LongTREC

Funded by the European Union

Bioinformatics Summer School

Long-reads Transcriptomics

Satrio Wibowo

University of Nottingham,
Nottingham, UK

LongTREC - The Long-reads TRanscriptome European Consortium
Marie Skłodowska-Curie grant agreement No 101072892

Course Contents

1. Library Preparations
2. Basecalling

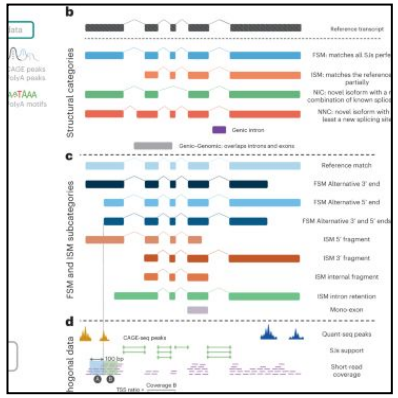
Section 1

Library preparation (lib-prep)

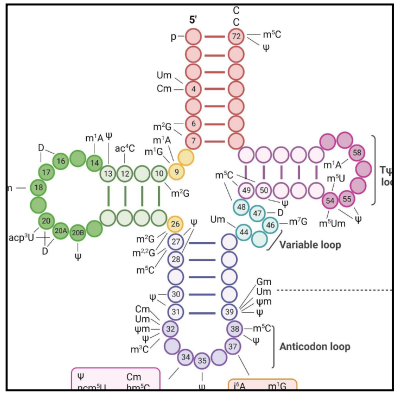
Library preparation strategies for long-reads RNA-seq experiments

Knowing your goal

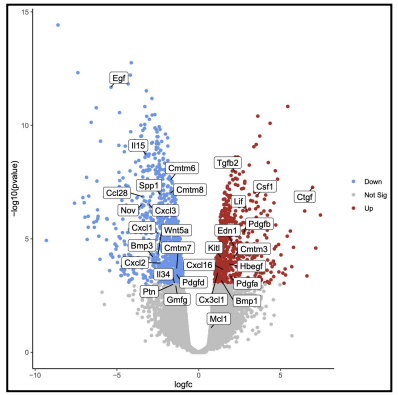
What is your goal?



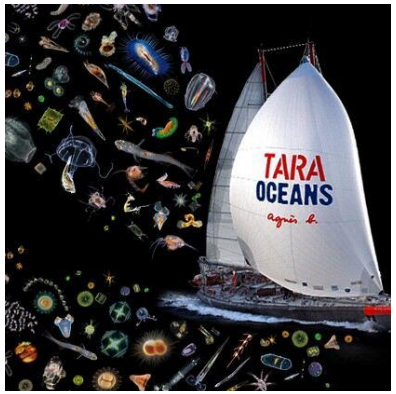
Transcriptomics



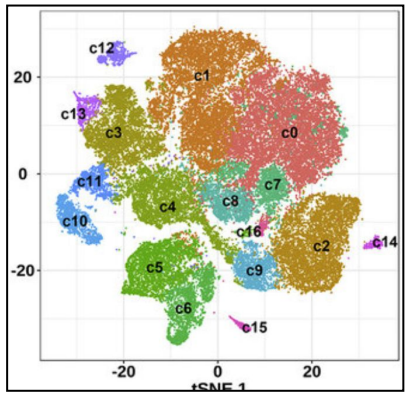
Epi-transcriptomics



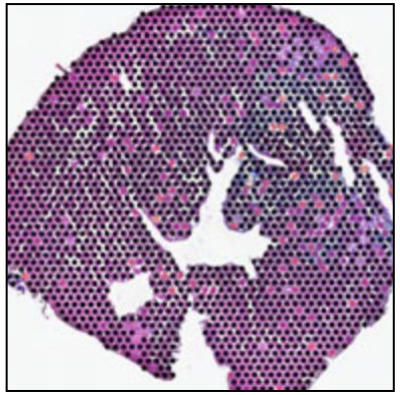
Bulk



Meta-transcriptomics



Single-cell



Spatial

RNA-seq

Main difference with short-reads lib-prep

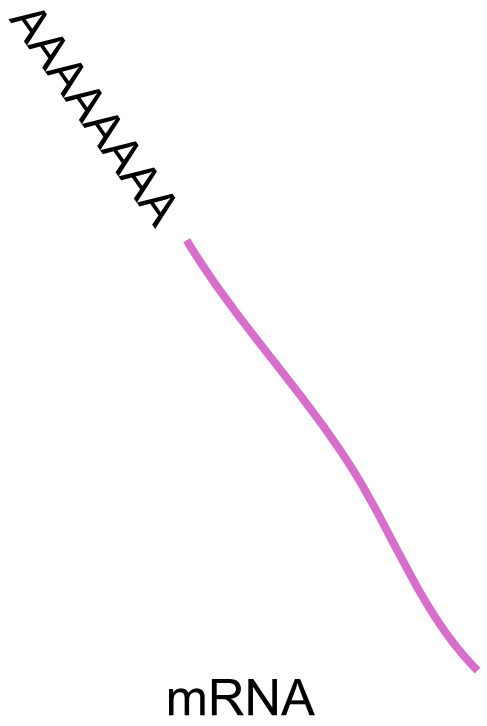


Image: PacBio

RNA-seq

Main difference with short-reads lib-prep

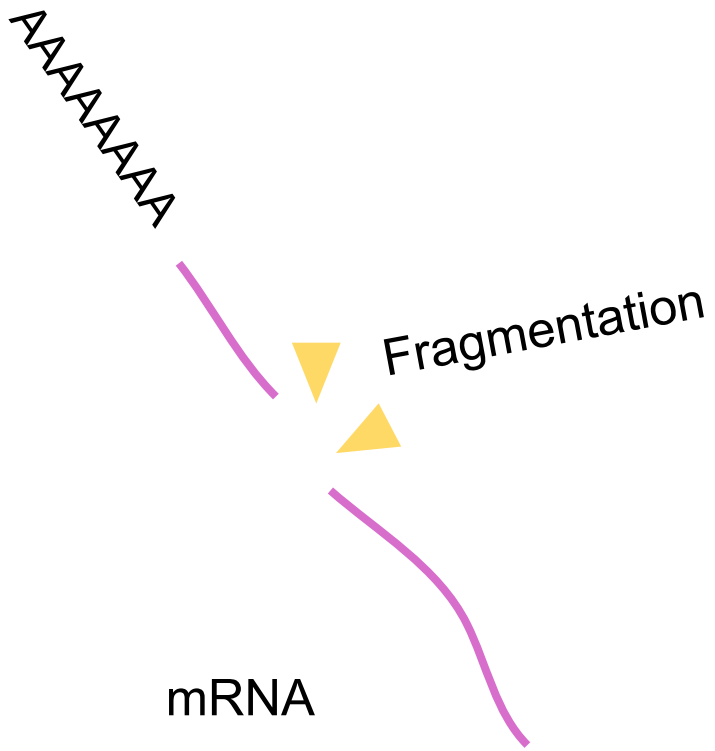
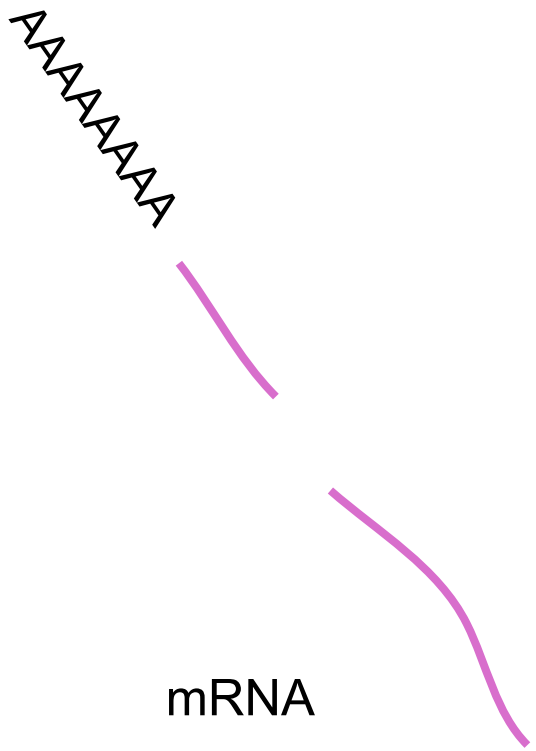


Image: Thermo

RNA-seq

Main difference with short-reads lib-prep



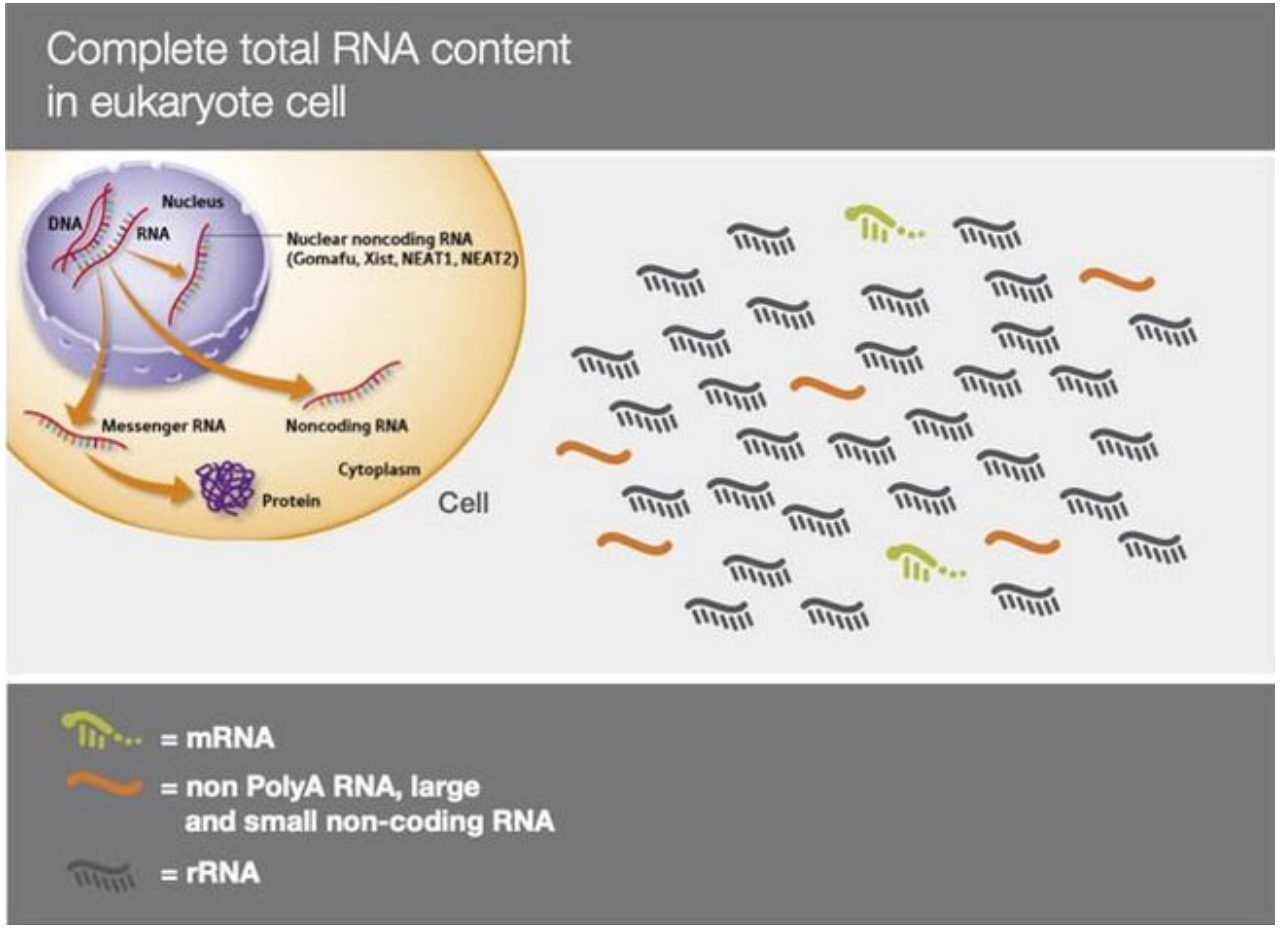
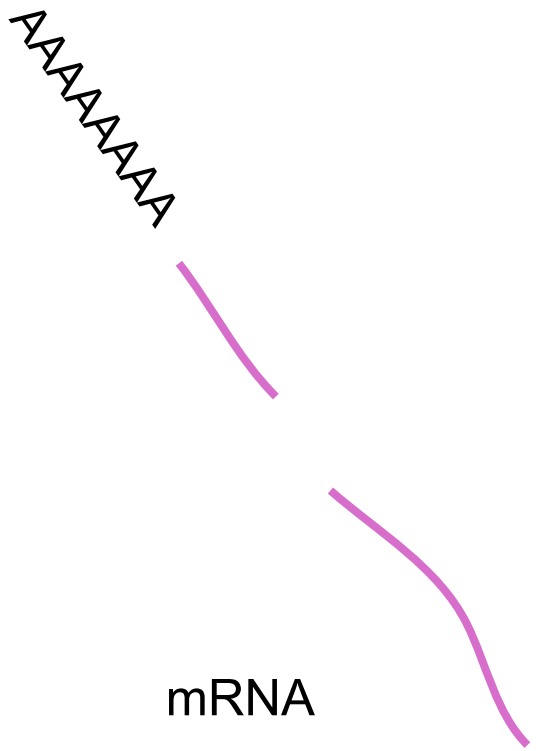


Image: Thermo

RNA-seq

Main difference with short-reads lib-prep



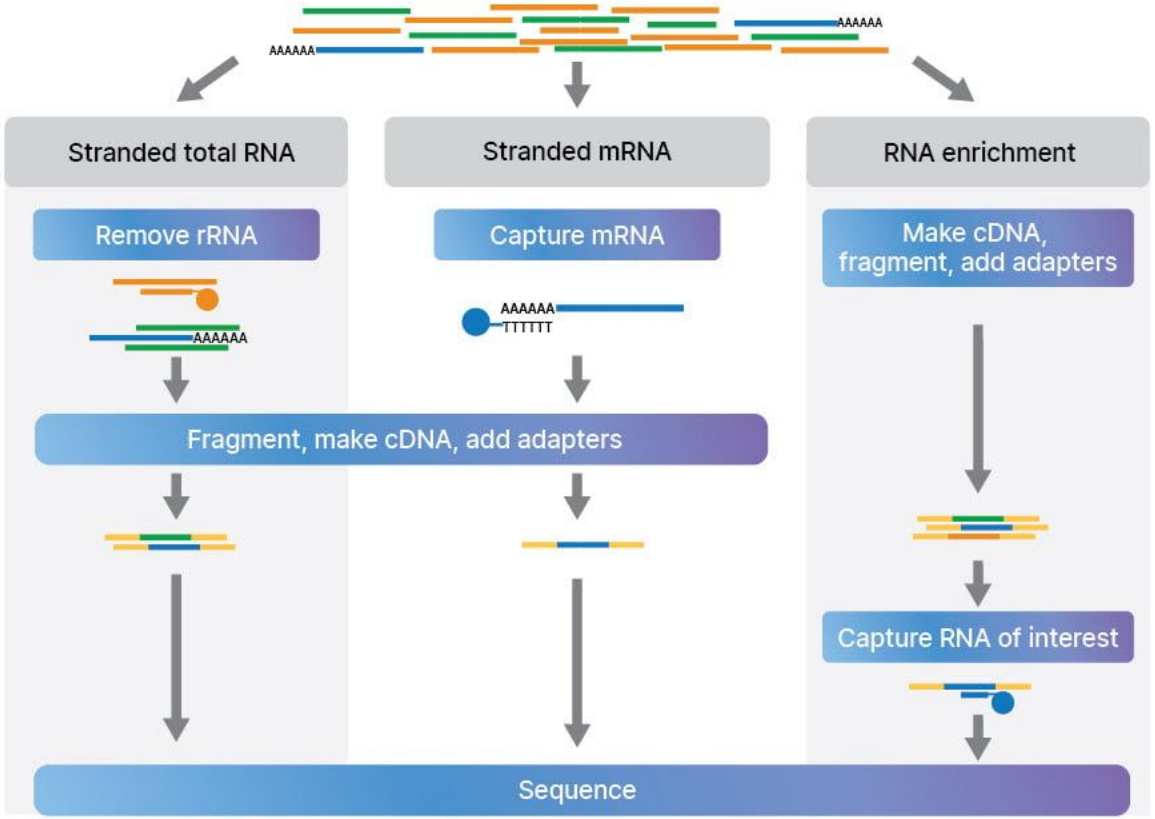


Image: Illumina

Consideration Working with RNA

Garbage in, garbage out

Consideration Working with RNA

Garbage in, garbage out

Consideration Working with RNA

Garbage in, garbage out

Consideration Working with RNA

Garbage in, garbage out

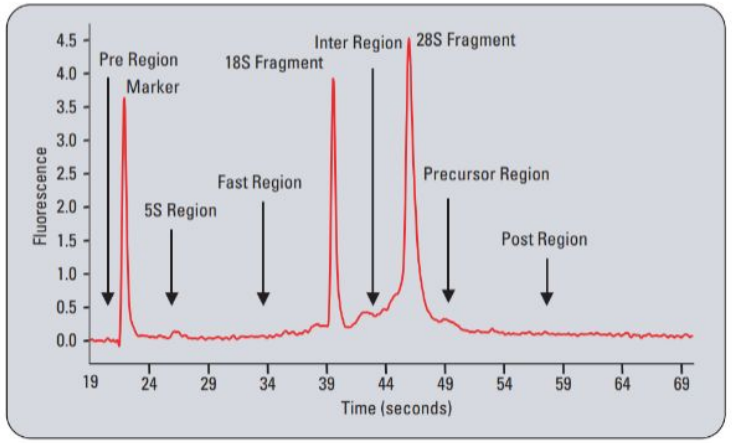
Consideration Working with RNA

Garbage in, garbage out

RIN < 7

Consideration Working with RNA

Garbage in, garbage out



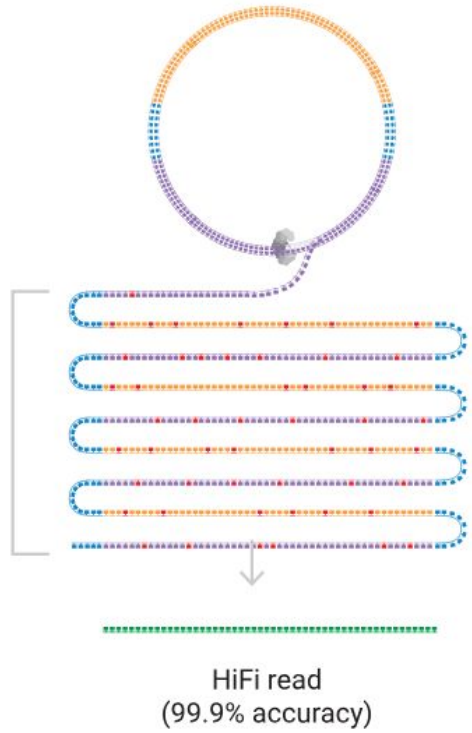
RIN < 7

cDNA

Iso-Seq

PacBio

Description	Full-length transcript isoform sequencing
Kit	SMRTbell prep kit 3.0
Multiplex	up to 12
Preparation time	8 Hours 2 Safe stop point
Input	300 ng Total RNA
Output	1M (Sequel I); 8M (Sequel II); Full-length HiFi transcripts



HiFi long reads generated from PacBio sequencing

Image: PacBio

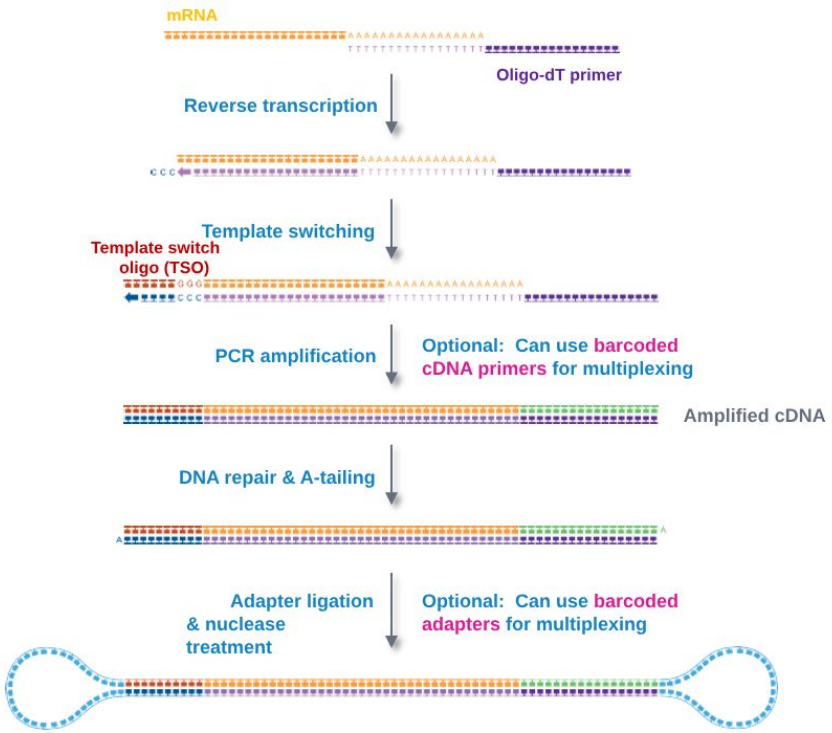
Library preparation

cDNA

Iso-Seq

PacBio

Description	Full-length transcript isoform sequencing
Kit	SMRTbell prep kit 3.0
Multiplex	up to 12
Preparation time	8 Hours 2 Safe stop point
Input	300 ng Total RNA
Output	1M (Sequel I); 8M (Sequel II); Full-length HiFi transcripts



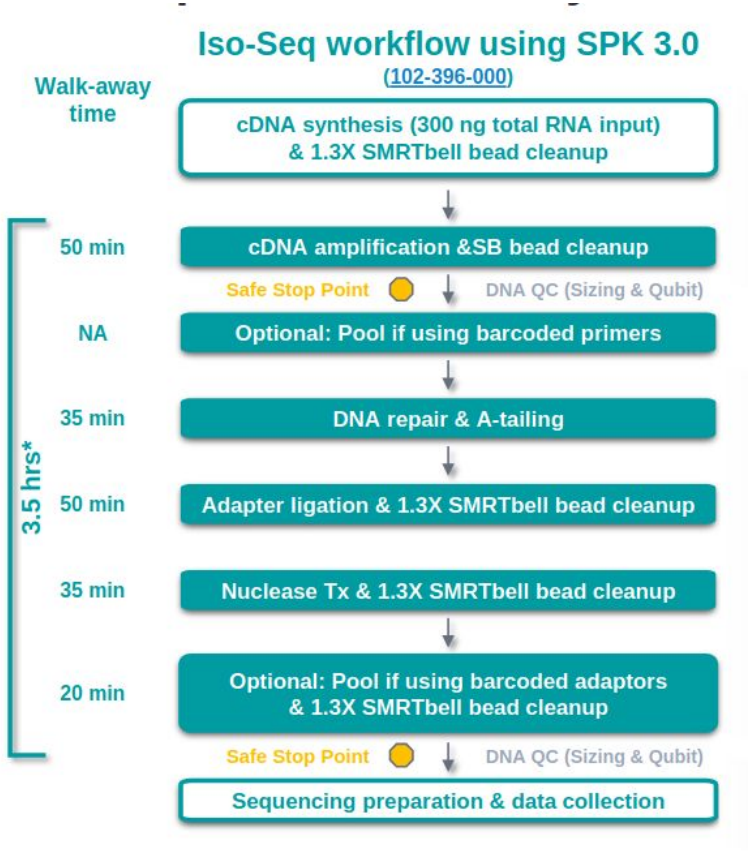
Completed Library

Image: PacBio

cDNA

Iso-Seq

PacBio



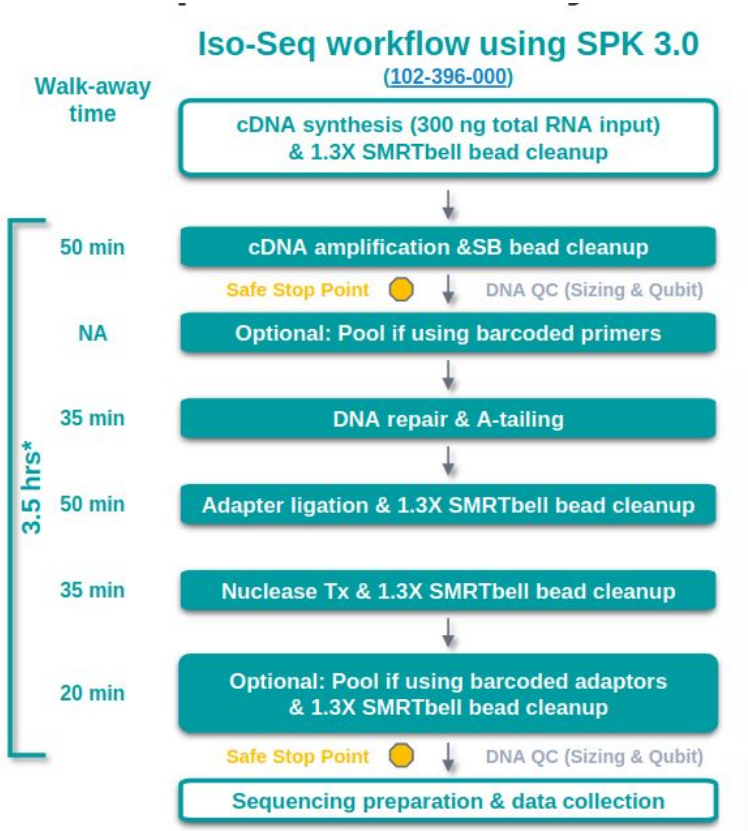
Complete Iso-Seq library preparation workflow for cDNA sequencing

Image: PacBio

cDNA

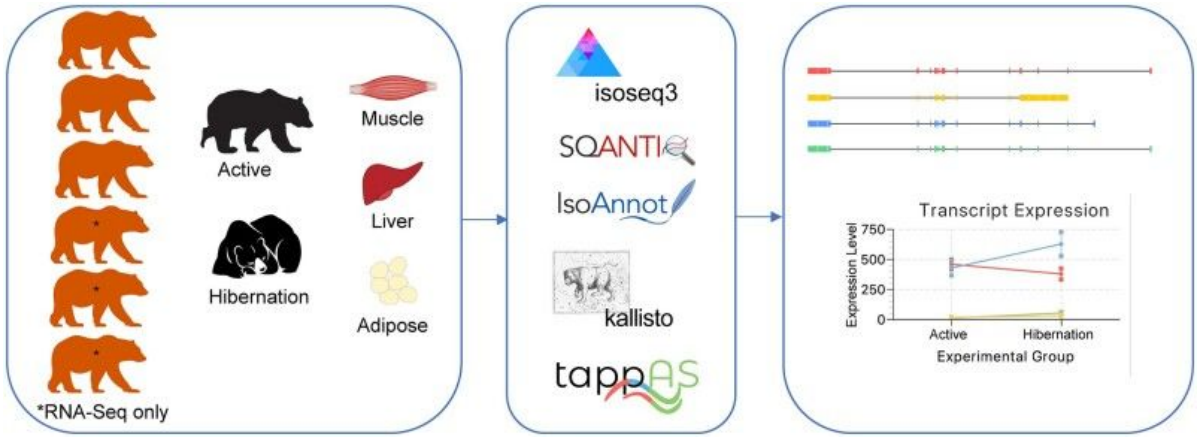
Iso-Seq

PacBio



Complete Iso-Seq library preparation workflow for cDNA sequencing

Image: PacBio



Adipose is the most dynamic tissue during hibernation with the highest number of genes with differential isoform usage and isoform switching as compared to liver and muscle.

Tseng et al.(2022). Long-read isoform sequencing reveals tissue-specific isoform expression between active and hibernating brown bears (Ursus arctos).


Iso-Seq is great
But it can be better
imgflip.com

cDNA

Kinnex (Commercial version of MAS-ISO-seq)

PacBio

Description	Full-length isoform discovery with flexible sample multiplexing
Kit	Kinnex full-length RNA kit
Multiplex	5M × 3-plex (Sequel II); 5M x 6-plex (Vega); 10M × 6-plex (Revio SPRQ)
Preparation time	1.5 days 4 Safe stop point
Input	300 ng Total RNA
Output	15-20M (Sequel II); 20-30M (Vega); 50-60M (Revio SPRQ) Full-length HiFi transcripts (Q40)
Notable Studies	Al Khafaji et al., 2023. Nature Portofolio

Completed Library with cDNA concatenated

Image: PacBio

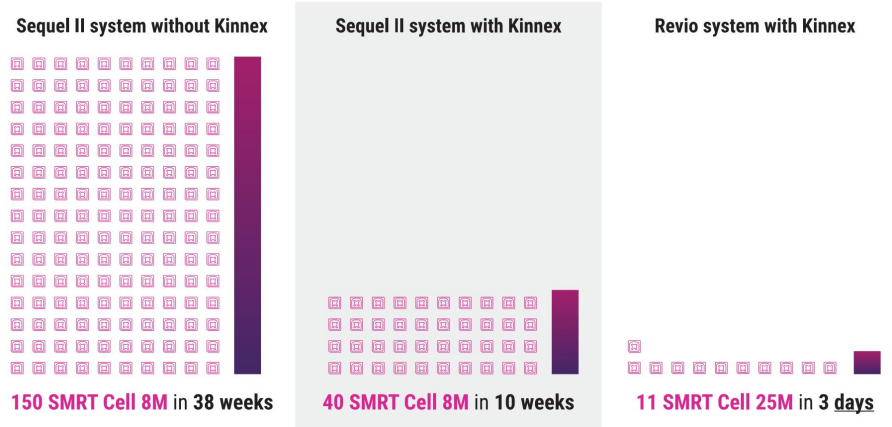
cDNA

Kinnex (Commercial version of MAS-ISO-seq)

PacBio

Description	Full-length isoform discovery with flexible sample multiplexing
Kit	Kinnex full-length RNA kit
Multiplex	5M × 3-plex (Sequel II); 5M x 6-plex (Vega); 10M × 6-plex (Revio SPRQ)
Preparation time	1.5 days 4 Safe stop point
Input	300 ng Total RNA
Output	15-20M (Sequel II); 20-30M (Vega); 50-60M (Revio SPRQ) Full-length HiFi transcripts (Q40)
Notable Studies	Al Khafaji et al., 2023. Nature Portofolio

Completed Library with cDNA concatenated



Kinnex increase throughput significantly

Image: PacBio

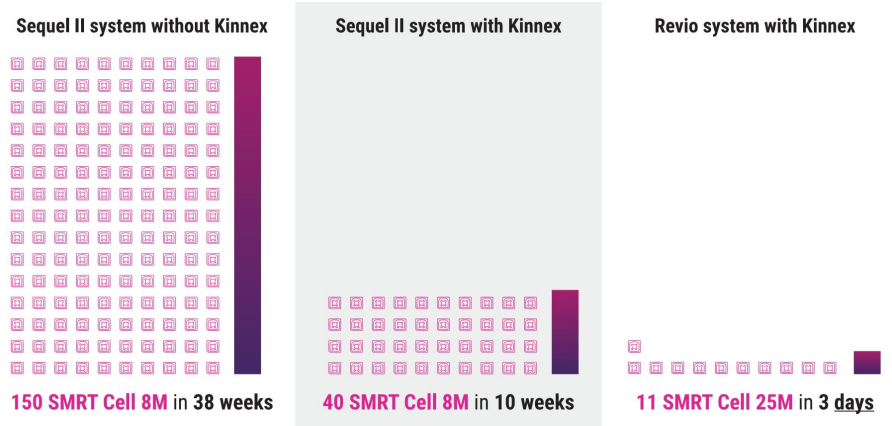
cDNA

Kinnex (Commercial version of MAS-ISO-seq)

PacBio

Description	Full-length isoform discovery with flexible sample multiplexing
Kit	Kinnex full-length RNA kit
Multiplex	5M × 3-plex (Sequel II); 5M x 6-plex (Vega); 10M × 6-plex (Revio SPRQ)
Preparation time	1.5 days 4 Safe stop point
Input	300 ng Total RNA
Output	15-20M (Sequel II); 20-30M (Vega); 50-60M (Revio SPRQ) Full-length HiFi transcripts (Q40)
Notable Studies	Al Khafaji et al., 2023. Nature Portofolio

Completed Library with cDNA concatenated



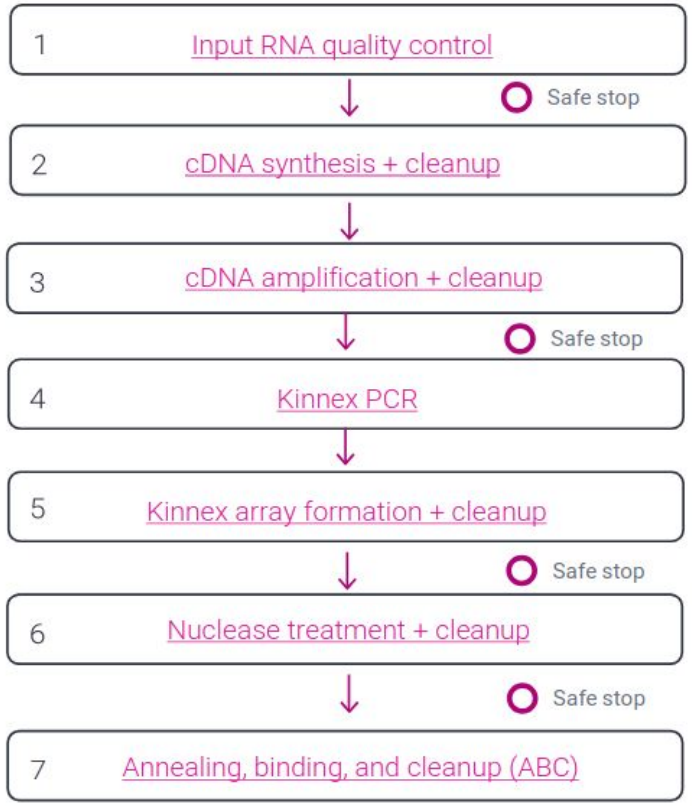
Kinnex increase throughput significantly

Image: PacBio

cDNA

Kinnex (Commercial version of MAS-ISO-seq)

PacBio



SMRTbell® library preparation	
Iso-Seq® Express 2.0 Kit	PacBio® 103-071-500
SMRTbell® cleanup beads	PacBio® 102-158-300*
Elution buffer (50 mL)	PacBio® 101-633-500*
Kinnex™ PCR 8-fold kit	PacBio® 103-071-600*
Revio® SPRQ™ polymerase kit ***or***	PacBio® 103-520-100
Vega™ polymerase kit ***or***	PacBio® 103-517-600
Revio® polymerase kit** ***or***	PacBio® 102-817-600
Sequel® II binding kit 3.2**	PacBio® 102-333-300

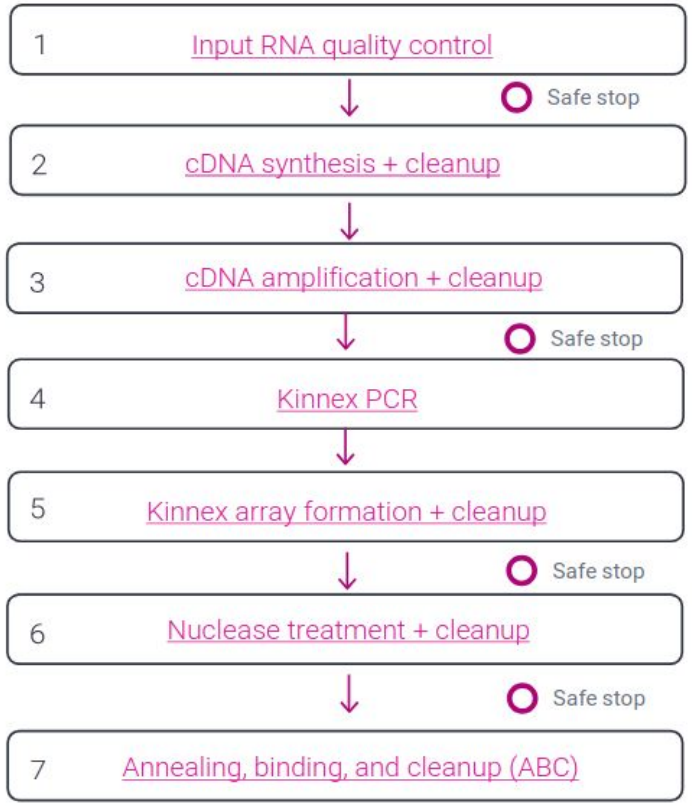
Complete Kinnex library preparation workflow for cDNA sequencing

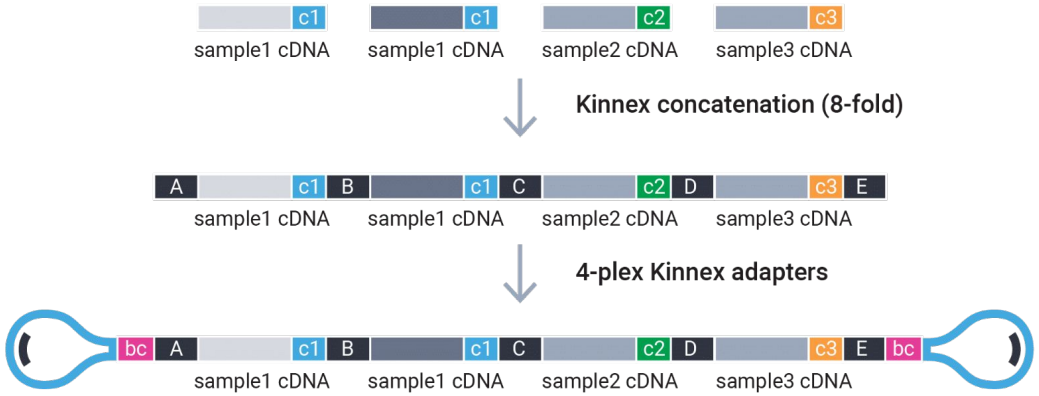
Image: PacBio

cDNA

Kinnex (Commercial version of MAS-ISO-seq)

PacBio





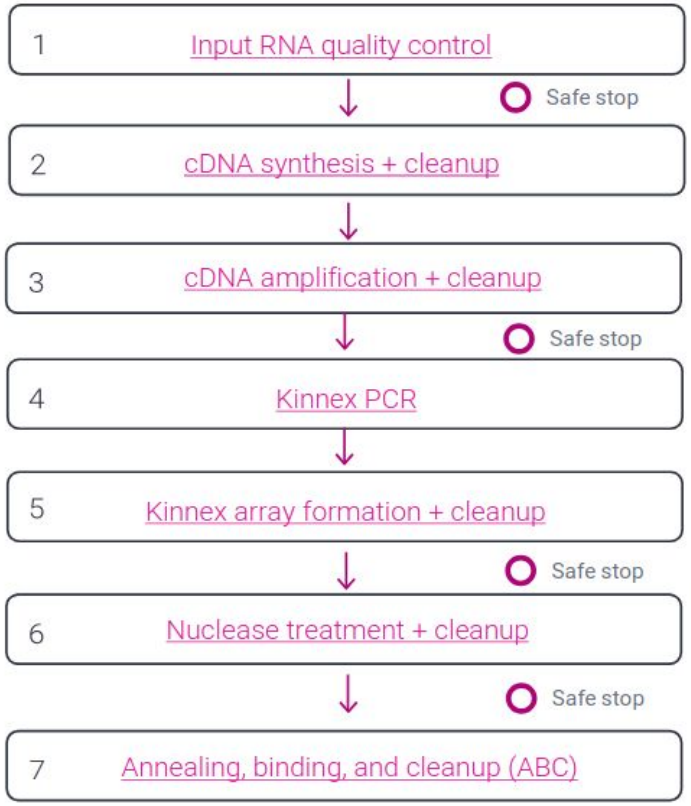
Complete Kinnex library preparation workflow for cDNA sequencing

Image: PacBio

cDNA

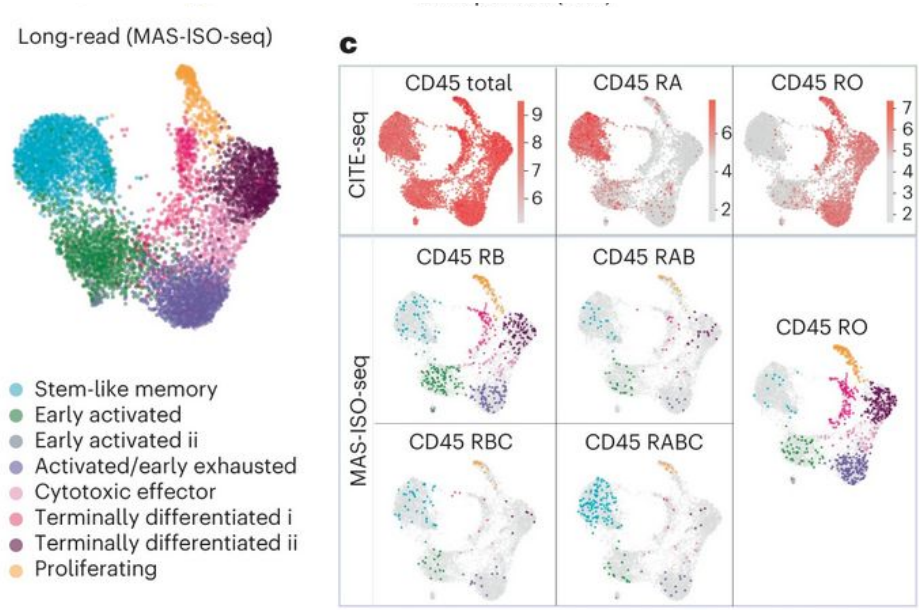
Kinnex (Commercial version of MAS-ISO-seq)

PacBio



Complete Kinnex library preparation workflow for cDNA sequencing

Image: PacBio



Increasing the throughput >15-fold to nearly 40 million cDNA reads per run. MAS-ISO-seq demonstrated a 12- to 32-fold increase in the discovery of differentially spliced genes.

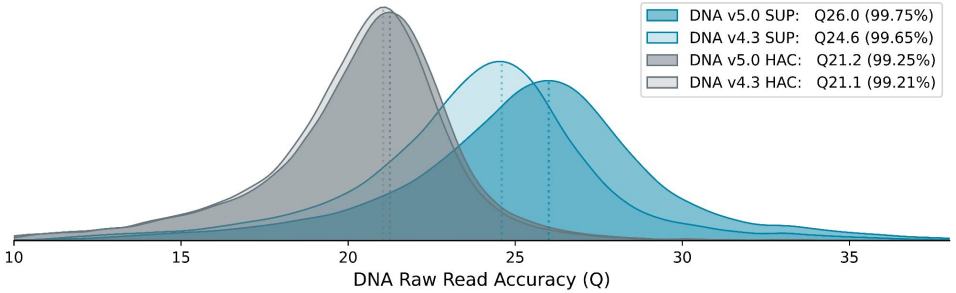
Al'Khafaji, A.M., Smith, J.T., Garimella, K.V. et al (2024). High-throughput RNA isoform sequencing using programmed cDNA concatenation. https://doi.org/10.1038/s41587-023-01815-7

cDNA

cDNA-PCR (R10.4.1)

Nanopore

Description	A sequencing kit optimised for identification and quantification of full-length transcripts with highest output.
Kit	cDNA-PCR Sequencing Kit V14
Multiplex	up to 24-plex (cDNA-PCR Barcoding Kit V14)
Preparation time	225 minutes + PCR
Input	10 ng poly(A)+ RNA or 500 ng total RNA
Output	10-15 million cDNA (MinION); 100 million cDNA (PromethION)



Latest transformer model and R10 pores deliver higher accuracy

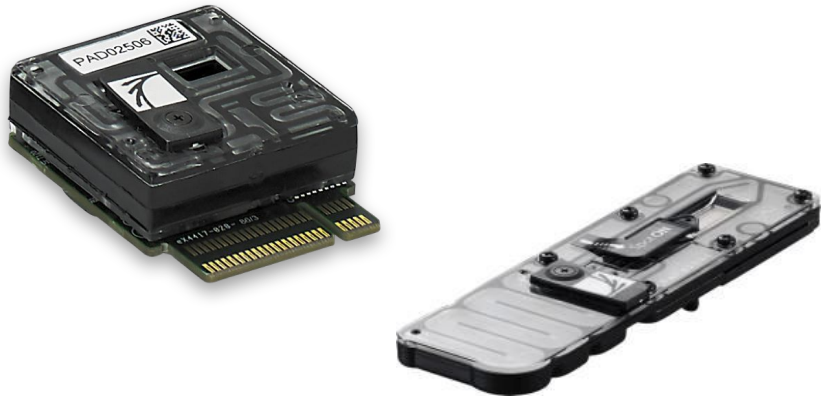
Image: Nanopore

cDNA

cDNA-PCR (R10.4.1)

Nanopore

Description	A sequencing kit optimised for identification and quantification of full-length transcripts with highest output.
Kit	cDNA-PCR Sequencing Kit V14
Multiplex	up to 24-plex (cDNA-PCR Barcoding Kit V14)
Preparation time	225 minutes + PCR
Input	10 ng poly(A)+ RNA or 500 ng total RNA
Output	10-15 million cDNA (MinION); 100 million cDNA (PromethION)



Two types of nanopore device with different sequencing throughput (Left: PromethION, Right: MinION)

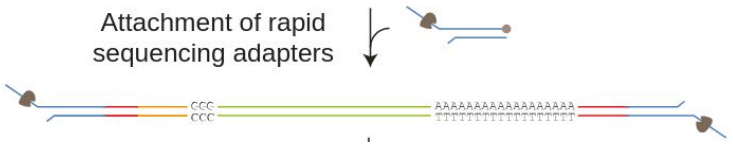
Image: Nanopore

cDNA

cDNA-PCR (R10.4.1)

Nanopore

Description	A sequencing kit optimised for identification and quantification of full-length transcripts with highest output.
Kit	cDNA-PCR Sequencing Kit V14
Multiplex	up to 24-plex (cDNA-PCR Barcoding Kit V14)
Preparation time	225 minutes + PCR
Input	10 ng poly(A)+ RNA or 500 ng total RNA
Output	10-15 million cDNA (MinION); 100 million cDNA (PromethION)



Completed Library

Common Gotchas

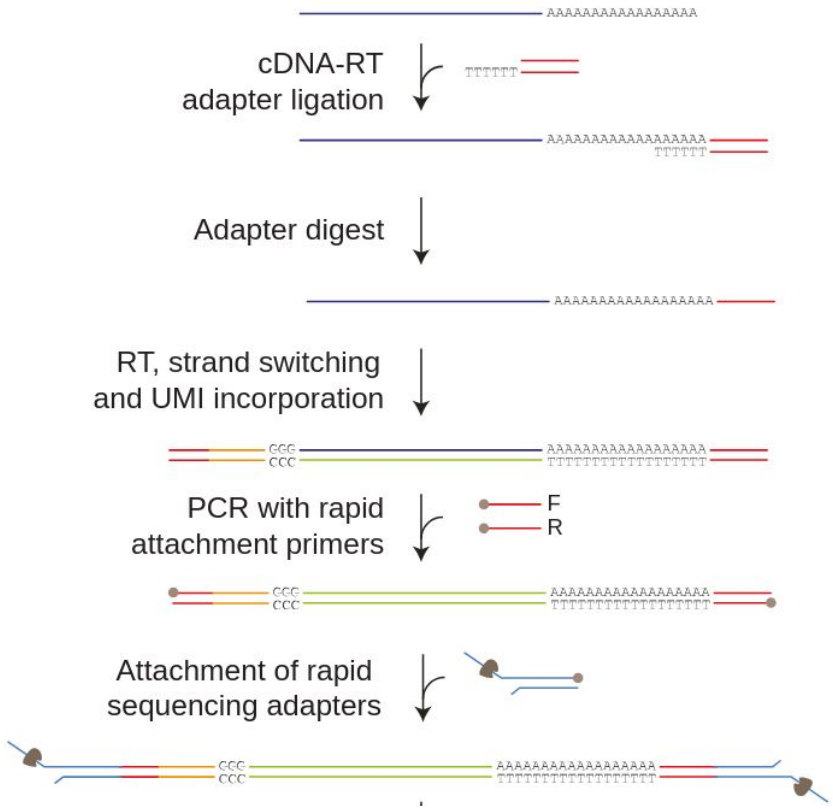
Always perform flow cell QC prior lib prep, ensure you have enough pores!

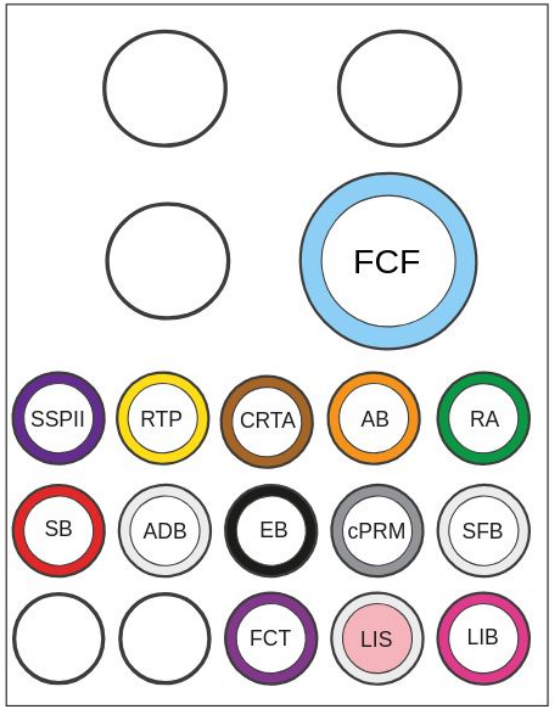
Image: Nanopore

cDNA

cDNA-PCR (R10.4.1)

Nanopore





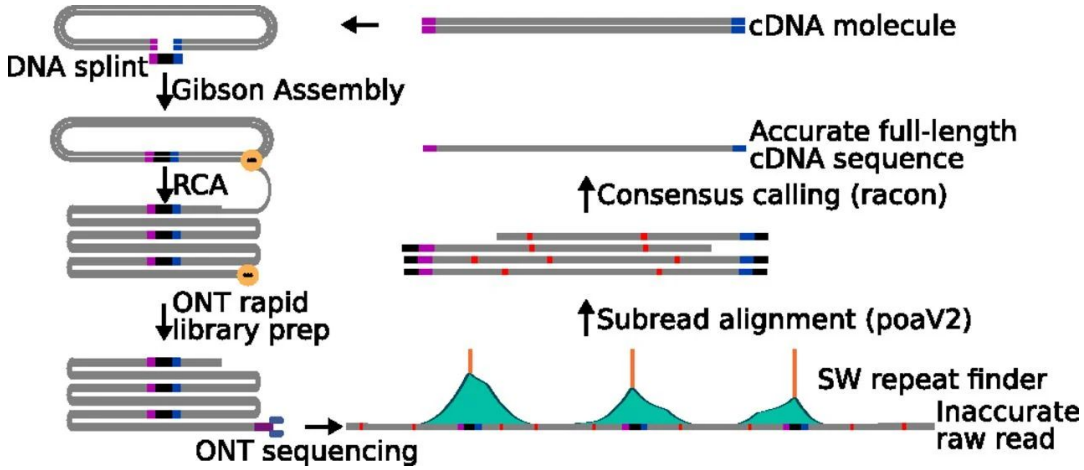
RTP: RT Primer
CRTA: cDNA RT Adapter
AB: Annealing Buffer
RA: Rapid Adapter
SB: Sequencing Buffer
ADB: Adapter Buffer
SFB: Short Fragment Buffer
SSPII: Strand Switching Primer II
EB: Elution Buffer
cPRM: cDNA Primer
LIB: Library Beads
LIS: Library Solution
FCF: Flow Cell Flush
FCT: Flow Cell Tether

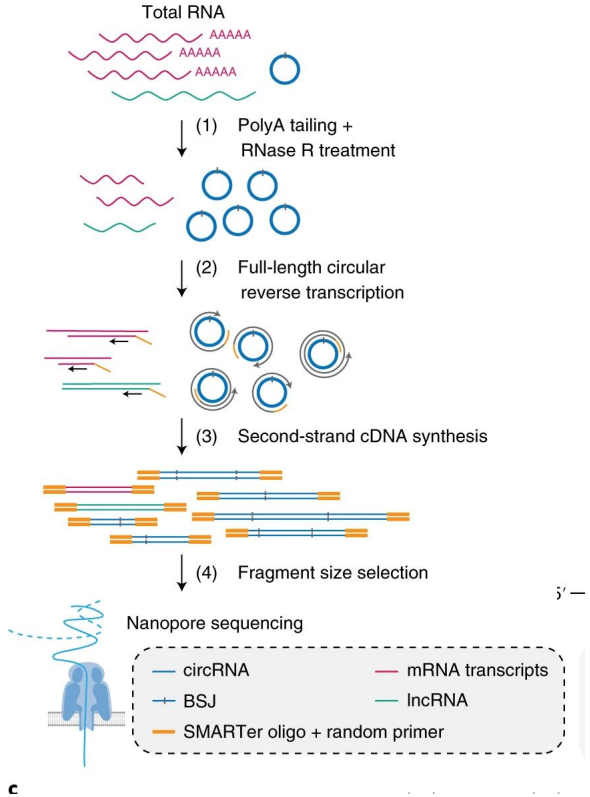
Complete library workflow for cDNA sequencing with cDNA-PCR kit

Image: Nanopore

cDNA

cDNA-PCR (R10.4.1)





Nanopore

Community

Mandalorion was used to analyze the R2C2 reads of the 96 B cells and identified several surface receptor isoforms expressed by 96 distinct single human B cells.

R. Volden, et. al. Improving nanopore read accuracy with the R2C2 method enables the sequencing of highly multiplexed full-length single-cell cDNA. https://doi.org/10.1073/pnas.1806447115 (2018).

Circular reverse transcription and size selection achieves a 20-fold higher enrichment of circRNAs from total RNA compared to previous methods.

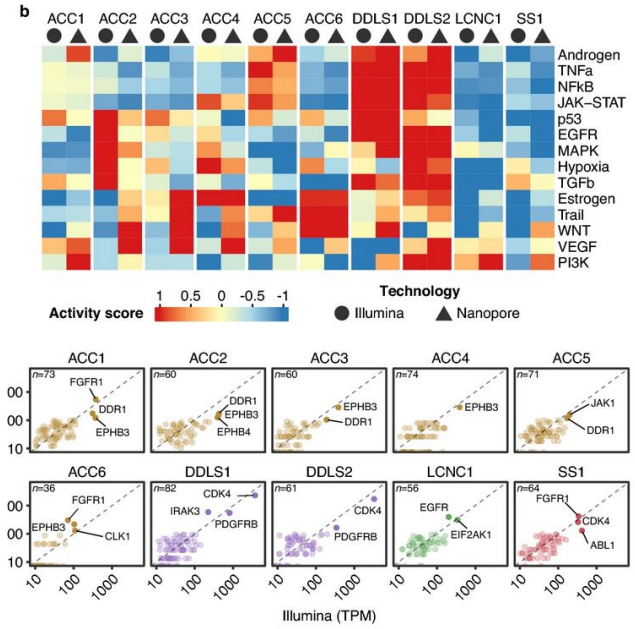
Zhang, J., Hou, L., Zuo, Z. et al. Comprehensive profiling of circular RNAs with nanopore sequencing and CIRI-long. Nat Biotechnol 39, 836–845 (2021). https://doi.org/10.1038/s41587-021-00842-6

cDNA

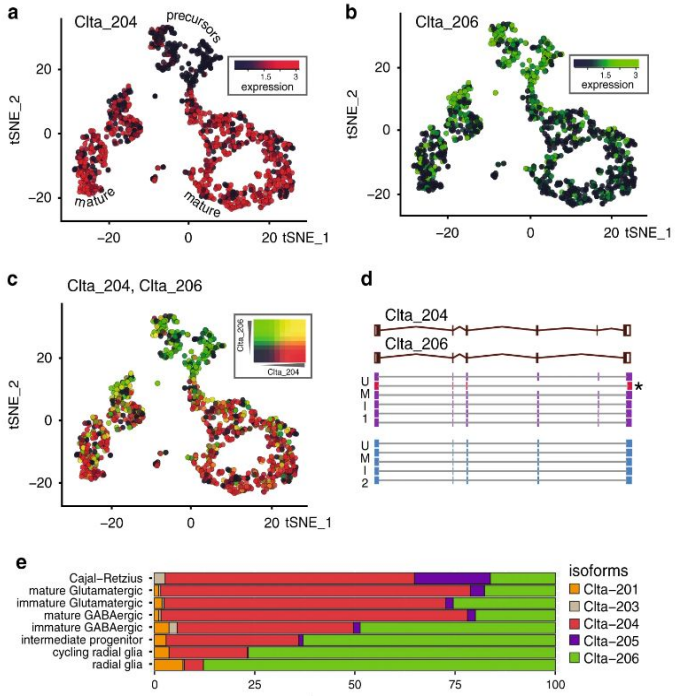
cDNA-PCR (R10.4.1)

Nanopore

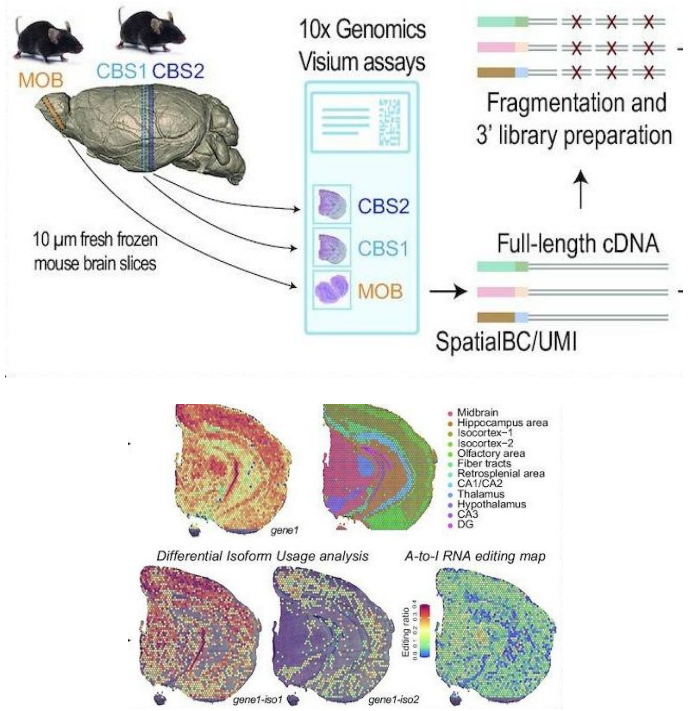
Community



Shallow nanopore RNA-seq enables rapid and biologically meaningful transcriptome profiling of tumors (Mock et al 2023)



Nanopore scRNA-seq reveals transcript isoform diversity (Lebrigand et al 2020).



SiT can be used to profile isoform expression and sequence heterogeneity in different areas of the tissue (Lebrigand et al 2023).

mRNA (Native RNA)

Direct RNA sequencing (RNA001, RNA002)

Nanopore

Description	A sequencing kit optimised for sequencing native RNA with improved output and accuracy.
Kit	Direct RNA Sequencing Kit
Multiplex	Under development
Preparation time	135 minutes
Input	300 ng poly(A)+ RNA or 1 ug total RNA
Output	2 GB cDNA (MinION); 30 million cDNA (PromethION)

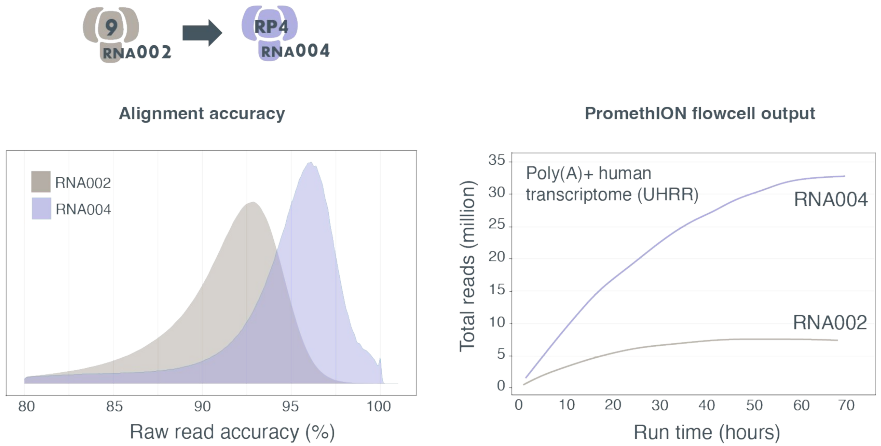
Image: Nanopore

mRNA (Native RNA)

Direct RNA sequencing (RNA004)

Nanopore

Description	A sequencing kit optimised for sequencing native RNA with improved output and accuracy.
Kit	Direct RNA Sequencing Kit
Multiplex	Under development
Preparation time	135 minutes
Input	300 ng poly(A)+ RNA or 1 ug total RNA
Output	2 GB cDNA (MinION); 30 million cDNA (PromethION)



Overall quality and throughput significantly improved in RNA004. Libprep step is the same.

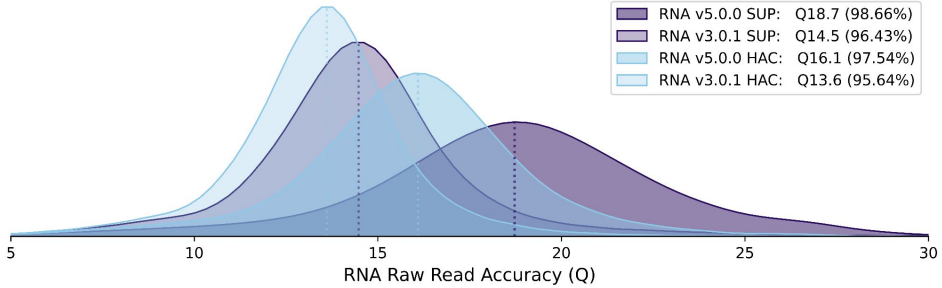
Image: Nanopore

mRNA (Native RNA)

Direct RNA sequencing (RNA004)

Nanopore

Description	A sequencing kit optimised for sequencing native RNA with improved output and accuracy.
Kit	Direct RNA Sequencing Kit
Multiplex	Under development
Preparation time	135 minutes
Input	300 ng poly(A)+ RNA or 1 ug total RNA
Output	2 GB cDNA (MinION); 30 million cDNA (PromethION)



Latest transformer model and RP4 pores deliver higher accuracy

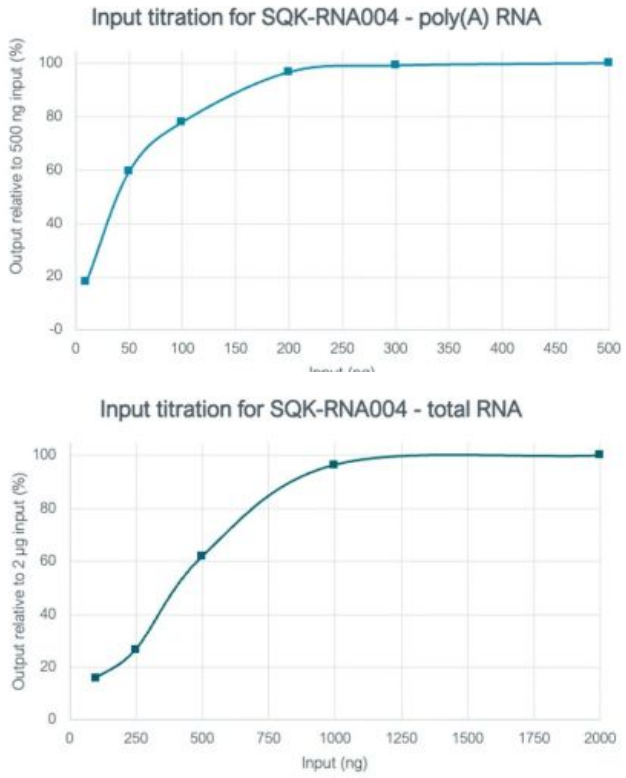
Image: Nanopore

mRNA (Native RNA)

Direct RNA sequencing (RNA004)

Nanopore

Description	A sequencing kit optimised for sequencing native RNA with improved output and accuracy.
Kit	Direct RNA Sequencing Kit
Multiplex	Under development
Preparation time	135 minutes
Input	300 ng poly(A)+ RNA or 1 ug total RNA
Output	2 GB cDNA (MinION); 30 million cDNA (PromethION)



Direct RNA kit support poly A+ RNA or total RNA

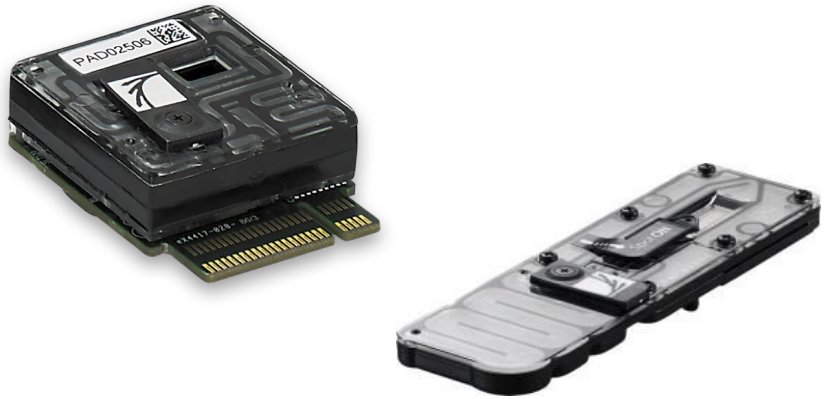
Image: Nanopore

mRNA (Native RNA)

Direct RNA sequencing (RNA004)

Nanopore

Description	A sequencing kit optimised for sequencing native RNA with improved output and accuracy.
Kit	Direct RNA Sequencing Kit
Multiplex	Under development
Preparation time	135 minutes
Input	300 ng poly(A)+ RNA or 1 ug total RNA
Output	2 GB cDNA (MinION); 30 million cDNA (PromethION)



Two types of nanopore device with different sequencing throughput (Left: PromethION, Right: MinION)

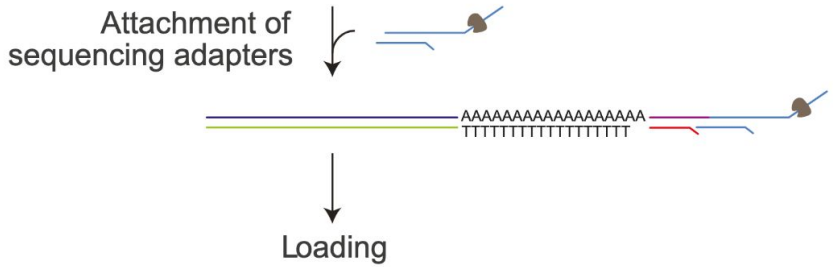
Image: Nanopore

mRNA (Native RNA)

Direct RNA sequencing (RNA004)

Nanopore

Description	A sequencing kit optimised for sequencing native RNA with improved output and accuracy.
Kit	Direct RNA Sequencing Kit
Multiplex	Under development
Preparation time	135 minutes
Input	300 ng poly(A)+ RNA or 1 ug total RNA
Output	2 GB cDNA (MinION); 30 million cDNA (PromethION)



Completed Library

Common Gotchas

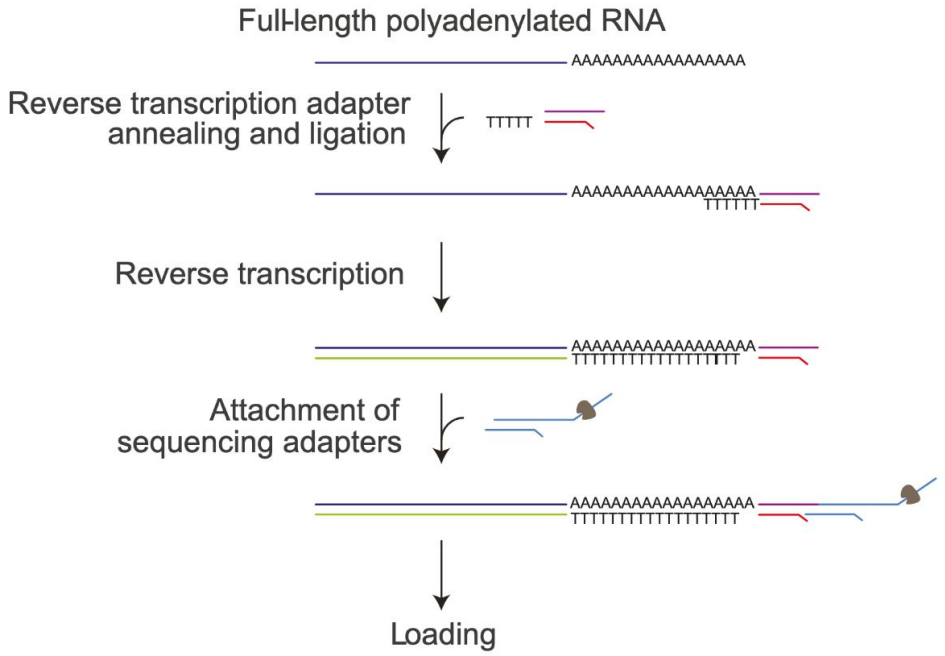
Always perform flow cell QC prior lib prep, ensure you have enough pores!

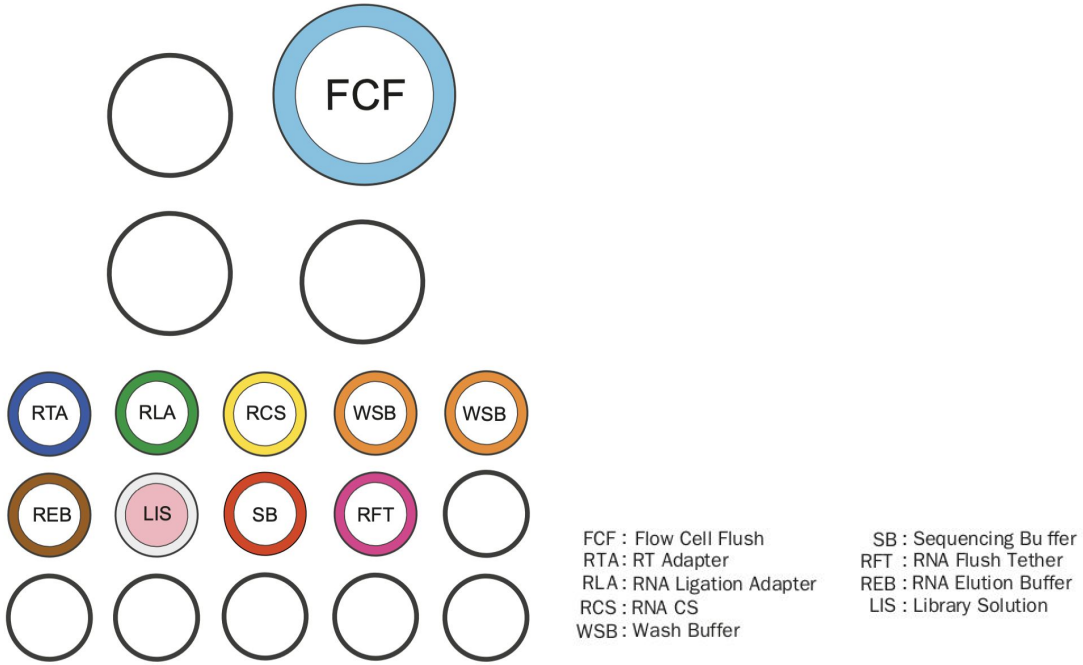
Image: Nanopore

mRNA (Native RNA)

Direct RNA sequencing (RNA004)

Nanopore





Complete library workflow for Direct RNA sequencing with Direct RNA Sequencing kit

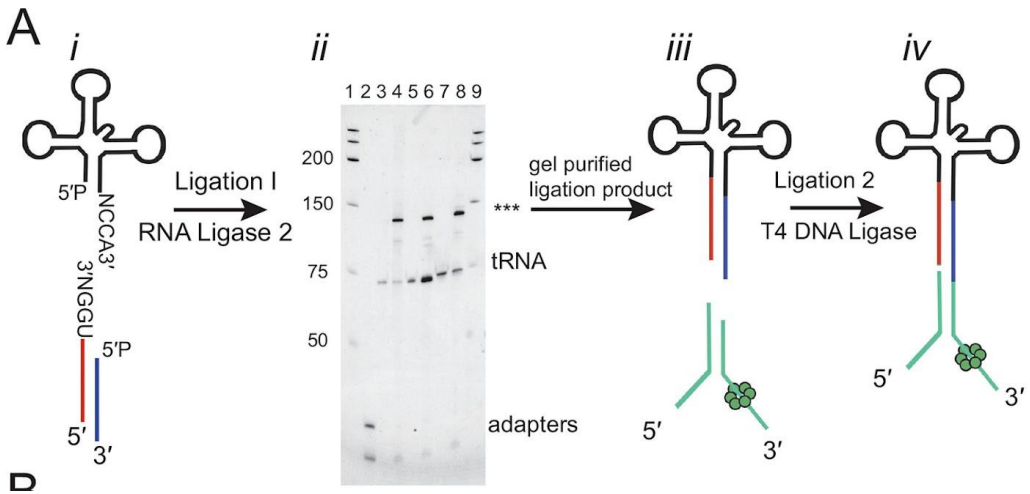
Image: Nanopore

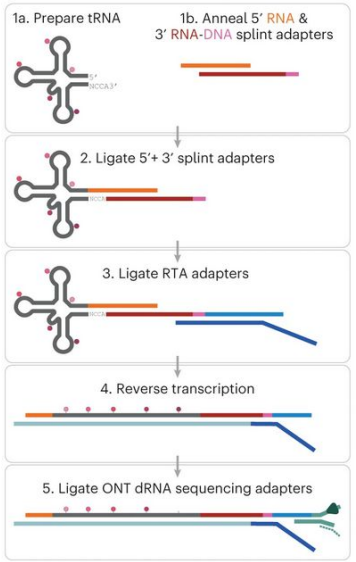
tRNA (Native RNA)

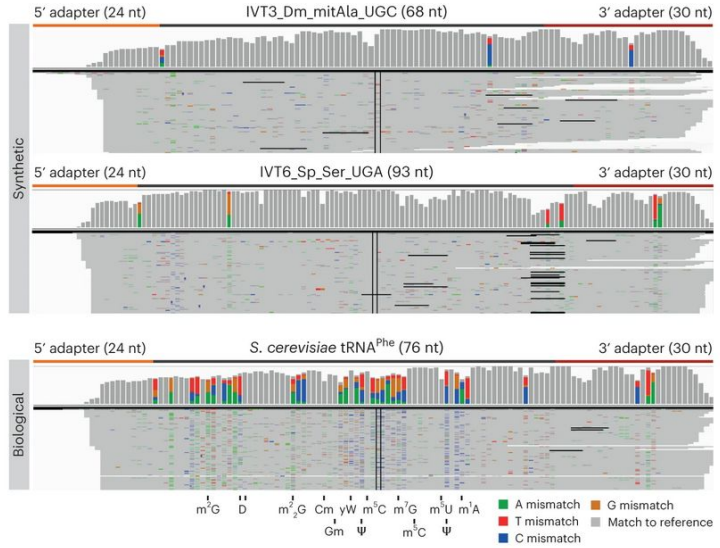
Direct RNA sequencing (RNA004)

Nanopore

Community







Nanopore sequencing detected all 43 expected isoacceptors in total E. coli MRE600 tRNA

Thomas, N. K., Poodari, V. C., Jain, M., Olsen, H. E., Akeson, M., & Abu-Shumays, R. L. (2021). Direct Nanopore Sequencing of Individual Full Length tRNA Strands. ACS nano, 15(10), 16642–16653. https://doi.org/10.1021/acsnano.1c06488

Nano-tRNAseq can efficiently sequence both IVT and native tRNA populations

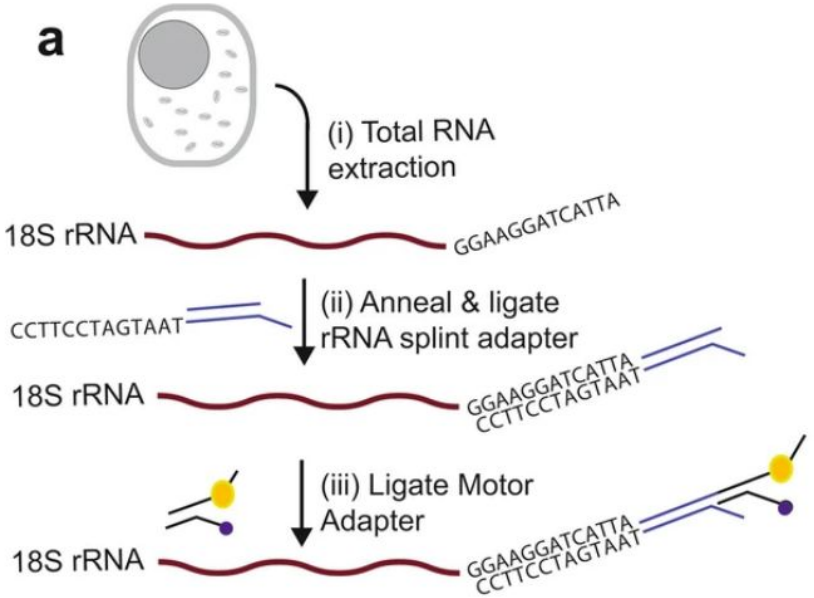
Lucas, M.C., Pryszcz, L.P., Medina, R. et al. Quantitative analysis of tRNA abundance and modifications by nanopore RNA sequencing. Nat Biotechnol 42, 72–86 (2024). https://doi.org/10.1038/s41587-023-01743-6

rRNA (Native RNA)

Direct RNA sequencing (RNA004)

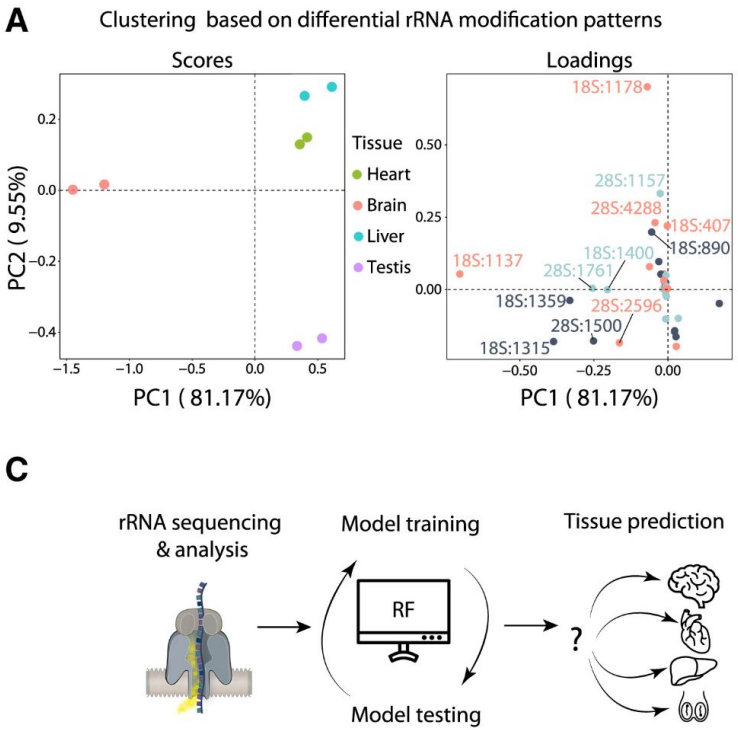
Nanopore

Community



Differential rRNA modification patterns can be used to predict tissue types and developmental stages

Jain, M., Olsen, H.E., Akeson, M., Abu-Shumays, R. (2021). Adaptation of Human Ribosomal RNA for Nanopore Sequencing of Canonical and Modified Nucleotides. Inhttps://doi.org/10.1007/978-1-0716-1374-0_4



Differential rRNA modification patterns can be used to predict tissue types and developmental stages

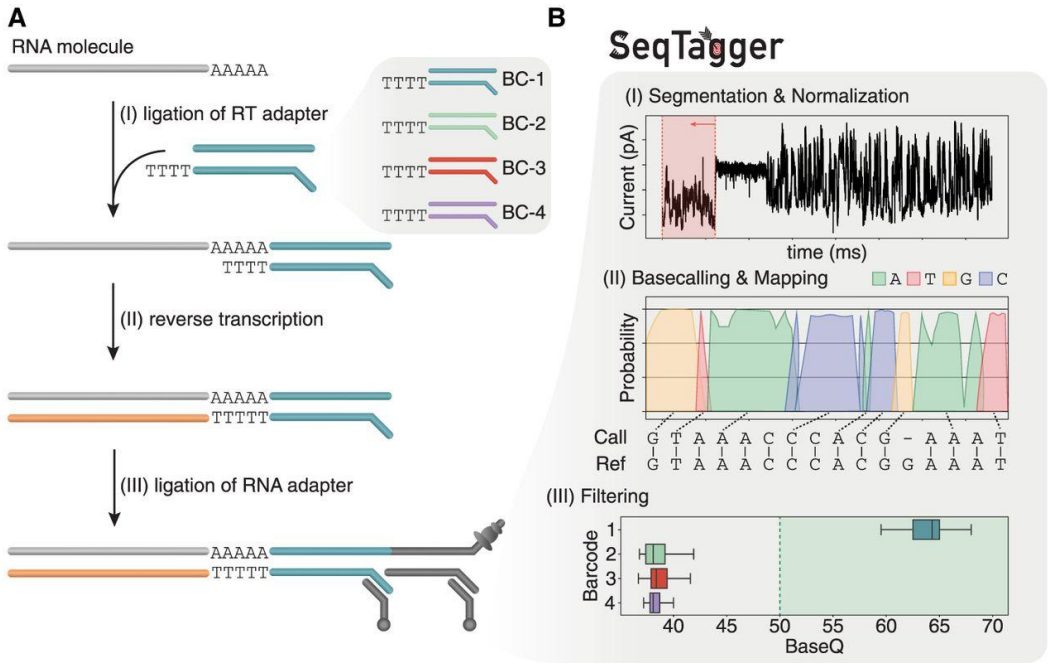
Milenkovic, et al. (2025). Epitranscriptomic rRNA fingerprinting reveals tissue-of-origin and tumor-specific signatures. Molecular cell, 85(1), 177–190.e7. https://doi.org/10.1016/j.molcel.2024.11.014

Multiplexing

Direct RNA sequencing (RNA004)

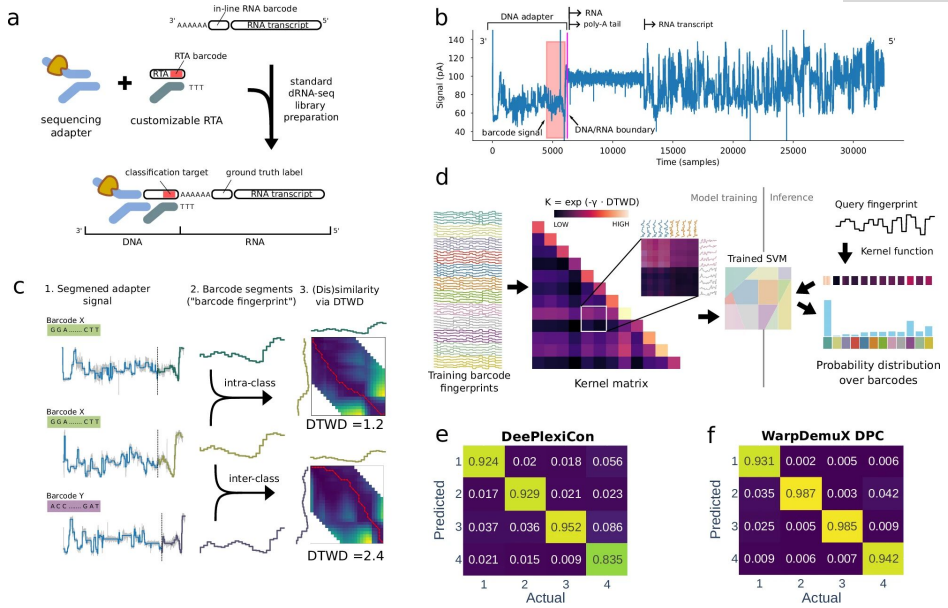
Nanopore

Community



SeqTagger, a rapid and robust method that can demultiplex DRS data sets with 99% precision and 95% recall.

Thomas, N. K., Poodari, V. C., Jain, M., Olsen, H. E., Akeson, M., & Abu-Shumays, R. L. (2021). Direct Nanopore Sequencing of Individual Full Length tRNA Strands. ACS nano, 15(10), 16642–16653. https://doi.org/10.1021/acsnano.1c06488



Rapid phenotypic profiling of different SARS-CoV-2 viruses through multiplexed sequencing of longitudinal samples on a single flowcell, identifying systematic differences in transcript abundance and poly(A) tail lengths

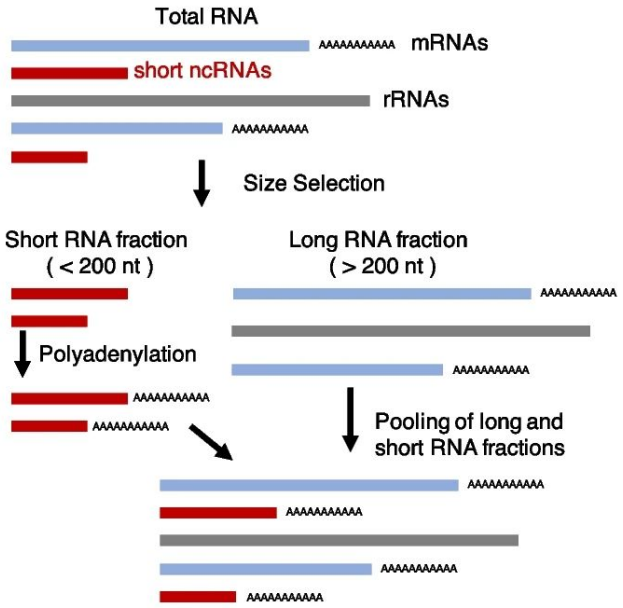
van der Toorn, W., Bohn, P., Liu-Wei, W. et al. Demultiplexing and barcode-specific adaptive sampling for nanopore direct RNA sequencing. Nat Commun 16, 3742 (2025). https://doi.org/10.1038/s41467-025-59102-9

Improving Representation

Direct RNA sequencing (RNA004)

Nanopore

Community



NERD-seq expands representation of frequently modified non-coding RNAs, such as snoRNAs, snRNAs, scRNAs, srpRNAs, tRNAs, and rRFs.

Saville, L., Wu, L., Habtewold, J. et al. NERD-seq: a novel approach of Nanopore direct RNA sequencing that expands representation of non-coding RNAs. Genome Biol 25, 233 (2024). https://doi.org/10.1186/s13059-024-03375-8

Section 1.1

Bonus

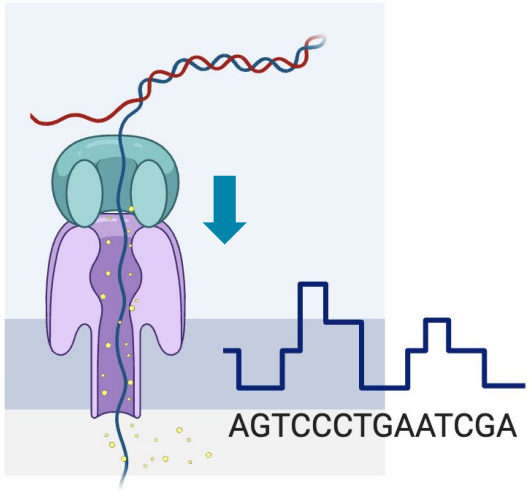
Library preparation strategies for long-reads RNA-seq experiments

Adaptive Sampling

Direct RNA sequencing (RNA004)

Nanopore

Community



Adaptive sampling can be used to enrich or depletes target transcripts.

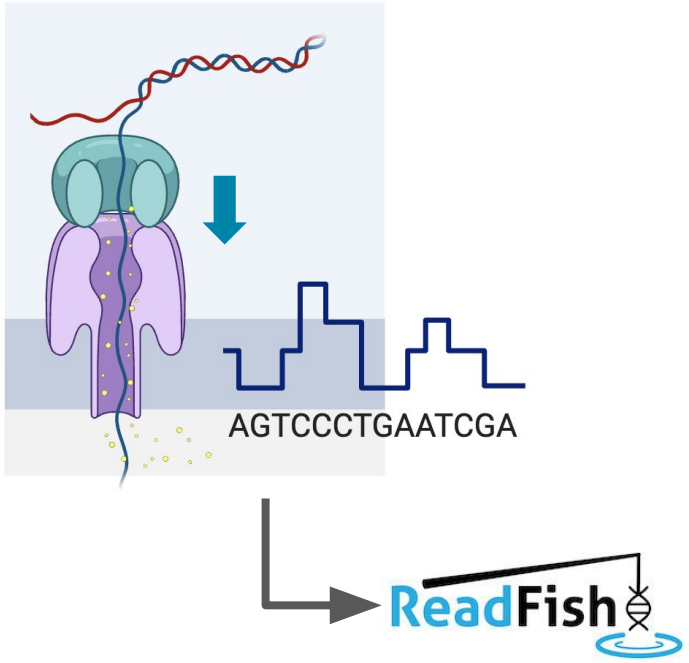
Image: Nanopore

Adaptive Sampling

Direct RNA sequencing (RNA004)

Nanopore

Community



Process data in realtime

Adaptive sampling can be used to enrich or depletes target transcripts.

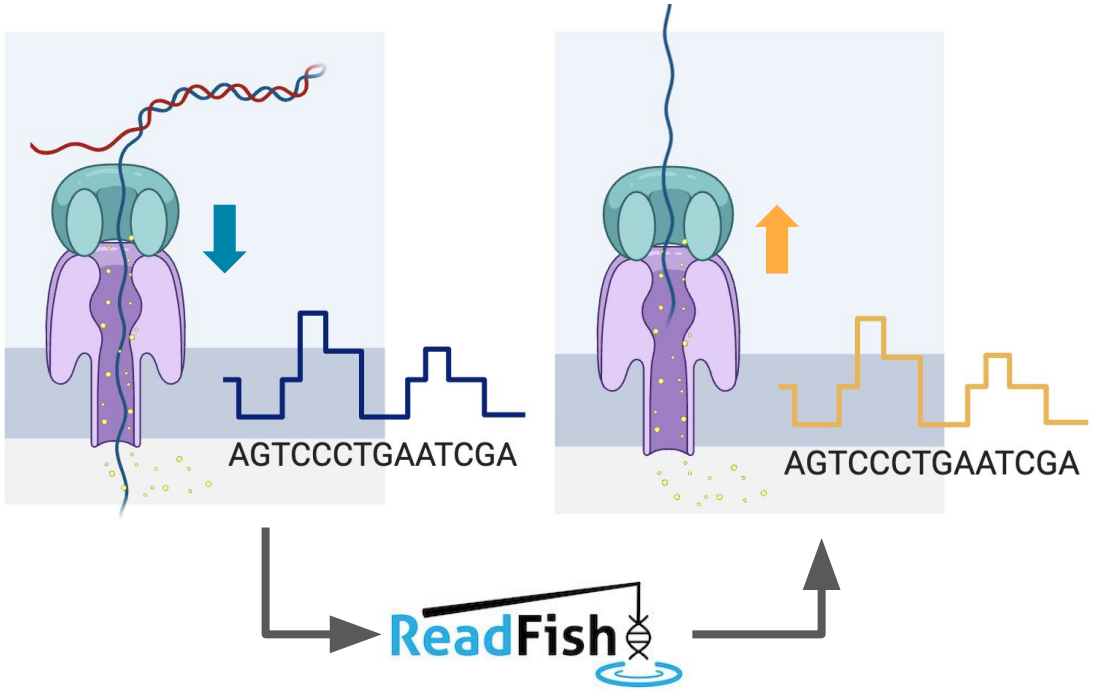
Image: Nanopore

Adaptive Sampling

Direct RNA sequencing (RNA004)

Nanopore

Community



Process data in realtime

Adaptive sampling can be used to enrich or depletes target transcripts.

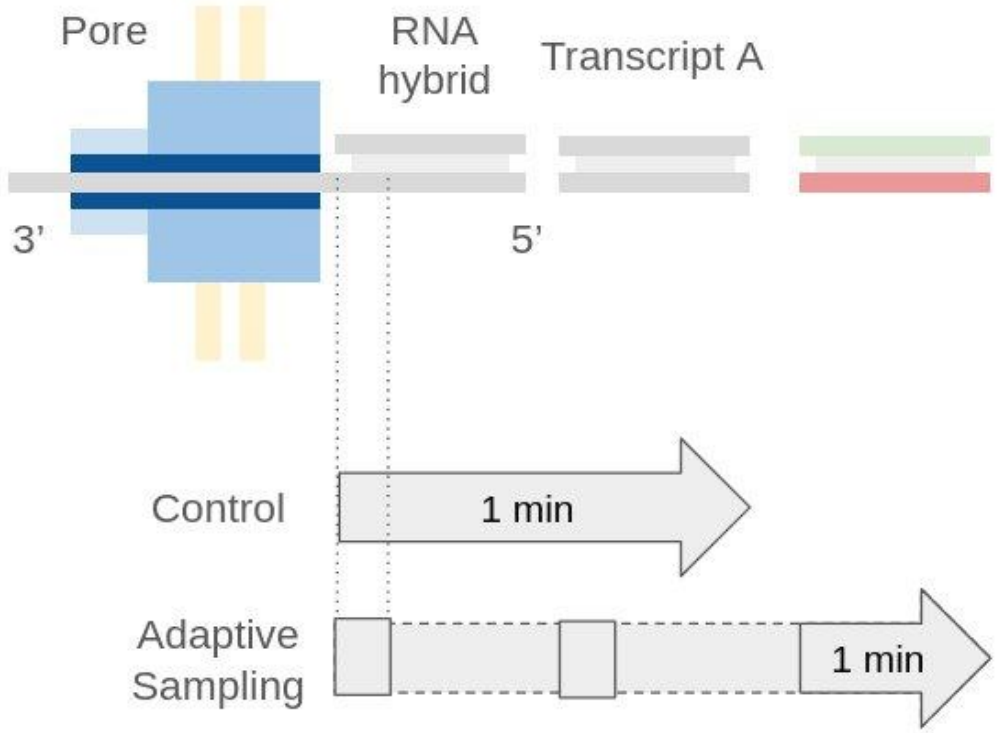


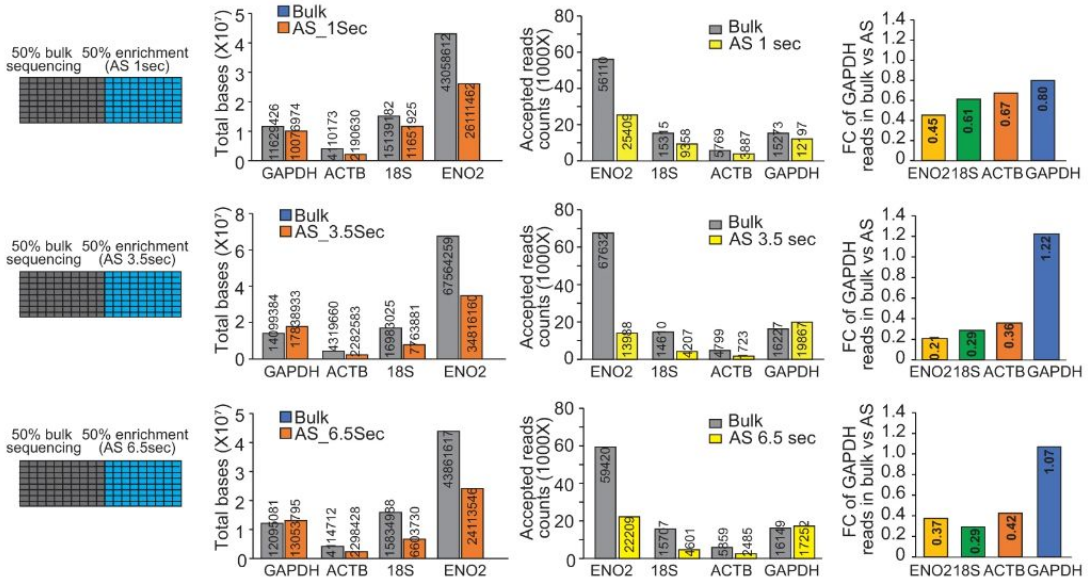
Image: Nanopore

Adaptive Sampling

Direct RNA sequencing (RNA004)

Nanopore

Community



Adaptive sampling performed on a pool of in vitro transcribed RNAs resulted in a net increase of 22-30% in the proportion of transcripts of interest in the population.

Wang, J., et al.. (2024). Direct RNA sequencing coupled with adaptive sampling enriches RNAs of interest in the transcriptome. https://doi.org/10.1038/s41467-023-44656-3

Image: Nanopore

Section 2

Basecalling

Raw data formats and Basecalling

PacBio

.

PacBio

Official tool	ICS v13.3 (Revio)
	Base calling HiFi read generation with DeepConsensus Methylation calling Barcode demultiplexing BAM file generation

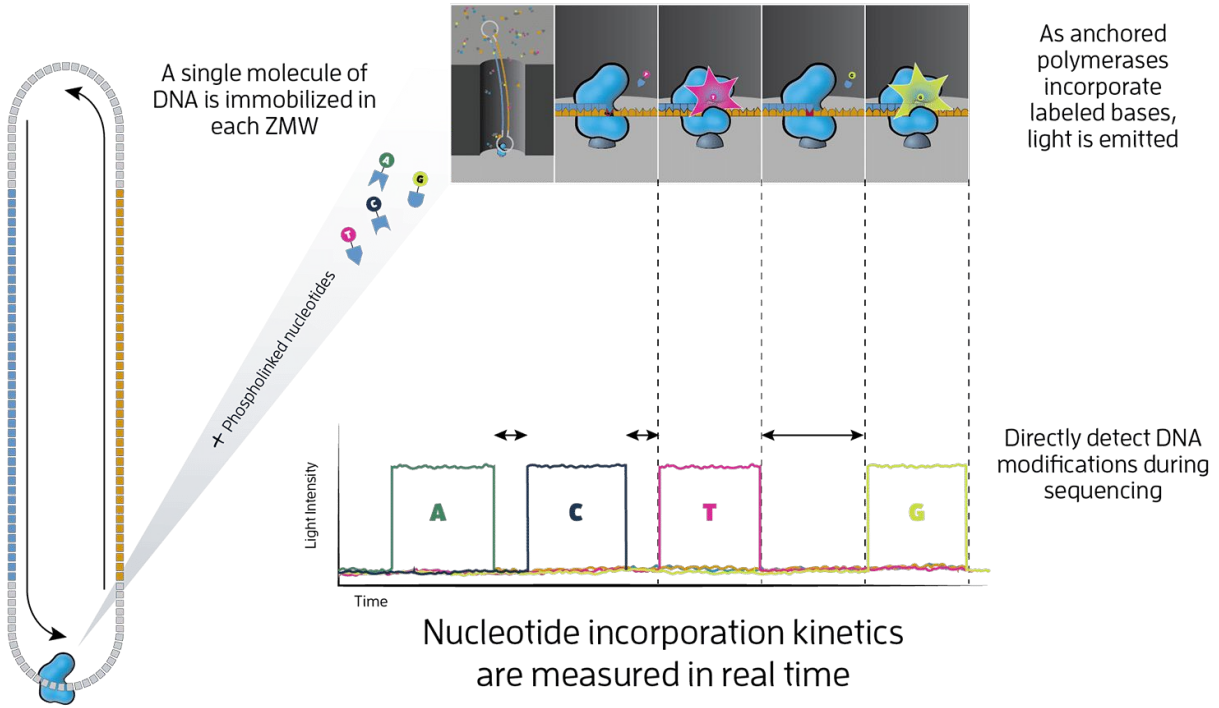


Image: Nanopore

PacBio

.

Official tool	ICS v13.3 (Revio)
	Base calling HiFi read generation with DeepConsensus Methylation calling Barcode demultiplexing BAM file generation

PacBio

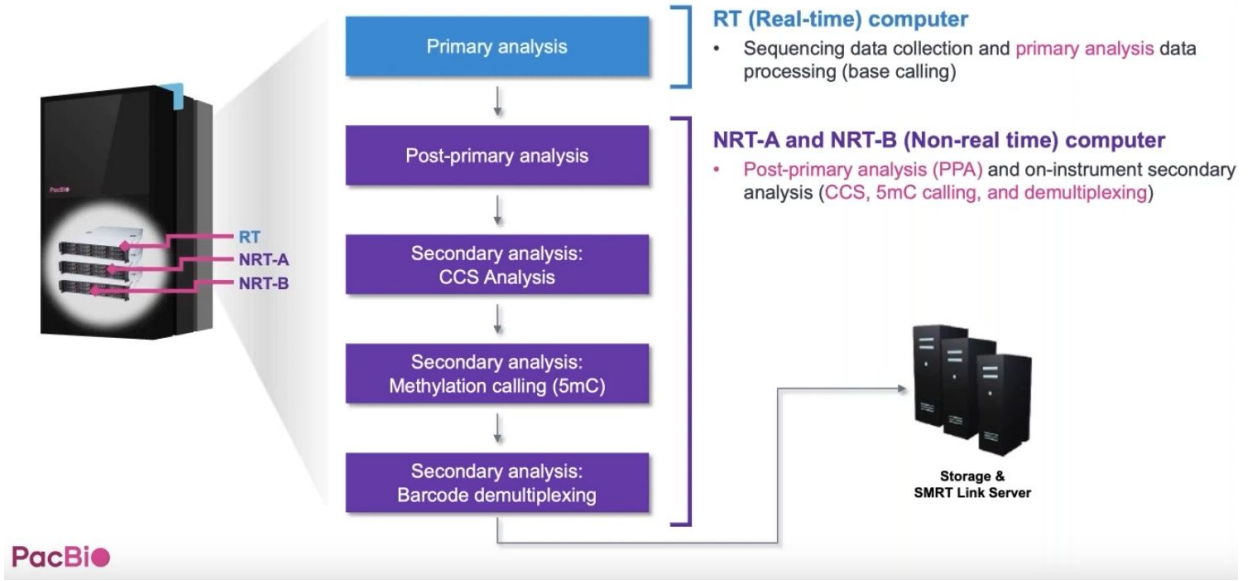


Image: Nanopore

Nanopore Basecalling

cDNA-PCR (R10.4.1)

Nanopore

Official tool	MinKNOW (dorado) Dorado standalone
Model	dna_r10.4.1_e8.2_400bps_fast@v5.2.0 dna_r10.4.1_e8.2_400bps_hac@v5.2.0 dna_r10.4.1_e8.2_400bps_sup@v5.2.0
Resource	PromethION Tower, P2I Professional/Gaming Laptop/Tower

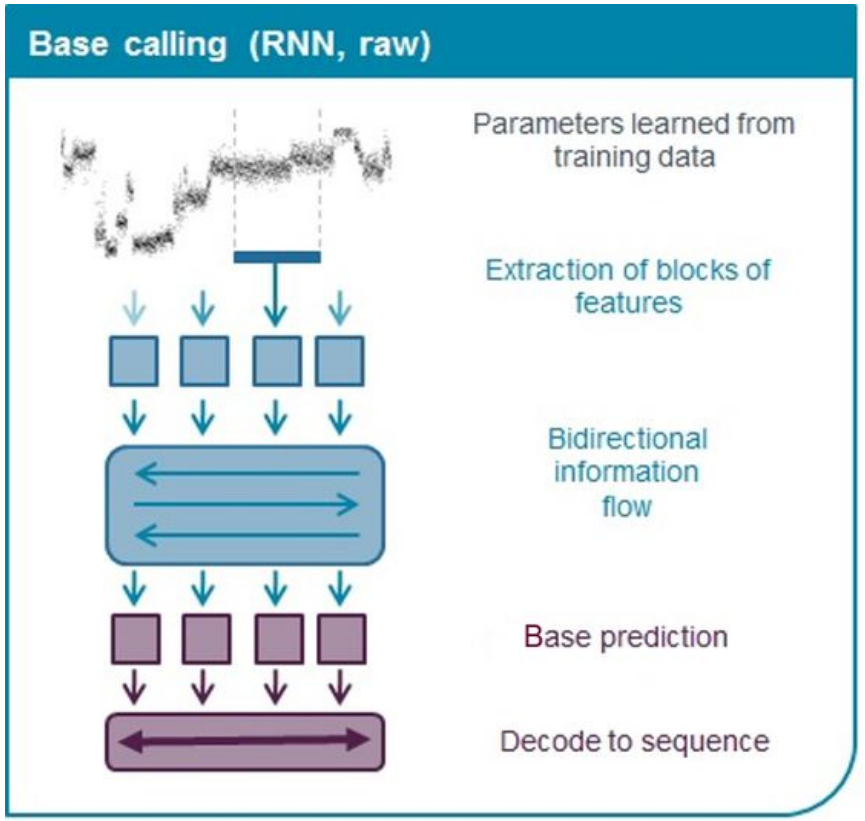


Image: Nanopore

Nanopore Basecalling

Direct RNA sequencing (RNA004)

Nanopore

Official tool	MinKNOW (dorado) Dorado standalone Remora
Model and mods	**rna004_130bps_fast@v5.2.0** **rna004_130bps_hac@v5.2.0** m5C m6A_DRACH inosine_m6A pseU **rna004_130bps_sup@v5.2.0** m5C_2OmeC m6A_DRACH inosine_m6A_2OmeA pseU_2OmeU 2OmeG
Resource	PromethION Tower, P2I Professional/Gaming Laptop/Tower

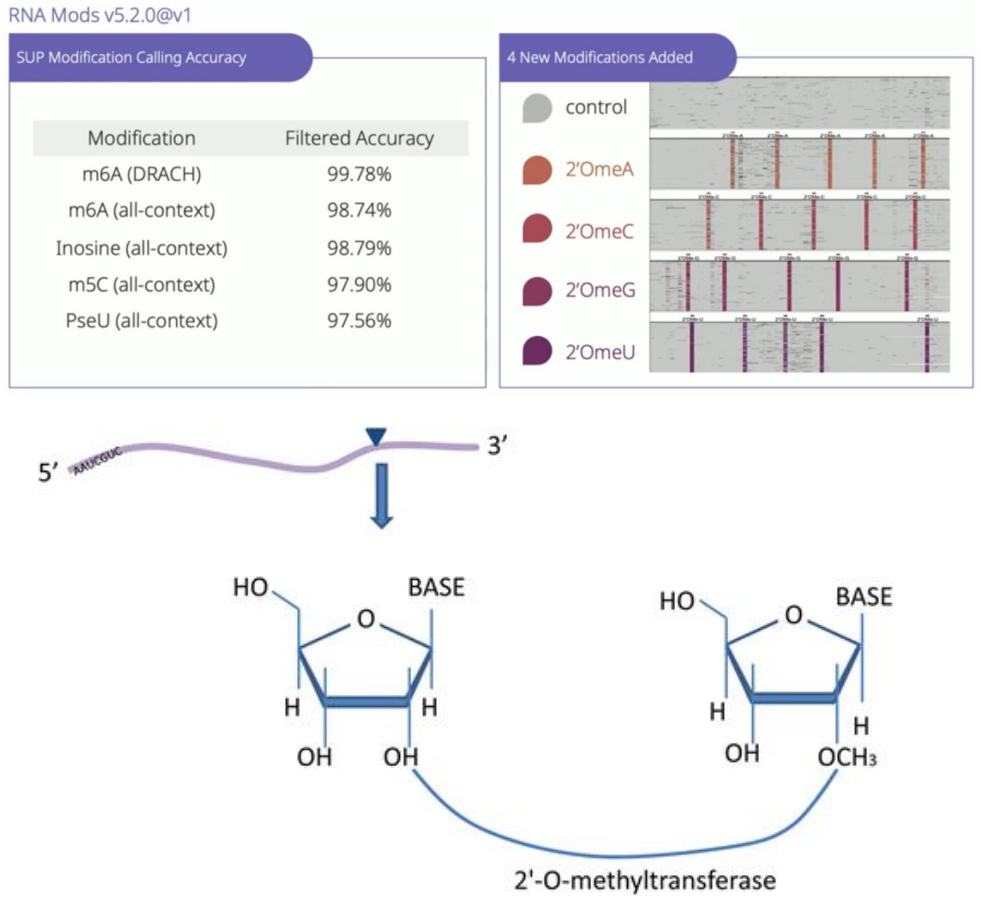


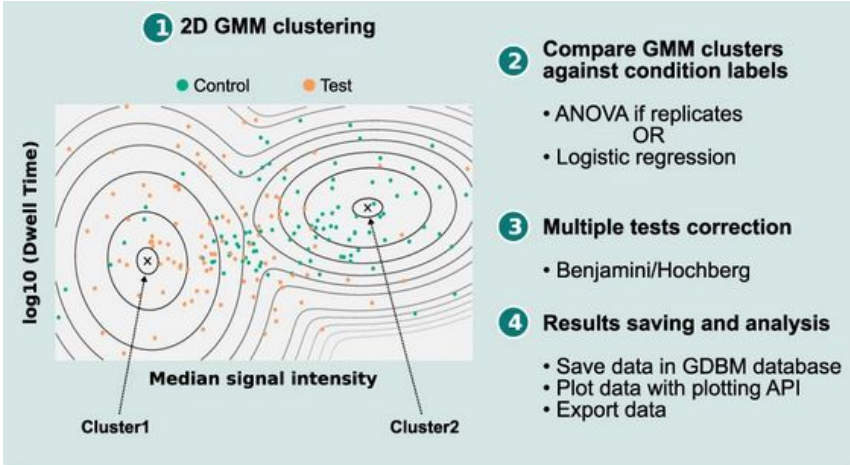
Image: Nanopore

Nanopore Modcalling

Direct RNA sequencing (RNA004)

Nanopore

Community tool	**EpiNano (Liu et al, 2019)** m6A **Nanocompore (Leger, 2021)** De novo in comparison with IVT **m6anet (Hendra et al, 2022)** m6A **NanoRMS (Begik, 2022)** De novo, modification sites **m6ABasecaller (Cruciani, 2023)** m6A **and many more …**



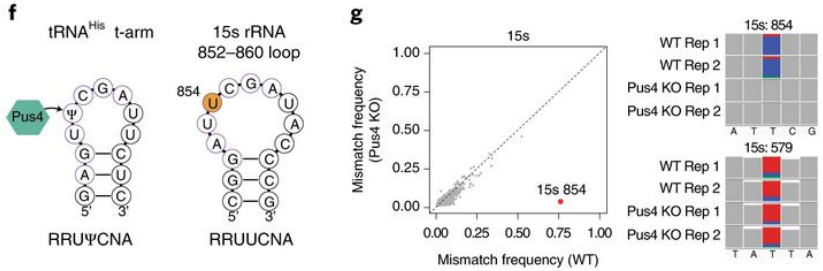


Image: Nanopore

Questions

1. What method kinnex was based on?
2. What improved throughput in Kinnex kit?
3. Why RT is performed in dRNA sequencing?
4. What mods supported by dorado?

Thank You!

For more information about the LongTREC Summer School:

https://longtrec.eu