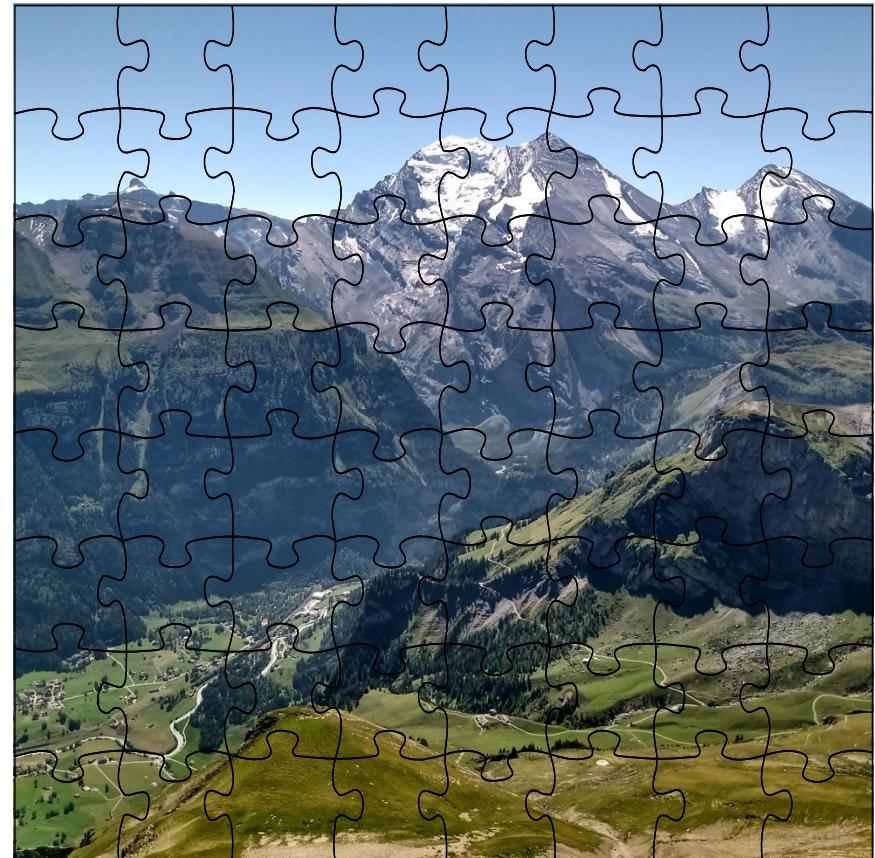
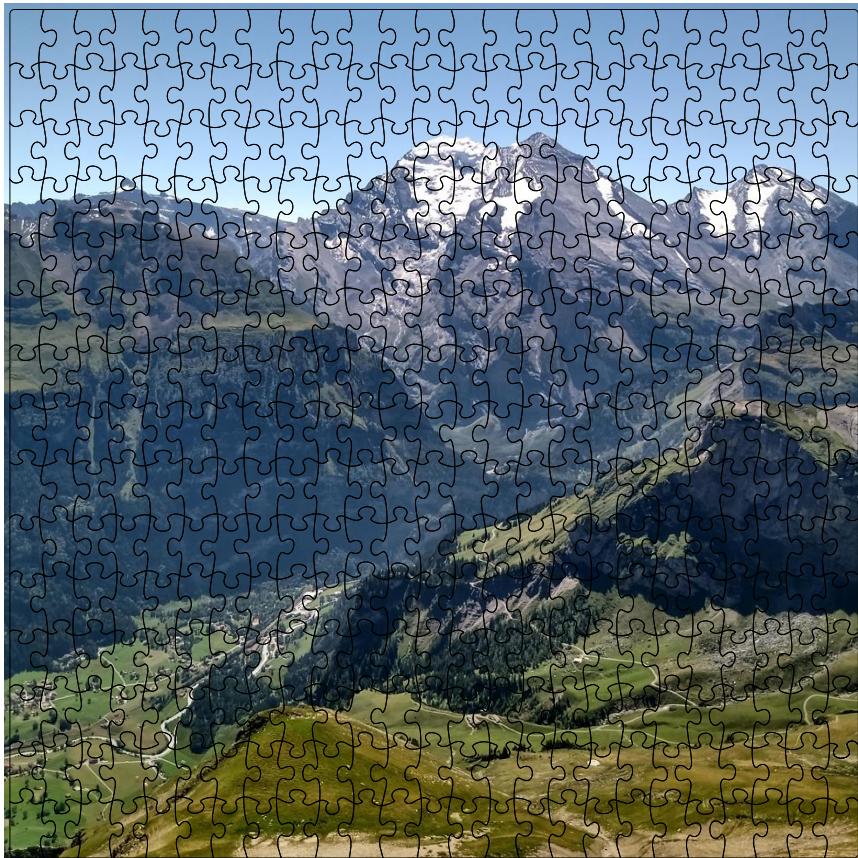


# Long-read sequence analysis

Applications

# Why long reads?



# Applications

- (Genome) assembly
- Variant analysis
- Transcriptome analysis
- Epigenetics
- Metagenomics

# Assembly

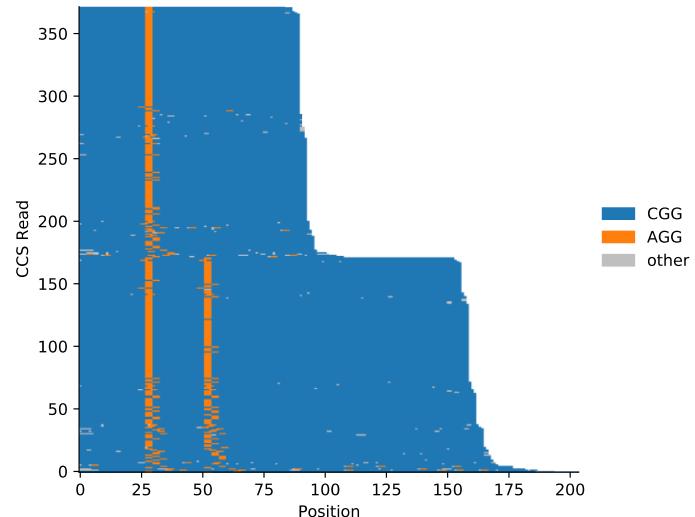
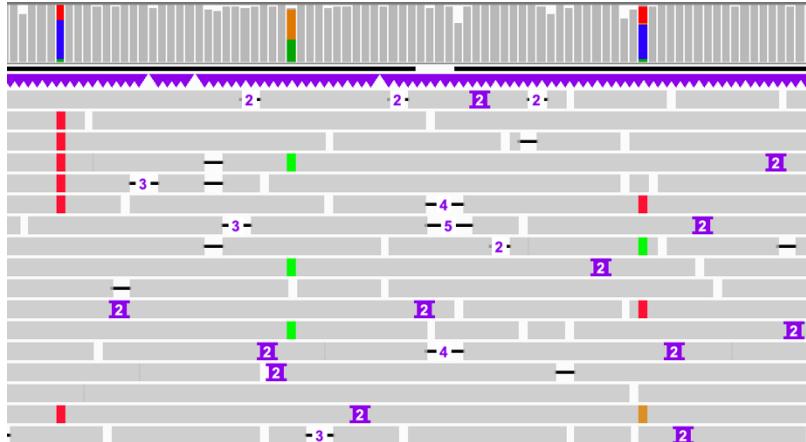
- Reconstructing a DNA sequence out of fragmented sequences:
  - Better understand variations between and within species
  - Reduces resources and increases accuracy for many applications
- Longer reads:
  - Less assembly errors
  - Higher contiguity (longer contigs)
  - Lower computational resources required

# LR assembly software

- Microbial:
  - Flye
  - Miniasm
  - Unicycler
  - Trycycler
- Large(r) genomes:
  - Shasta (ONT only)
  - Falcon (PacBio only)
  - Canu
  - Flye
  - Hifiasm
  - IPA
  - Peregrine
  - Verkko (integration PacBio and ONT)

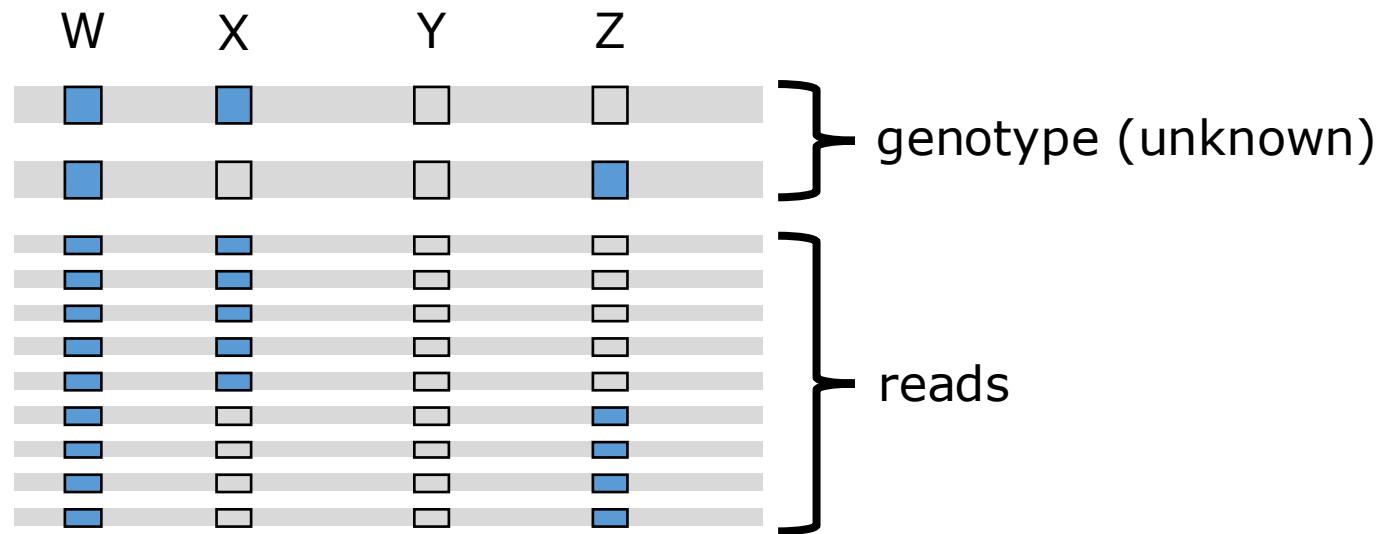
# Variant analysis

- Low accuracy long reads: not particularly good at variant analysis
- HiFi: good for large and short variation
- Major applications
  - Structural variation/large repeats
  - Phasing of variants

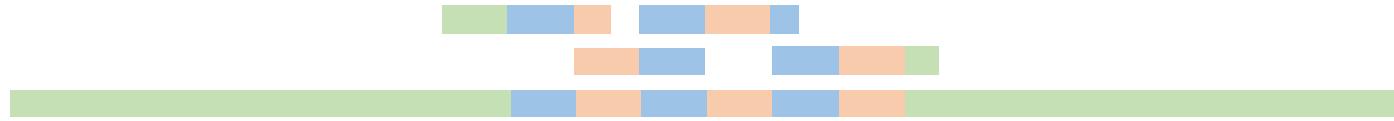


# Phasing

- Variant calling:
  - DeepVariant (Pacbio + ONT)
  - Clair3 (ONT)
- Phasing: Whatshap

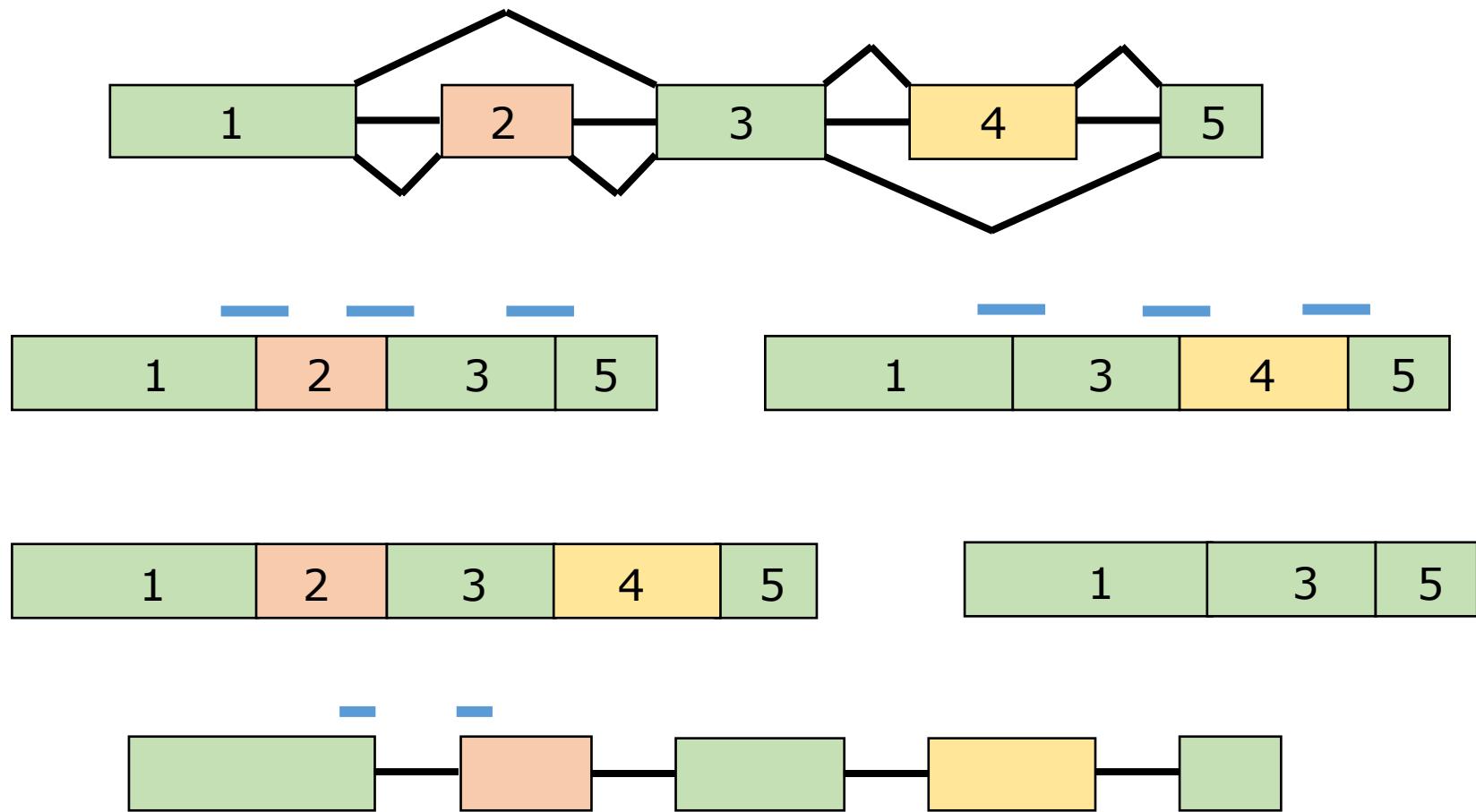


# Repeat expansion



- Medaka or PathSTR: ONT
- TRGT: PacBio

# Transcriptome analysis

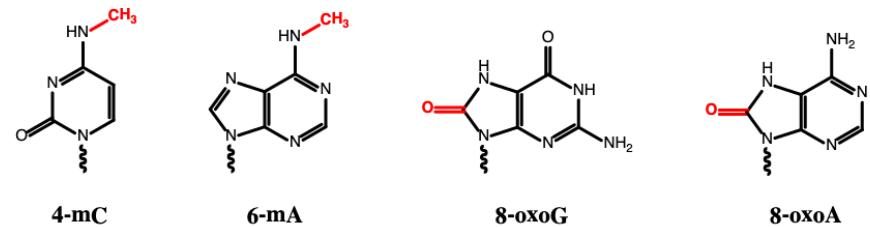
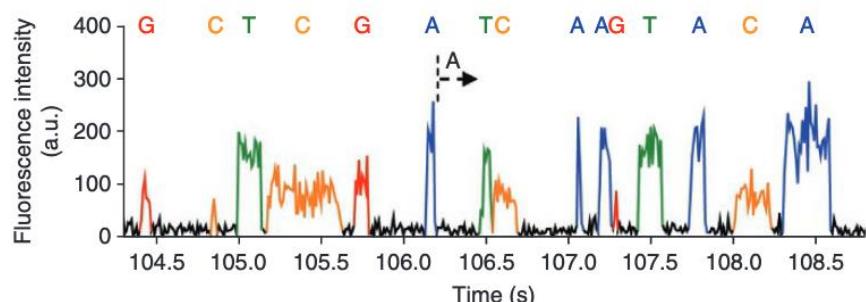
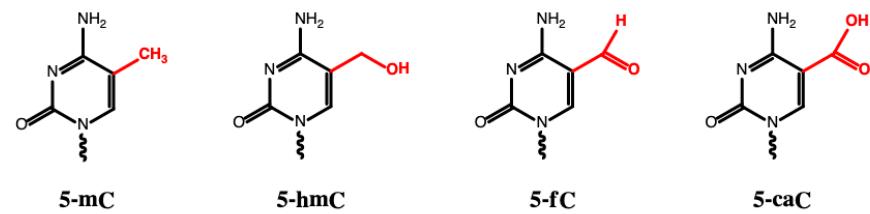
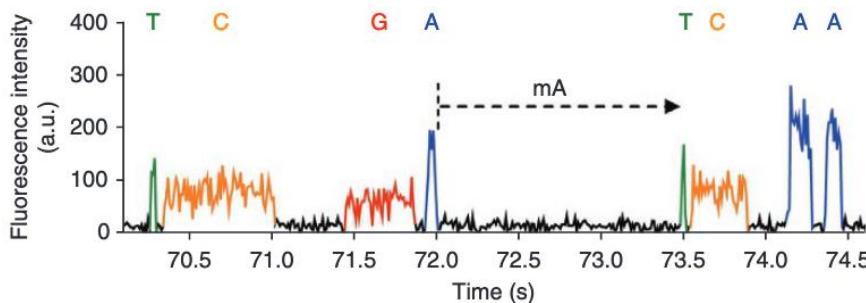


# Transcriptome analysis

- Full transcripts: less ambiguous identification + quantification of transcript isoforms
- Better gene model prediction (e.g. for genome annotation)
- ONT direct RNA-seq:
  - No PCR amplification
  - No reverse transcription
  - Base modification identification
- Relevance of RNA integrity
  - RIN > 8 (PacBio and ONT recommendations)
  - RNA degradation can bias results towards 3'

# Epigenetics

- Base modification in non-amplified libraries (Pacbio and ONT)
- No multiplexing
- Also possible in direct RNA-seq (ONT)



# Metagenomics

- Full length 16S genes: better classification (PacBio CCS)
- Metagenome Assembled Genomes (MAGs) -> flye --meta

<https://bmcmicrobiol.biomedcentral.com/articles/10.1186/s12866-021-02094-5>  
<https://bmcmicrobiol.biomedcentral.com/articles/10.1186/s12866-016-0891-4>  
<https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-021-07607-0>

# Long read tools

- Evolving field: new tools and updates

## Mini-Review

### A Hitchhiker's Guide to long-read genomic analysis

Medhat Mahmoud,<sup>1,4</sup> Daniel P. Agustinho,<sup>1,4</sup> and Fritz J. Sedlazeck<sup>1,2,3</sup>

<sup>1</sup>Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA; <sup>2</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA; <sup>3</sup>Department of Computer Science, Rice University, Houston, Texas 77005, USA

**Table 1.** List of methods for long-read analysis and its function

Function	Tool	Technology	GitHub	Citation
Basecalling (A)	CCS	PacBio	<a href="https://github.com/PacificBiosciences/ccs">https://github.com/PacificBiosciences/ccs</a>	
	Dorado	ONT	<a href="https://github.com/nanoporetech/dorado">https://github.com/nanoporetech/dorado</a>	
	Google Deep Consensus	PacBio	<a href="https://github.com/google/deepconsensus">https://github.com/google/deepconsensus</a>	Baid et al. 2023
Reads QC (B)	LongReadSum	ONT/PacBio	<a href="https://github.com/WGLab/LongReadSum">https://github.com/WGLab/LongReadSum</a>	
	LongQC	ONT/PacBio	<a href="https://github.com/yukasawa/LongQC">https://github.com/yukasawa/LongQC</a>	Fukasawa et al. 2020
	NanoPack	ONT/PacBio	<a href="https://github.com/wdecoster/nanopack">https://github.com/wdecoster/nanopack</a>	De Coster et al. 2018
Alignment (C)	LRA	ONT/PacBio	<a href="https://github.com/ChaissonLab/LRA">https://github.com/ChaissonLab/LRA</a>	Ren and Chaisson 2021
	minimap2	ONT/PacBio	<a href="https://github.com/lh3/minimap2">https://github.com/lh3/minimap2</a>	Li 2018, 2021
	NGMLR	ONT/PacBio	<a href="https://github.com/phires/ngmlr">https://github.com/phires/ngmlr</a>	Sedlazeck et al. 2018b
	pbmm2	PacBio	<a href="https://github.com/PacificBiosciences/pbmm2">https://github.com/PacificBiosciences/pbmm2</a>	
	VACmap	ONT/PacBio	<a href="https://github.com/micahvista/VACmap">https://github.com/micahvista/VACmap</a>	Ding et al. 2024
Alignment QC (D)	Vulcan	ONT/PacBio	<a href="https://gitlab.com/treangenlab/vulcan">https://gitlab.com/treangenlab/vulcan</a>	Fu et al. 2021
	Winnowmap2	ONT/PacBio	<a href="https://github.com/marbl/Winnowmap">https://github.com/marbl/Winnowmap</a>	Jain et al. 2022
	NanoPack	ONT/PacBio	<a href="https://github.com/wdecoster/nanopack">https://github.com/wdecoster/nanopack</a>	De Coster et al. 2018
	Sambamba	ONT/PacBio	<a href="https://github.com/biod/sambamba">https://github.com/biod/sambamba</a>	Tarasov et al. 2015
	SAMtools	ONT/PacBio	<a href="https://github.com/samtools/samtools">https://github.com/samtools/samtools</a>	Li et al. 2009
SNV & indels calling (E)	Clair3	ONT/PacBio	<a href="https://github.com/HKU-BAL/Clair3">https://github.com/HKU-BAL/Clair3</a>	Zheng et al. 2022
	DeepSomatic	Illumina/ONT/PacBio	<a href="https://github.com/google/deepsomatic">https://github.com/google/deepsomatic</a>	Park et al. 2024
	DeepVariant	ONT/PacBio	<a href="https://github.com/google/deepvariant">https://github.com/google/deepvariant</a>	Poplin et al. 2018
	Longshot	ONT/PacBio	<a href="https://github.com/pjedge/longshot">https://github.com/pjedge/longshot</a>	Edge and Bansal 2019
	cuteSV	ONT/PacBio	<a href="https://github.com/tjiangHT/cuteSV">https://github.com/tjiangHT/cuteSV</a>	Jiang et al. 2020
SV calling (F)	DELLY	ONT/PacBio/Illumina	<a href="https://github.com/dellytools/delly?tab=readme-ov-file">https://github.com/dellytools/delly?tab=readme-ov-file</a>	Rausch et al. 2012
	NanomonoSV	ONT/PacBio	<a href="https://github.com/friend1ws/nanomonsv">https://github.com/friend1ws/nanomonsv</a>	Shiraishi et al. 2023
	pbsv	PacBio	<a href="https://github.com/PacificBiosciences/pbsv">https://github.com/PacificBiosciences/pbsv</a>	
	Sawfish	PacBio	<a href="https://github.com/PacificBiosciences/sawfish">https://github.com/PacificBiosciences/sawfish</a>	Saunders et al. 2024
	Severus	ONT/PacBio	<a href="https://github.com/KolmogorovLab/Severus">https://github.com/KolmogorovLab/Severus</a>	Keskus et al. 2024
Copy number variants (G)	Sniffles2	ONT/PacBio	<a href="https://github.com/fritzsedlazeck/Sniffles">https://github.com/fritzsedlazeck/Sniffles</a>	Smolka et al. 2024
	SAVANA	ONT/PacBio	<a href="https://github.com/cortes-ciriano-lab/savana">https://github.com/cortes-ciriano-lab/savana</a>	Elrick et al. 2024
	SVision	ONT/PacBio	<a href="https://github.com/xjtu-omics/SVision">https://github.com/xjtu-omics/SVision</a>	Lin et al. 2022b
	HifICNV	PacBio	<a href="https://github.com/PacificBiosciences/HifICNV">https://github.com/PacificBiosciences/HifICNV</a>	
	Spectre	ONT/PacBio	<a href="https://github.com/fritzsedlazeck/Spectre">https://github.com/fritzsedlazeck/Spectre</a>	
Tandem repeat (TRs) (H)	Medaka	ONT	<a href="https://github.com/nanoporetech/medaka">https://github.com/nanoporetech/medaka</a>	Lee et al. 2021
	pathSTR	ONT	<a href="https://github.com/wdecoster/pathSTR">https://github.com/wdecoster/pathSTR</a>	De Coster et al. 2024
	Straglr	ONT	<a href="https://github.com/bcgsc/straglr">https://github.com/bcgsc/straglr</a>	Chiu et al. 2021
	StrSpy	ONT	<a href="https://github.com/unique379/strspy">https://github.com/unique379/strspy</a>	Hall et al. 2022
	TRGT	PacBio	<a href="https://github.com/PacificBiosciences/trgt">https://github.com/PacificBiosciences/trgt</a>	Dolzhenko et al. 2024
Genotyping (I)	kanpig	ONT/PacBio	<a href="https://github.com/ACEnglish/kanpig">https://github.com/ACEnglish/kanpig</a>	English et al. 2024a
	SVJedi	ONT/PacBio	<a href="https://github.com/llecompte/SVJedi">https://github.com/llecompte/SVJedi</a>	Romain and Lemaitre 2023
Downstream analysis—phasing (J)	HapCUT2	ONT/PacBio	<a href="https://github.com/vibansal/HapCUT2">https://github.com/vibansal/HapCUT2</a>	Edge et al. 2017
	HiPhase	PacBio	<a href="https://github.com/PacificBiosciences/HiPhase">https://github.com/PacificBiosciences/HiPhase</a>	Holt et al. 2024
	WhatsHap	ONT/PacBio	<a href="https://github.com/whatshap/whatsHap">https://github.com/whatshap/whatsHap</a>	Martin et al. 2023

(continued)

# Long read tools

- Evolving field: new tools and updates

nature reviews genetics

<https://doi.org/10.1038/s41578-025-00828-z>

Review article

Check for updates

## Transcriptomics in the era of long-read sequencing

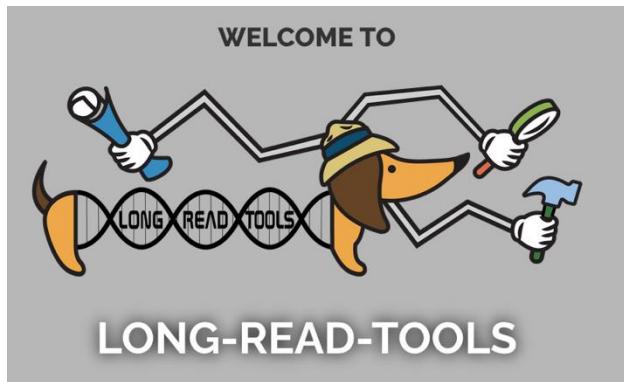
Carolina Monzó <sup>1,2</sup>, Tianyuan Liu <sup>1,2</sup> & Ana Conesa <sup>1,2,4</sup>

**Table 2 | Software tools for long-read RNA sequencing data processing and downstream analysis**

Tool	Usage	Platform	Refs.
DeepConsensus	Basecalling	PacBio	82
Dorado	Basecalling	ONT	<a href="https://nanoporetech.com/platform/accuracy/">https://nanoporetech.com/platform/accuracy/</a>
Minimap2	Mapping	PacBio, ONT	84
uLTRA	Mapping	PacBio, ONT	85
Graphmap2	Mapping	PacBio, ONT	86
deSALT	Mapping	PacBio, ONT	87
2passtools	Mapping	PacBio, ONT	88
Splam	Mapping	PacBio, ONT	94
Magic-BLAST	Mapping	PacBio, ONT	90
FLAIR	Read clustering Transcript identification and quantification	PacBio, ONT	92
Mandalorion	Read clustering Transcript identification and quantification	PacBio, ONT	31
IsoQuant	Splicing and intron graph construction Alignment correction Transcript identification and quantification Gene-level visualization	PacBio, ONT	32
IsoTools	Splicing and intron graph construction Transcript identification and quantification Differential isoform expression Differential splicing analysis Quality control visualization Gene-level visualization	PacBio, ONT	37
Isosceles	Splicing and intron graph construction Alignment correction Transcript identification and quantification at single-cell, pseudo-bulk and bulk resolution	PacBio, ONT	93
StringTie2	Splicing and intron graph construction Transcript identification and quantification	PacBio, ONT	94
ESPRESSO	Read classification Splice junction correction Transcript identification and quantification Gene-level visualization	PacBio, ONT	95
TALON	Read classification Transcript identification and quantification	PacBio, ONT	96
FLAMES	Read classification Transcript identification and quantification	PacBio, ONT	97
Bambu	Read classification Machine-learning filtering for true novel transcripts Transcript identification and quantification	PacBio, ONT	33
Freddie	Read clustering Transcript identification and quantification	PacBio, ONT	98
LyRIC	Transcript identification and quantification	PacBio, ONT	<a href="https://github.com/guigolab/LyRIC">https://github.com/guigolab/LyRIC</a>
IsoSplitter	Read clustering Transcript identification and quantification	PacBio, ONT	99
RNA-Bloom2	Splicing and intron graph construction Transcript identification and quantification	PacBio, ONT	100
rnaSPAdes	Splicing and intron graph construction Transcript identification and quantification	PacBio, ONT	101

# Long read tools

<https://long-read-tools.org/>



# Downstream analysis

- ONT:
  - Workflows on [EPI2ME](#)
- PacBio:
  - Pbbioconda
- nf-core
  - <https://nf-co.re/bacass>
  - <https://nf-co.re/viralrecon>
  - <https://nf-co.re/isoseq>
  - <https://nf-co.re/mag>
  - <https://nf-co.re/nanoseq>
  - <https://nf-co.re/ampliseq>

