube. The aliquots are stored at -80 °C. The TSO ribonucleotides are prone to degradation. Loss of the ribonucleotides will lead in considerable reduction/absence of cDNA yield.

b. The PolyT primer is reconstituted in nuclease free H_2O at a concentration of 1200 μM . *The information sheet from the manufacturer usually provides a dilution volume for a solution with a 100 μM oligo concentration. To create the solution with the 1200 μM oligo concentration, adjust the dilution volume accordingly by reducing 12X times the recommended volume presented on the information sheet.* Then 1 μl is diluted in 99 μl of nuclease free H_2O and stored in aliquots at 7 μl per tube. The aliquots can be stored at -80 °C.

RNA quantification

Total RNA can be quantified using the “Qubit RNA HS Assay Kit” according to manufacturer instructions.

Assess DNA contamination in the RNA extraction

DNA contamination can be measured using the Qubit dsDNA HS Reagent.

Removal of DNA contamination from total RNA

Use the DNA-free DNA Removal according to manufacturer instructions for the removal of DNA from RNA samples.

Assess the profile of the extracted RNA

Total RNA profile is determined using the Agilent RNA Screentape following manufacturer instructions except that the samples are not heated at 72 °C.

Spike-In RNA

ERCC (ERCC RNA Spike-In Mix 1) is added during the cDNA synthesis step. Aim to obtain a final percentage of 5 % of the reads assigned to ERCCs assuming that ploy(A) fraction of total RNA is 5 %. Target the sequenced reads of the spiked-in RNA to be 5% of the total amount of sequenced reads). The amount of spiked RNA ($mass_{\text{spiked RNA}}$) that is going to be added in the reaction mix can be calculated as follows:

$$mass_{\text{(spiked RNA)}} = (fraction_{\text{(spiked reads)}} \times fraction_{\text{(target RNA)}} \times mass_{\text{RNA input}}) / Total_RNA_extracted$$

where:

mass_{spiked RNA}: mass (ngs) of spike-in RNA (SIRVs or ERCC) to be added in the sample.

fraction_{spiked reads}: desired fraction of sequenced spike-in RNA reads relative to the total amount of sequenced reads.

fraction_{target RNA}: fraction of the total RNA used in the sample, that is going to be synthesized into cDNA molecules.

mass_{RNA input}: mass (ngs) of RNA input per sample.

Then the volume (µl) of spike-in RNA to be used is calculated as follows:

$$\text{volume}_{(\text{spike-in RNA})} = \text{mass}_{(\text{spike-in RNA})} / (\text{concentration}_{(\text{spike-in RNA})})$$

where

concentration_(spike-in RNA): concentration (ngs/µl) of the spike-in RNAs solution.

volume_{spike-in RNA}: volume (µl) from the spike-in RNAs solution to be added into the sample.

The value for the “mass_{RNA input}” is mass_{RNA input} = 300 ngs (300 ngs of total RNA is used in the cDNA synthesis reactions).

For the ERCC RNA Spike-In Mix 1, the **mass**_{spiked RNA} = 0.45 ngs

The concentration of the stock solutions are:

- The “ERCC RNA Spike-In Mix 1” tube contains 10 µl of ERCC RNAs at a concentration of 103.515 fmoles/ul or 30.3 ng/µl .
- Prepare the appropriate dilution of each Spike-In Mix needed. In the new diluted solution the “mass_{spiked RNA}” for either the “ERCC RNA Spike-In Mix 1” or the “Spike-in RNA Variant (SIRVs) Control set 3 kit” should correspond, *if possible*, to 0.1 ul of the final diluted volume. So we need to have the following dilutions:

For the “ERCC RNA Spike-In Mix 1” we are going to dilute 6.72 times the stock solution. So, in 5.72 µl of “THE RNA solution” add 1 µl from the “ERCC RNA Spike-In Mix 1” stock solution (new concentration= 4.5 ng/ul).

Afterwards we will have to take Volume_{spike-in RNA} = ((0.45 ngs)/(4.5 ng/µl))=0.1 µl of the diluted solution.

cDNA Library generation and sequencing on MinION

Generally, follow the ONT “1D Strand switching cDNA by ligation (SQK-LSK108)” protocol but with custom cDNA synthesis protocol (as described below), and the end repair and d(A) tailing steps are performed separately. An overview of the protocol as follows:

- 1. cDNA synthesis and amplification
- 2. End-repair of cDNA molecules
- 3. dA-tail of cDNA molecules
- 4. Adapter ligation
- 5. Sequencing
- 6. Base-calling

cDNA synthesis

Our cDNA synthesis protocol involves a customized version of the Smart-seq protocol¹. The protocol is based on the terminal deoxynucleotidyl transferase activity of the wild-type MMLV (Moloney murine leukemia virus) reverse transcriptase².

Preparation of Master Mixes

- 1. Thaw and vortex all reagents and keep master mixes on ice until use.
- 2. Label three 1.5 ml eppendorf tubes: “**pre-RT**”, “**RT**”, “**PCR**”
- 3. Always use fresh TSO primer as it is prone to degradation.
- 4. Prepare the “**pre-RT mix**” according to Table 1 below.

Table 1: **pre-RT mix**

pre-RT mix	Total RNA (µl /sample)
------------	------------------------

1	ERCC RNA Spike-In Mix 1	x
2	RNase Inhibitor (40 U/μL * 125 μL = 5000U)	0.05
3	Poly-T primer (stock: 12 μM)	0.7
4	Superscript IV first-strand buffer (5×)	0.4
5	Nuclease free water	0.19
6	dNTP Mix (stock: 10 mM each)	0.56
Total =		2

- Pipette 2 μL of **pre-RT mix** to a PCR tube and add 1uL of sample (**300 ng of total RNA**). Include a negative control (**1 μL of water/RNA buffer**).
- Incubate the samples in a thermocycler set according to Table 2 below.

Table 2: pre-RT incubation

Temperature	Time	Purpose
72°C	3 min	Unfolding of RNA secondary structures, Poly-T primer binding
4°C	10 min	Poly-T primer binds
25°C	1 min	Poly-T primer binds more specifically
4°C	Hold	

- Prepare the “**RT mix**” according to Table 3

	RT mix	μl /sample
1	Nuclease free H ₂ O	0.85
2	Superscript IV first-strand buffer (5×)	0.8
3	DTT (stock: 100 mM)	0.175
4	TSO (stock: 12 μM)	0.7

5	RNAse inhibitor (stock: 40 U/ μ l)	0.175
6	SuperScript IV reverse transcriptase (stock: 200 U/ μ l)	0.35
7	Betaine (stock: 5 M)	0.7
8	MgCl ₂ (stock: 100 mM)	0.25
	Total =	4

8. Following pre_RT incubation, add 4 μ l of RT mix to each sample, mix and briefly spin down.
9. Incubate the samples in a thermocycler set according to Table 4 below

Table 4: **SSIV RT protocol**

Temperature	Time	Cycle	Purpose
50°C	10 min	1	RT and template-switching
55°C	30 sec	10	Unfolding of RNA secondary structures
50°C	30 sec		Completion/continuation of RT
60°C	30 sec	5	Unfolding of RNA secondary structures
55°C	30 sec		Completion/continuation of RT
50°C	30 sec	1	Finish template switching
65°C	30 sec	5	Unfolding of RNA secondary structures
60°C	30 sec		Completion/continuation of RT
50°C	30 sec	1	Finish template switching
70°C	30 sec	5	Unfolding of RNA secondary structures
65°C	30 sec		Completion/continuation of RT
50°C	30 sec	1	Finish template switching
75°C	30 sec	5	Unfolding of RNA secondary structures

70°C	30 sec		Completion/continuation of RT
50°C	1 min	1	Final finish template switching
80°C	10 min	1	Enzyme inactivation
4°C	Hold	1	

10. Prepare the **PCR master mix** according to Table 5

Table 5: PCR master mix

	PCR Mix	(μ l per 7 μ l of RT reaction)
1	PCR-Grade Water	47.6
2	10X Advantage 2 PCR Buffer (Advantage 2 PCR Kit)	7
3	50X dNTP Mix (Advantage 2 PCR Kit)	2.8
4	PCR primer (stock: 12 μ M)	2.8
5	50X Advantage 2 Polymerase Mix (Advantage 2 PCR Kit)	2.8
	Total =	63

11. Following RT incubation, add 63 μ l of PCR mix to each sample, mix and briefly spin down

12. Incubate the samples in a thermocycler set according to Table 6 below

Table 6: PCR protocol

Temperature	Time	Cycle
95°C	1 min	1
95°C	20 sec	5
58°C	4 min	
68°C	6 min	

95°C	20 sec	11 or 12 cycles , aim for ~1-2 µg of cDNA per 70 µl of PCR amplification reaction
64°C	30 sec	
68°C	6 min	
72°C	10 min	1
4°C	Hold	1

13. The amplified product is subsequently cleaned with Agencourt AMPure XP beads as is described below.

Agencourt AMPure XP cleanup of cDNA amplification products

- a. Allow AMPure XP beads to equilibrate to room temperature for at least 30 minutes.
- b. Vortex the beads until evenly mixed, then add 0.9X sample volume of Agencourt AMPure XP beads to the sample in the same tube as used for PCR.
- c. Pipet the entire volume up and down to mix thoroughly. Place the sample tubes on a roler mix for 5 - 8 minutes to let the DNA bind to the beads. Briefly spin the samples to collect the liquid from the side of the tube.
- d. Place the sample tubes on the magnetic separation device for ~2 minutes until the liquid appears completely clear, and there are no beads left in the supernatant.
- e. While the samples are on the magnetic separation device, pipette out the supernatants. Keep the samples on the magnetic separation device. Add 200 µl of freshly made 80% ethanol to each sample without disturbing the beads. Wait for 30 seconds and carefully pipette out the supernatant containing contaminants.
- f. DNA will remain bound to the beads during the washing process. Repeat step 4 once more. Briefly spin the samples to collect the liquid from the side of the wall.
- g. Place the samples on the magnetic device for 30 seconds, then remove all the remaining ethanol with a pipette.
- h. Place the samples at room temperature until the pellet appears dry (~ 5 minutes). You may see a tiny crack in the pellet when it is dry.
- i. Once the beads are dry, add 51 µl of TE buffer to cover the bead pellet.

- j. Remove the samples from the magnetic separation device and mix thoroughly to resuspend the beads. Incubate the sample with rotation at room temperature for 5 – 8 minutes.
- k. Put the tubes on the magnet and after ~2 minutes recover the supernatant which should contain the cleaned amplified cDNA. Determine the quantity of the cDNA and profile using Qubit HS DNA Assay Kit and Agilent D5000 Tapestation, respectively, following manufacturer instructions.

End-repair of DNA

End repair of 1 µg of amplified cDNA is carried out using NEBNext End Repair Module (New England Biolabs, E6050S) following manufacturer instructions. This is followed by 0.9X Ampure XP beads cleanup (described above).

dA-tailing reaction

d(A) tailing of the recovered end-repaired cDNA is carried out using NEBNext dA-Tailing Module (New England Biolabs, E6053S) following manufacturer instructions. This is followed by 0.9X Ampure XP beads cleanup (described above).

Adapter ligation

Ligation of ONT sequencing adapters onto recovered d(A)-tailed cDNA (up to 1 µg) is carried out following ONT SQK-LSK-108 protocol. However, you can increase the incubation time from 10 minutes to 1 - 4 hours at room temperature.

ONT MinION sequencing kit

ONT SQK-LSK-108 protocol is followed for the sequencing part.

Basecalling

Basecalling can be done off-line using Albacore from ONT

Data analysis example commandline arguments

Basecalling

Albacore (ONT, version 2.0.2)

```
read_fast5_basecaller.py -r --flowcell SQK-LSK108 --kit SQK-LSK108 --input %s --save_path %s --
worker_threads 23 -o fastq" %(input_dir,save_path))
```

Minionqc³ (version 1.0)

```
Rscript ~/MinionQC.R -p 23 -i $('pwd')/files -o $('pwd')/results
```

Pauvre (version 0.1.2, <https://github.com/conchoecia/pauvre>)

```
pauvre marginplot --no-transparent --fastq ../Bo_E_1H_C010_10_pass.fastq > pauvre.out 2> pauvre.out
```

Porechop (version 0.2.3, <https://github.com/rrwick/Porechop>)

```
~/porechop --format fasta -t 47 -i $read5.fasta -o $read5.choped.fasta > porechop.stdout 2>
porechop.stdout
```

Cutadapt⁴ poly(A) trimming from read ends (version 1.15)

```
~/local/bin/cutadapt --info-file=trim_info -f fasta -a "A[100]" -o $read2.cutadapt.fasta $read2.fasta
```

GMAP⁵ (version 2018-03-25)

GMAP for alignment QC

```
~/gmap -t 23 -D $dirc -f samse -d $ref $read1 > $outsam.sam
```

GMAP for transcriptome assembly

```
~/gmap -t 23 -D $dirc --cross-species --max-intronlength-ends=10000 -n 1 -z sense_force -f samse -d $ref
$read1 > $outsam1.sam 2> gmap.stdout
```

Minimap2⁶ (version 2.9 (r720))


```
~/minimap2 -ax splice -t 23 $ref $reads1 > $outsam1.sam
```

Samtools⁷ (version 1.3.2)

AlignQC⁸ (version 1.2)

```
~/alignqc analyze $outsam1.sort.bam --specific_tmpdir $dirc/tmp1 -r $ref -a $annotation -o  
alignqc.xhtml --output_folder $dirc/alignQC.ouput_b4_correction > alignqc.stdout
```

Canu⁹ (Canu 1.7)

```
canu useGrid=false -correct gnuplotImageFormat=png corOutCoverage=10000 corMhapSensitivity=high  
corMinCoverage=0 correctedErrorRate=0.16 overlapper=minimap ovsMethod=sequential  
minReadLength=200 minOverlapLength=100 genomeSize=1500000000 -p Bo_E_all_pass_edited -d  
Bo_E_all_pass_edited -nanopore-raw Bo_E_all_pass_edited.fasta
```

LoRDEC¹⁰ (v0.8, using GATB v1.4.1)

```
~/lordec-correct -2 $illumina_reads -T 47 -p -k 19 -s 3 -i $nanopore.fasta -o  
"$nanopore"_lordec_corrected.fasta
```

GFOLD¹¹ (v1.1.4)

```
gfold diff -norm NO -s1 Bo.E.2H -s2 Bo.E.1H -suf .abs_cnt3 -o Bo.E.2HvsBo.E.1H.abs.diff >  
Bo.E.2HvsBo.E.1H.abs.diff.stdout
```

cDNA_Cupcake (version 5.3, https://github.com/Magdoll/cDNA_Cupcake/wiki)

```
~/collapse_isoforms_by_sam.py --input $read1 -s $outsam.sorted.sam --dun-merge-5-shorter -o $pref  
~/filter_by_count.py $pref.collapsed --min_count=2 >filter_by_count.stdout
```

```
~/filter_away_subset.py $pref.collapsed >filter_away_subset.stdout
```

```
~/filter_away_subset.py $pref.collapsed.min_fl_2
```

cDNA_Cupcake for assembly evaluation using 5-hour timepoint

```
~/collapse_isoforms_by_sam.py -c 0.95 -i 0.95 -input $read1 -s $sortedsam --dun-merge-5-shorter -o $pref
```

TAMA (version tc0.0, <https://github.com/GenomeRIK/tama>)

```
~/tama_collapse.py -d merge_dup -s $sortedsam -f $ref -p $pref -x no_cap -c 95 -i 95
```

TAPIS¹² (1.2.1)

```
alignPacBio.py -p 22 -v -K 10000 -o tapis_output $indexesDir $indexName $reference $reads
```

```
run_tapis.py -p -t 30 -o run_tapis_output $annotation tapis_output/$bamfile
```

SQANTI¹³ (version 1.2)

```
sqanti_qc.py -z -t 47 -fl $fl_abundance -c $sj_covillumina -e $isoExpression -x $gmapindex -o $output -d qc_output $isoforms.fa $gtf $ref
```

```
sqanti_filter.py -d filter_output -i "$isoforms"_corrected.fasta "$output"_classification.txt
```

Troubleshooting

Time Taken

Anticipated Results

References

1. Ramskold, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol* **30**, 777-782, doi:10.1038/nbt.2282 (2012).
2. Zajac, P., Islam, S., Hochgerner, H., Lonnerberg, P. & Linnarsson, S. Base preferences in non-templated nucleotide incorporation by MMLV-derived reverse transcriptases. *PLoS One* **8**, e85270, doi:10.1371/journal.pone.0085270 (2013).
3. Lanfear, R., Schalamun, M., Kainer, D., Wang, W. & Schwessinger, B. MinIONQC: fast and simple quality control for MinION sequencing data. *Bioinformatics (Oxford, England)*, doi:10.1093/bioinformatics/bty654 (2018).
4. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal; Vol 17, No 1: Next Generation Sequencing Data Analysis* DO - 10.14806/ej.17.1.200 (2011).
5. Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics (Oxford, England)* **21**, 1859-1875, doi:10.1093/bioinformatics/bti310 (2005).
6. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics (Oxford, England)*, doi:10.1093/bioinformatics/bty191 (2018).
7. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
8. Weirather, J. L. *et al.* Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research* **6**, 100, doi:10.12688/f1000research.10571.2 (2017).
9. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome research* **27**, 722-736, doi:10.1101/gr.215087.116 (2017).
10. Salmela, L. & Rivals, E. LoRDEC: accurate and efficient long read error correction. *Bioinformatics (Oxford, England)* **30**, 3506-3514, doi:10.1093/bioinformatics/btu538 (2014).
11. Feng, J. *et al.* GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics (Oxford, England)* **28**, 2782-2788, doi:10.1093/bioinformatics/bts515 (2012).
12. Abdel-Ghany, S. E. *et al.* A survey of the sorghum transcriptome using single-molecule long reads. *Nature communications* **7**, 11706, doi:10.1038/ncomms11706 (2016).
13. Tardaguila, M. *et al.* SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. *Genome research*,

doi:10.1101/gr.222976.117 (2018).