Sep 03, 2024

Real-time and programmable transcriptome sequencing with PROFIT-seq

DOI
dx.doi.org/10.17504/protocols.io.5jyl8p19rg2w/v1

lingling hou[1], Jinyang Zhang[1]

[1]Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

Jinyang Zhang
Institute of Zoology

OPEN ACCESS

DOI: dx.doi.org/10.17504/protocols.io.5jyl8p19rg2w/v1

Protocol Citation: lingling hou, Jinyang Zhang 2024. Real-time and programmable transcriptome sequencing with PROFIT-seq. **protocols.io https://dx.doi.org/10.17504/protocols.io.5jyl8p19rg2w/v1**

Manuscript citation:
Real-time and programmable transcriptome sequencing with PROFIT-seq

Protocol status: Working
We use this protocol and it's working

Created: October 17, 2023

Last Modified: September 03, 2024

Protocol Integer ID: 89376

Keywords: non-coding RNA, nanopore adaptive sequencing

Funders Acknowledgement:
National Key R&D Project
Grant ID: 2021YFC2301300

Disclaimer

DISCLAIMER – FOR INFORMATIONAL PURPOSES ONLY; USE AT YOUR OWN RISK

The protocol content here is for informational purposes only and does not constitute legal, medical, clinical, or safety advice, or otherwise; content added to **protocols.io** is not peer reviewed and may not have undergone a formal approval of any kind. Information presented in this protocol should not substitute for independent professional judgment, advice, diagnosis, or treatment. Any action you take or refrain from taking using or relying upon the information presented here is strictly at your own risk. You agree that neither the Company nor any of the authors, contributors, administrators, or anyone else associated with **protocols.io**, can be held responsible for your use of the information contained in or linked to this protocol or any of our Sites/Apps and Services.

Abstract

PROgrammable Full-length Isoform Transcriptome sequencing (PROFIT-seq) is a method that enriches target transcripts while maintaining unbiased quantification of the whole transcriptome. PROFIT-seq employs combinatorial reverse transcription to capture polyadenylated, non-polyadenylated, and circular RNAs, coupled with nanopore adaptive sampling that selectively enriches target transcripts during sequencing.

Guidelines

- Handle RNA and enzymes on ice whenever possible.
- All reagents must be kept nuclease-free.
- Use only molecular grade nuclease-free water.
- Perform the incubation steps in the thermal cycler.
- When performing multiple reactions, use a master mix containing 10% excess.
- If not mentioned otherwise, use a 1:1 ratio of beads to sample to select nucleic acids >150-200 nt in clean-ups step using magnetic beads.

Materials

General materials:
- RNaseZap (Thermo Fisher Scientific, cat. No. AM9782)
- AMPure XP (Beckman, cat. No. A63882)
- RNAClean XP (Beckman, cat. No. A63987)
- Nuclease-free water (UltraPure DNase/RNase-free water, Thermo Fisher Scientific, cat. No. 10977015)
- Ethanol, pure 200 proof, for molecular biology (Sigma-Aldrich, cat. No. E7023-500mL)
- 25-ml reagent reservoir for eight-channel pipettes (VWR, cat. No. BITXSR-0025-BWM)
- 1.5ml DNA LoBind tubes (Eppendorf, cat. Nos. 0030108051)
- 1.5 mL tubes (Axygen, cat. No. MCT-150-C)
- 0.2 mL PCR tubes (Axygen, cat. No. PCR-02-C)
- 0.2 mL 8-strip PCR Tubes and Domed Cap Strips (Axygen, cat. No. PCR-0208-CP-C)
- 10-, 100-, 200- and 1000µl filter barrier tips pipette tip, filter, sterile (Axygen, cat. nos. TF-300-R-S, TF-100-R-S, TF-200-R-S, TF-1000-L-R-S)
- Eppendorf ThermoMixer C with Thermo top (Eppendorf, mod no. 5382000023)
- Tube Rotator (MACSmix, mod no.130-090-753)
- DynaMag-2 magnet for 1.5-ml microtube (Thermo Fisher Scientific, mod no. 12321D)
- DynaMag-96 Magnets (Thermo Fisher Scientific, mod no. 12331D)
- Refrigerated centrifuge (e.g.: Eppendorf, mod no. 5430R)
- Bench top centrifuge (e.g.: Eppendorf, mod no. minispin)
- Thermocycler (e.g.: Biometra mod no. TRIO 48)
- Minicentrifuge (e.g.: Kelyn-Bell, mod no. LX-200)
- Deoxynucleotide (dNTP) Solution Mix (10mM) (New England Biolabs, cat. No. N0447L)

For ribosomal RNA depletion:
- KAPA RiboErase Kit (HMR) Human/Mouse/Rat (KAPA Biosystems, cat. No. KK8481/2)

For combinatorial reverse transcription:
- SMARTer™ PCR cDNA Synthesis Kit (Clonetech, cat. No. 634925 & 634926)
- NEBNext Quick Ligation Module (NEB, cat. No. E5056)

For cDNA amplification:
- LongAmp™ Taq 2X Master Mix (NEB, cat. No. M0287)

For splint-based circularization:
- NEBuilder HiFi DNA Assembly Reaction (NEB, cat. No. E2621S)
- Exonuclease I (E. coli) (NEB, cat. No. M0293L)
- Exonuclease III (E. coli) (NEB, cat. No. M0206L)

For rolling circle amplification (RCA):
- phi29 DNA Polymerase (NEB, cat. No. M0269L)

For Debranching of the RCA products:
- T7 Endonuclease I (NEB, cat. No. M0302L)

For BluePippin size selection:
- BluePippin reagents and Agarose Cassettes(0.75% (w/v) agarose gel cassette, dye-free, S1 marker, low range, 1-10 kb)，SageScience, cat. No. BLF7510)

For nanopore sequencing library construction
- Nanopore ligation sequencing kit (Nanopore, cat. No. LSK109+NBD104/114)
- Nanopore sequencer (Oxford Nanopore Technologies, MinION / GridION / PromethION)

For quality control:
- Qubit double-stranded DNA HS assay kit (Thermo Fisher Scientific, cat. No. Q32854)
- Qubit RNA HS assay kit (Thermo Fisher Scientific, cat. No. Q32852)
- Qubit assay tubes (Thermo Fisher Scientific, cat. No. Q32856)
- HS NGS Fragment Kit (1-6000bp) (Agilent, cat. No. DNF-474-0500)
- HS Genomic DNA Kit (Agilent, cat. No. Agilent DNF-488-0500)
- HS RNA Kit (15NT) (Agilent, cat.no. DNF-472-0500)
- Qubit 3.0 fluorometer (Thermo Fisher Scientific, mod no. Q33216) with kit or Qubit 4.0 fluorometer, Thermo Fisher Scientific, mod no Q33238)
- 5200 Fragment Analyzer systems (Agilent, mod no. M5310AA), Bioanalyzer 2100 (Agilent, mod no. M5310AA), Tapestation (Agilent, mod no. M5310AA), or other equivalent systems.
- Nanodrop 2000 (Thermo Fisher Scientific, mod no ND2000)

Primer used in this protocol:

	Primer name	Primer sequence
	RT.dT	5'-CTACACGACGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTVN-3'
	RT.N6	5'-CTACACGACGCTCTTCCGATCTNNNNNN-3'
	RT.ds.rv	5'-AGATCGGAAGAGCGTCGTGTAG-3'
	R.P	5'-CTACACGACGCTCTTCCGATCT-3'
	F.P	5'-AAGCAGTGGTATCAACGCAGAGTAC-3'
	TSO (SMARTer kit)	5'–AAGCAGTGGTATCAACGCAGAGTACXXXXX–3'
	RT.linker	5'-AGATCGGAAGAGCGTCGTGTAGTGAGGCTGATGAGTTCCATANNNNNTATATNNNNNNATCACTACTTAGTTTTTTGATAGCTTCAAGCCAGAGTTGTCTTTTTCTCTTTGCTGGCAGTAAAAG-3'
	TSO.linker	5'-CTCTGCGTTGATACCACTGCTTAAAGGGATATTTTCGATCGCNNNNNATATANNNNNTTAGTGCATTTGATCCTTTTACTCCTCCTAAAGAACAACCTGACCCAGCAAAA

	Primer name	Primer sequence
		GGTACACAATACTTTTACTGCCAGCAAAGAG-3'
	Linker.R	5'-CTCTGCGTTGATACCACTGCTT-3'
	N6R 3'-terminal phosphorothioate (PTO) modifications	5'-NNNN*N*N-3'

Before start

All experimental procedures including reagent preparation should be performed under RNase- and DNase-free conditions. Bleach the workbench with RNaseZap before starting the experiment. Change gloves frequently to avoid RNase contamination.

Primer:
- Order 100 nmol of DNA oligo, PAGE purification.
- Dissolve the primer powder in nuclease-free water to a final concentration of 50 μM.
- The dissolved primers can be stored at -20°C for months.

- RT.dT.ds(1.4uM) can be obtained by annealing RT.dT (50 μM) and RT.ds.rv (50 μM) 1:1 at 1.4 μM in anealing buffer (10 mM Tris-HCl pH 7.5, 50 mM NaCl).
- RT.N6.ds(1.4uM) can be obtained by annealing RT.N6 (50 μM) and RT.ds.rv (50 μM) 1:1 at 1.4 μM in anealing buffer (10 mM Tris-HCl pH 7.5, 50 mM NaCl).

DNA Splint preparation:
- The DNA splint was generated by primer extension of four oligos: RT.linker RT.linker, TSO Linker.R (50 μM), TSO and Linker.R.

Ribosomal RNA depletion

2h 30m

1 **Remove ribosomal RNA from extracted RNA samples:**

0.2 Use 1 µg total RNA (RNA Quality Number (RQN) ≥ 7.0) as starting material.

0.3 Equilibrate the RNAClean XP beads for at least 00:30:00 at Room temperature, and vortex thoroughly to resuspend the beads. 30m

0.4 Remove ribosomal RNA (rRNA) using an RNase H-based commercial kit – RiboErase kit (human/mouse/rat) (KAPA Biosystems, KK8481) according to the manufacturer's user guide, and elute RNA in 15 µL nuclease-free water.

0.5 Aspirate 2 µL RNA for quantification with a Qubit fluorometer to assess the recovery after rRNA
depletion.

0.6 Aspirate another 1 µL RNA for detecting whether mRNA is degraded or not with 5200 Fragment
Analyzer systems (Agilent).

Reverse transcription

1 **Combinatorial reverse transcription using RT.dT.ds / RT.N6.ds / RT.N6.ss primers:** 3h 30m
Capture polyadenylated / non-polyadenylated and circular transcriptome using combinatorial RT primers

1.1 Adjust the volume of 1 µg RNA to 10 µL with nuclease-free water.

1.2 Mix by pipetting and spin down briefly in a microcentrifuge.

1.3 Incubate in a hot-lid thermal cycler at 65 °C for 00:05:00, then place on ice immediately. 5m

1.4 Prepare the following reagents in a 0.2 mL RNase/DNase-free PCR tube:

	Reagent	Volume (ul)
	rRNA depleted RNA	10
	10X NEBNext Quick Ligation Reaction Buffer	3
	RT.dT.ds oligods (1.4uM)	0.5
	T4 DNA Ligase	1.5
	Total volume	15

1.5 Mix by pipetting and spin down briefly in a microcentrifuge.

1.6 Incubate at 25 °C for 00:10:00 . 10m

1.7 Add 0.5 µL of [M] 1.4 micromolar (µM) **RT.N6.ds** oligos to the mixture. 10m

1.8 Mix by pipetting, and spin the tubes briefly in a microcentrifuge.

1.9 Incubated at 25 °C for 00:10:00 . 10m

1.10 Prepare the following mix containing the components listed below in a 0.2 mL RNase/DNase-free PCR tube:

	Reagent	Volume (ul)
	Nuclease-free water	10
	5x first-strand buffer	8
	DTT (100mM)	2
	dNTP (10mM)	2
	RNase inhibitor	1
	TSO	1
	SMARTer reverse transcriptase (100U)	1

	Reagent	Volume (ul)
	Total volume	25

Total volume: 25uM

1.11 Mix by pipetting, and spin the tubes briefly in a microcentrifuge.

1.12 Place the tube in a thermal cycler and Incubate at 42 °C for 01:00:00 . 1h

1.13 Add 0.5 µL of RT.N6.ss ([M] 50 micromolar (µM)) to the reaction

1.14 Mix by pipetting, and spin the tubes briefly in a microcentrifuge.

1.15 Place the tube in the thermal cycler and incubate at 25 °C for 00:10:00 , 42 °C for 01:00:00 , and then 70 °C for 00:10:00 , and bring the sample to 4 °C before proceeding to the next step. 1h 20m

2 **Clean-up:** 30m

2.1 Vortex thoroughly to resuspend the AMPure XP beads before use.

2.2 Add 40 µL AMPure XP (Beckman, A63880) beads to the 40 µL RNA sample from step 2.15 and mix the sample well by pipetting up and down and spinning down briefly.

2.3 Incubate at Room temperature for 00:05:00 to bind RNA to the beads. 5m

2.4 Place the 0.2 mL tube on the magnetic stand until the solution is clear (~ 00:02:00). Keep the tube on the magnetic stand. Remove supernatant carefully, taking care not to disturb the beads. 2m

2.5 Add 200 µL of freshly prepared [M] 75 % (v/v) ethanol to the tube. Wait for 00:00:30, then discard the entire supernatant. 30s

▲ **CRITICAL STEP** Beads should always be kept on the magnetic stand while washing with ethanol, and should not be resuspended.

2.6 Wash beads once more with [M] 75 % (v/v) ethanol by repeating step 3.5 go to step #2.5 .

2.7 Quickly spin the tubes in a minicentrifuge and remove all residual liquid.

2.8 Air-dry the beads on the magnetic rack for 00:00:30 . 30s

▲**CRITICAL STEP** Beads should not be overdried during the process, because cracked beads will decrease the RNA recovery.

2.9 Resuspend beads with 10 µL nuclease-free water.

2.10 Incubate at room temperature for 00:02:00 to elute RNA from the beads. 2m

2.11 Put the tube back on the magnetic rack for 00:01:00 to pellet the beads. 1m

2.12 Transfer 8 µL supernatant to a new 0.2 mL tube. Take care not to disturb the bead pellet. The purified cDNA can be stored at -20 °C for several weeks.

3 **Pre-circularization amplification:** 1h

3.1 Add the following components to a 0.2 mL PCR tube on ice:

	Reagent	Volume (ul)	Final concentration
	LongAmp Taq 2x Master Mix (NEB: M0287)	25	1X
	10 uM TSO.F	2	0.4 uM

	Reagent	Volume (ul)	Final concentration
	10 uM 10x.R	2	0.4 uM
	Template cDNA	5	
	Nuclease-free water	16	
	Total volume:	50	

3.2 Mix by pipetting, and spin the tubes briefly in a microcentrifuge.

3.3 Transfer the PCR tubes to a thermocycler and begin the program below:

	Number of cycles	Denature	Anneal	Extend
	1	95 °C, 3 mins	-	-
	14-15	95 °C, 10 s	60 °C, 20 s	65 °C, 60 s
	1	-	-	65 °C, 1 min
	Hold	4-10 °C		

4 **Clean-up** 30m

4.1 Purification the amplified cDNA with AMPure XP beads as described before using a 0.8:1 ratio (40 µL beads). Elute cDNA in 10 µL .

4.2 The cDNA can be stored at -20 °C for several weeks.

Circularization & rolling circle amplification

5 **Splint-based circularization:** 1h
The cDNA libraries were constructed using the **10X R2C2 protocol**.

CITATION

Volden R, Vollmers C (2022). Single-cell isoform analysis in human immune cells..
LINK
https://doi.org/10.1186/s13059-022-02615-z

5.1 Set up the following reaction mix on ice:

	Reagent	Volume (ul)
	Amplified cDNA (100ng)	10
	DNA splint (100ng)	
	Deionized H2O	
	2X NEBuilder HIFI DNA Assembly Master Mix (NEB: E2621S)	10
	Total volume	20

5.2 Incubate at 50 °C for 01:00:00 . 1h

6 **Digestion of non-circularized products:** 1h

6.1 Add 1 µL Exonuclease I (20 U / µl) and 0.3 µL Exonuclease III (100 U / µl) to digest non-circular DNA.

6.2 Mix by pipetting, and spin the tubes briefly in a microcentrifuge.

6.3 Incubate at 37 °C for 01:00:00 . 1h

7 **Clean-up:** 30m

7.1 Purification cDNA with AMPure XP beads as described before.

7.2 Elute in 10 µL of elution buffer ([M] 10 millimolar (mM) Tris, pH 8.0).

7.3 The cDNA library can be stored at -20 °C for several weeks.

8 **Rolling circle amplification (RCA):** 13h

8.1 Set up the following reaction mix on ice:

	Reagent	Volume (ul)
	10X Phi29 Buffer	20
	10 mM dNTP	10
	exonuclease-resistant random hexamers NNNN*N*N (100 µM) with two 3'-terminal phosphorothioate (PTO) modifications	4
	BSA	2
	DNA	8
	H2O	152
	Phi29 DNA polymerase (NEB: M0269L)	4
	Total volume	200

8.2 Divide the mixture into 4 tubes of 50 µL reaction.

8.3 Incubated at 30 °C Overnight , then 65 °C for 00:20:00 . 13h 20m

9 **Clean-up:** 30m

9.1 Mix each two tubes of sample into one tube.

9.2 Purification cDNA with AMPure XP beads as described before using a 0.5:1 ratio (50 µl beads).

9.3 Elute in T7 Endonuclease I reaction mix below.

10 **Debranching of the RCA products:** 30m

10.1 Add the T7 Endonuclease I (NEB: M0302L) reaction mix to the beads above to debranch and elute the RCA product:

	Reagent	Volume (ul)
	ultrapure water	52
	NEB buffer 2	6
	T7 Endonuclease I (10 units/µL)	2
	Total volume	60

10.2 Incubate on a thermal shaker at 37 °C for 00:30:00 under constant agitation at 1800 rpm. 30m

10.3 Place the tubes on a magnet, and aspirate the supernatant into a new tube.

11 **Clean-up** 30m

11.1 Purification the DNA in supernatant with AMPure XP beads as described before using a 0.5:1 ratio (30 µL beads).

11.2 Elute in 15 µL of elution buffer ([M] 10 millimolar (mM) Tris, pH 8.0).

11.3 The debranched RCA products can be stored at -20 °C for several weeks.

12 **Size selection of the RCA products**

12.1 Warm the agarose gel cassettes and reagents to room temperature, vortex and spin the loading solution briefly, and flick and briefly spin the S1 marker.

12.2 Add 40 µL running buffer to the elution wells following the general guidelines for preparing samples and cassettes as described in the BluePippin User Guide. Samples can be vortexed briefly to properly mix them with the loading solution.

12.3 Set the program to recover DNA fragments larger than 7 kb using the BluePippin (Sage Science)
and BLF7510 cassette with the '0.75DF 1-10 kb Marker S1-Improved Recovery' cassette definition.

13 **Nanopore sequencing library construction**

13.1 Prepare the cDNA libraries following the Oxford Nanopore protocol "**Ligation sequencing amplicons - native barcoding (SQK-LSK109 with EXP-NBD104 and EXP-NBD114**)". The specific steps include end-prep, native barcode ligation and nanopore adapter ligation, and flow cell priming and loading. The updated barcoding kit (SQK-LSK114) can replace SQK-LSK109 for this experiment.

Note

Make sure the DNA meets the quantity and quality requirements specified by the manufacturers. Using too little or too much DNA will affect your library preparation. 100–200 fmol input DNA is required for each sample when using R9.4.1 flow cells. The online tool NEBioCalculator (https://nebiocalculator.neb.com/#!/dsdnaamt) is recommended to convert dsDNA mass to moles.

Using different barcodes for different samples is critical for demultiplexing pooled sequencing reads.

A 0.5:1 AMPure beads-to-sample ratio is recommended to select against short fragments after sequencing adapter ligation.

The wash buffer SFB, not LFB which is designed to enrich DNA fragments >3 kb, must be used to purify all fragments equally.

Load 30-50 fmol DNA library for R9.4.1 flow cells.

4. Nanopore sequencing

14 Install and configuration MinKNOW

14.1 Install MinKNOW following the **instructions from nanopore community**

14.2 Backup the original sequencing configuration file:

```
cd /opt/ont/minknow/conf/package/sequencing
cp sequencing_MIN106_DNA.toml  sequencing_MIN106_DNA.toml.bak
```

14.3 Edit the following line to change reads break intervals:

```
[analysis_configuration.read_detection]
mode = "transition"
minimum_delta_mean = 80.0
look_back = 2
break_reads_after_events = 250
break_reads_after_seconds = 0.4 # Change it from 1.0 to 0.4
break_reads_on_mux_changes = true
open_pore_min = 150.0
open_pore_max = 250.0
```

15 Install PROFIT-seq

15.1 Install anaconda3

```
# Install system libraries
sudo apt install zlib1g-dev libncurses5-dev libbz2-dev liblzma-dev

# Install Mambaforge3
wget https://github.com/conda-
forge/miniforge/releases/latest/download/Mambaforge-pypy3-Linux-
x86_64.sh
bash Mambaforge-pypy3-Linux-x86_64.sh
```

15.2 Create virtual environment for PROFIT-seq

```
mamba create -p /home/biols/envs/python3.8.10 python==3.8.10
mamba activate /home/biols/envs/python3.8.10
```

15.3 Install PROFIT-seq

```
pip install --recursive https://github.com/bioinfo-biols/PROFIT-
seq.git
cd PROFIT-seq

# Install python dependencies
pip install -r requirements.txt
```

PROFIT-seq is based on ReadUntil API

CITATION

Payne A, Holmes N, Clarke T, Munro R, Debebe BJ, Loose M (2021). Readfish enables targeted nanopore sequencing of gigabase-sized genomes..

LINK

https://doi.org/10.1038/s41587-020-00746-x

CITATION

Loose, M., Malla, S., & Stout, M. (2016). Real-time selective sequencing using nanopore technology. Nature methods.
LINK
https://doi.org/10.1038/nmeth.3930

15.4 Install other dependencies:

```
mamba install bioconda::porechop
mamba install bioconda::minimap2
mamba install bioconda::samtools
mamba install bioconda::stringtie
mamba install bioconda::gffread
mamba install bioconda::salmon
```

CITATION

Li H (2018). Minimap2: pairwise alignment for nucleotide sequences..
LINK
https://doi.org/10.1093/bioinformatics/bty191

CITATION

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H (2021). Twelve years of SAMtools and BCFtools..
LINK
https://doi.org/10.1093/gigascience/giab008

CITATION

Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2..
LINK
https://doi.org/10.1186/s13059-019-1910-1

CITATION

Pertea G, Pertea M (2020). GFF Utilities: GffRead and GffCompare..
LINK
https://doi.org/pii:ISCBCommJ-304.10.12688/f1000research.23297.2

CITATION

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017). Salmon provides fast and bias-aware quantification of transcript expression..
LINK
https://doi.org/10.1038/nmeth.4197

16 **Prepare reference genome and annotation:**

Prepare the minimap2 index of reference genome

```
# For example, for human samples:
wget
https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_
37/GRCh38.p13.genome.fa.gz
gunzip GRCh38.p13.genome.fa.gz
minimap2 -t 6 -x splice -d GRCh38.p13.genome.fa.mmi
GRCh38.primary_assembly.genome.fa
wget
https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_
37/gencode.v37.annotation.gtf.gz
```

17 **Connect MinION Mk1B and insert the FLO-MIN106D flow cell correctly:**

Open the MinKNOW software to make sure flow cell have been successfully recognized.

18 **Run PROFIT-seq**

18.1 Switch to minknow user to start PROFIT-seq process:

```
# Switch to minknow user
sudo su - minknow
bash && source /home/biols/.bashrc

# Activate PROFIT-seq environment
mamba activate /home/biols/envs/python3.8.10

# Run PROFIT-seq
python3 PROFIT-seq.py \
    --mm_idx
/home/biols/data/hg38/GRCh38.primary_assembly.genome.fa.mmi \
    --gtf /home/biols/data/hg38/gencode.v37.annotation.gtf
```

Note

Always use the fast basecalling config when running PROFIT-seq to avoid performance issues.

Note

The PROFIT-seq need to run with minknow user to grant the right permission.

Note

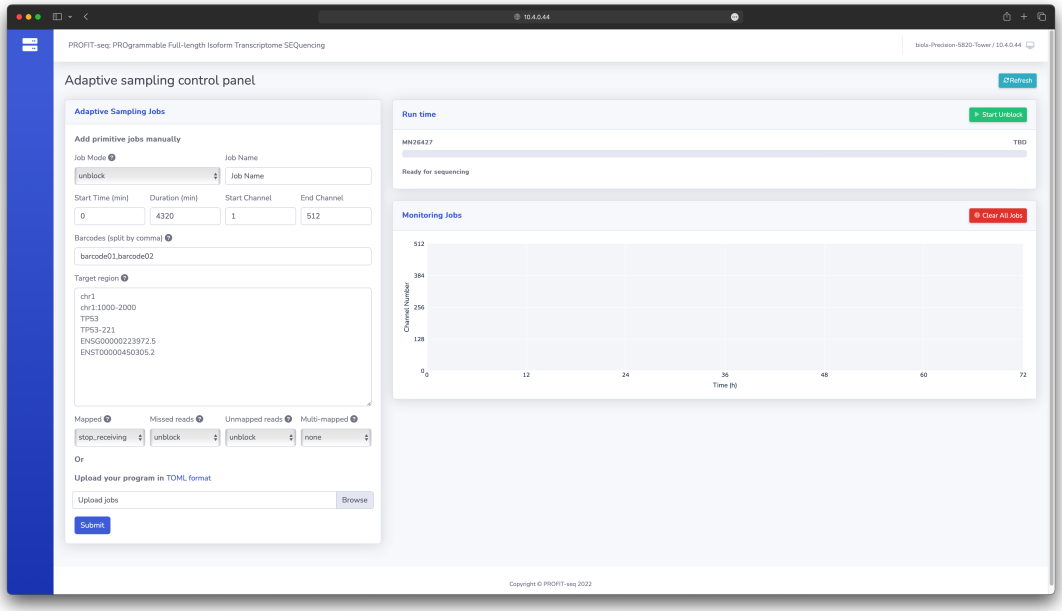
Usage: PROFIT-seq.py [OPTIONS]

Options:
 --minknow_host TEXT ip address for MinKNOW host. Defaults to to 127.0.0.1.
 --minknow_port TEXT port for MinKNOW service. Defaults to 8000.
 --guppy_address TEXT address for guppy server. Defaults to ipc:///tmp/.guppy/5555.
 --guppy_config TEXT guppy basecalling config. Defaults to dna_r9.4.1_450bps_fast.
 --dashboard_port TEXT guppy basecalling config. Defaults to 55280.
 --mm_idx TEXT Minimap2 index of reference sequences [required]
 --version Show the version and exit.
 --help Show this message and exit.

18.2 If everything works fine, the prompt of url for dashboard will appear on your screen. For example:

```
(python3.8.10) minknow@biols-Precision-5820-
Tower:/home/biols/git/PROFIT-seq$  python3 PROFIT-seq.py --mm_idx
/home/biols/data/hg38/GRCh38.primary_assembly.genome.fa.mmi --gtf
/home/biols/data/hg38/gencode.v37.annotation.gtf
[Mon 2024-09-02 19:57:16] [INFO ] Connected to MinKNOW server
[Mon 2024-09-02 19:57:16] [INFO ] Connected to Guppy server
[Mon 2024-09-02 19:57:47] [INFO ] Loaded reference index
[Mon 2024-09-02 19:57:54] [INFO ] Loaded annotation gtf
[Mon 2024-09-02 19:57:54] [INFO ] Starting PROFIT-seq dashboard
 * Serving Flask app 'app.server'
 * Debug mode: off
WARNING: This is a development server. Do not use it in a
production deployment. Use a production WSGI server instead.
 * Running on all addresses (0.0.0.0)
 * Running on http://127.0.0.1:55280
 * Running on http://10.4.0.44:55280
Press CTRL+C to quit
```

18.3 Access the control panel from the given URL (in this case: http://10.4.0.44:55280)



18.4 Configure the adaptive sampling jobs using the "Adaptive Sampling Jobs" panel or upload jobs with the TOML format.

Note

Example of a valid TOML config:

```
[[jobs]]
name = "Unblock_mt"time = [0, 240]
ch = [129, 256]
bc = "all"target = [
    {region = ["chrM", "0", "16569"], action="enrich"},
    {region = "multi", action="unblock"},
    {region = "miss", action="unblock"},
    {region = "unmapped", action="wait"},
]
```

Basic options:

```
- name: User-specified name for each target job.
- time: range of start and end time (minutes) for the job.
- ch: range of start and end channel for the job. (0-512 for
MinION).
- bc: barcode for the job.
- target: specify what actions should be performed when
aligning to specific region.
```

Available barcode options:

```
- barcode01,barcode02 (comma-seperated list of barcode names,
only reads with these barcodes will be processed)
- classified (all reads with classified barcodes will be
processed)
- unclassified (all reads with unclassified barcodes will be
processed)
- all (all reads will be processed)
```

Available target options:

```
# Priority: all > unmapped > mapped > multi > region
- chrom:start-end (reads that mapped to spefici region will be
processed)
- multi (reads that are multi-mapped will be processed)
- mapped (all reads mapped to the reference index will be
processed)
- miss (reads mapped to the reference index, but missed any
target regions will be processed)
- unmapped (all reads that could not be aligned to the
reference index will be processed)
- all (all reads with be processed)
```

Available action options:

```
# action
- stop_receiving (finish sequencing this read)
- unblock (reject this read)
- wait (wait for decision in the next chunk)
- balance (balance coverage for all target regions with action
`balance`)

At lease one of the following combination of actions are
required for a valid job
1. unmapped + mapped
2. unmapped + regions + miss
```

18.5 Start sequencing protocol in MinKNOW.

18.6 Click 'Run unblock' in PROFIT-seq control panel to start pore manipulation.

5. Data analysis

19 **Raw data basecalling:**
Re-basecall the sequenced reads with hac model

```
# Switch to minknow user
sudo su - minknow

# User guppy for basecalling
/opt/ont/guppy/bin/guppy_basecall_client \
    -r --input_path path_to_input --save_path path_to_output \
    -c dna_r9.4.1_450bps_hac.cfg \
    --port ///tmp/.guppy/5555 \
    --barcode_kits "EXP-NBD114" \
    --compress_fastq

# Demultiplex different barcodes
/opt/ont/guppy/bin/guppy_barcoder \
    -i path_to_output/pass \
    -s path_to_output/barcoded \
    --compress_fastq \
    --disable_pings
```

20 **Channel demultiplex (optional):**

If various adaptive sampling jobs are assigned to different channels, use scripts/step1_demultiplex.py to demultiplex reads from different sequencing channels.

```
python3 scripts/step1_demultiplex.py -i input.fastq.gz -o
sample1.fastq.gz --start 1 --end 256
```

Note

Usage: step1_demultiplex.py [OPTIONS]

Options:
-i, --infile TEXT input gzipped fastq file. [required]
-o, --outfile TEXT output gzipped fastq file. [required]
-st, --start INTEGER start channel number. [required]
-en, --end INTEGER end channel number. [required]
-t, --threads INTEGER number of threads.

21 **Adapter trimming & consensus calling:**

21.1 Trim sequencing barcodes using porechop:

```
porechop -i sample1.fastq.gz -o sample1.trimmed.fastq.gz --threads
32 --check_reads 1000
```

21.2 Get consensus reads using scripts/step2_consensus.py

```
python3 scripts/step2_consensus.py \
    -i sample1.trimmed.fastq.gz \
    -s sequencing_summary_FAQ85160_399ee876.txt \
    -o ./output_sample1 \
    -p sample1 \
    -t 16 \
    --trimA
```

Note

Usage: step2_consensus.py [OPTIONS]

Options:
 -i, --input PATH input trimmed fastq. [required]
 -s, --summary PATH input sequencing summary generate by MinKNOW. [required]
 -o, --outdir PATH output directory. [required]
 -p, --prefix TEXT output prefix name. [required]
 -r, --adapter PATH Adapter sequences file. Defaults to embedded splint adapter sequences.
 -t, --threads INTEGER number of threads. Defaults to number of cpu cores.
 --trimA trim 3' poly(A) sequences

22 **Isoform assembly & quantification:**

Perform downstream analysis using scripts/step3_analysis.py:

```
python step3_analysis.py \
    -i ./output_sample1 \
    -p sample1 \
    -r GRCh38.primary_assembly.genome.fa \
    -a gencode.v37.annotation.gtf \
    -t 16 \
    --assemble \
    --bed ../cancer_panel.bed
```

Note

Usage: step3_analysis.py [OPTIONS]

Options:
-i, --workspace PATH directory of step2_consensus.py output [required]
-p, --prefix TEXT sample prefix for step2_consensus.py [required]
-r, --genome PATH reference genome fasta. [required]
-a, --gtf PATH gene annotation gtf. [required]
-b, --bed PATH bed file for target regions. [required]
-t, --threads INTEGER number of threads. Defaults to number of cpu cores.
--assemble perform transcript isoform assemble.
--help Show this message and exit.

23 **Output files:**

	Filename	Description
	workspace/prefix.fl.fa	Full-length consensus reads
	workspace/prefix_isoforms.genes.sf	Gene-level quantification results
	workspace/prefix_isoforms.gtf	Isoform annotations
	workspace/prefix_isoforms.transcripts.sf	Transcript-level quantification results
	workspace/prefix.recovered.fa	Non-full-length sequence

Citations

Step 15.3

Payne A, Holmes N, Clarke T, Munro R, Debebe BJ, Loose M. Readfish enables targeted nanopore sequencing of gigabase-sized genomes.
https://doi.org/10.1038/s41587-020-00746-x

Step 15.3

Loose, M., Malla, S., & Stout, M. . Real-time selective sequencing using nanopore technology
https://doi.org/10.1038/nmeth.3930

Step 15.4

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression.
https://doi.org/10.1038/nmeth.4197

Step 15.4

Li H. Minimap2: pairwise alignment for nucleotide sequences.
https://doi.org/10.1093/bioinformatics/bty191

Step 15.4

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools.
https://doi.org/10.1093/gigascience/giab008

Step 15.4

Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2.
https://doi.org/10.1186/s13059-019-1910-1

Step 15.4

Pertea G, Pertea M. GFF Utilities: GffRead and GffCompare.
https://doi.org/pii:ISCBCommJ-304.10.12688/f1000research.23297.2

Step 5

Volden R, Vollmers C. Single-cell isoform analysis in human immune cells.
https://doi.org/10.1186/s13059-022-02615-z