



**University of Michigan
Medical School**

Long-read RNA sequencing data

Dingjie Wang, PhD

Department of Computational Medicine and Bioinformatics

University of Michigan

March 15, 2024

Database	Data type	Species	Platform	Library	# of Samples
GTEx	long read RNA-seq	Home sapiens (GTEx tissue and cell line)	ONT	PCR-cDNA	86
GTEx	long read RNA-seq	Home sapiens (GTEx tissue and cell line)	ONT	direct-cDNA	2
GTEx	long read RNA-seq	Home sapiens (K562 cell line)	ONT	direct-cDNA	4
GTEx	long read RNA-seq	Home sapiens (Heart)	ONT	direct-cDNA	4
ENCODE4	long read RNA-seq	Home sapiens (human tissues and cell lines)	PacBio: Sequel II	cDNA	95
ENCODE	long read RNA-seq	Home sapiens	ONT	Direct RNA	6
ENCODE	long read RNA-seq	Home sapiens	ONT	CapTrap	6
ENCODE	long read RNA-seq	Home sapiens	ONT	PCR-cDNA	3
ENCODE	long read RNA-seq	Home sapiens	ONT	R2C2	3
ENCODE	long read RNA-seq	Home sapiens	PacBio	SMRTbell	98
LRGASP	long read RNA-seq	Home sapiens	ONT: MinION	cDNA	4
LRGASP	long read direct RNA-seq	Home sapiens	ONT: MinION	Direct RNA	4
LRGASP	long read RNA-seq	Home sapiens	ONT: MinION	R2C2	4
LRGASP	long read RNA-seq	Home sapiens	ONT: MinION	CapTrap	4
LRGASP	long read RNA-seq	Home sapiens	PacBio:Sequel II	cDNA	4
LRGASP	long read RNA-seq	Home sapiens	PacBio:Sequel II	CapTrap	4
SG-Nex	long read RNA-seq	Home sapiens	ONT(MinION, GridION, PromethION)	PCR cDNA	8
SG-Nex	long read RNA-seq	Home sapiens	ONT(MinION, GridION, PromethION)	Direct cDNA	7
SG-Nex	long read direct RNA-seq	Home sapiens	ONT(MinION, GridION, PromethION)	Direct RNA	14
DirectRMDb	long read direct RNA-seq	Home sapiens	ONT	Direct RNA	43

Database	Data type	Species	Platform	Library	# of Samples
ENCODE	long read scRNA-seq	Mus musculus	ONT	Parse Single Cell Whole Transcriptome Kit	36
ENCODE	long read scRNA-seq	Mus musculus	PacBio	Parse Single Cell Whole Transcriptome Kit	28
ENCODE4	long read RNA-seq	Mus musculus (mouse tissues and cell lines)	PacBio: Sequel II	cDNA	64
ENCODE	long read RNA-seq	Mus musculus	ONT	Direct RNA	2
ENCODE	long read RNA-seq	Mus musculus	ONT	CapTrap	1
ENCODE	long read RNA-seq	Mus musculus	ONT	PCR-cDNA	1
ENCODE	long read RNA-seq	Mus musculus	ONT	R2C2	1
ENCODE	long read RNA-seq	Mus musculus	PacBio	SMRTbell	66
LRGASP	long read RNA-seq	Mus musculus	ONT: MinION	cDNA	1
LRGASP	long read direct RNA-seq	Mus musculus	ONT: MinION	Direct RNA	1
LRGASP	long read RNA-seq	Mus musculus	ONT: MinION	R2C2	1
LRGASP	long read RNA-seq	Mus musculus	ONT: MinION	CapTrap	1
LRGASP	long read RNA-seq	Mus musculus	PacBio:Sequel II	cDNA	1
LRGASP	long read RNA-seq	Mus musculus	PacBio:Sequel II	CapTrap	1
DirectRMDB	long read direct RNA-seq	Mus musculus	ONT	Direct RNA	18
LRGASP	long read RNA-seq	manatee	ONT: MinION	cDNA	1
LRGASP	long read direct RNA-seq	manatee	ONT: MinION	Direct RNA	1
LRGASP	long read RNA-seq	manatee	ONT: MinION	R2C2	1
LRGASP	long read RNA-seq	manatee	ONT: MinION	CapTrap	1
LRGASP	long read RNA-seq	manatee	PacBio:Sequel II	cDNA	1
LRGASP	long read RNA-seq	manatee	PacBio:Sequel II	CapTrap	1
DirectRMDB	long read direct RNA-seq	Drosophila melanogaster	ONT	Direct RNA	5
DirectRMDB	long read direct RNA-seq	Saccharomyces cerevisiae	ONT	Direct RNA	8
DirectRMDB	long read direct RNA-seq	Zea mays	ONT	Direct RNA	5
DirectRMDB	long read direct RNA-seq	Arabidopsis thaliana	ONT	Direct RNA	3
DirectRMDB	long read direct RNA-seq	Other	ONT	Direct RNA	42

Long-read RNA sequencing in GTEx consortium project

Article | Published: 03 August 2022

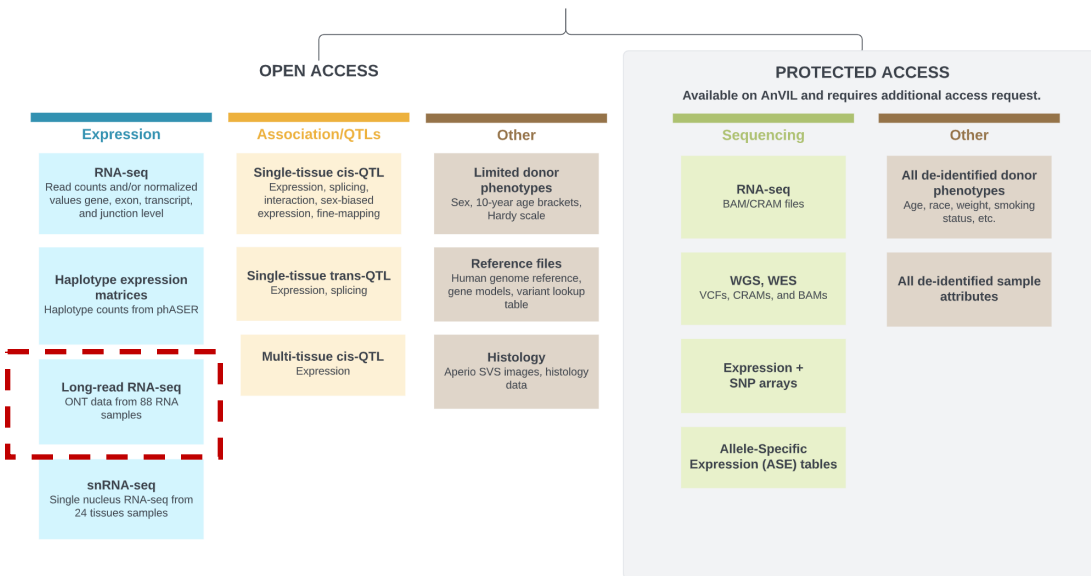
Transcriptome variation in human tissues revealed by long-read sequencing

[Dafni A. Glinos](#) ✉, [Garrett Garborcauskas](#) ✉, [Paul Hoffman](#), [Nava Ehsan](#), [Lihua Jiang](#), [Alper Gokden](#), [Xiaoguang Dai](#), [François Aguet](#), [Kathleen L. Brown](#), [Kiran Garimella](#), [Tera Bowers](#), [Maura Costello](#), [Kristin Ardlie](#), [Ruiqi Jian](#), [Nathan R. Tucker](#), [Patrick T. Ellinor](#), [Eoghan D. Harrington](#), [Hua Tang](#), [Michael Snyder](#), [Sissel Juul](#), [Pejman Mohammadi](#), [Daniel G. MacArthur](#), [Tuuli Lappalainen](#) ✉ & [Beryl B. Cummings](#) ✉

[Nature](#) **608**, 353–359 (2022) | [Cite this article](#)

34k Accesses | **63** Citations | **277** Altmetric | [Metrics](#)

Adult GTEx Data and Resources



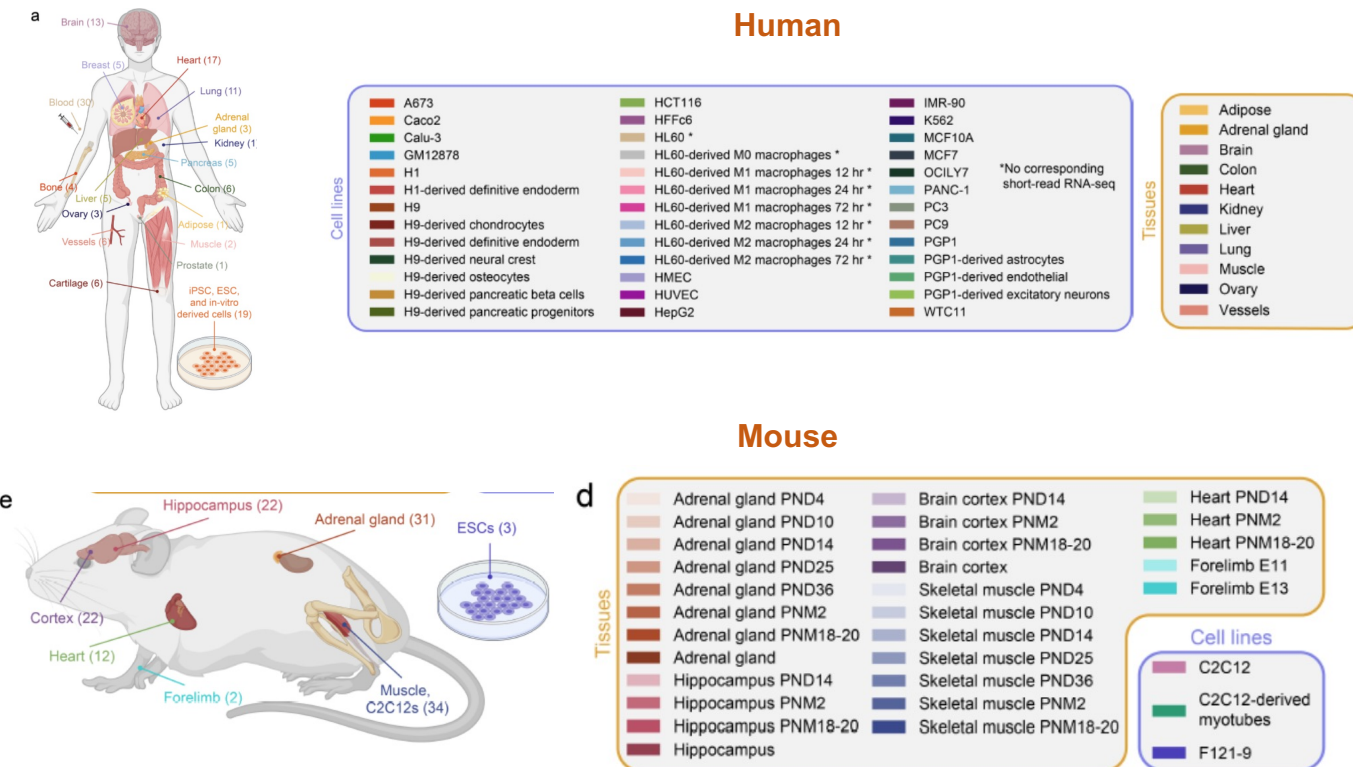
GTEx long-read RNA sequencing data

- **Platforms**
 - ONT: MinION and GridION
- **Protocols**
 - PCR-cDNA and direct-cDNA protocols
- **Reference genome**
 - hg38
- **Reference annotation**
 - GENCODE v26
- **# of samples**
 - **88** GTEx tissue and cell line samples
 - **86** PCR-cDNA and **2** direct-cDNA
 - **4** K562 cell line samples (PCR-cDNA)
 - **4** Heart - Left Ventricle (PCR-cDNA)

Long-read RNA sequencing in ENCODE4

The ENCODE4 long-read RNA-seq collection reveals distinct classes of transcript structure diversity

Fairlie Reese^{1,2}, Brian Williams³, Gabriela Balderrama-Gutierrez^{1,2}, Dana Wyman², Muhammed Hasan Çelik², Elisabeth Rebboah², Narges Rezaie², Diane Trout³, Milad Razavi-Mohseni^{4,5}, Yunzhe Jiang^{6,7}, Beatrice Borsari^{6,7,8}, Samuel Morabito², Heidi Yahan Liang², Cassandra J. McGill^{1,2}, Sorena Rahmanian², Jasmine Sakr^{2,9}, Shan Jiang^{1,2}, Weihua Zeng^{1,2}, Klebea Carvalho², Annika K. Weimer¹⁰, Louise A. Dionne¹¹, Ariel McShane^{12,13}, Karan Bedi^{14,15}, Shaimae I. Elhajjajy¹⁶, Sean Upchurch³, Jennifer Jou¹⁰, Ingrid Youngworth¹⁰, Idan Gabdank¹⁰, Paul Sud¹⁰, Otto Jolanki¹⁰, J. Seth Strattan¹⁰, Meenakshi S. Kagda¹⁰, Michael P. Snyder¹⁰, Ben C. Hitz¹⁰, Jill E. Moore¹⁶, Zhiping Weng¹⁶, David Bennett^{18,19}, Laura Reinholdt¹¹, Mats Ljungman^{15,20}, Michael A. Beer^{4,5}, Mark B. Gerstein^{6,7,21,22,23}, Lior Pachter^{3,24}, Roderic Guigó^{8,25}, Barbara J. Wold³, and Ali Mortazavi^{1,2}



Human long-read RNA sequencing data

- Platforms
 - PacBio: Sequel II
- Protocols
 - cDNA
- Human genome
 - hg38
- Reference annotation
 - GENCODE v29
- # of samples
 - 95 samples from 50 human tissues and cell lines

Mouse long-read RNA sequencing data

- Platforms
 - PacBio: Sequel II
- Protocols
 - cDNA
- Reference genome
 - mm10
- Reference annotation
 - GENCODE vM21
- # of samples
 - 64 samples from 31 mouse tissues and cell lines

Raw data are available at the following links:

• Human: <https://www.encodeproject.org/carts/829d339c-913c-4773-8001-80130796a367/>

• Mouse: <https://www.encodeproject.org/carts/55367842-f225-45cf-bf5e-5ba5e4182768/>

Long-read RNA sequencing in ENCODE (Include ENCODE4, collection from Aifu)

Long-read single-cell RNA-seq summary in ENCODE

Database	Data type	Species	Platform	Library	# of Samples
ENCODE	long read scRNA-seq	Mus musculus	ONT	Parse Single Cell Whole Transcriptome Kit	36
ENCODE	long read scRNA-seq	Mus musculus	PacBio	Parse Single Cell Whole Transcriptome Kit	28

Long-read RNA-seq summary in ENCODE

Database	Data type	Species	Platform	Library	# of Samples
ENCODE	long read RNA-seq	Home sapiens	ONT	Direct RNA	6
ENCODE	long read RNA-seq	Home sapiens	ONT	CapTrap	6
ENCODE	long read RNA-seq	Home sapiens	ONT	PCR-cDNA	3
ENCODE	long read RNA-seq	Home sapiens	ONT	R2C2	3
ENCODE	long read RNA-seq	Home sapiens	PacBio	SMRTbell	98
ENCODE	long read RNA-seq	Mus musculus	ONT	Direct RNA	2
ENCODE	long read RNA-seq	Mus musculus	ONT	CapTrap	1
ENCODE	long read RNA-seq	Mus musculus	ONT	PCR-cDNA	1
ENCODE	long read RNA-seq	Mus musculus	ONT	R2C2	1
ENCODE	long read RNA-seq	Mus musculus	PacBio	SMRTbell	66

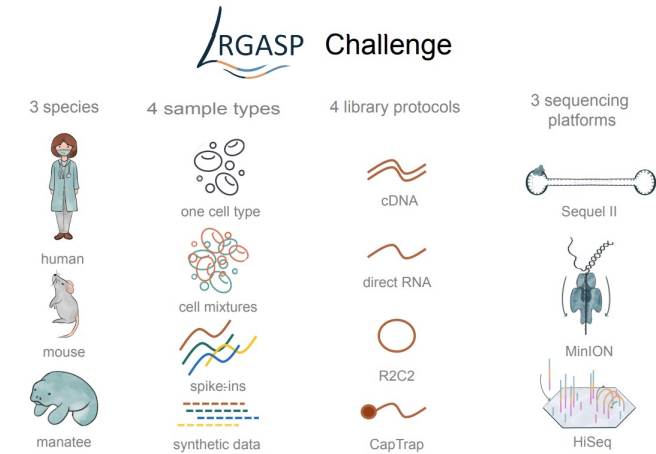
Long-read RNA sequencing in LRGASP consortium project

New Results [Follow this preprint](#)

Systematic assessment of long-read RNA-seq methods for transcript identification and quantification

Francisco J. Pardo-Palacios, Dingjie Wang, Fairlie Reese, Mark Diekhans, Sílvia Carbonell-Sala, Brian Williams, Jane E. Loveland, Maite De María, Matthew S. Adams, Gabriela Balderrama-Gutierrez, Amit K. Behera, Jose M. Gonzalez, Toby Hunt, Julien Lagarde, Cindy E. Liang, Haoran Li, Marcus Jerryd Meade, David A. Moraga Amador, Andrey D. Prijibelski,  Inanc Birol, Hamed Bostan, Ashley M. Brooks, Muhammed Hasan Çelik, Ying Chen, Mei R.M. Du, Colette Felton,  Jonathan Göke, Saber Hafezqorani, Ralf Herwig, Hideya Kawaji, Joseph Lee,  Jian-Liang Li,  Matthias Lienhard, Alla Mikheenko, Dennis Mulligan, Ka Ming Nip, Mihaela Persea, Matthew E. Ritchie, Andre D. Sim, Alison D. Tang, Yuk Kei Wan, Changqing Wang, Brandon Y. Wong, Chen Yang, If Barnes, Andrew Berry,  Salvador Capella,  Namrita Dhillon, Jose M. Fernandez-Gonzalez, Luis Ferrández-Peral, Natália Garcia-Reyero, Stefan Goetz, Carles Hernández-Ferrer, Liudmyla Kondratova, Tianyuan Liu, Alessandra Martinez-Martin, Carlos Menor, Jorge Mestre-Tomás, Jonathan M. Mudge, Nedka G. Panayotova, Alejandro Paniagua, Dmitry Repchevsky,  Eric Rouchka,  Brandon Saint-John, Enrique Sapena, Leon Sheynkman, Melissa Laird Smith, Marie-Marthe Suner,  Hazuki Takahashi, Ingrid Ashley Youngworth,  Piero Carninci, Nancy D. Denslow, Roderic Guigó, Margaret E. Hunter, Hagen U. Tilgner, Barbara J. Wold, Christopher Vollmers, Adam Frankish,  Kin Fai Au, Gloria M. Sheynkman,  Ali Mortazavi, Ana Conesa,  Angela N. Brooks

doi: <https://doi.org/10.1101/2023.07.25.550582>



Data summary in LRGASP consortium project

Database	Data type	Species	Platform	Library	# of Samples
LRGASP	long read RNA-seq	Home sapiens	ONT: MinION	cDNA	4
LRGASP	long read direct RNA-seq	Home sapiens	ONT: MinION	Direct RNA	4
LRGASP	long read RNA-seq	Home sapiens	ONT: MinION	R2C2	4
LRGASP	long read RNA-seq	Home sapiens	ONT: MinION	CapTrap	4
LRGASP	long read RNA-seq	Home sapiens	PacBio:Sequel II	cDNA	4
LRGASP	long read RNA-seq	Home sapiens	PacBio:Sequel II	CapTrap	4
LRGASP	long read RNA-seq	Mus musculus	ONT: MinION	cDNA	1
LRGASP	long read direct RNA-seq	Mus musculus	ONT: MinION	Direct RNA	1
LRGASP	long read RNA-seq	Mus musculus	ONT: MinION	R2C2	1
LRGASP	long read RNA-seq	Mus musculus	ONT: MinION	CapTrap	1
LRGASP	long read RNA-seq	Mus musculus	PacBio:Sequel II	cDNA	1
LRGASP	long read RNA-seq	Mus musculus	PacBio:Sequel II	CapTrap	1
LRGASP	long read RNA-seq	manatee	ONT: MinION	cDNA	1
LRGASP	long read direct RNA-seq	manatee	ONT: MinION	Direct RNA	1
LRGASP	long read RNA-seq	manatee	ONT: MinION	R2C2	1
LRGASP	long read RNA-seq	manatee	ONT: MinION	CapTrap	1
LRGASP	long read RNA-seq	manatee	PacBio:Sequel II	cDNA	1
LRGASP	long read RNA-seq	manatee	PacBio:Sequel II	CapTrap	1

Challenge 1: Transcript isoform detection with a high-quality genome

Challenge 2: Transcript isoform quantification

Challenge 3: *De novo* transcript isoform identification

Long-read RNA sequencing in SG-NEx consortium project

New Results

 [Follow this preprint](#)

A systematic benchmark of Nanopore long read RNA sequencing for transcript level analysis in human cell lines

 Ying Chen, Nadia M. Davidson, Yuk Kei Wan, Harshil Patel, Fei Yao, Hwee Meng Low, Christopher Hendra, Laura Watten, Andre Sim, Chelsea Sawyer, Viktoriia Iakovleva, Puay Leng Lee, Lixia Xin, Hui En Vanessa Ng, Jia Min Loo, Xuewen Ong, Hui Qi Amanda Ng, Jiaxu Wang, Wei Qian Casslynn Koh, Suk Yeah Polly Poon, Dominik Stanojevic, Hoang-Dai Tran, Kok Hao Edwin Lim, Shen Yon Toh, Philip Andrew Ewels, Huck-Hui Ng, N.Gopalakrishna Iyer, Alexandre Thiery, Wee Joo Chng, Leilei Chen, Ramanuj DasGupta, Mile Sikic, Yun-Shen Chan, Boon Ooi Patrick Tan, Yue Wan, Wai Leong Tam, Qiang Yu, Chiea Chuan Khor, Torsten Wüstefeld, Ploy N. Pratanwanich, Michael I. Love, Wee Siong Shoh Goh, Sarah B. Ng, Alicia Oshlack,  Jonathan Göke, SG-NEx consortium

doi: <https://doi.org/10.1101/2021.04.21.440736>

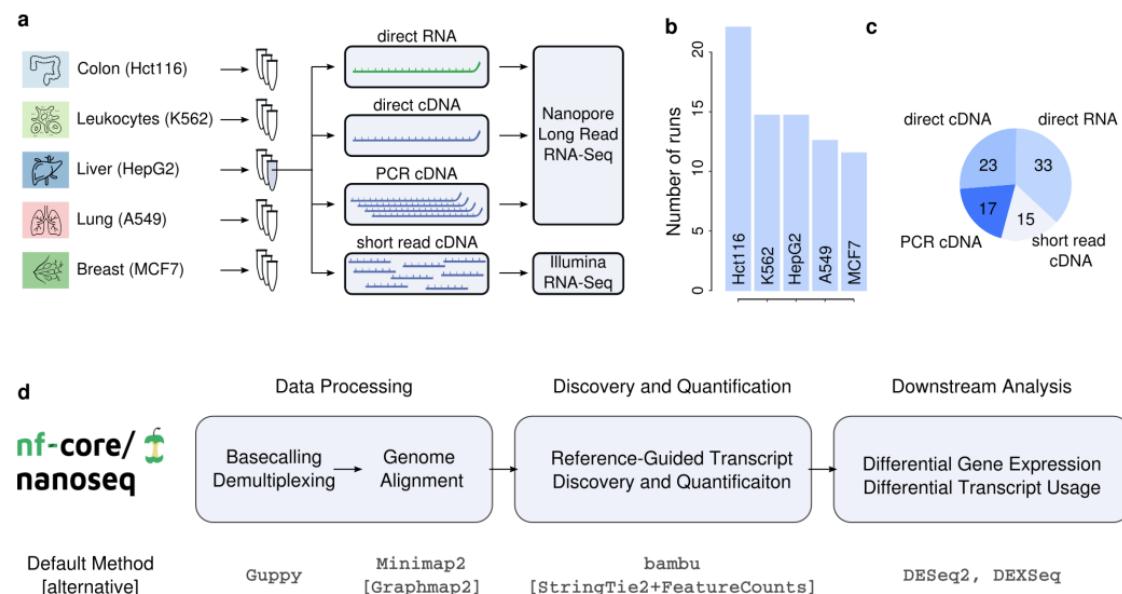


Figure 1. Overview of the Singapore Nanopore Expression (SG-NEx) datasets and processing pipeline

(a) 5 cancer cell lines were sequenced with multiple replicates using 4 different RNA-Seq protocols. (b) Shown is the number of sequencing runs generated for each SG-NEx cell line. (c) Shown is the number of sequencing runs for each of the sequencing technologies. (d) Illustration of the nf-core nextflow pipeline (nanoseq) for streamlined processing of Nanopore long read RNA-Seq data.

Data summary in SG-NEx consortium project

Database	Data type	Species	Platform	Library	# of runs	# of Samples
SG-Nex	long read RNA-seq	Home sapiens	ONT(MinION, GridION, PromethION)	PCR cDNA	27	8
SG-Nex	long read RNA-seq	Home sapiens	ONT(MinION, GridION, PromethION)	Direct cDNA	30	7
SG-Nex	long read direct RNA-seq	Home sapiens	ONT(MinION, GridION, PromethION)	Direct RNA	55	14

Github: <https://github.com/GoekelLab/sq-nex-data?tab=readme-ov-file>

Data are listed in the sample spreadsheet:
<https://github.com/GoekelLab/sq-nex-data/blob/master/docs/samples.tsv>

Long-read RNA sequencing in directRMDB database (Collection from Aifu)

JOURNAL ARTICLE

DirectRMDB: a database of post-transcriptional RNA modifications unveiled from direct RNA sequencing technology

Yuxin Zhang, Jie Jiang, Jiongming Ma, Zhen Wei, Yue Wang, Bowen Song, Jia Meng, Guifang Jia, João Pedro de Magalhães, Daniel J Rigden, Daiyun Hang, Kunqi Chen

Author Notes

Nucleic Acids Research, Volume 51, Issue D1, 6 January 2023, Pages D106–D116,
<https://doi.org/10.1093/nar/gkac1061>

Published: 16 November 2022 Article history

The overall design of DirectRMDB.

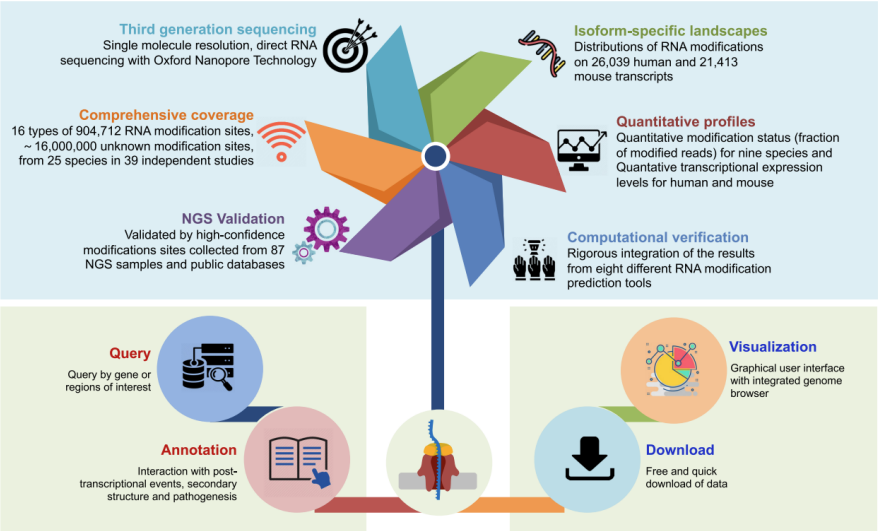


Table 1. brief description and comparison of modification calling tools

	Modification	Input	Isoform-level?	Control sample?	Algorithm
nanoPsu	Ψ	BAM	No	No	Sequencing error
ELIGOS	Mixed		No	No	
the work of Parker <i>et al.</i>	/		No	Yes	
Drummer		Nanopolish output	Yes	Yes	Current signal
xpore			Yes	Yes	
m6Anet	m ⁶ A	Tombo output	Yes	No	
MINES			No	No	
Nanom6A			No	No	

Note: ‘/’ means that detected modification types depend on the modification-free samples. For example, if an m⁶A-free sample is used as a control, reported sites are expected to be m⁶A methylation.

Table 2. The data statistics for DirectRMDB

Species	RNA modifications						Total
	m ⁶ A	Ψ	2'-O-Me	m ⁵ C	m ¹ A	Other	
Human	195 871	134 834	1506	26 033	2979	3803	365 026
Mouse	186 175	45 397	/	970	693	/	233 235
Yeast	148	59	/	/	/	19	226
22 other species	203 973	102 241	/	/	/	11	306 225

Note: The numbers in the table indicate the total count of each modification type. In human, ‘Other’ refers to m⁷G, m⁵U, m⁶Am, and AtoI modifications, while in yeast, ‘Other’ refers to ac⁴c, D, Y and f⁵C. Please refer to Supplementary Table S4 for more details.

Data summary in directRMDB database

Database	Data type	Species	Platform	Library	# of Samples
DirectRMDB	long read direct RNA-seq	Home sapiens	ONT	Direct RNA	43
DirectRMDB	long read direct RNA-seq	Mus musculus	ONT	Direct RNA	18
DirectRMDB	long read direct RNA-seq	Drosophila melanogaster	ONT	Direct RNA	5
DirectRMDB	long read direct RNA-seq	Saccharomyces cerevisiae	ONT	Direct RNA	8
DirectRMDB	long read direct RNA-seq	Zea mays	ONT	Direct RNA	5
DirectRMDB	long read direct RNA-seq	Arabidopsis thaliana	ONT	Direct RNA	3
DirectRMDB	long read direct RNA-seq	Other	ONT	Direct RNA	42

Thanks for your attention!