Funded by
the European Union

Bioinformatics Summer School

Long-reads *Epi*-transcriptomics

Leda Katopodi

Centre for Genomic Regulation,
Barcelona, Spain

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

Course Contents

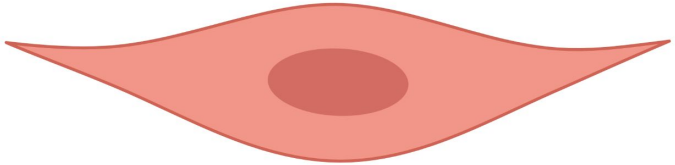
1. Introduction to Epitranscriptomics
2. Hands-on session: Theory & rationale
3. Hands-on session: Analysing Nano-tRNAseq data with AMaNITA
4. Discussion

Epitranscriptomics w/ Long-Read Sequencing Technologies

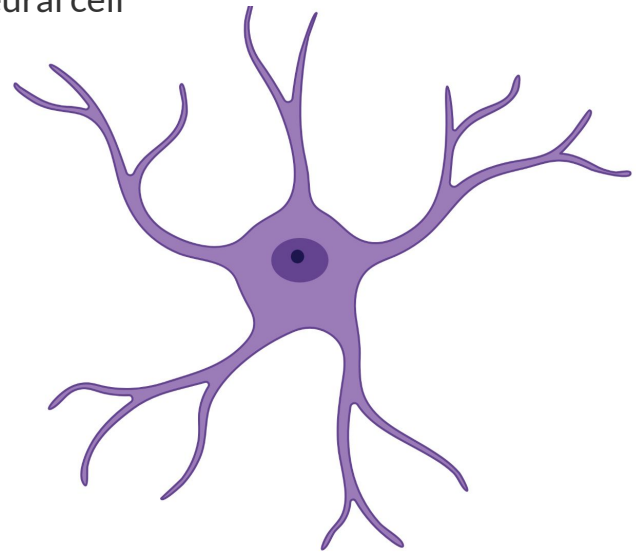
The promise of native RNA sequencing
for the detection of RNA modifications

Why do we care about epitranscriptomics?

muscle cell



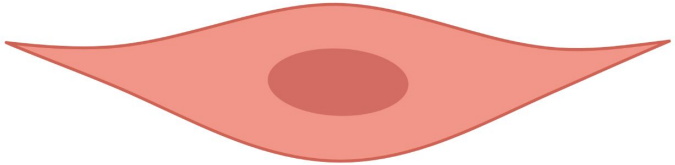
neural cell



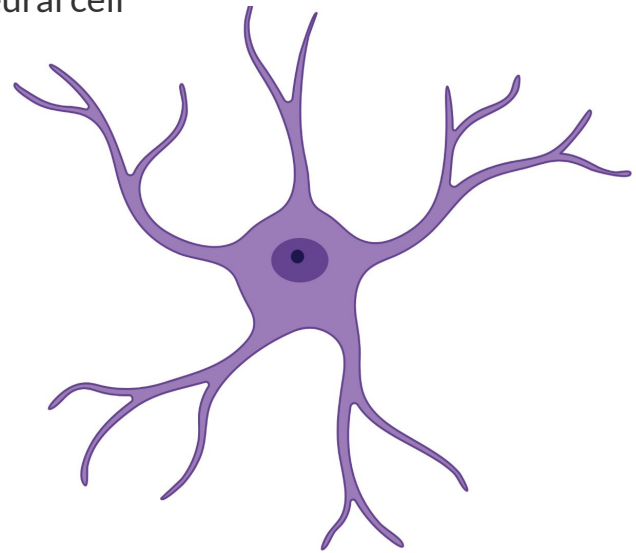
How can the same DNA
give rise to different cell types?

Why do we care about epitranscriptomics?

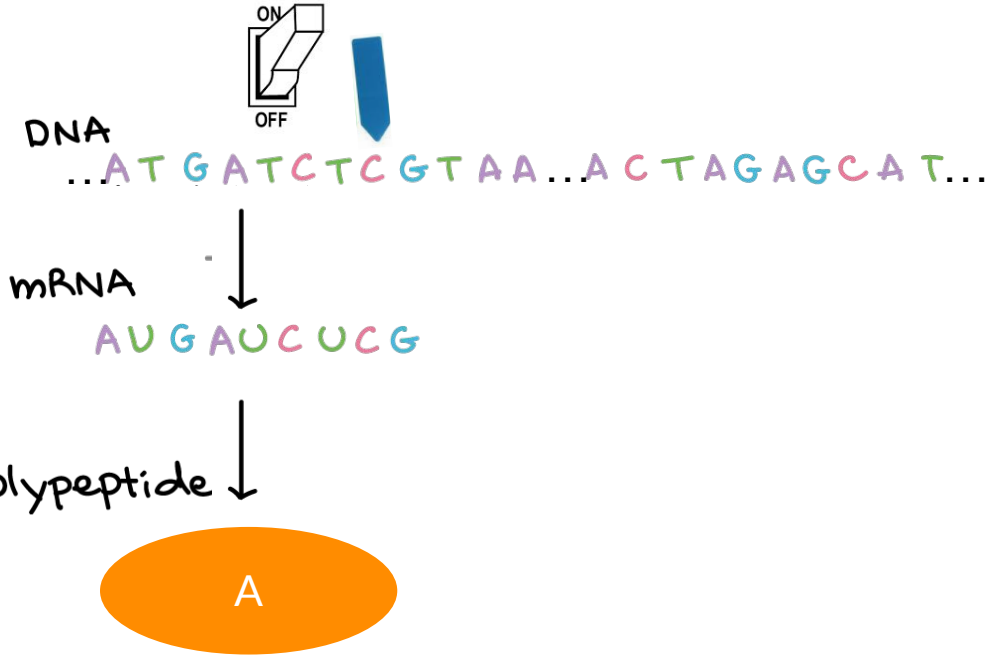
muscle cell



neural cell

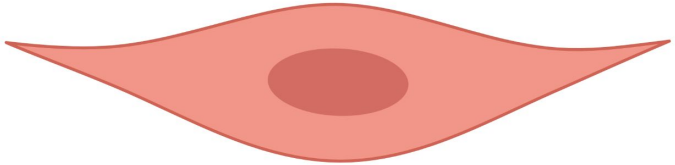


How can the same DNA
give rise to different cell types?

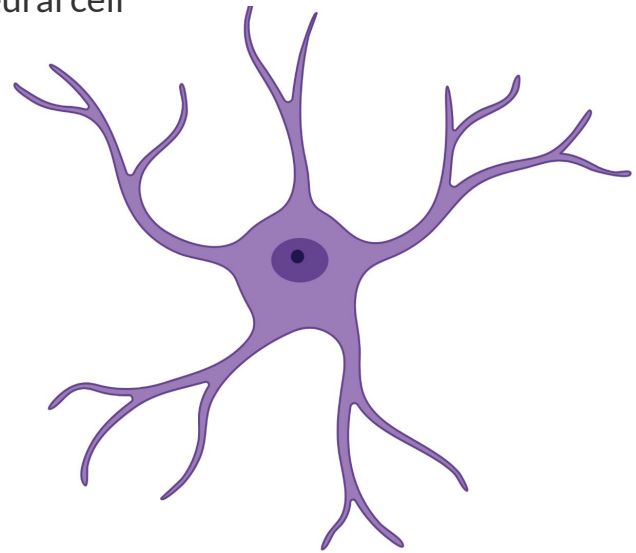


Why do we care about epitranscriptomics?

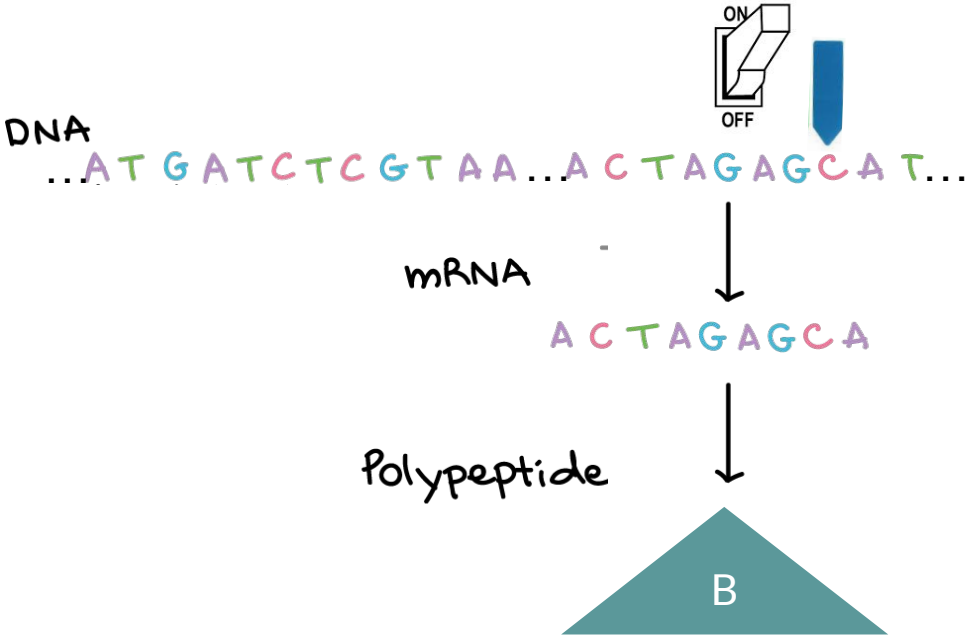
muscle cell



neural cell

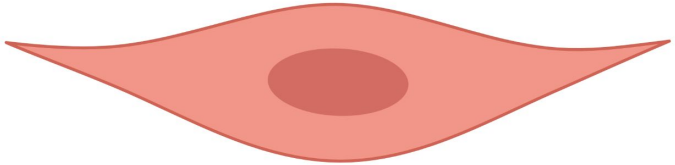


How can the same DNA
give rise to different cell types?

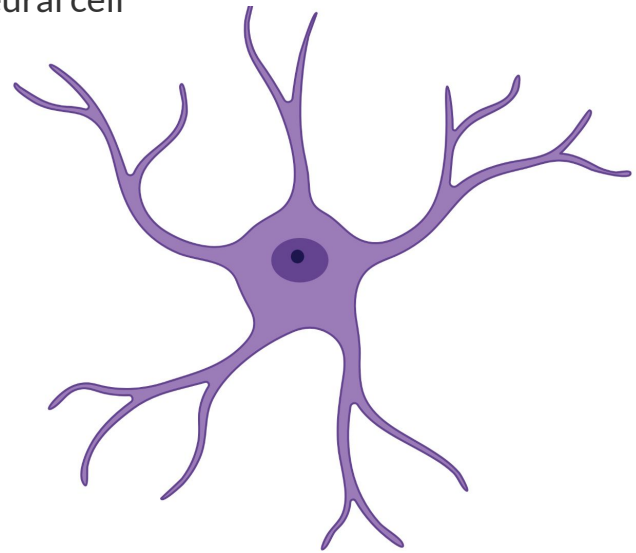


Why do we care about epitranscriptomics?

muscle cell

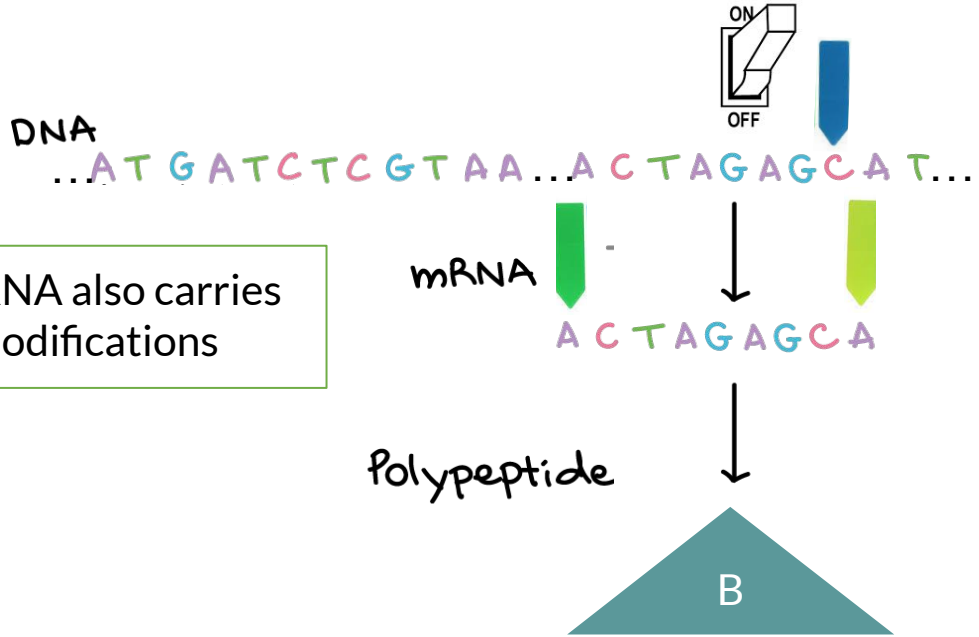


neural cell



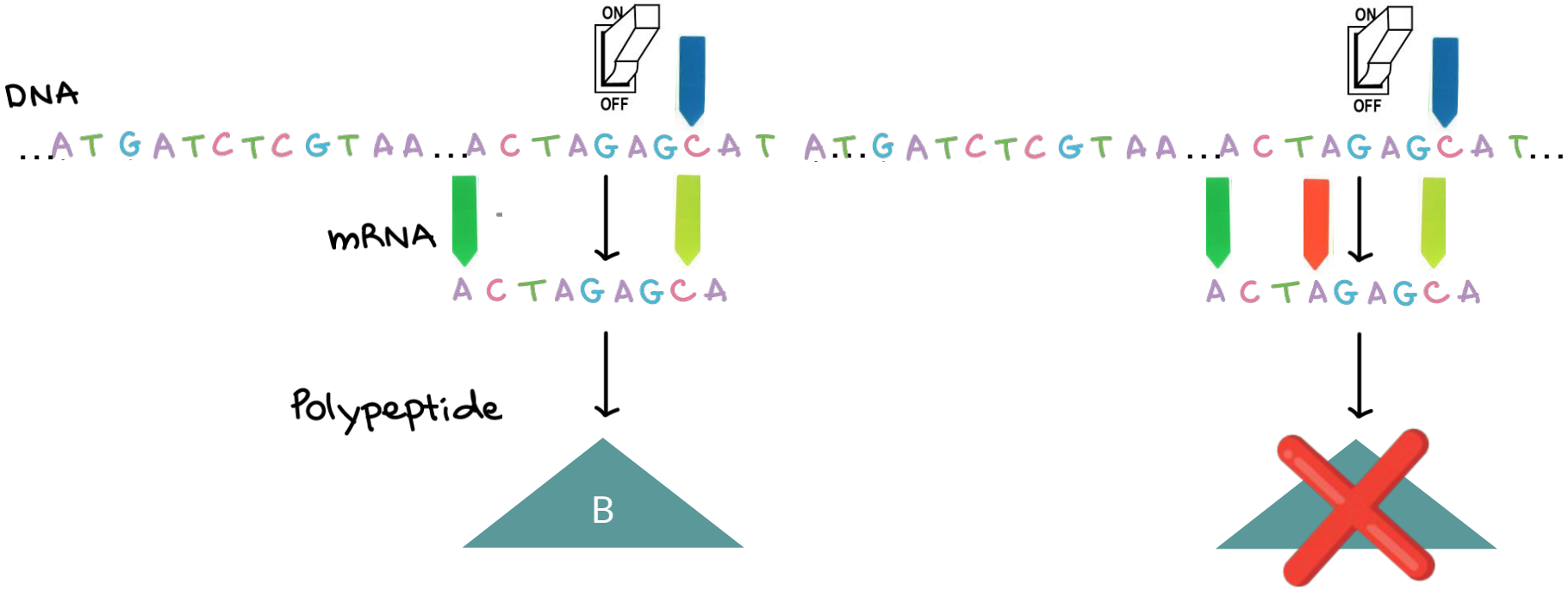
How can the same DNA
give rise to different cell types?

the RNA also carries modifications



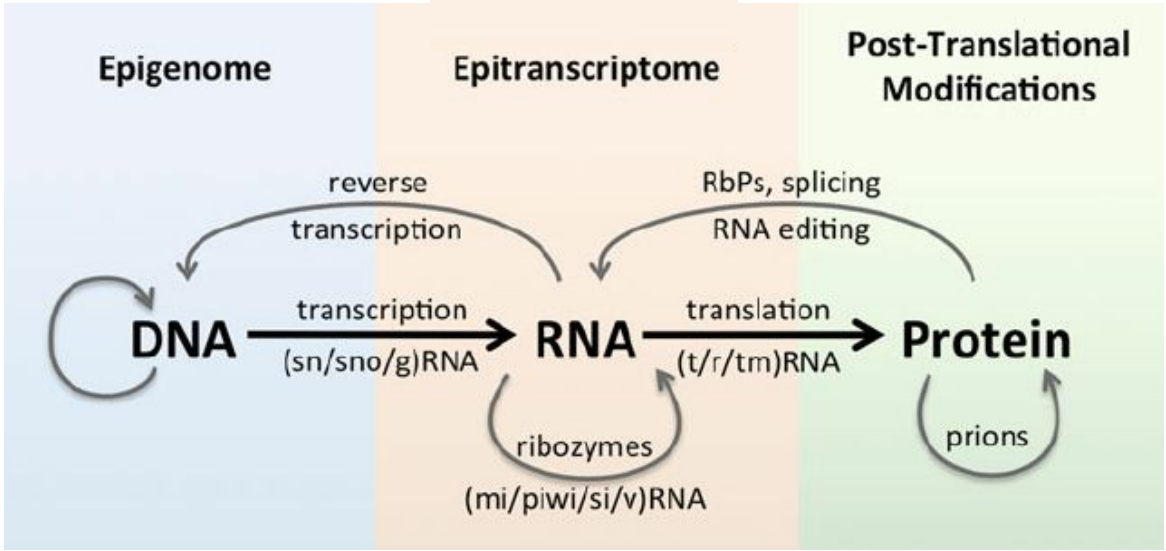
Why do we care about epitranscriptomics?

and those RNA modifications may overwrite DNA modifications



The central dogma of biology is a lie

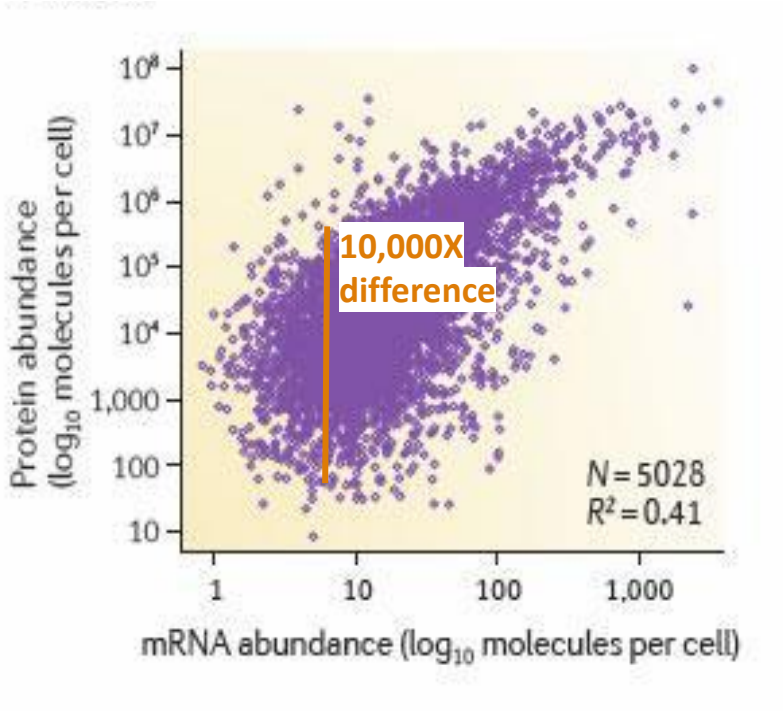
The central dogma of biology is a lie… or just an omission of the actual truth



Saletore et al., *Genome Biology* 2012

A largely disregarded player: *the post-transcriptional regulatory layer*

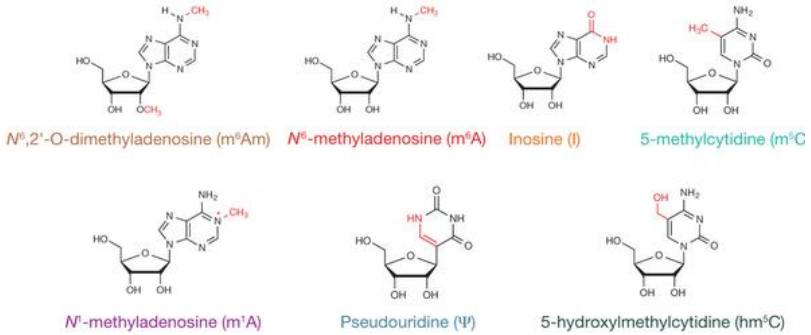
- Translation efficiency
- tRNA availability
- mRNA half life
- Codon usage
- RNA structure



- miRNA activity
- RNA binding proteins
- 5’ RNA degradation
- Ribosome specialization
- **RNA modifications (epitranscriptome)**

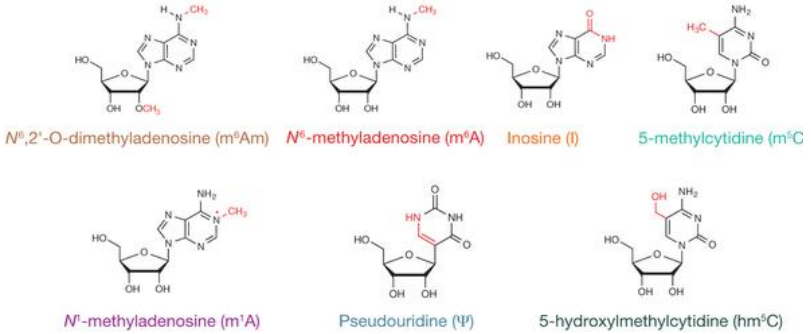
A historical overview: 1950s-2010s

1950s: RNA modifications fine-tune structure/function

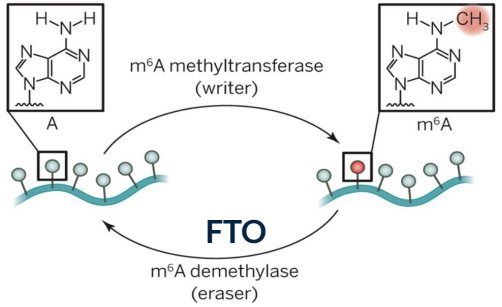


A historical overview: 1950s-2010s

1950s: RNA modifications fine-tune structure/function



2011: RNA modifications are reversible!

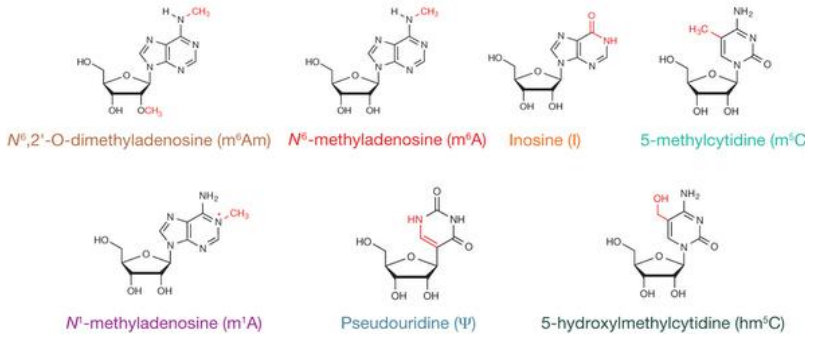


>> **function**

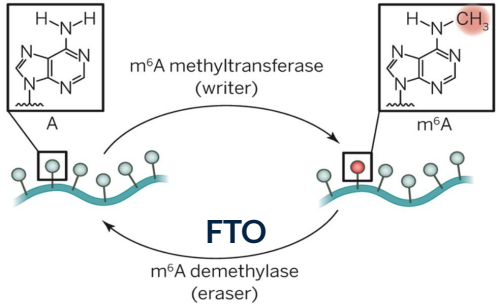
Jia et al., *Nature Chemical Biology* 2011

A historical overview: 1950s-2010s

1950s: RNA modifications fine-tune structure/function



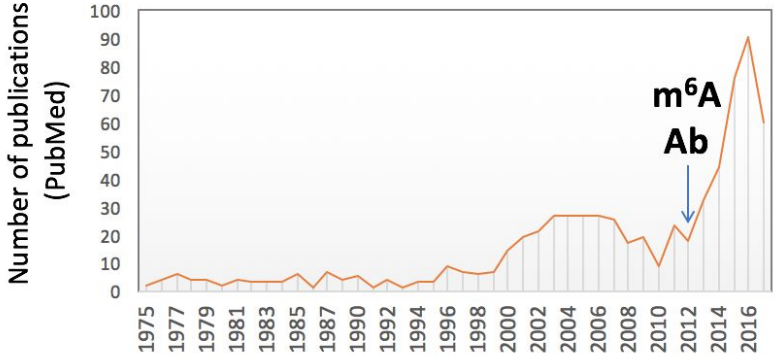
2011: RNA modifications are reversible!



>> **function**

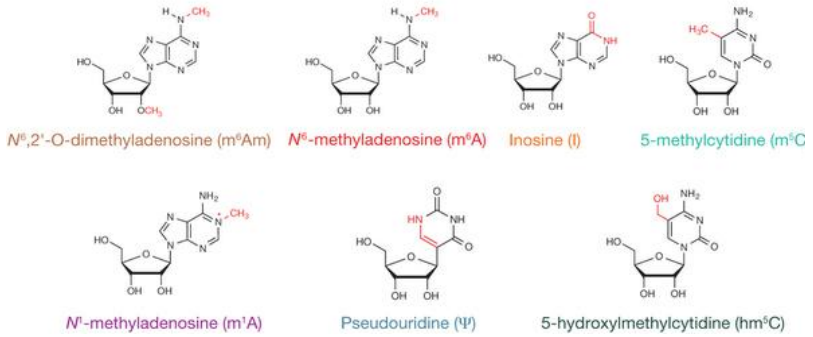
Jia et al., *Nature Chemical Biology* 2011

2012: First genome-wide method (m6A-Seq)

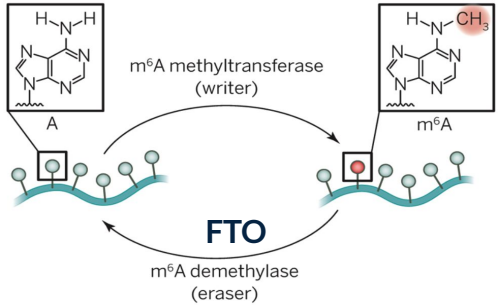


A historical overview: 1950s-2010s

1950s: RNA modifications fine-tune structure/function



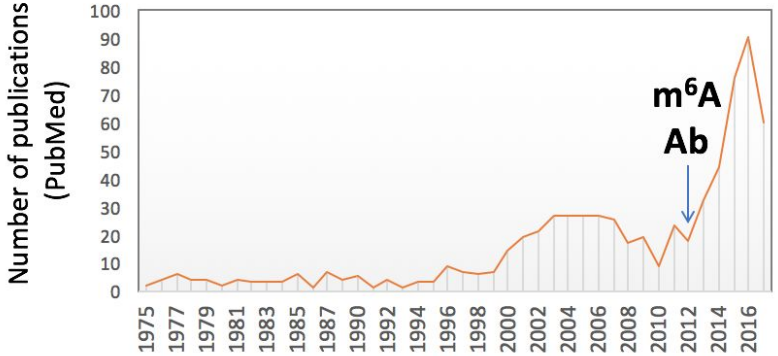
2011: RNA modifications are reversible!



>> **function**

Jia et al., *Nature Chemical Biology* 2011

2012: First genome-wide method (m6A-Seq)



2013+: Pivotal roles of m6A in cellular functions

- Cell differentiation (*2014*)
- Stress responses (*2015*)
- mRNA half lives /RNA stability (*2013*)
- Sex determination (*2016*)
- Embryonic development (*2017*)
- Splicing (*2024*)

But… what about other modifications?

How can RNA modifications be detected with sequencing-based methods?

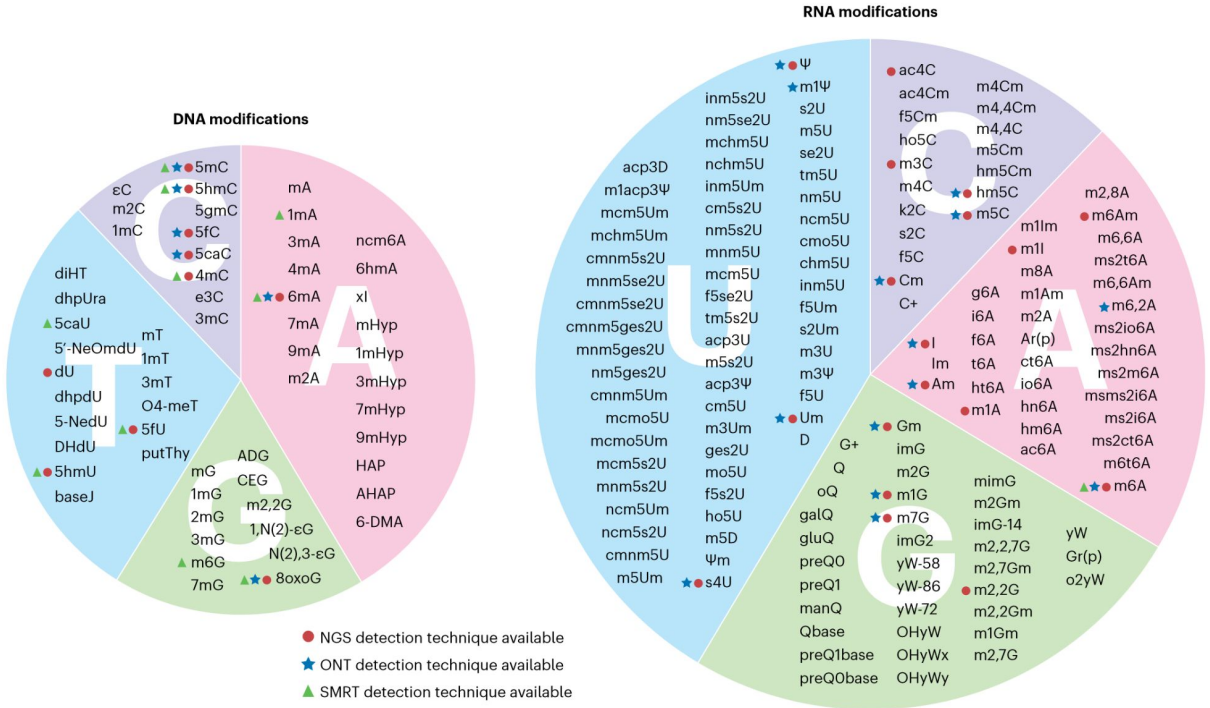
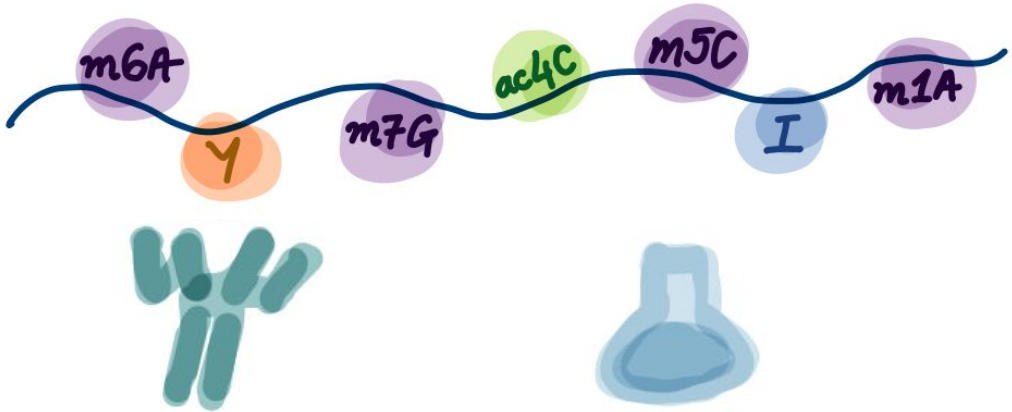


Fig. 1 | List of DNA and RNA modifications, and sequencing-based methods used to detect them. Known DNA (left) and RNA (right) modifications classified by their reference nucleotide; sequencing methods used to detect them are indicated. Adapted from ref. [6].

Jonkhout et al., *RNA* 2017
Lucas et al., *Nature Methods* 2023

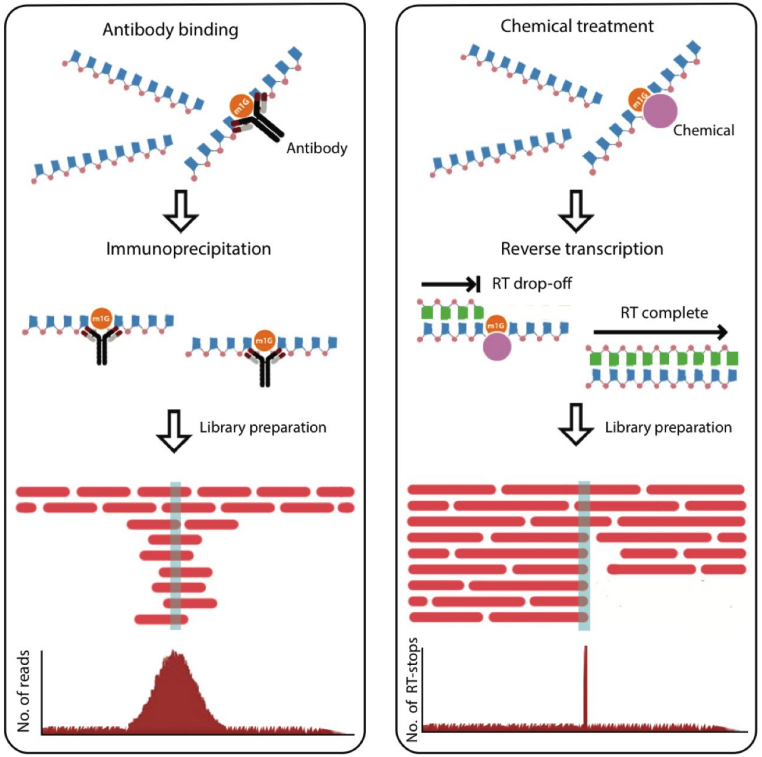
How can RNA modifications be detected with short-read sequencing?



How can RNA modifications be detected with short-read sequencing?

Antibody-based
RIP-Seq

e.g. m6A, m1A

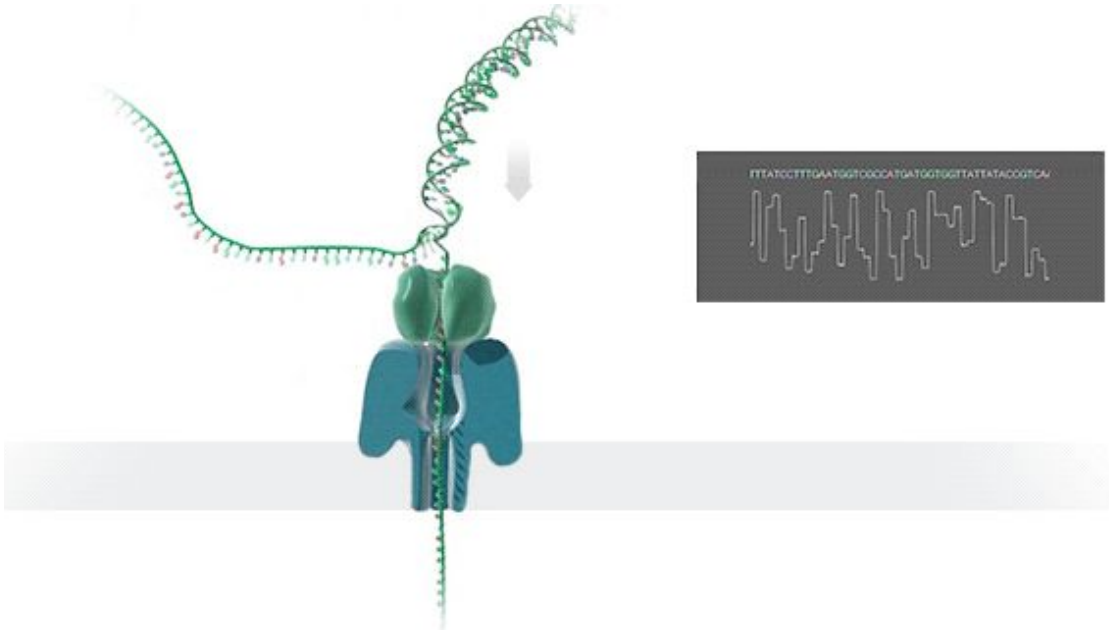


Chemical-based
Chem-Seq

e.g. m5C, Y

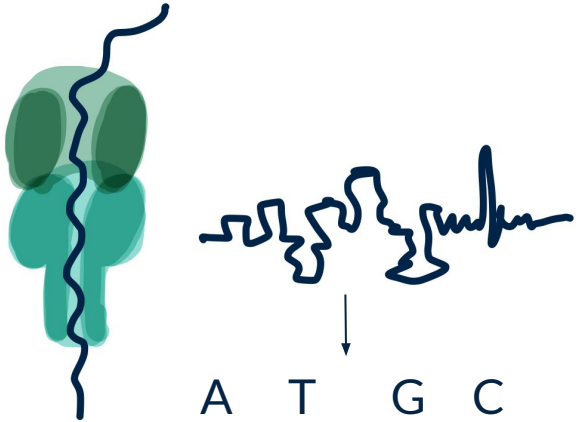
Jonkhout et al., *RNA* 2017

The promise of DRS



- No PCR bias
- **In principle, all RNA modifications**
- No custom protocols
- Stoichiometry information (% modified)
- Isoform-specific RNA modifications
- Multiple RNA modification types (m6A, m1A, m5C, etc) within the same molecule
- PolyA tail lengths
- **Single molecule resolution**

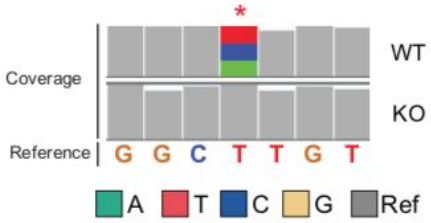
The promise of DRS



The promise of DRS

1

Differential
Basecalling 'errors'

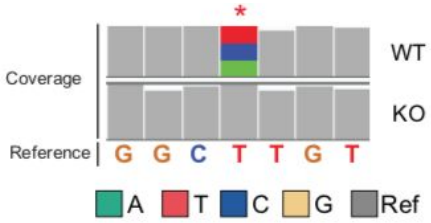


Liu et al., *Nature Communications* 2019
(Epinano)

DRS Strategies to detect RNA modifications

1

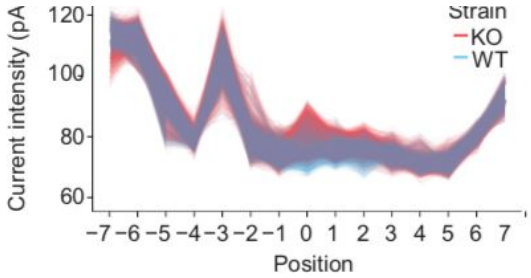
Differential
Basecalling 'errors'



Liu et al., *Nature Communications* 2019
(Epinano)

2

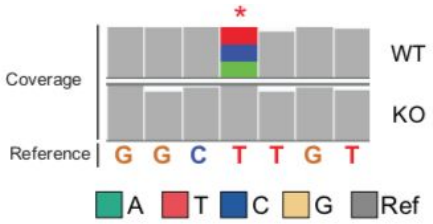
Alterations in
Current intensity



Begik et al., *Nature Biotechnology* 2021
(NanoRMS)

1

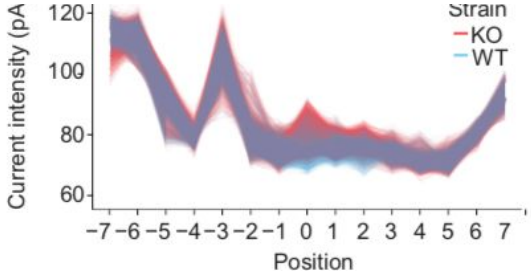
Differential
Basecalling 'errors'



Liu et al., *Nature Communications* 2019
(Epinano)

2

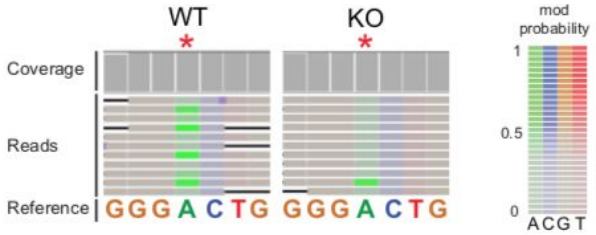
Alterations in
Current intensity



Begik et al., *Nature Biotechnology* 2021
(NanoRMS)

3

Modification-aware
Basecalling models



Cruciani et al., *Genome Biology* 2025
(m6ABasecaller)

Recap questions

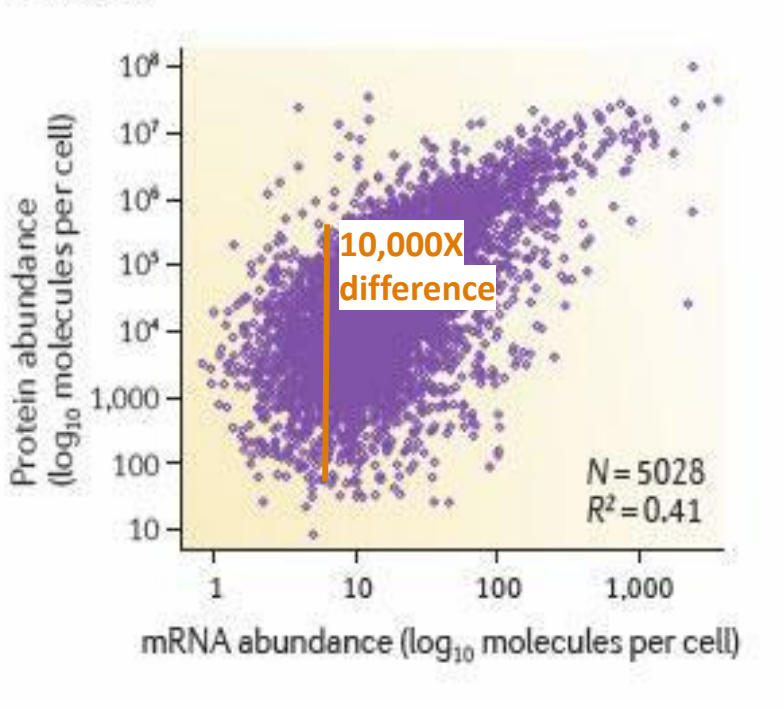
1. Why do we care about RNA modifications?
2. What is the most well-studied RNA modification and what are some with which biological functions has it been associated with?
3. What are the advantages of ONT DRS over short-read sequencing for modification detection?
4. What are the 3 main strategies in RNA modification detection? Which is the significantly superior one?

Hands-on Session: Theory & Rationale

Working with DRS tRNA modification data

A largely disregarded player: *the post-transcriptional regulatory layer*

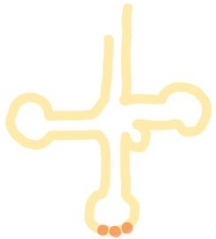
- Translation efficiency
- tRNA availability
- mRNA half life
- Codon usage
- RNA structure



- miRNA activity
- RNA binding proteins
- 5’ RNA degradation
- Ribosome specialization
- RNA modifications (epitranscriptome)

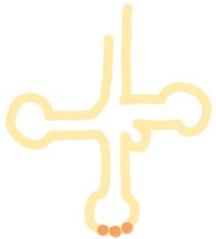
An even more largely disregarded player: *transfer RNAs*

tRNA



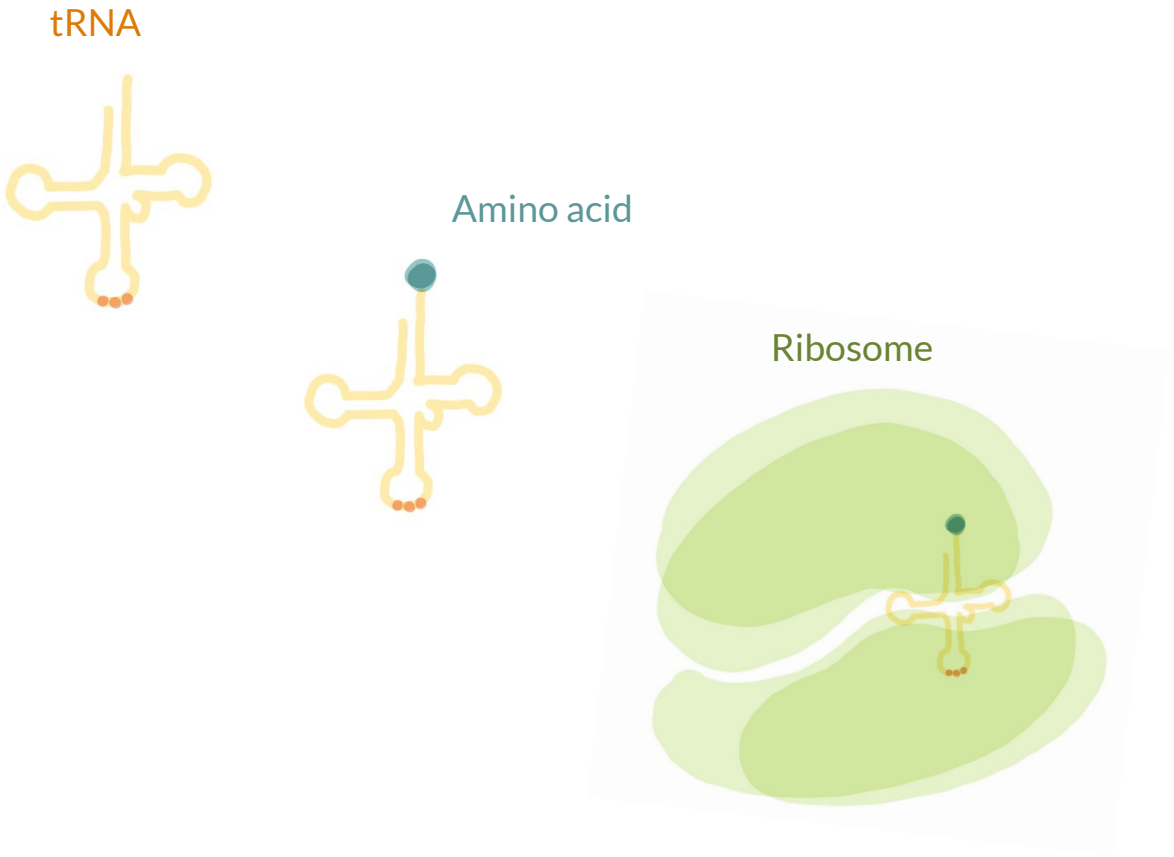
An even more largely disregarded player: *transfer RNAs*

tRNA

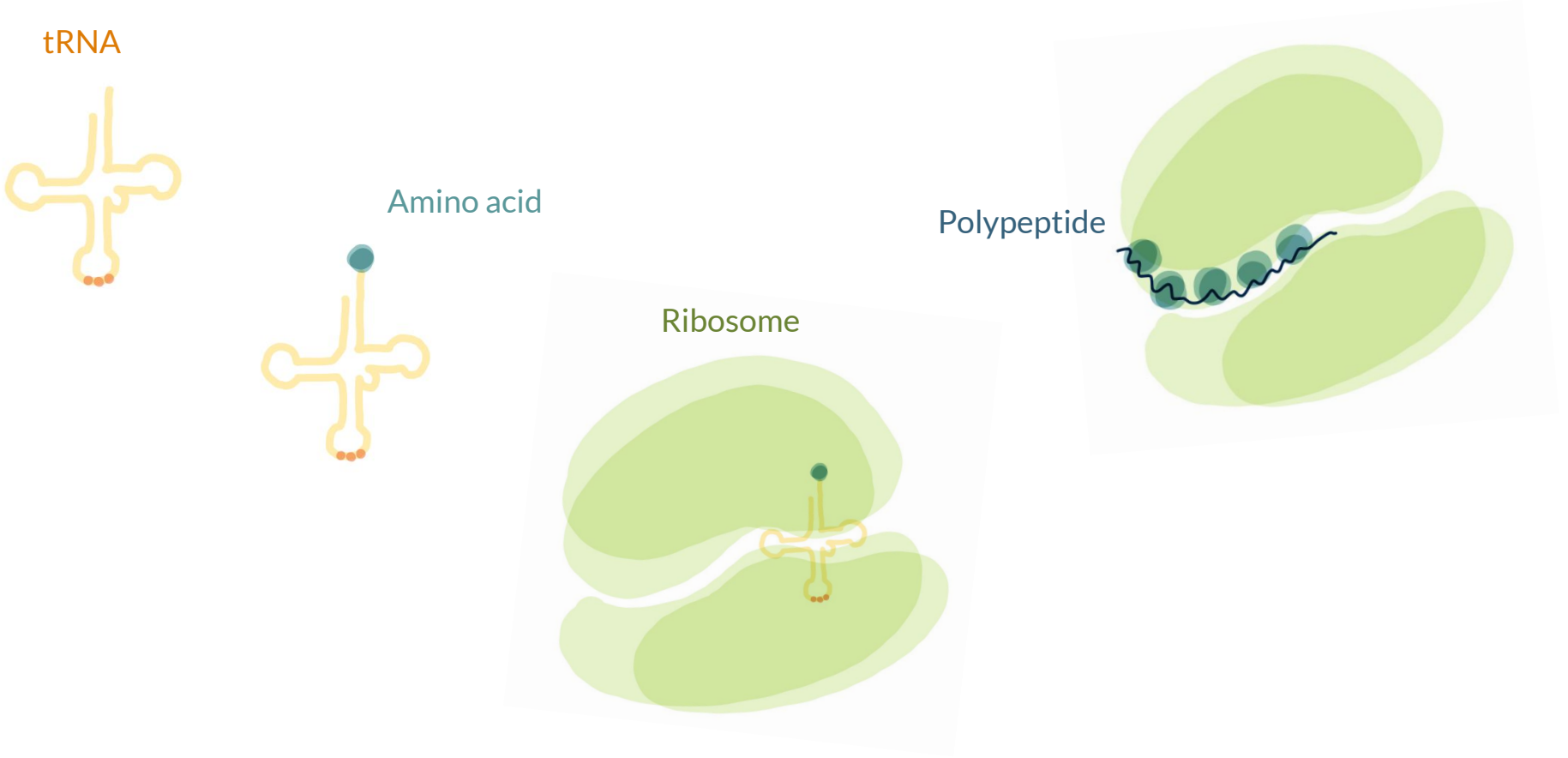


Amino acid

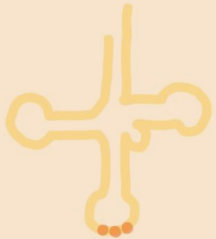
An even more largely disregarded player: *transfer RNAs*



An even more largely disregarded player: *transfer RNAs*



An even more largely disregarded player: *transfer RNAs*



tRNA

Amino acid

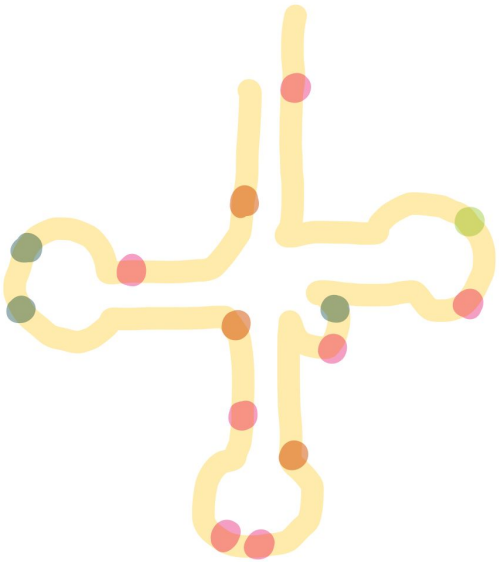
Polypeptid

Ribosome

THE END

An even more largely disregarded player: *transfer RNAs*

tRNA



- typically **70-90** nts long
- most **abundant** RNA species
- heavily post-transcriptionally modified **~13 mods/molecule**

Further reading:

Suzuki et al., *Nature Reviews Molecular Cell Biology* 2021

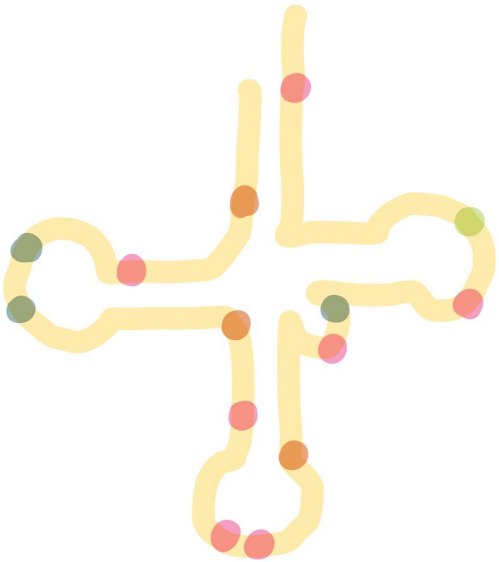
Pan et al., *Cell Research* 2018

Goodarzi et al., *Cell* 2016

CLose et al., *Cancer and Noncoding RNAs* (Chapter 10) 2018

An even more largely disregarded player: *transfer RNAs*

tRNA



- typically 70-90 nts long
- most abundant RNA species
- heavily post-transcriptionally modified ~13 mods/molecule

- chemical modifications affect **translational dynamics**
  - modifications affect structure and function of tRNA
  - modifications modulate direct tRNA–rRNA interactions

Further reading:

Suzuki et al., *Nature Reviews Molecular Cell Biology* 2021

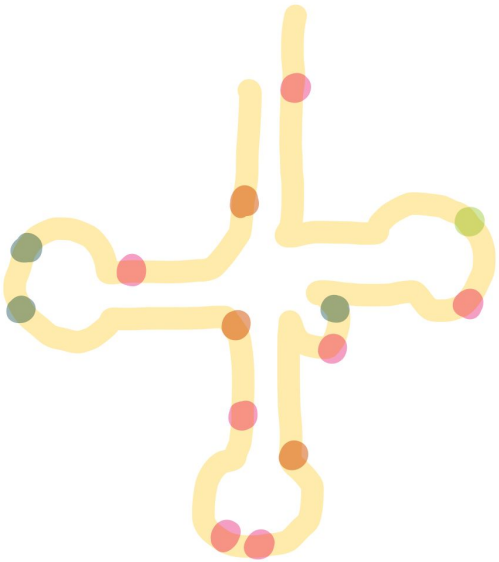
Pan et al., *Cell Research* 2018

Goodarzi et al., *Cell* 2016

CLose et al., *Cancer and Noncoding RNAs* (Chapter 10) 2018

An even more largely disregarded player: *transfer RNAs*

tRNA



- typically 70-90 nts long
- most abundant RNA species
- heavily post-transcriptionally modified ~13 mods/molecule
- chemical modifications affect translational dynamics

- tRNA abundance and modification dysregulation involved in
  - mitochondrial diseases
  - neurological disorders
  - cancer

Further reading:

Suzuki et al., *Nature Reviews Molecular Cell Biology* 2021

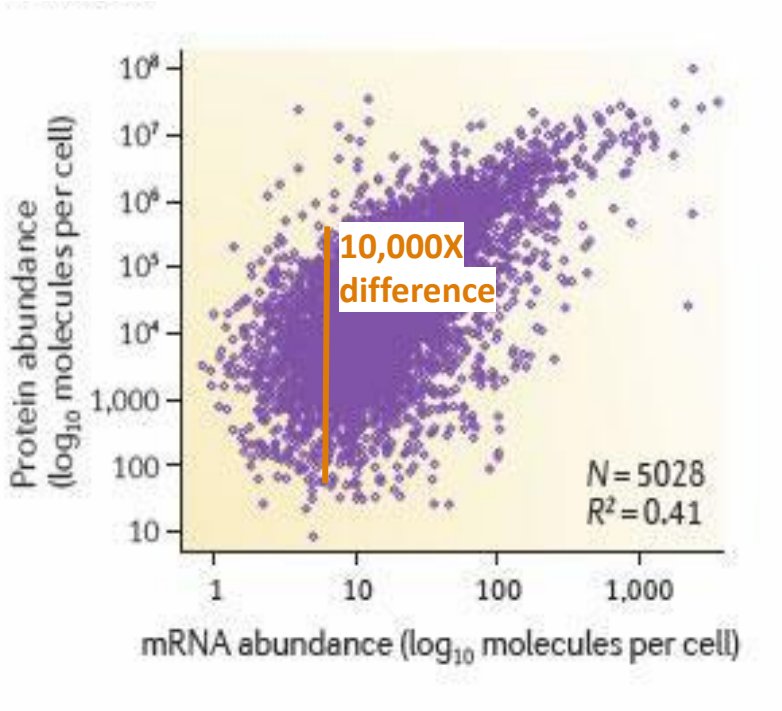
Pan et al., *Cell Research* 2018

Goodarzi et al., *Cell* 2016

CLose et al., *Cancer and Noncoding RNAs* (Chapter 10) 2018

Studying the post-transcriptional regulatory layer through tRNAs

- **Translation efficiency**
- **tRNA availability**
- mRNA half life
- **Codon usage**
- **RNA structure**



- miRNA activity
- RNA binding proteins
- 5’ RNA degradation
- Ribosome specialization
- **RNA modifications (epitranscriptome)**

the challenges:

- DRS caters towards longer reads; inefficient capturing of <200nt-long transcripts, unable at <100nt
- first 15nt at 5' typically lost due to an increase in RNA translocation speed
- clover-like secondary structure of tRNA molecules

ACS Nano > Vol 15/Issue 10 > Article

Open Access

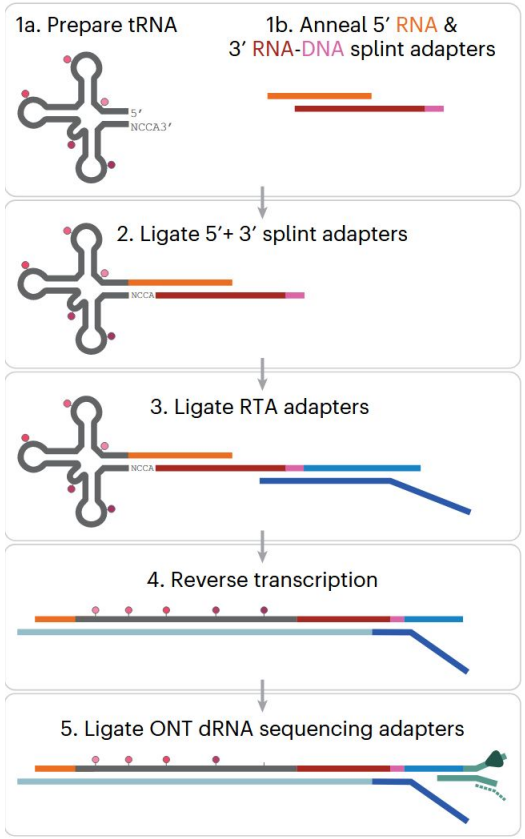
ARTICLE | October 7, 2021

Direct Nanopore Sequencing of Individual Full Length tRNA Strands

Niki K. Thomas, Vinay C. Poodari, Miten Jain, Hugh E. Olsen, Mark Akeson, and Robin L. Abu-Shumays*

Thomas et al., *ACS Nano* 2021

Capturing the tRNAome with DRS: the Nano-tRNAseq protocol



🧪 Nano-tRNAseq protocol

Article | Open access | Published: 06 April 2023

Quantitative analysis of tRNA abundance and modifications by nanopore RNA sequencing

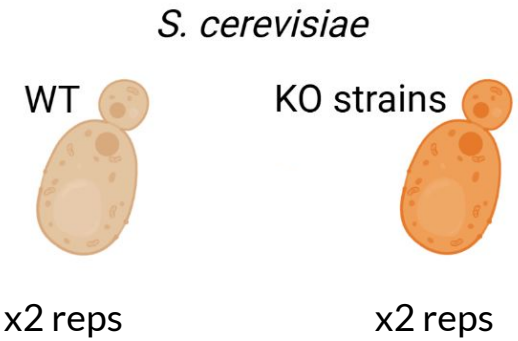
Morghan C. Lucas, Leszek P. Pryszcz, Rebeca Medina, Ivan Milenkovic, Noelia Camacho, Virginie Marchand, Yuri Motorin, Lluís Ribas de Pouplana & Eva Maria Novoa

Nature Biotechnology **42**, 72–86 (2024) | Cite this article

Lucas & Pryszcz et al., *Nature Biotechnology* 2023

Hands-on session rationale

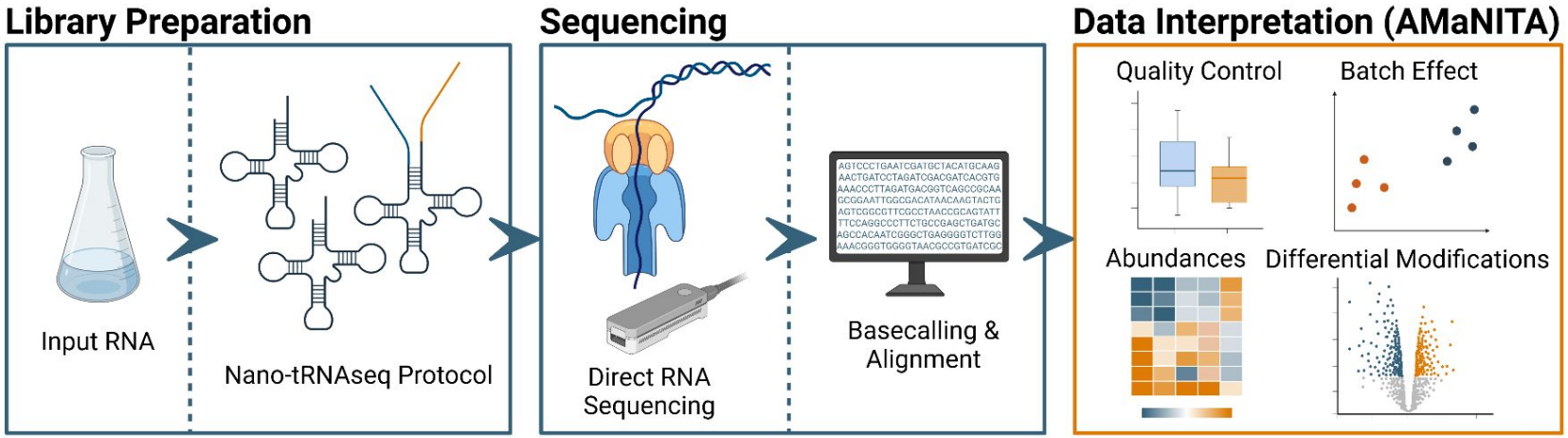
- 5 teams of 5-6 people
- 6 ***secret*** datasets



🎯

- get familiarized with the data and tool
- run the analysis on your selected dataset
- interpret and discuss your results with teammates
- have me judge you

Hands-on session rationale



Recap questions

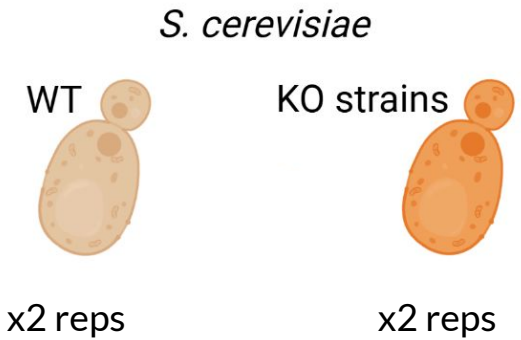
1. Why do we care about tRNAs?
2. Why is ONT DRS better for the sequencing of tRNAs?
3. What are the current limitations of studying tRNAs with ONT DRS?

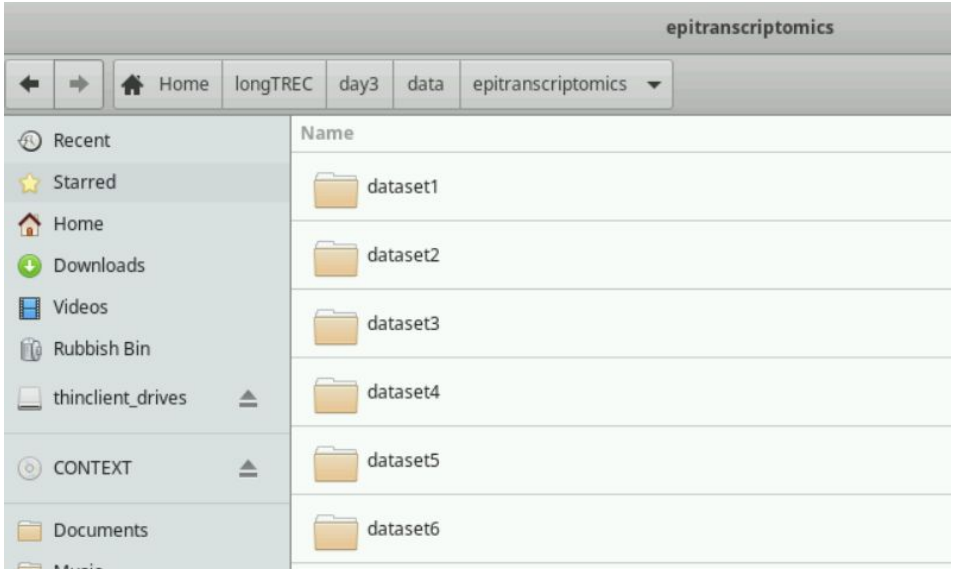
Hands-on Session: Downstream Interpretations with AMaNITA

Analyzing yeast Nano-tRNAseq data

Organising the Hands-on session: datasets

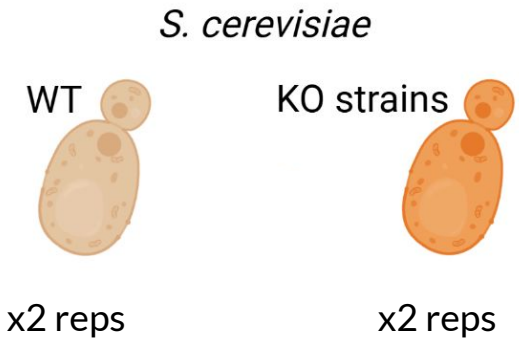
- 5 teams of 5-6 people
- 6 ***secret*** datasets

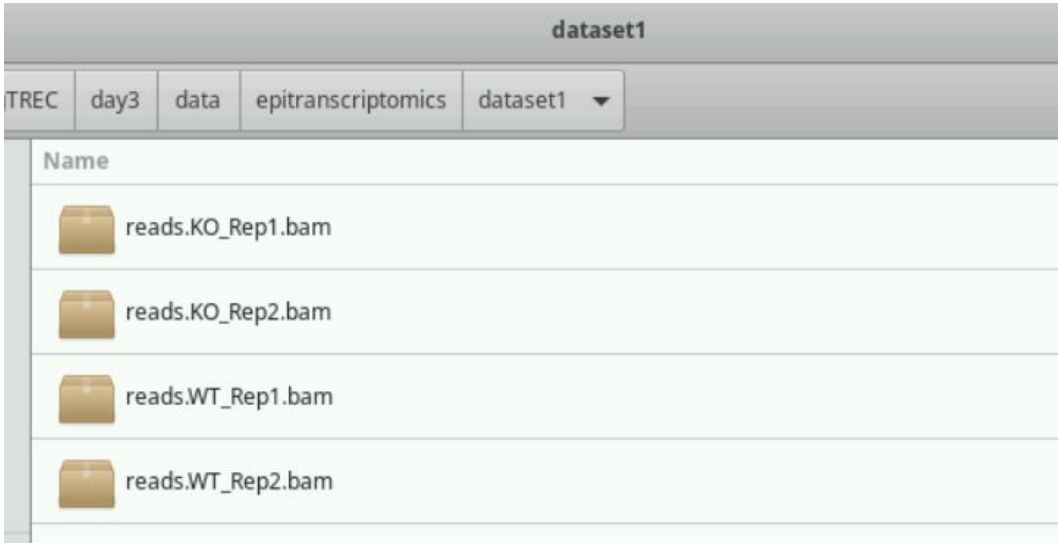


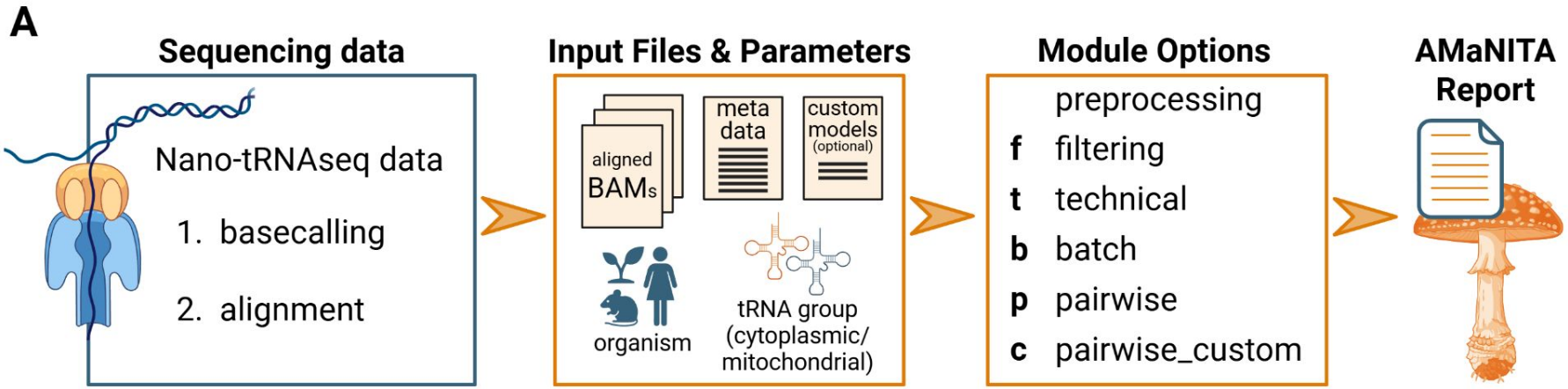


Organising the Hands-on session: file name conventions

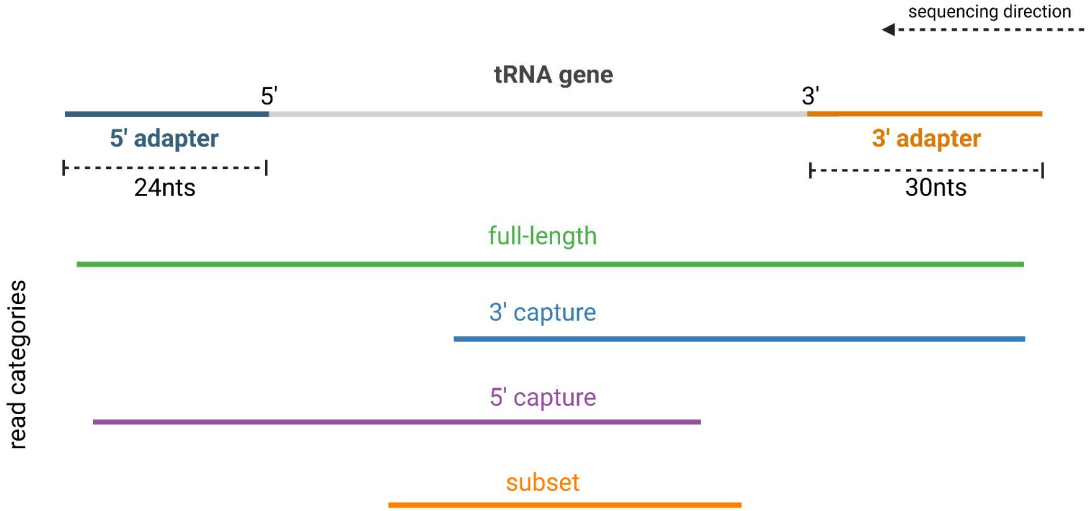
- 5 teams of 5-6 people
- 6 ***secret*** datasets






A
Sequencing data
Nano-tRNAseq data
1. basecalling
2. alignment
Input Files & Parameters
aligned BAMs
meta data
custom models (optional)
organism
tRNA group (cytoplasmic/ mitochondrial)
Module Options
preprocessing
f filtering
t technical
b batch
p pairwise
c pairwise_custom
AMaNITA Report

filtering module



Filters

1. All **5'** and **subset** reads are discarded
2. All **3'** reads spanning < 25b of the tRNA gene are discarded
3. Any type of reads with deletions/non-matches > 10b are discarded

batch module

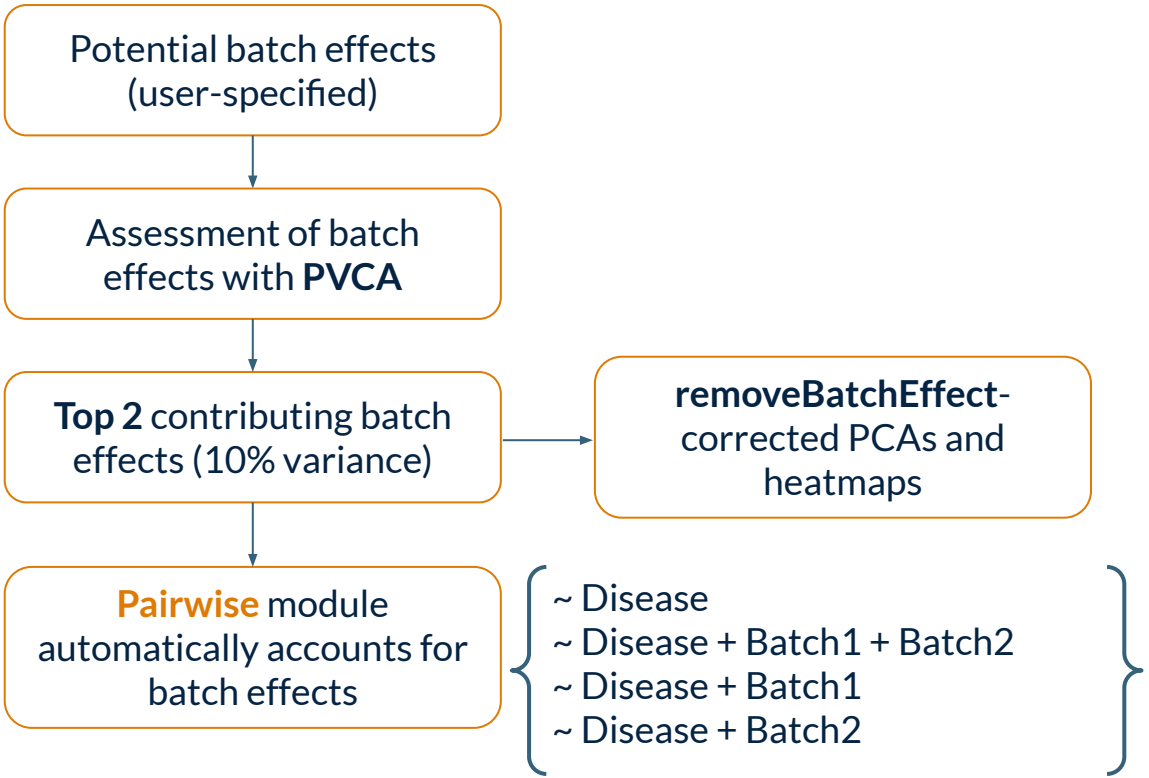
1. PVCA > identify potential batch effects,
   based on the variance explained by each variable

2. limma::removeBatchEffect > plotting ONLY (heatmaps and PCA plots)

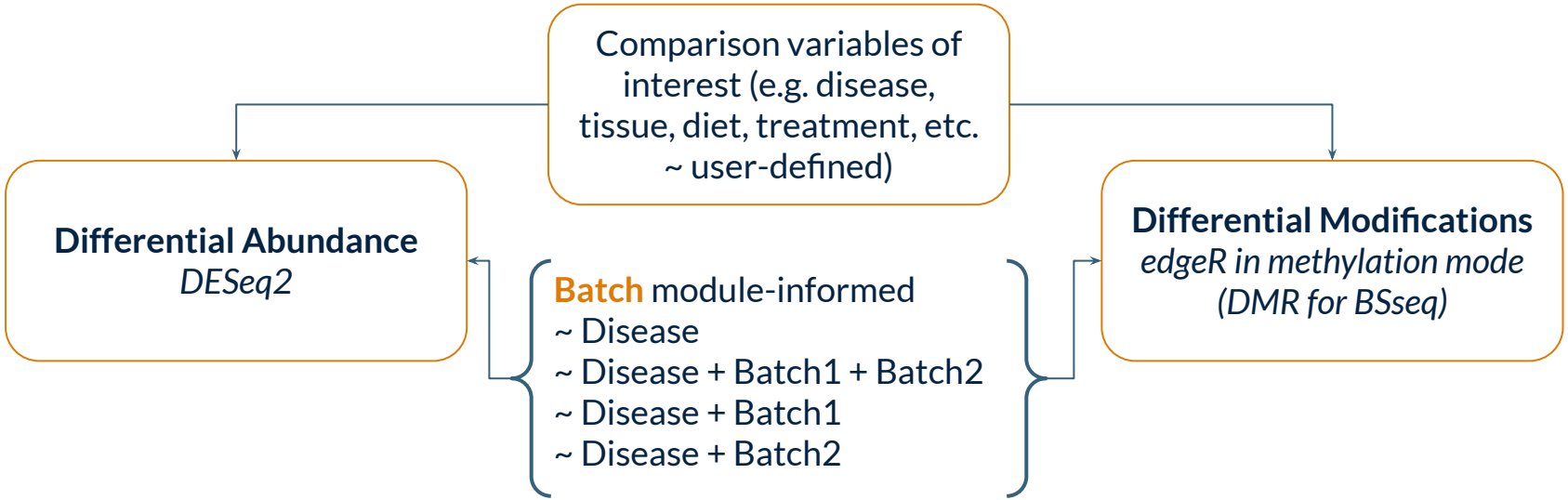
3. Linear model accounting for batch variables > differential analyses
   e.g. **~ Tissue** vs **~ Tissue + RunDate**

⚠️Batch effect correction should be treated with caution!
Unbalanced batches and/or using batch corrected data for purposes other than intended is very dangerous and can lead to wrong conclusions!⚠️
Nygaard et al., *Biostatistics* 2016

batch module



pairwise module



AMaNITA documentation on the GitHub repo

summer_school / practicals / day3 / epitranscriptomics / **AMaNITA_docs** /

LedaKatopodi Add files via upload ✓

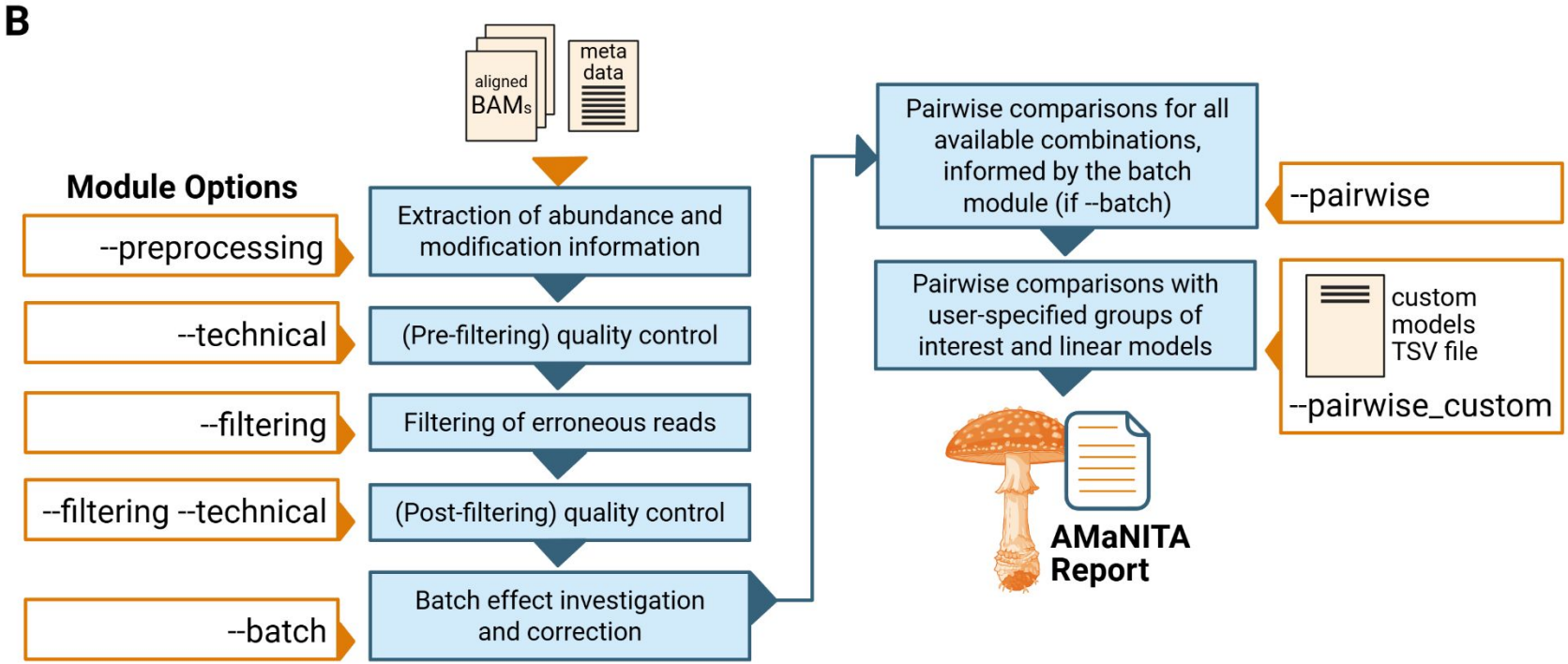
Name	Last commit message
..	
wiki	Add files via upload
README.md	Create README.md

README.md

🍄 AMaNITA 🍄

Abundance, Modifications, and Nanopore Intensity Toolbox/Application

AMaNITA workflow



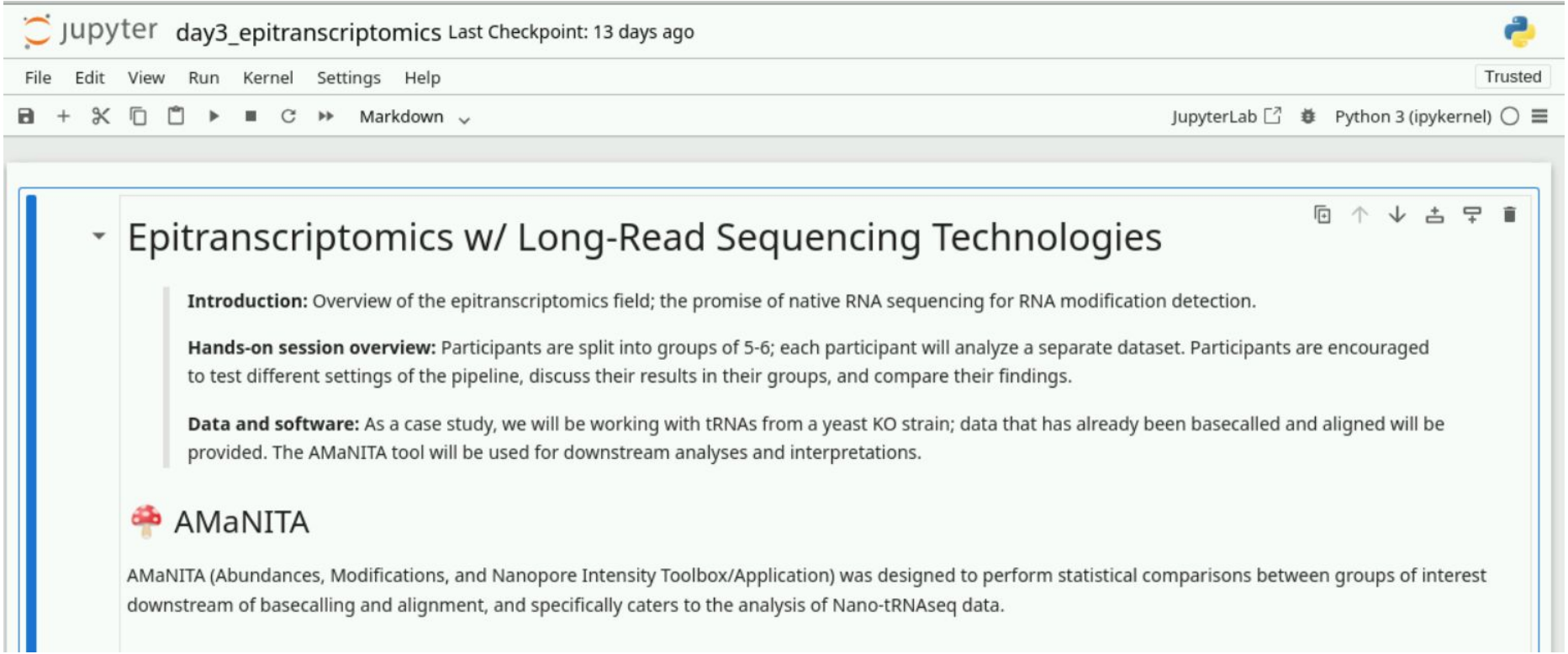
Building the metadata file

How to properly build your Metadata file

AMaNITA is built around a properly defined metadata table that provides all information required for an end-to-end analysis of the input data. This is a walkthrough on how to build your metadata table, things you should be careful, tips and tricks.

> **Important**
>
> The metadata file is a tab-separated file, expected to be named `{project_id}.metadata.tsv`.
>
> The metadata file is expected to be under the project directory, e.g. `{wrk_dir}/{project_id}/{project_id}.metadata.tsv`.
>
> Each **line** contains information on **one** sample.
>
> When creating the metadata table, make sure that the newline character is Unix-based (i.e. not Windows- or MacOS-based)
>
> Expected columns:
>
> - 1: `Project`
> - 2: `Run`
> - 3: `BAMID`
> - 4: `BarcodeID`
> - 5: `SampleID`
> - 6: (optional) `ReplicateID`
> - 6/7+: variables of interest, tagged with either of the following: `comp_`, `batch_`, or `extra_`
>
> ⚠️ Columns 1-6 (when ReplicateID is included) require the column names to be as shown above (`Project`, `Run`, `BAMID`, `BarcodeID`, `SampleID`, `ReplicateID`)

Running AMaNITA



Challenge 1: What are the KO?

Challenge 2: What are the differences between same-KO datasets?

Challenge 3: Let’s run batch effect analysis

Recap questions: Lessons learned

1. How sensitive is modification detection with DRS and to which factors?
2. What are the challenges, advantages, and limitations of using basecalling errors as a proxy for modification detection?

Discussion

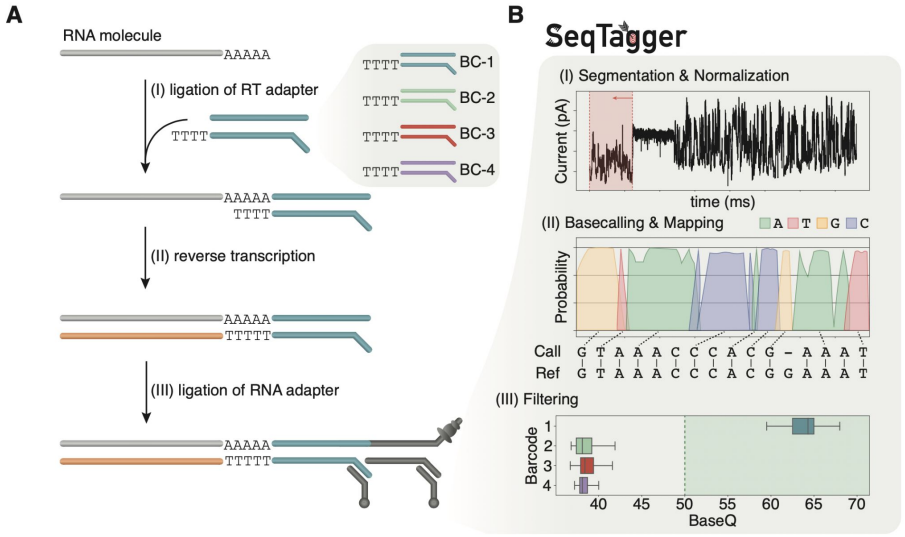
Last comments; take-home messages; Q&A

⚠️Batch effect correction should be treated with caution!
Unbalanced batches and/or using batch corrected data for purposes other than intended is very dangerous and can lead to wrong conclusions!⚠️

Nygaard et al., *Biostatistics* 2016

Priorities when dealing with batches:

1. Don't introduce batches! >> **multiplexing**
   (Pryszcz & Diensthuber et al., *Genome Research* 2025)



Batch effect correction is a blessing and a curse

⚠️Batch effect correction should be treated with caution!
Unbalanced batches and/or using batch corrected data for purposes other than intended is very dangerous and can lead to wrong conclusions!⚠️
Nygaard et al., *Biostatistics* 2016

Priorities when dealing with batches:

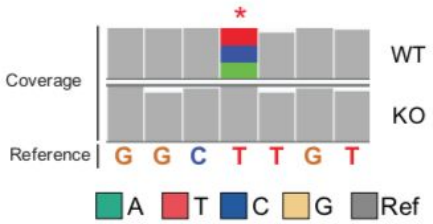
1. Don't introduce batches! >>
   multiplexing
   (Pryszcz & Diensthuber et al., *Genome Research* 2025)

2. Make sure your batches are balanced during the experimental design step

3. When performing batch effect correction, make sure you are not over-correcting (and therefore losing biological information), and that you are using batch-corrected data for its intended purposes

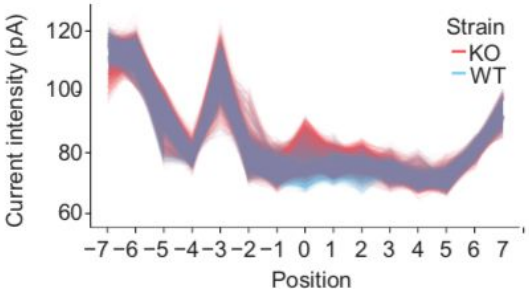
1

Differential
Basecalling 'errors'



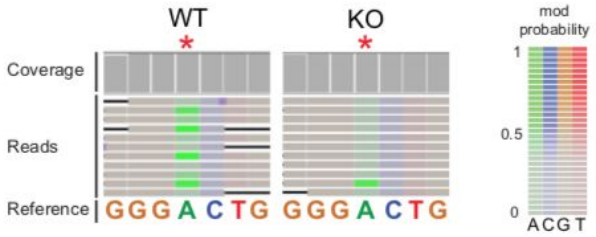
2

Alterations in
Current intensity



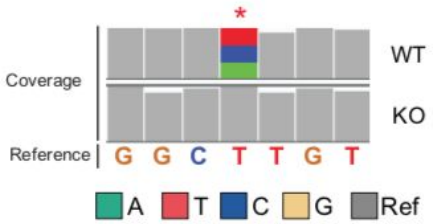
3

Modification-aware
Basecalling models



1

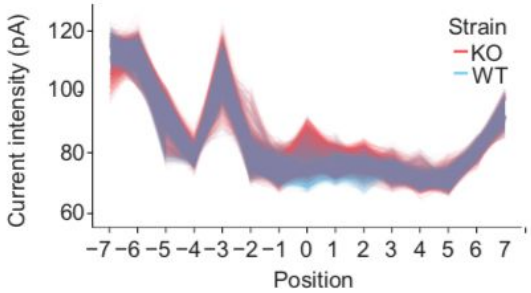
Differential
Basecalling 'errors'



Epinano Liu et al., 2019
DiffErr Parker et al., 2020
Eligos Jenjaroenpun et al., 2021
Drummer Price et al., 2020

2

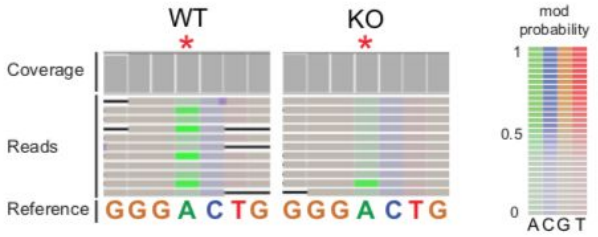
Alterations in
Current intensity



Nanopolish Simpson et al., 2017
Tombo Stoiber et al., 2017 (*)
Nanocompore Leger et al., 2021
NanoRMS Begik et al., 2021

3

Modification-aware
Basecalling models



Dorado modification-aware settings
m6ABasecaller Cruciani et al., 2025
Uncalled4 Kovaka et al., 2025

NanoConsensus Delgado-Tejedor et al., 2023

Comprehensive review: Furlan et al., *RNA Biology* 2021

Master Of Pores: a Nextflow pipeline for end-to-end analysis of DRS data

Cozzuto et al., *Frontiers in Genetics* 2020

Questions?

Feel free to reach out if you have any questions!

Leda Katopodi
(Xanthi Lida Katopodi)

lida.katopodi@crg.eu

LedaKatopodi (ledak)

Leda Katopodi

@secuenciasdeciencia

Thank You!

For more information about the LongTREC Summer School:

https://longtrec.eu