TRAINING COURSE IN Computational Methods for Epitranscriptomics

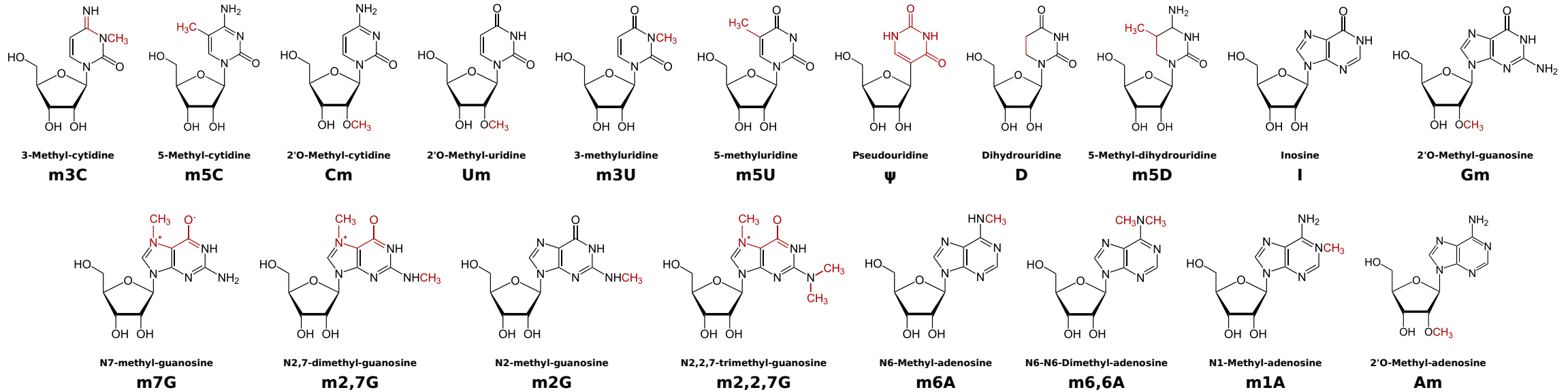## Using Nanopore sequencing to detect RNA modifications

Theoretical introduction to the various tools and techniques to identify RNA modifications from Nanopore data.
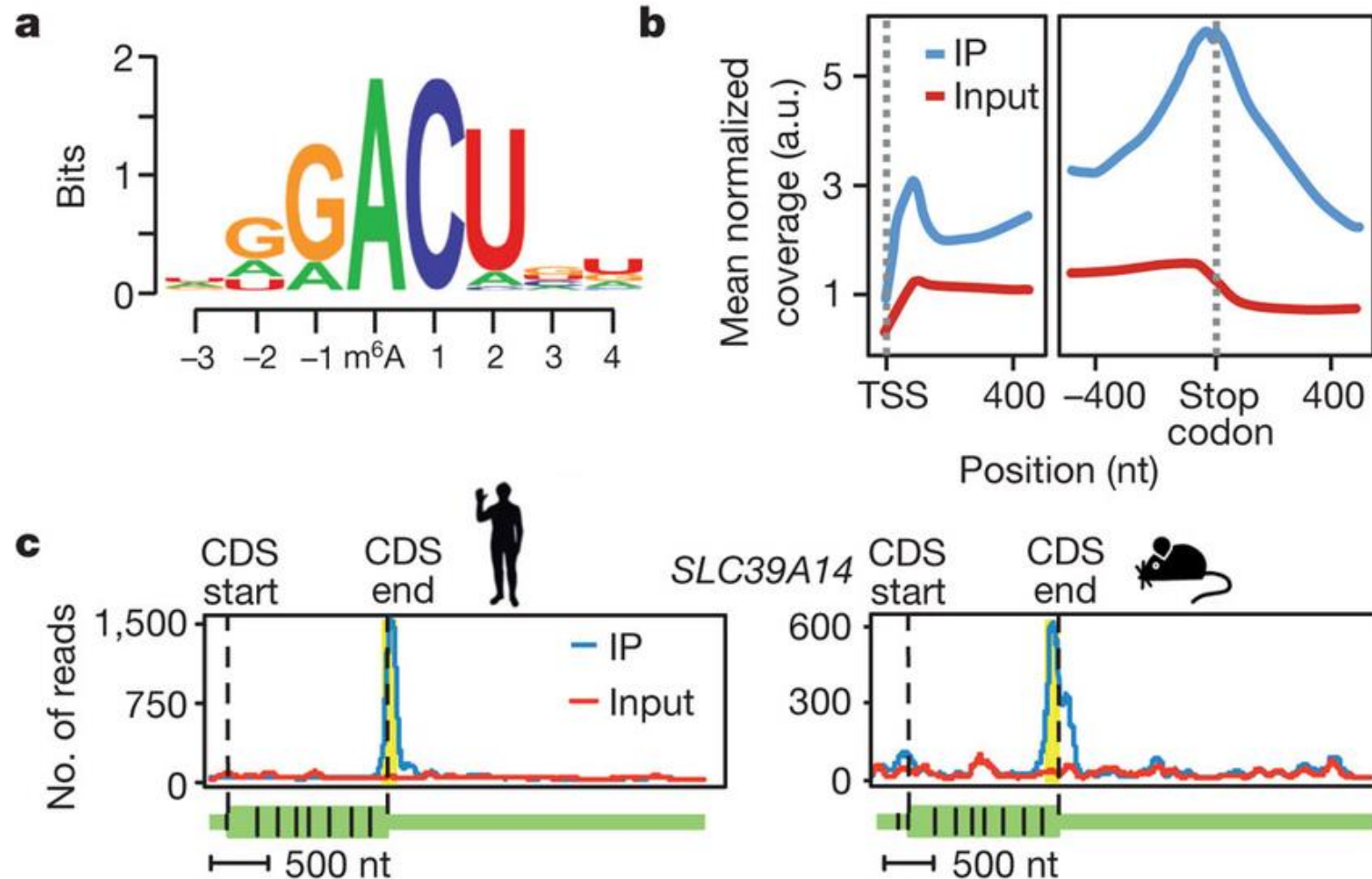
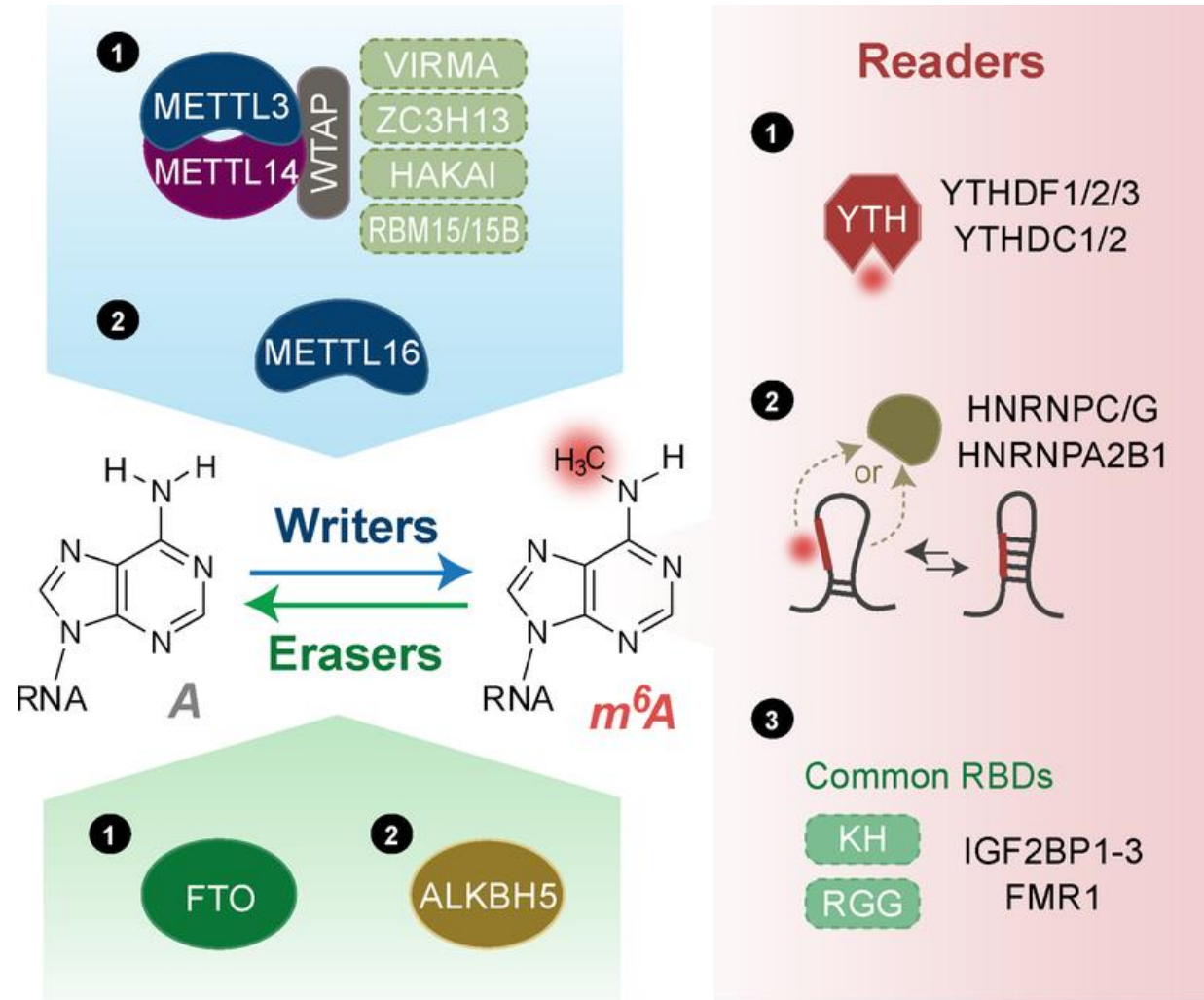Logan Mulroney

The world of RNA mods

- Highly conserved features found in archaea, bacteria, and eukarya
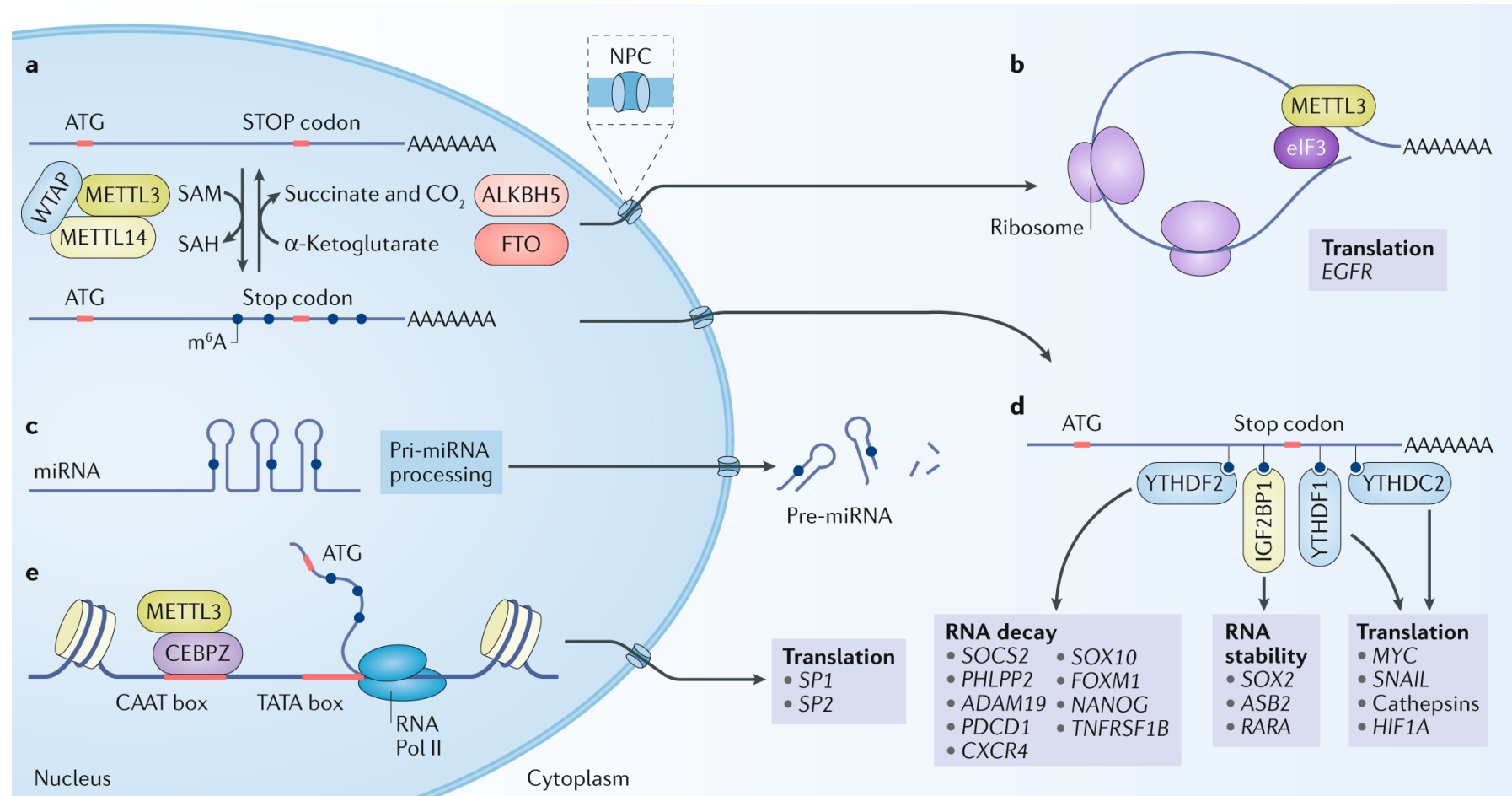- Impact on RNA structure and interaction properties

3-Methyl-cytidine	5-Methyl-cytidine	2'O-Methyl-cytidine	2'O-Methyl-uridine	3-methyluridine	5-methyluridine	Pseudouridine	Dihydrouridine	5-Methyl-dihydrouridine	Inosine	2'O-Methyl-guanosine
m3C	**m5C**	**Cm**	**Um**	**m3U**	**m5U**	**ψ**	**D**	**m5D**	**I**	**Gm**

N7-methyl-guanosine	N2,7-dimethyl-guanosine	N2-methyl-guanosine	N2,2,7-trimethyl-guanosine	N6-Methyl-adenosine	N6-N6-Dimethyl-adenosine	N1-Methyl-adenosine	2'O-Methyl-adenosine
m7G	**m2,7G**	**m2G**	**m2,2,7G**	**m6A**	**m6,6A**	**m1A**	**Am**

Modifications are widespread in mRNAs

Dominissini et al., Nature (2012)

RNA modifications are dynamic

Hailing Shi, Jiangbo Wei, and Chuan He, *Mol Cell* (2020)

RNA modifications have been implicated in diseases, such as cancer

Barbieri, and Kouzarides, *Nature Reviews* (2020)

There are ~170 known naturally occuring RNA modifications
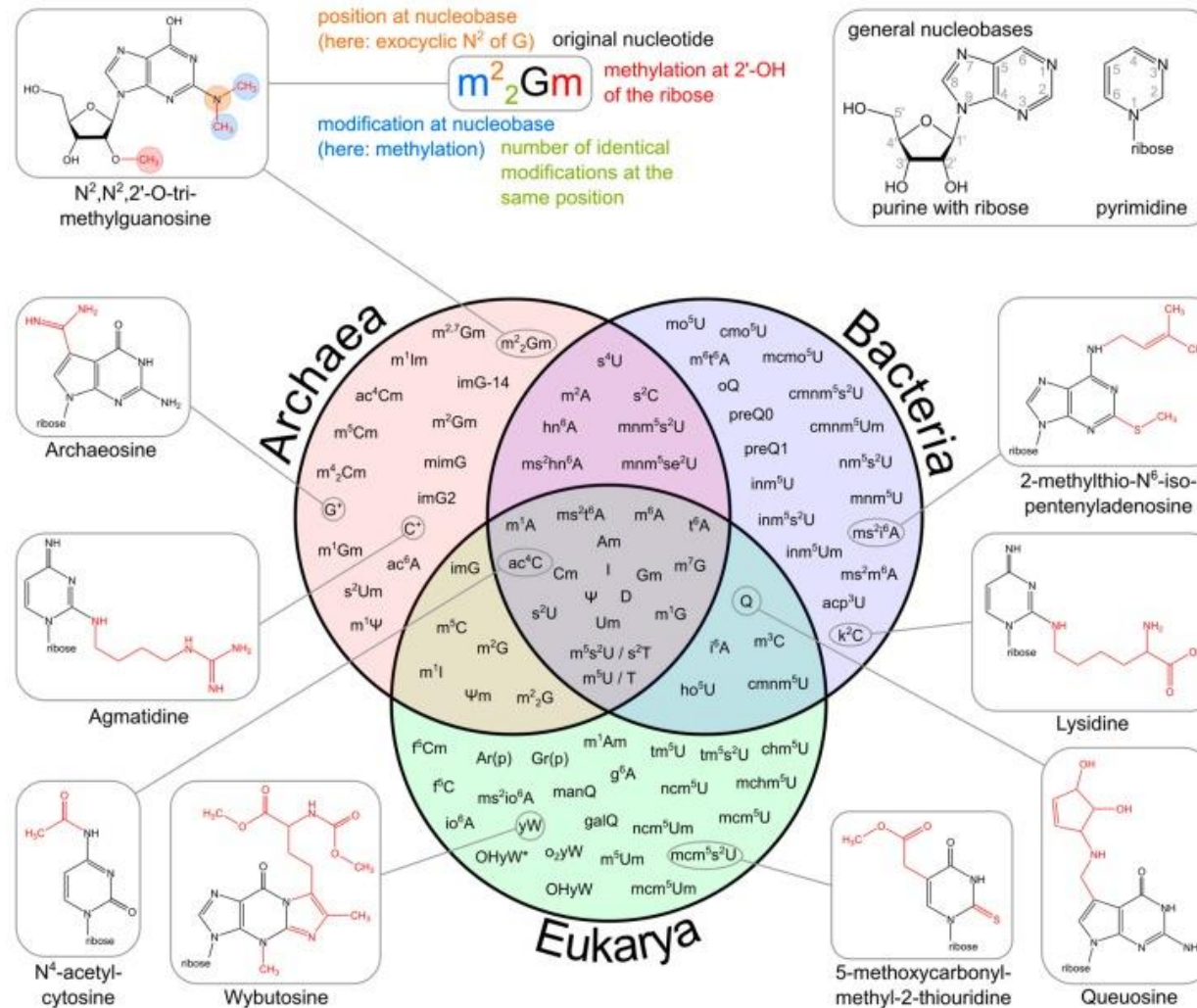
Lorenz C, Lünse CE, Mörl M., *Biomolecules (*2017)

Boccaletto P et al., *Nucleic Acids Research* (2017)

Established techniques for detecting RNA modifications

Mass Spectrometry
- Highly quantitative
- Highly specific
- Hard to set up
- Low throughput
- Hard to obtain sequence specificity

NGS based methods
- Transcriptome-wide
- Cross reactivity (antibody, chemical treatment)
- Need for specific assays/antibody for each modification
- Variable resolution

Jonkout N et al., *RNA* (2017)
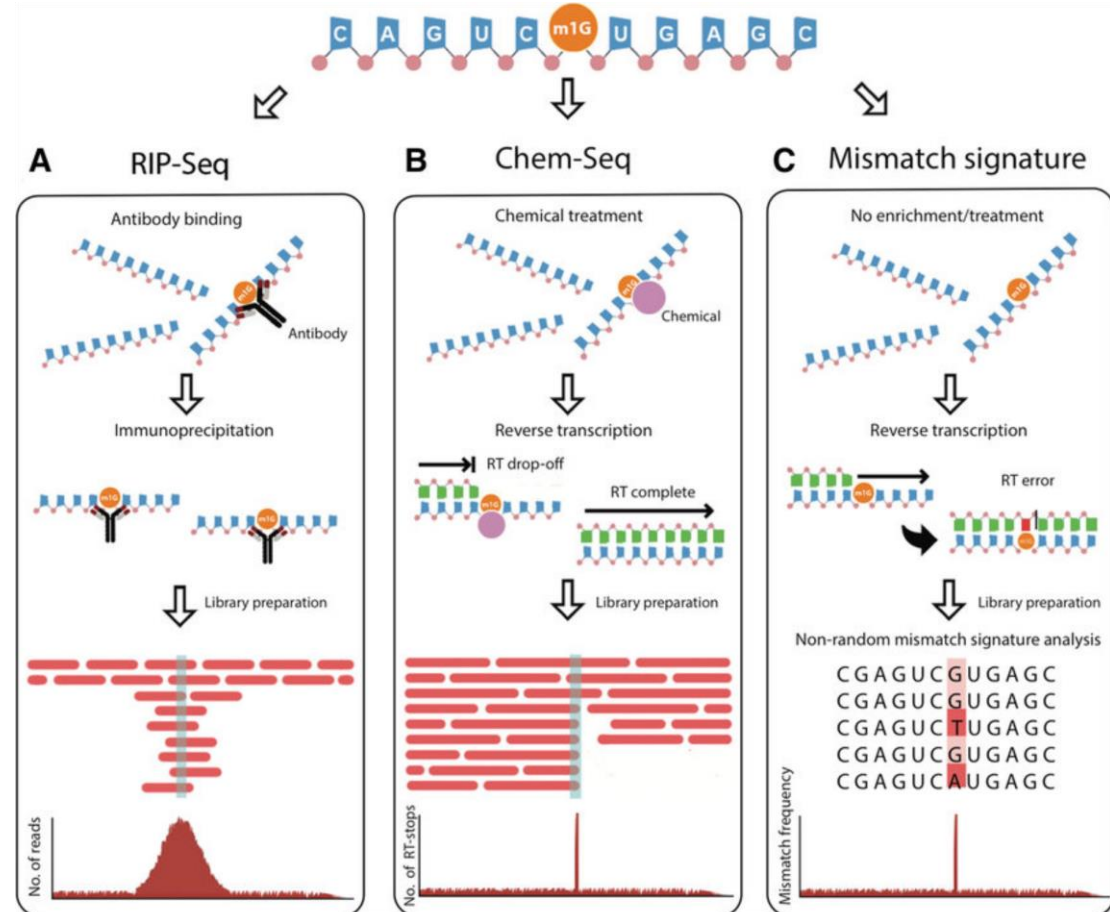
An emerging approach: nanopore sequencing

- Direct RNA nanopore sequencing (DRS) reads RNA molecules directly, bypassing cDNA conversion

- This avoids RT bias

- This avoids PCR bias

- The RNA modifications remain intact on the molecules during sequencing

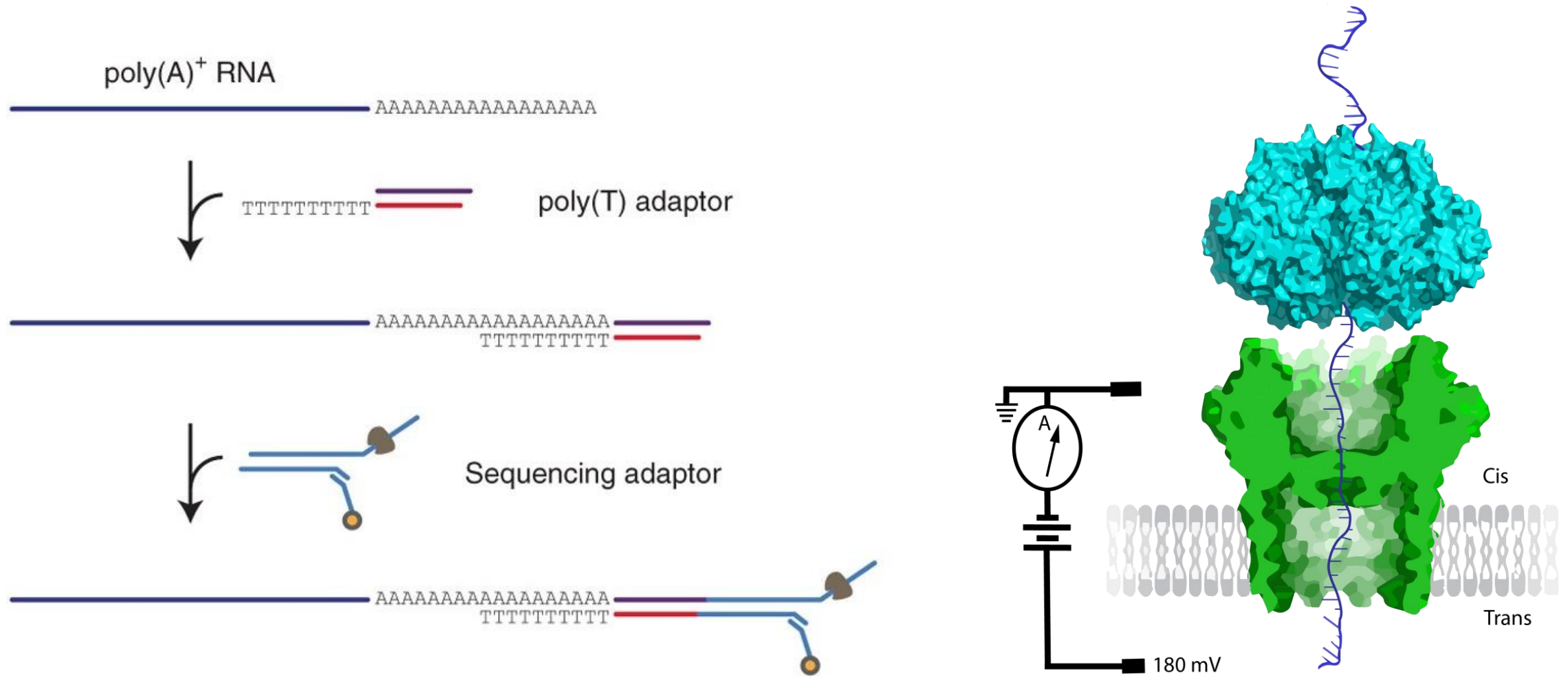- The RNA molecules are sequenced as they existed in the cell
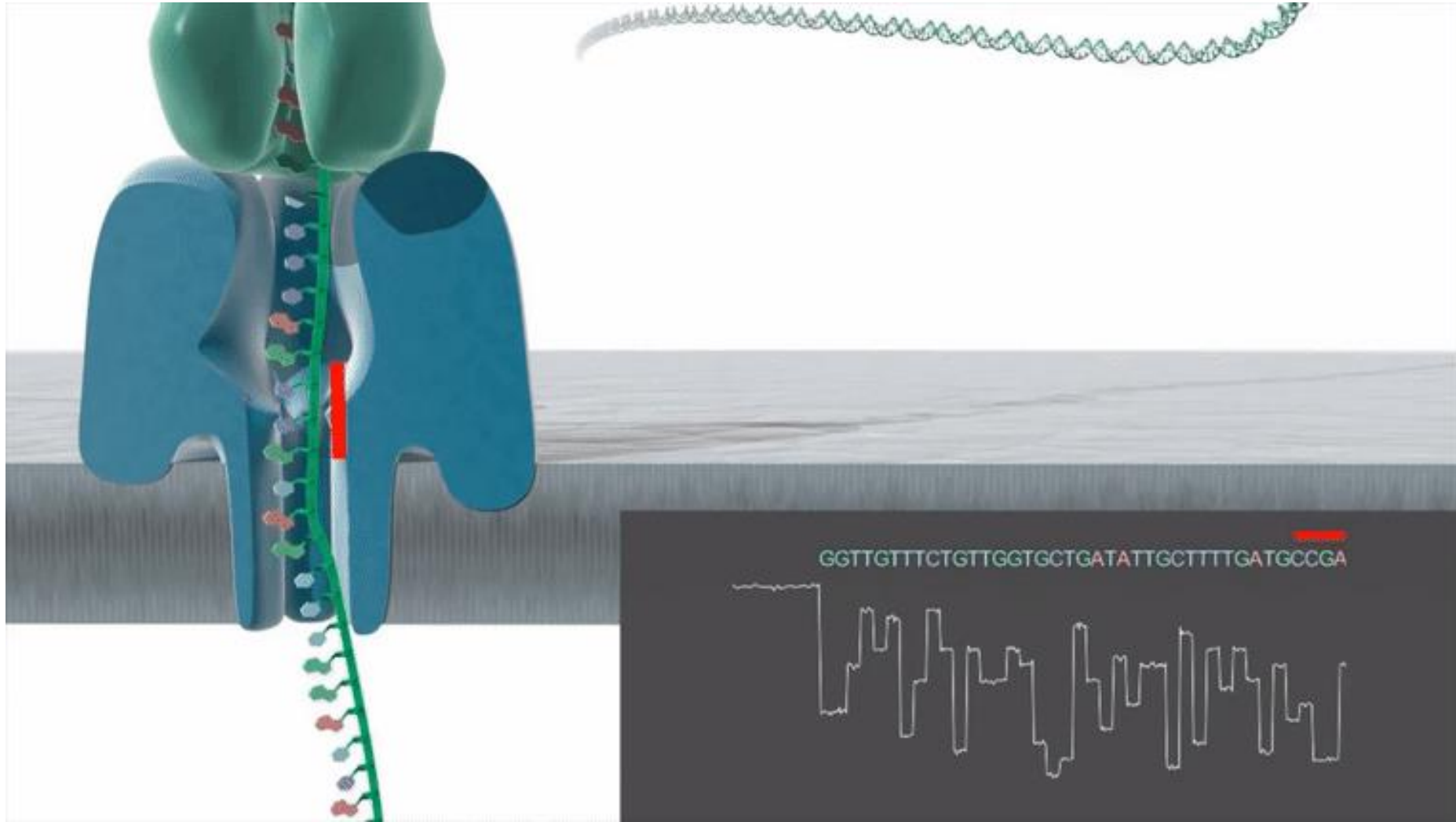
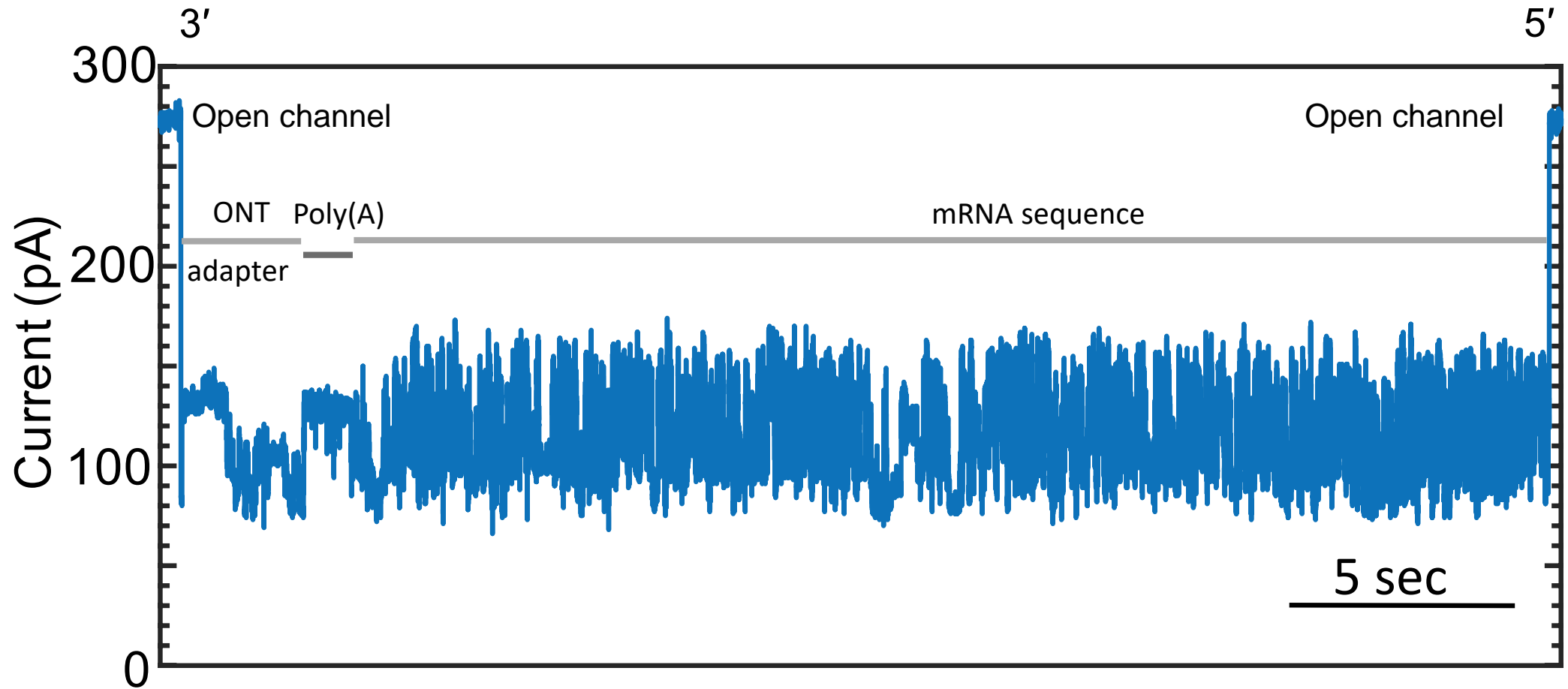Science 2016 Epitranscriptomics
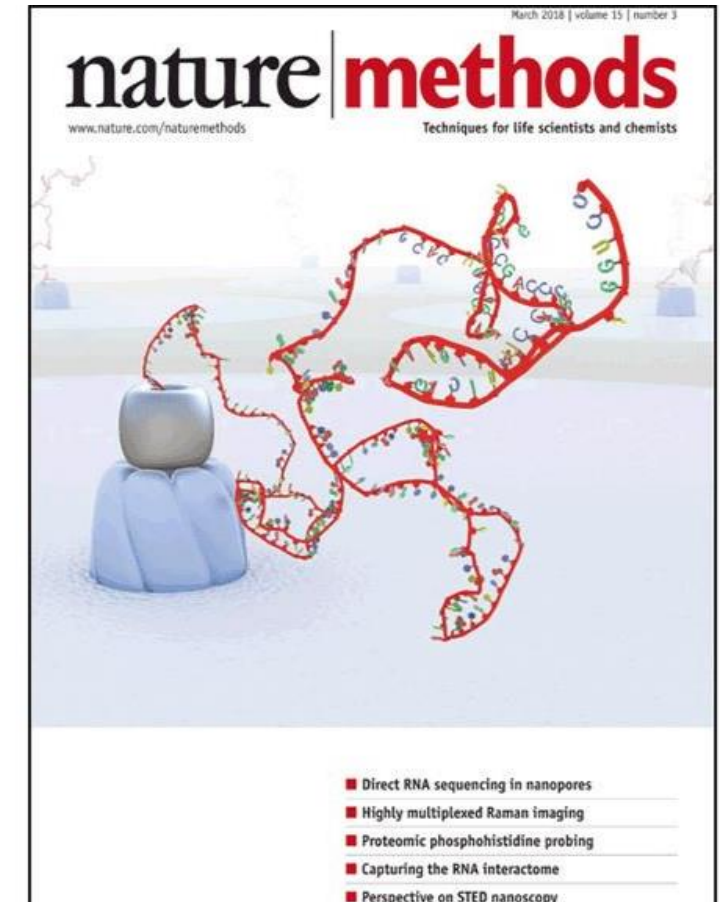
Nat.Med 2018 Direct RNA Sequencing

Direct RNA nanopore sequencing

The nanopore sensor directly interacts with the nucleotides

Oxford Nanopore Technologies animation

Typical direct RNA nanopore ionic current trace

Workman, et. al. *Nature Methods* (2019)

Modified nucleotides alter the ionic current in nanopore traces relative to canonical nucleotides
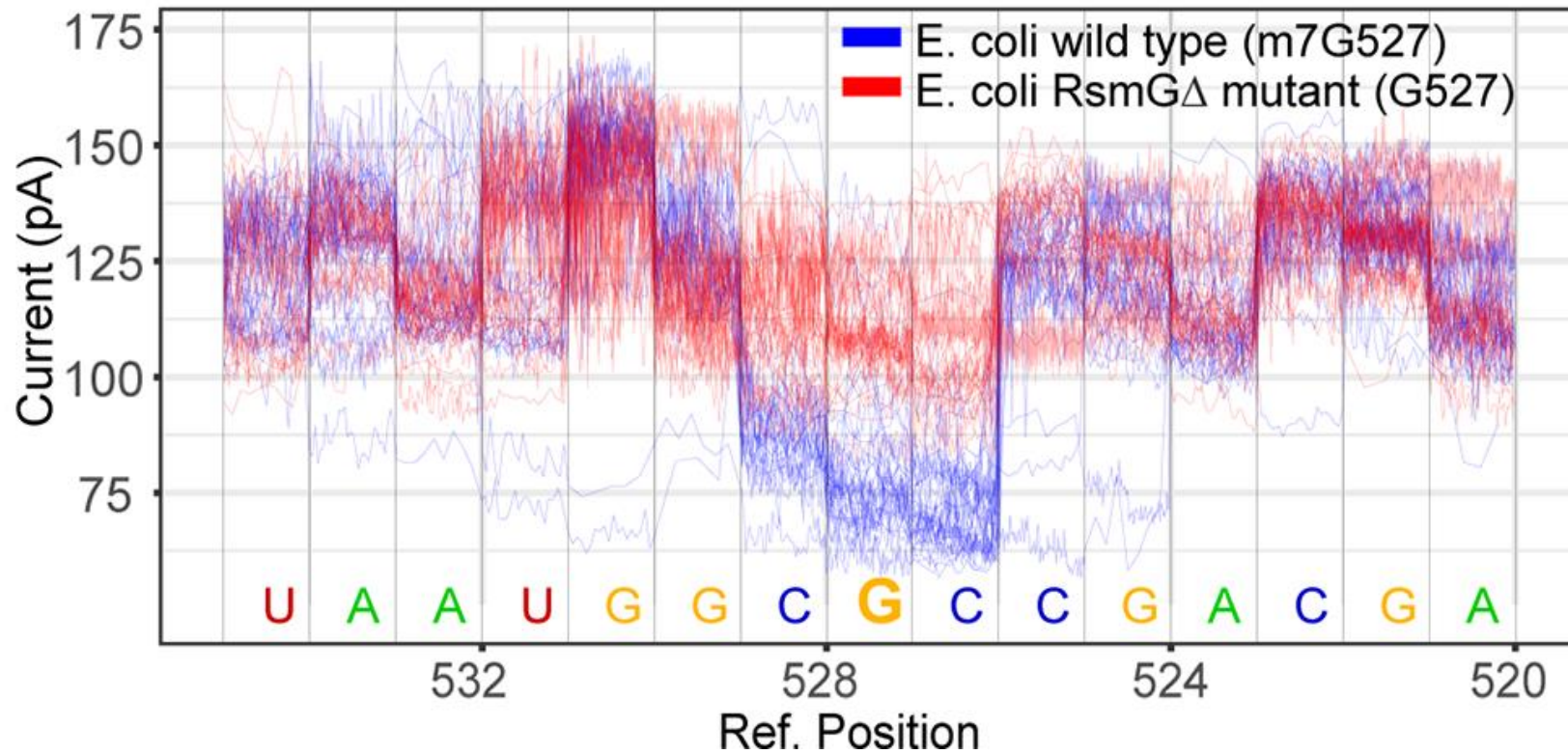
Two general strategies to detect RNA modifications from direct RNA nanopore data
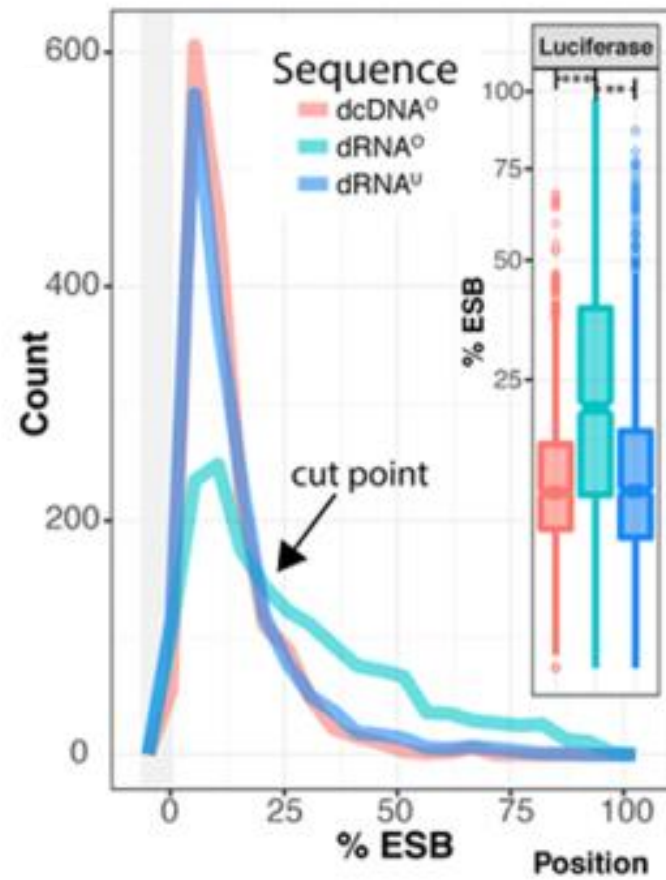

Furlan, Delgado-Tejedor, Mulroney et al., *RNA Biology* (2021)
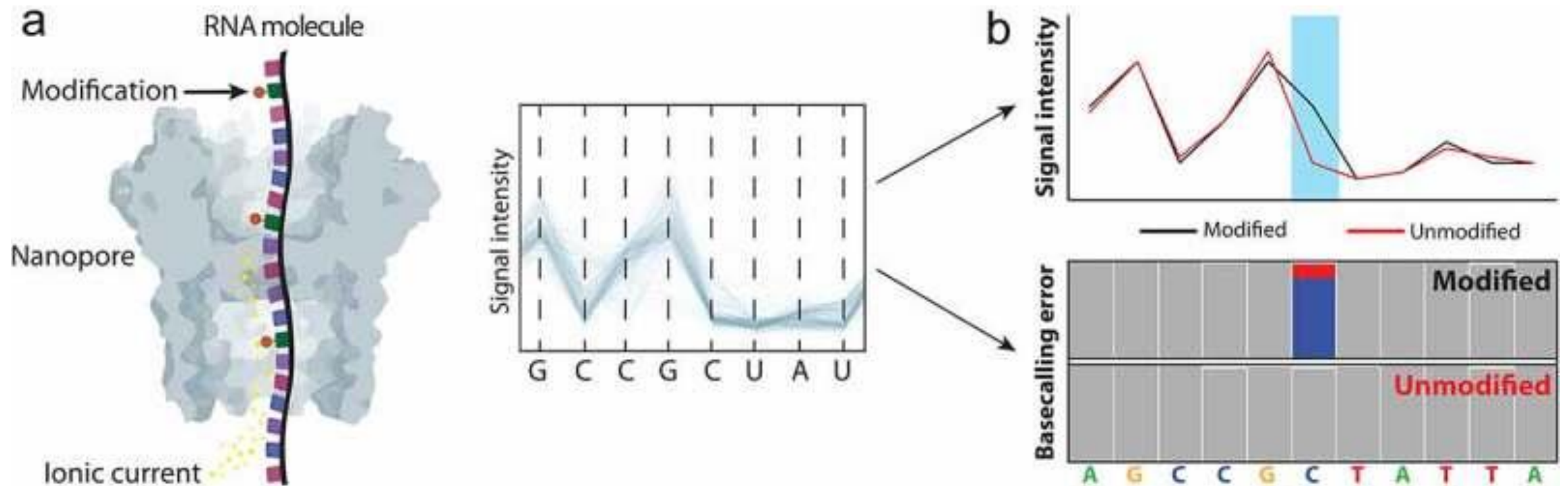
More than 15 software tools exist to detect RNA modifications from direct RNA nanopore data
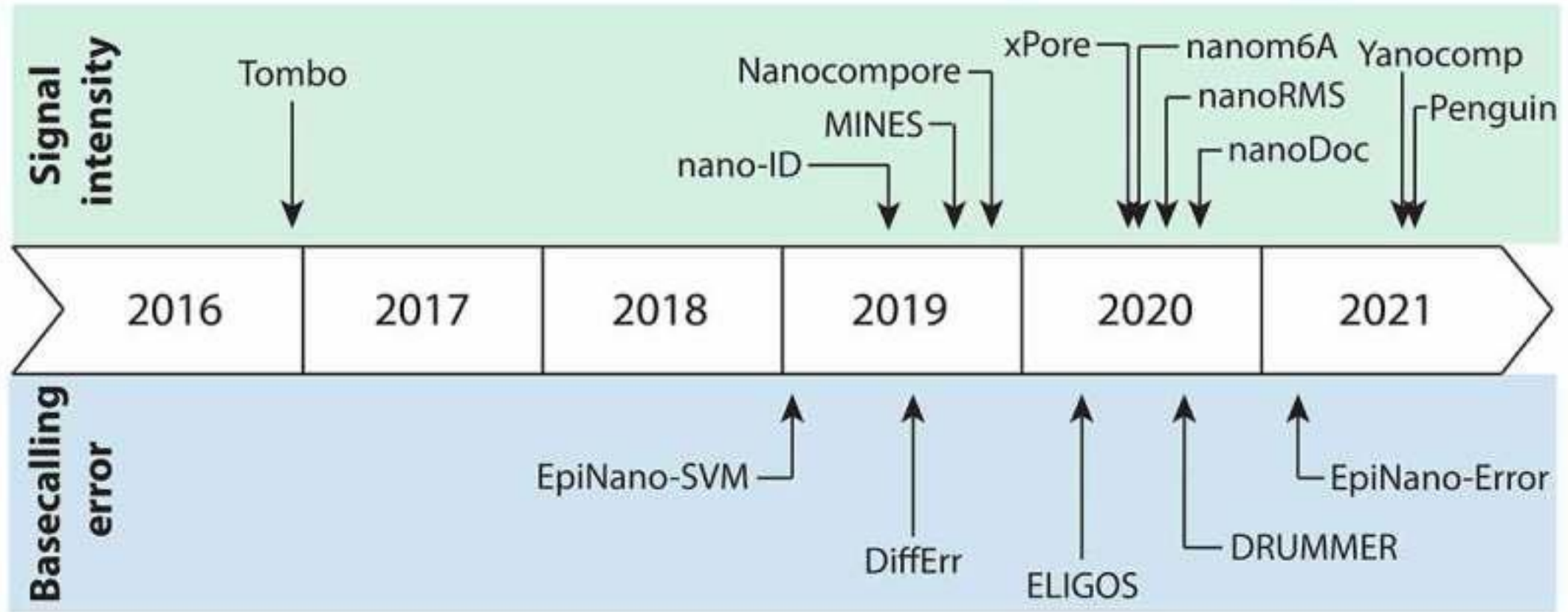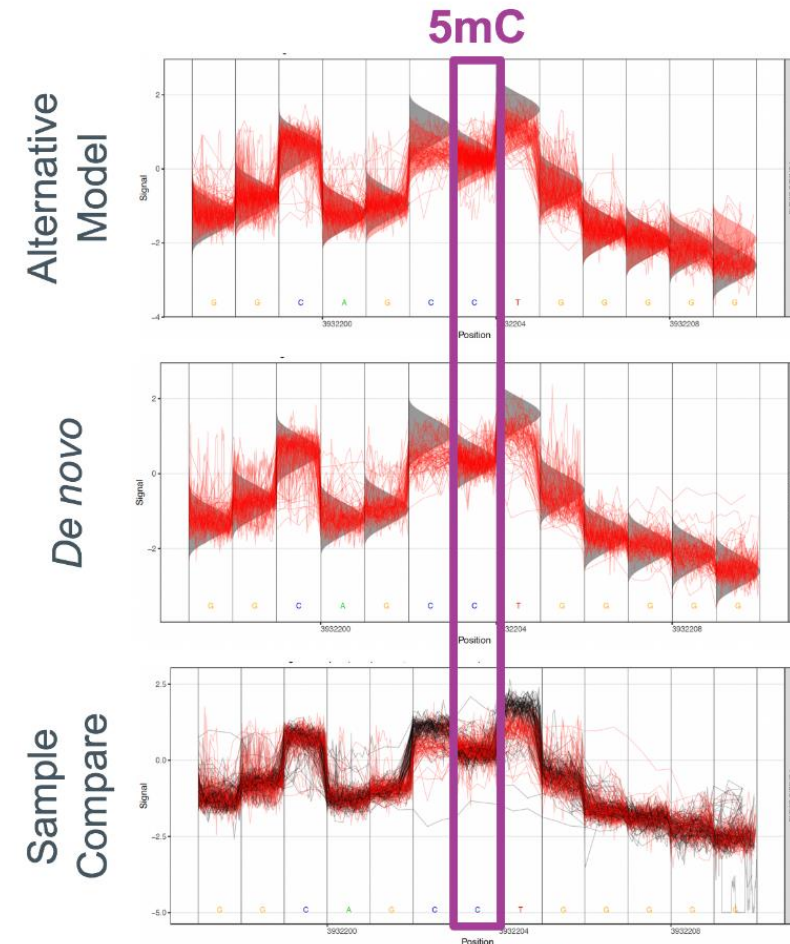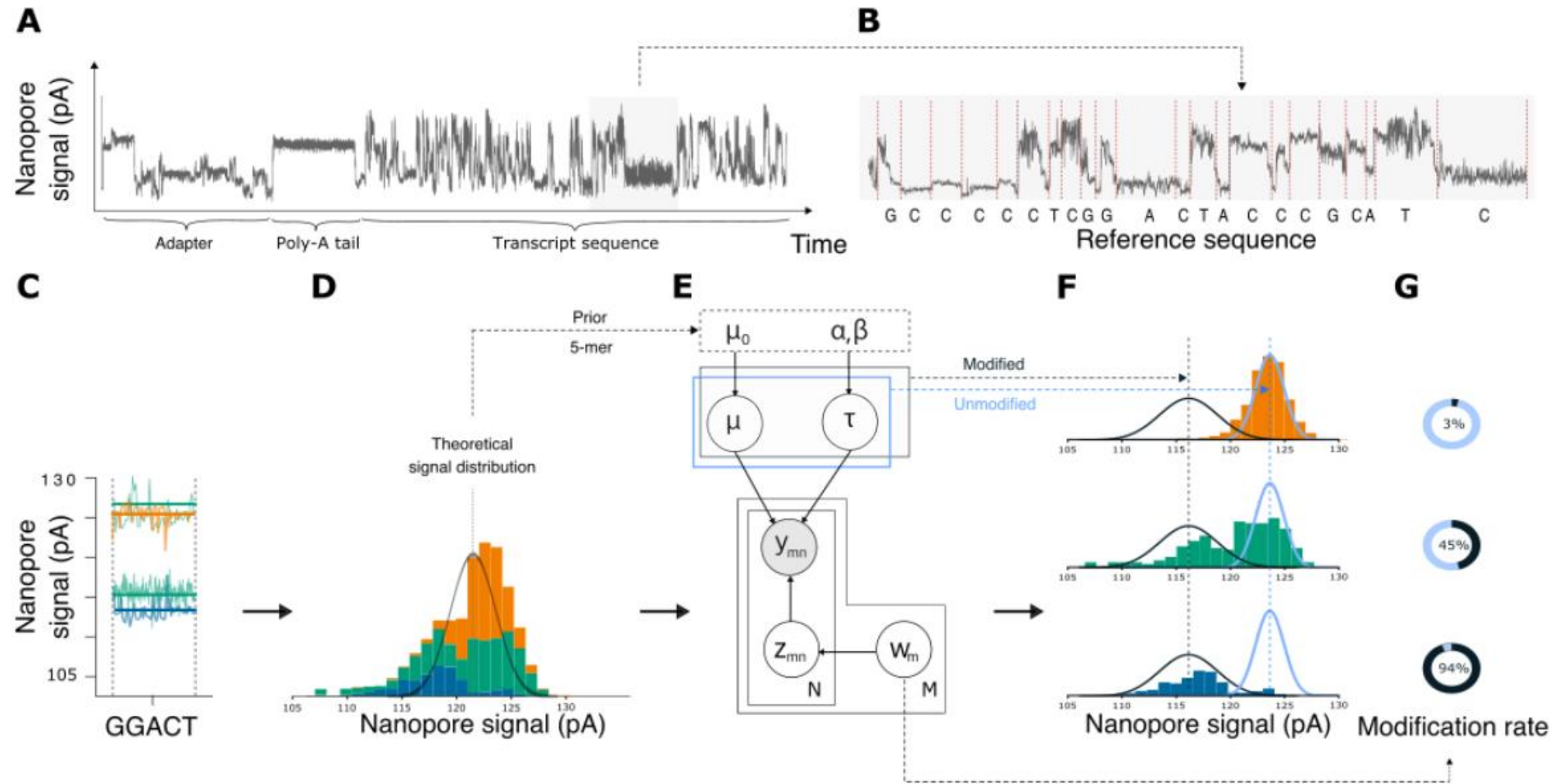
EpiNano
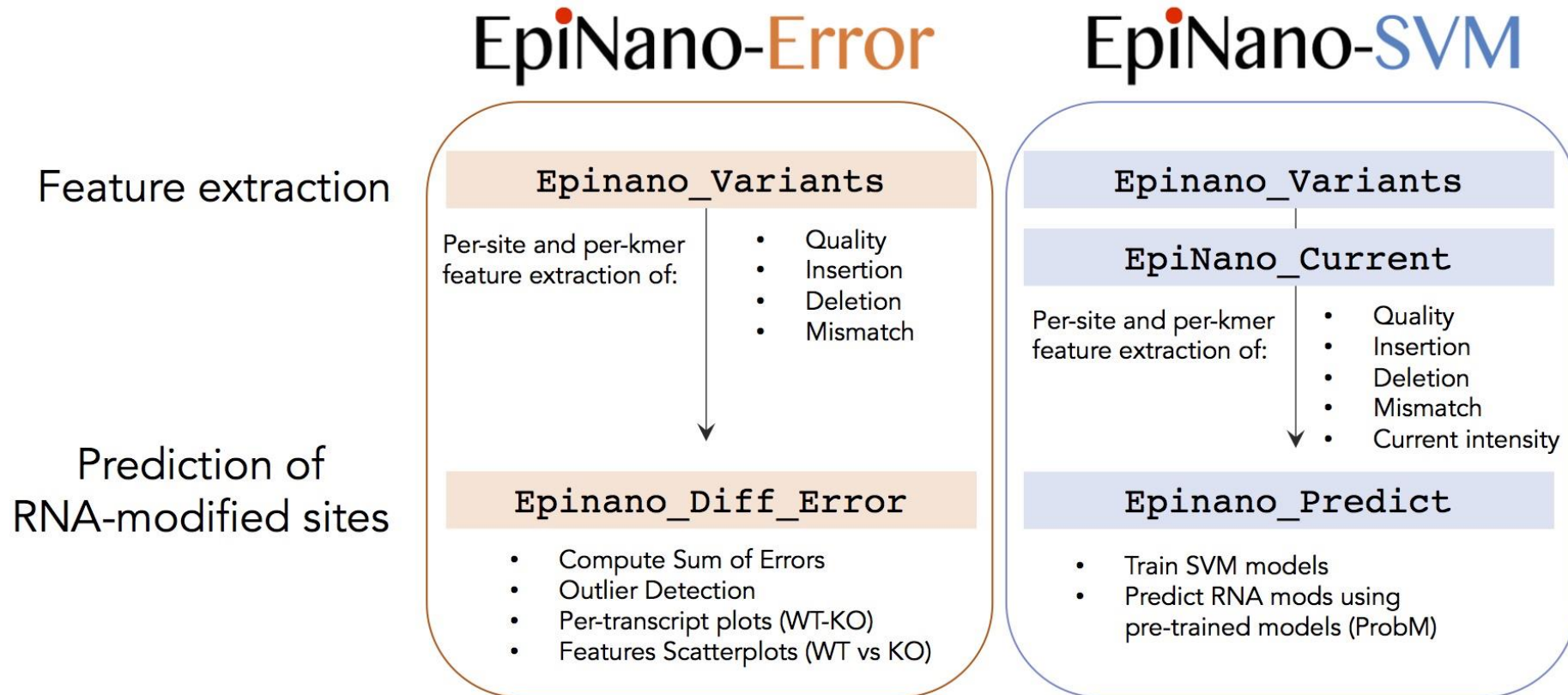
ELIGOS

Tombo

MODIFIED BASE DETECTION

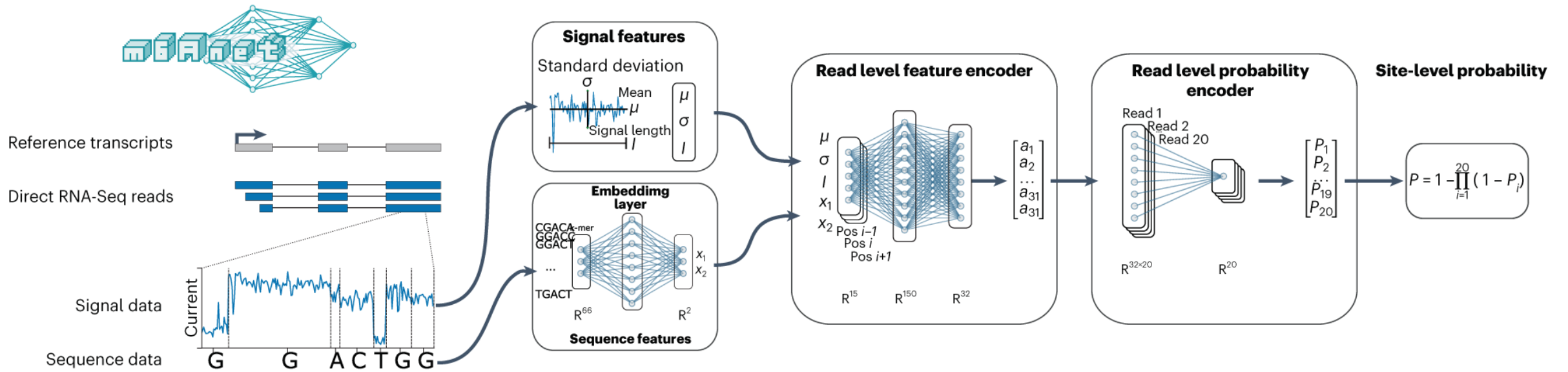Tombo provides three methods for the identification of non-standard bases

	Advantages	Disadvantages
Alternative Model	• Known alt. base • Exact alt. loc. • Good accuracy	• Requires alt. model estimation
De novo	• Apply to any sample	• High error rate • Inexact location • Alt. base unknown
Sample Compare	• Best AUC • Most robust	• Inexact location • Alternative base unknown

1 | Nanopore Community Meeting 2017 | @NanoporeConf #NanoporeConf

[nanoporetech/tombo](#) (doi:10.1101/094672)

Xpore

m6Anet

Nanocompore

- Identifies signal-level differences between conditions (e.g. WT vs IVT/KD/KO)

- Based on Nanopolish resquiggling

- Robust and flexible statistical framework

- Takes into account biological variability

- Allows for complex statistical designs (e.g. multi-factor designs, batch effects, etc.)
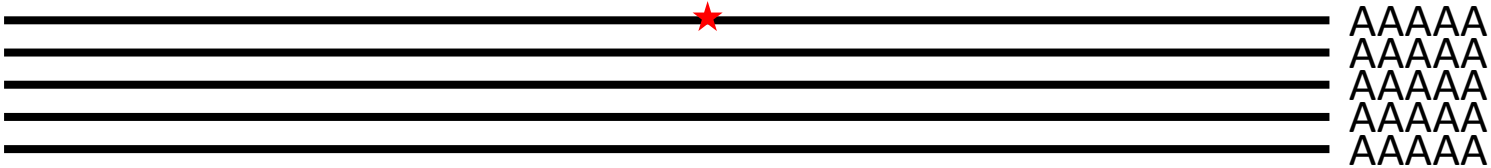
Nanocompore analysis pipeline
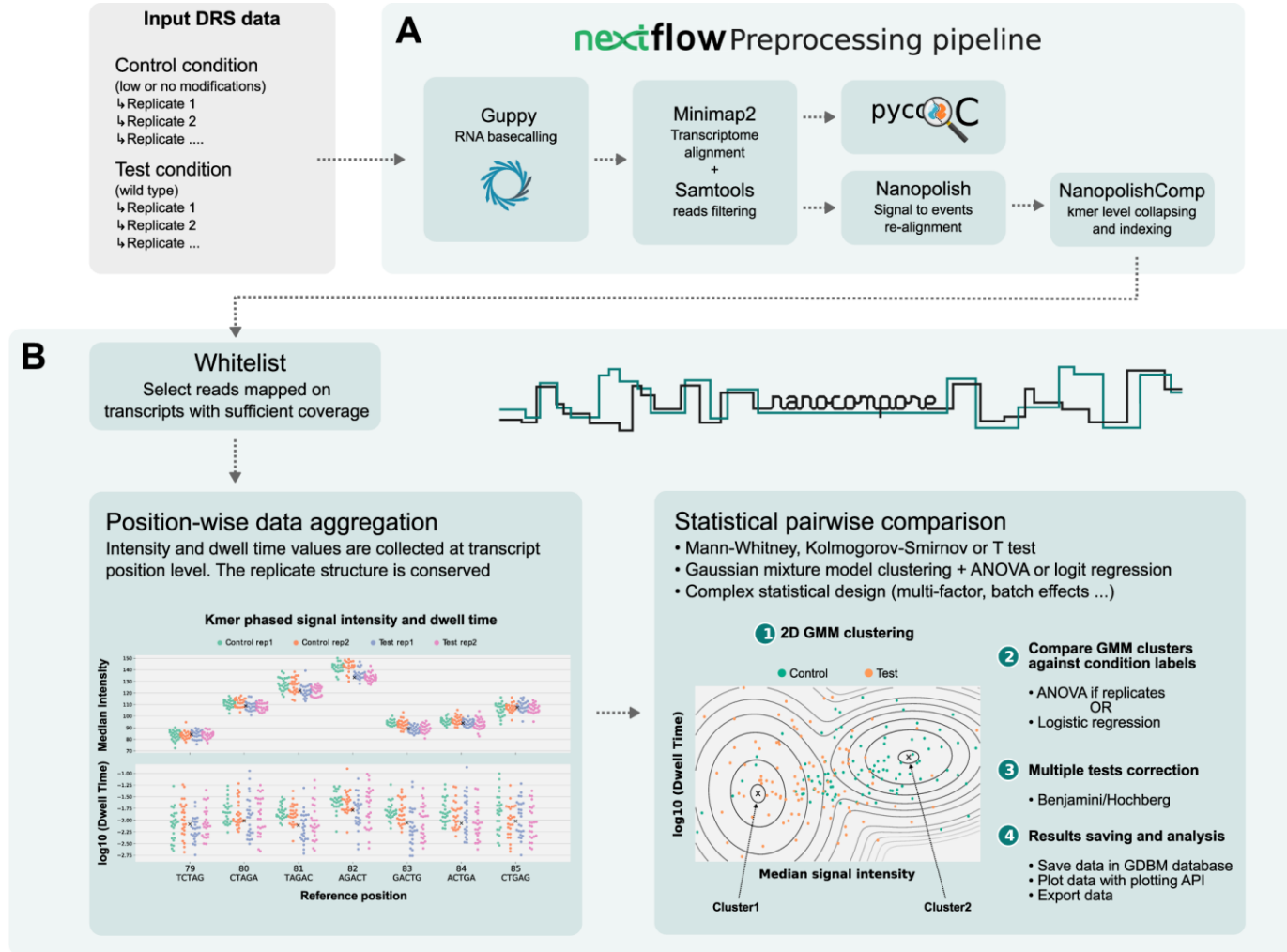
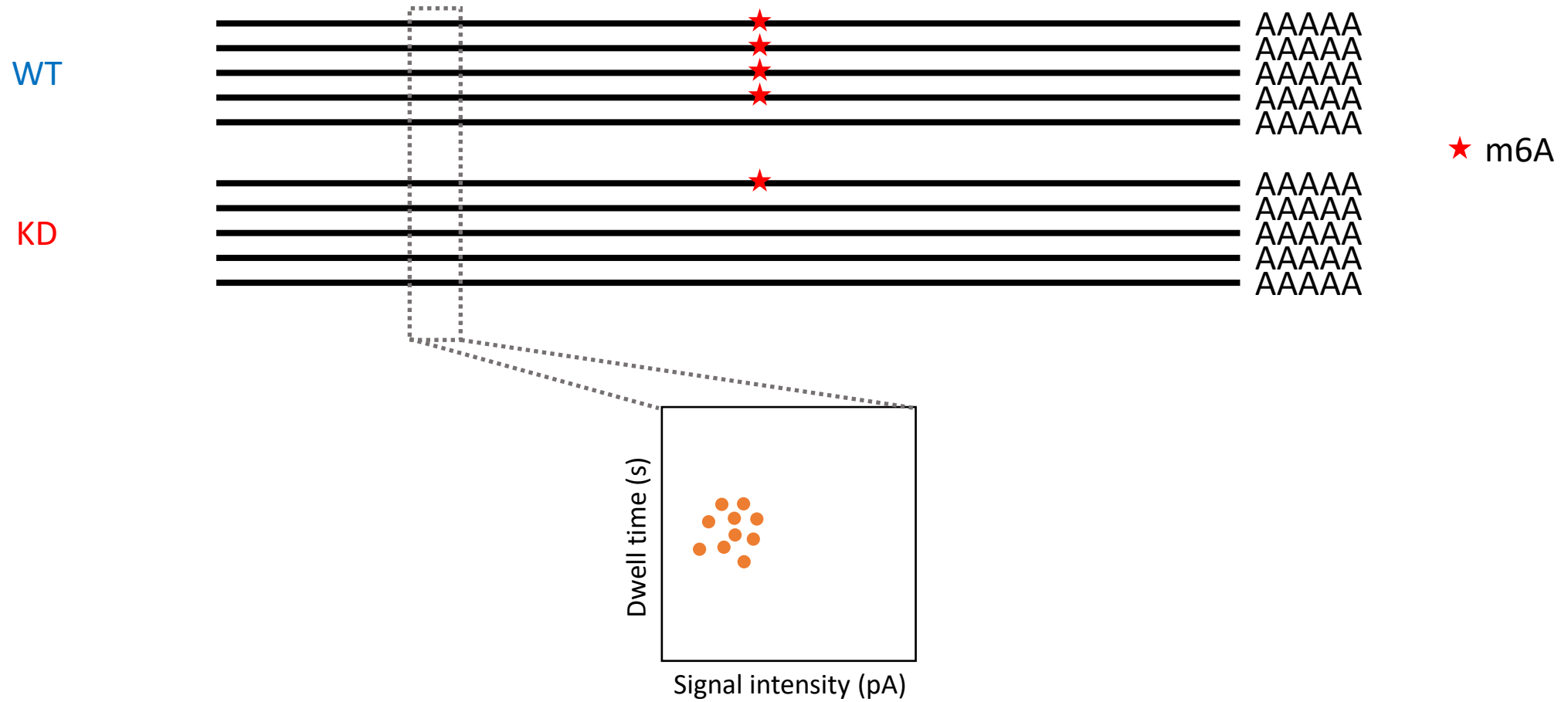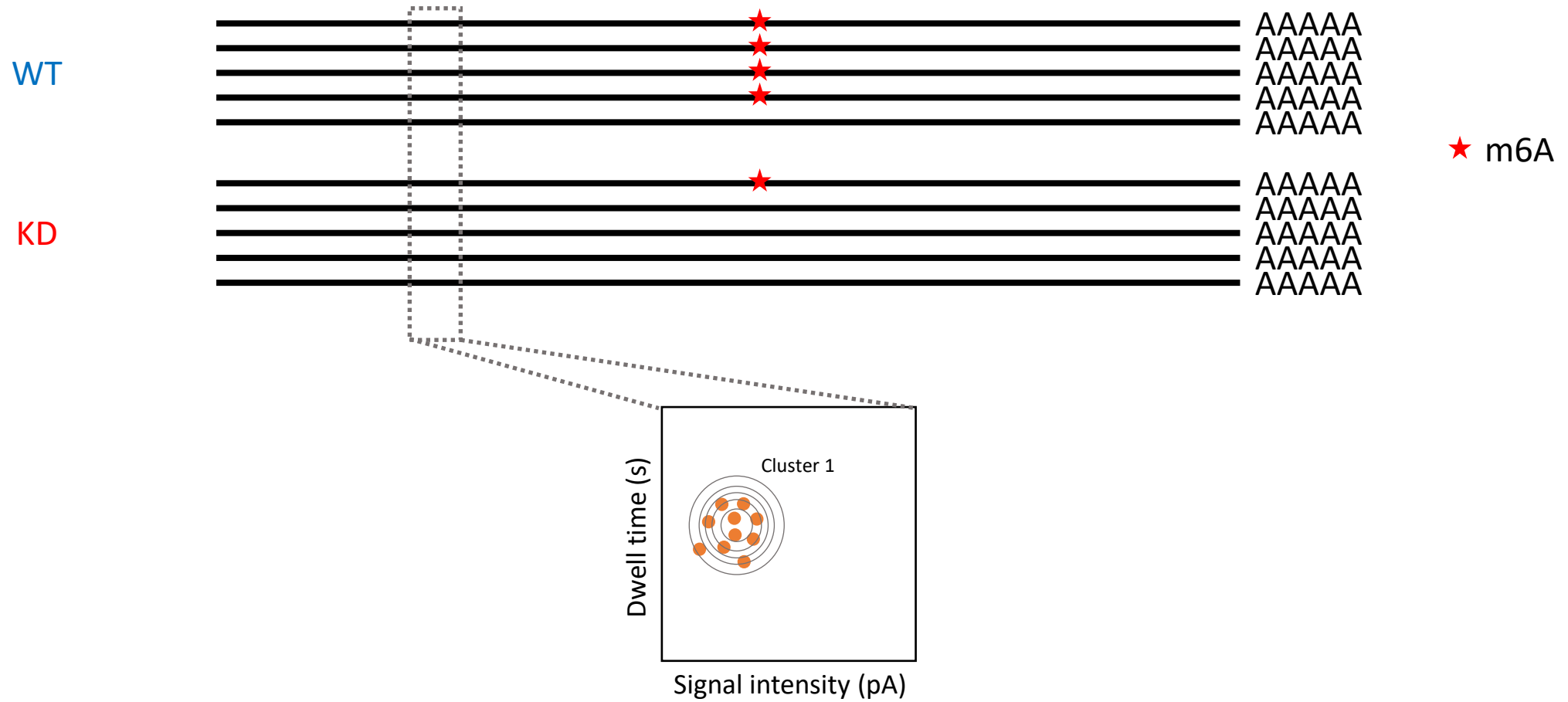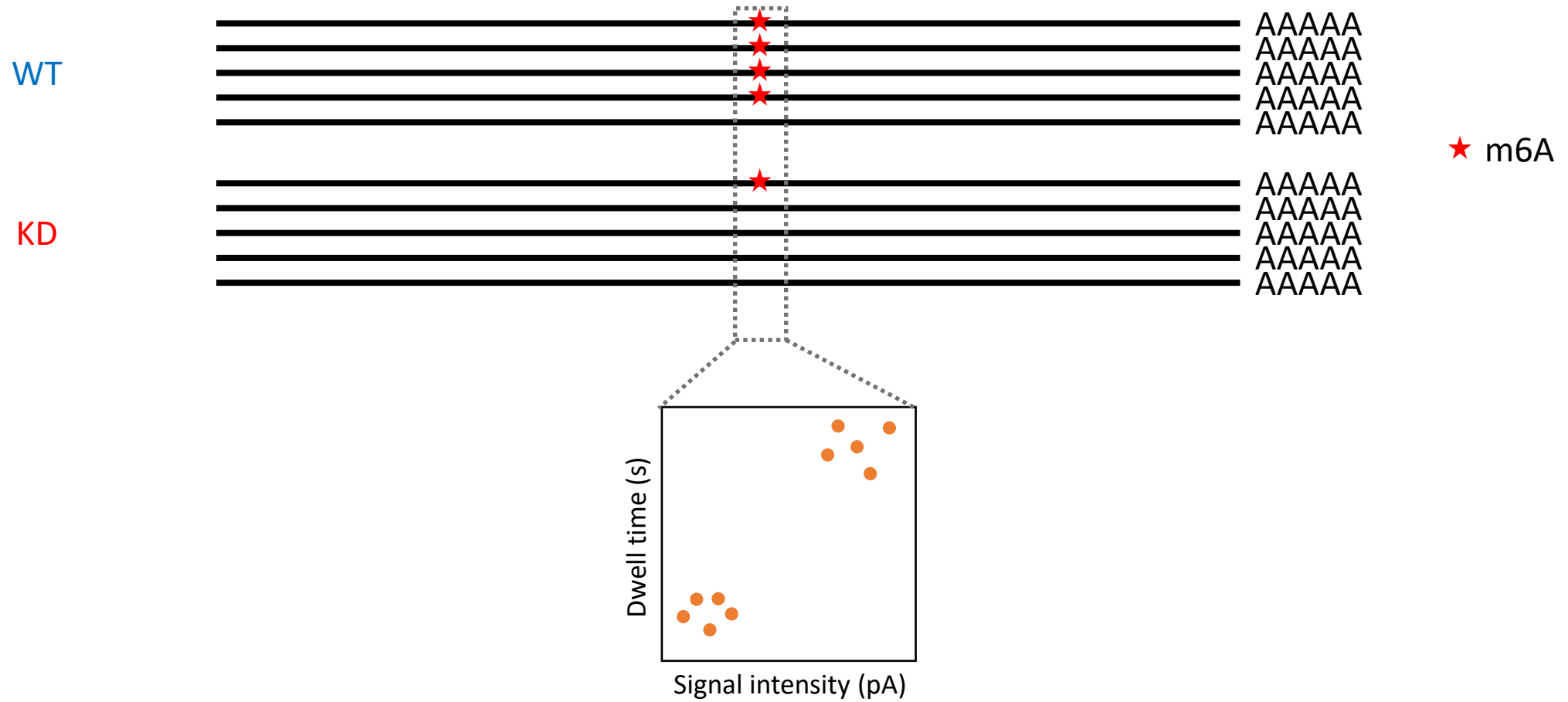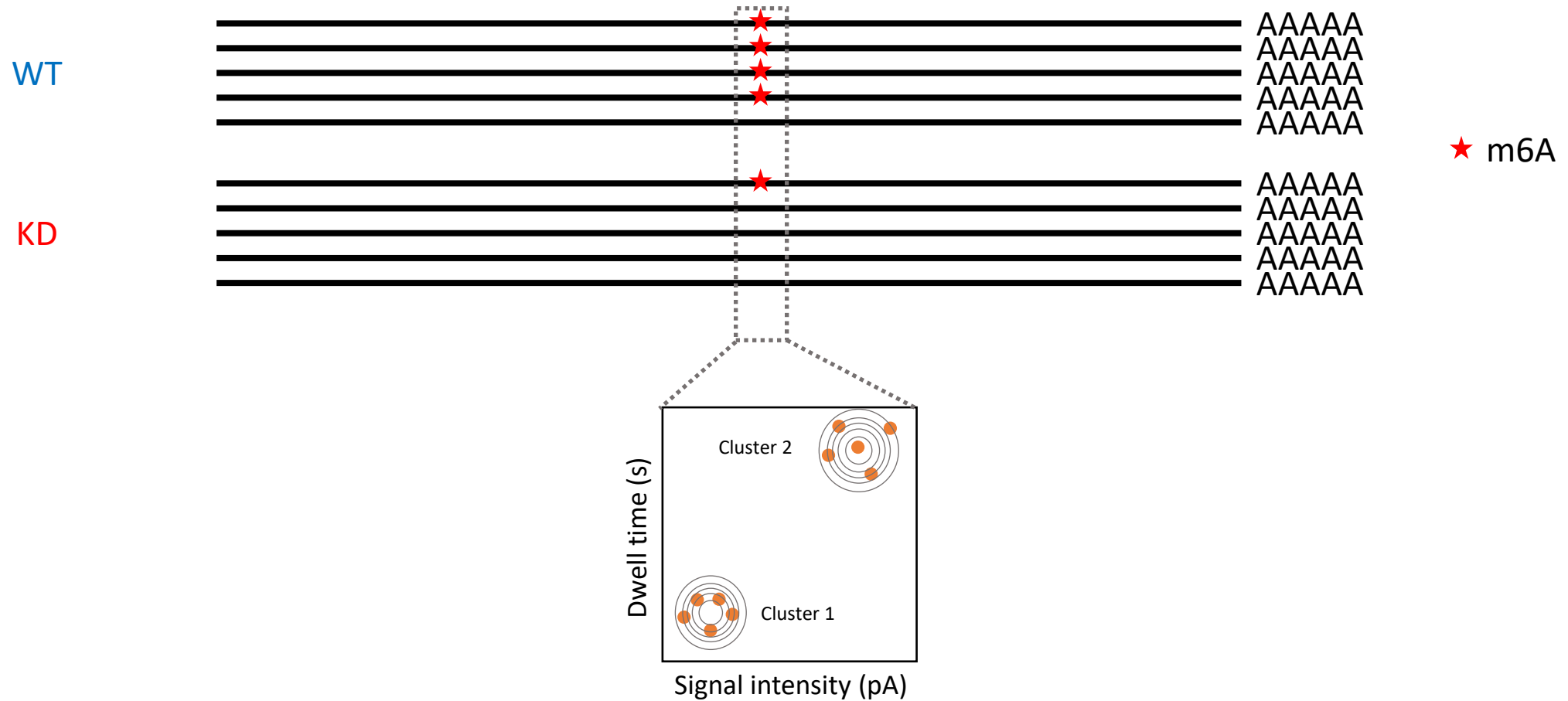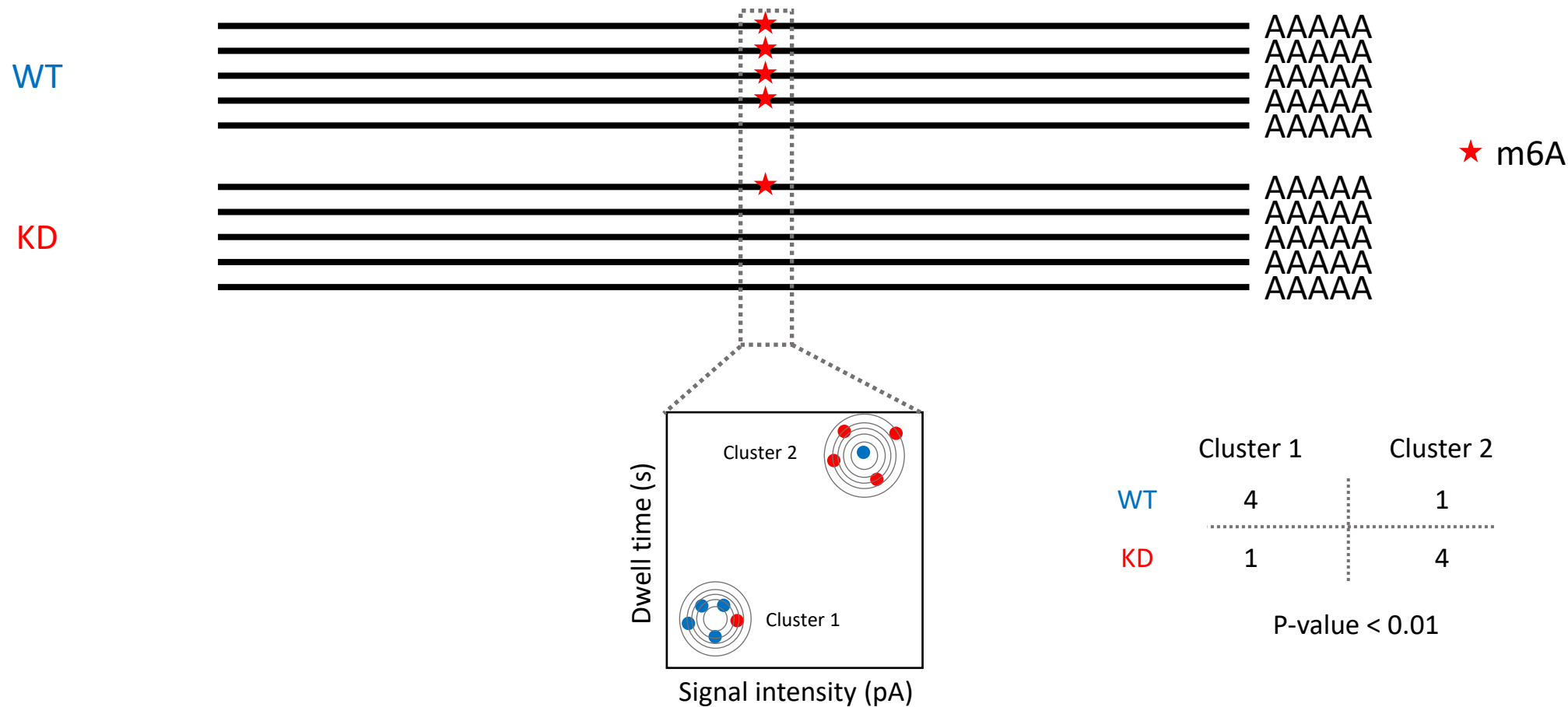Leger et. al. *Nature Comm* (2021)

GMM for modification detection

GMM for modification detection

GMM for modification detection

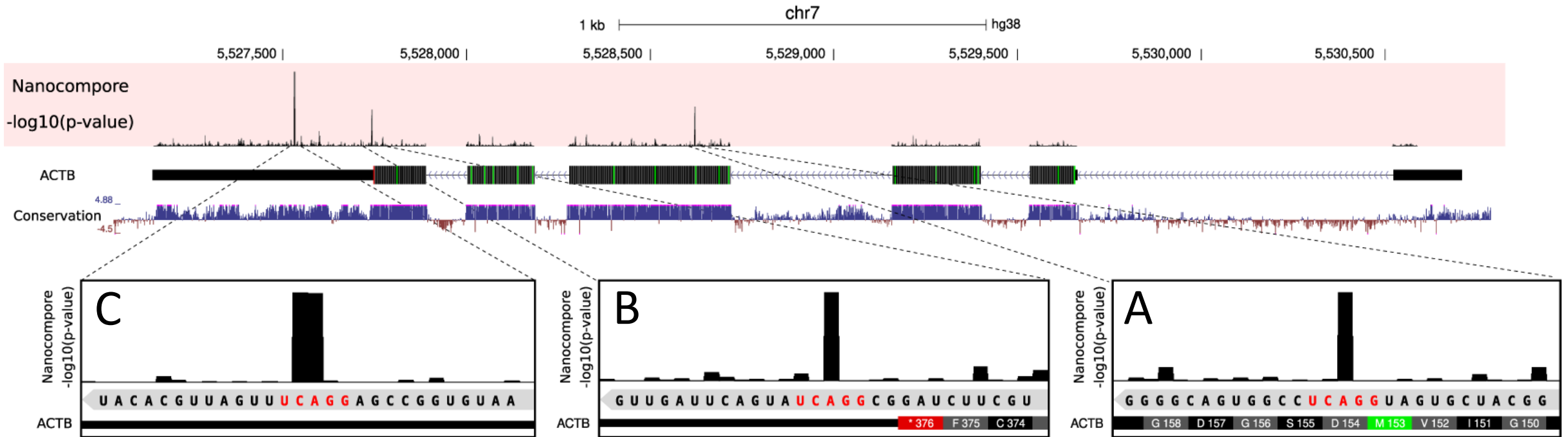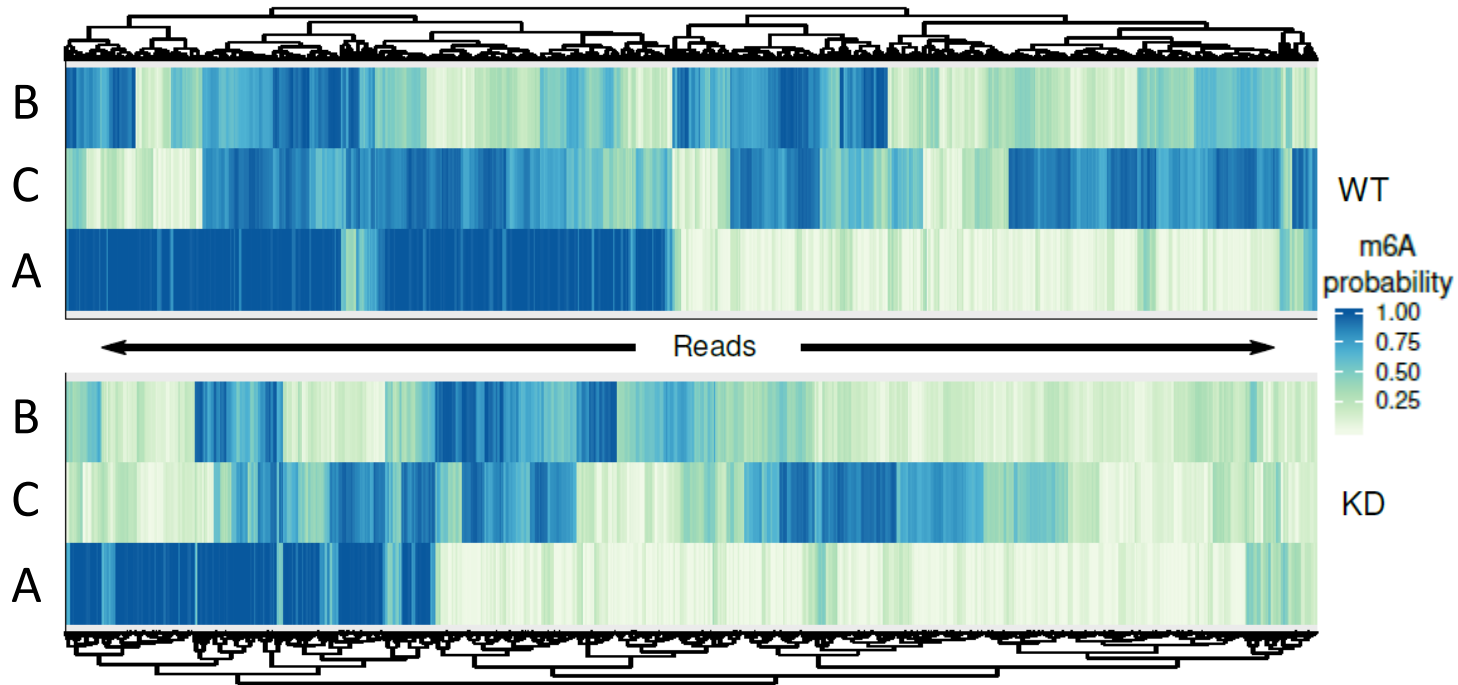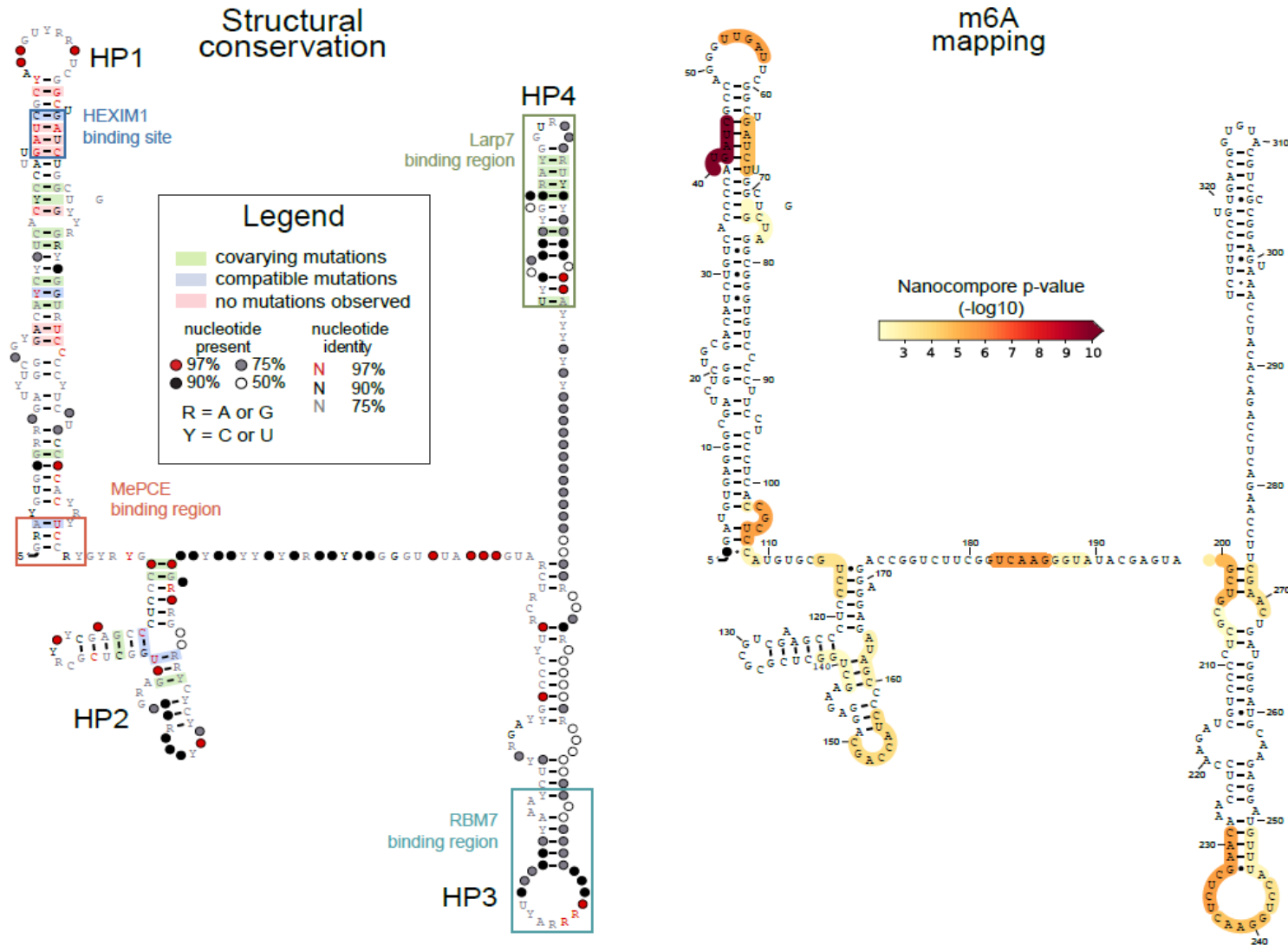# GMM for modification detection

GMM for modification detection

GMM for modification detection

Detecting m⁶A on ACTB transcripts

Detecting m^{6}A on ACTB transcripts

Leger, et al. *Nature Comm* (2021)
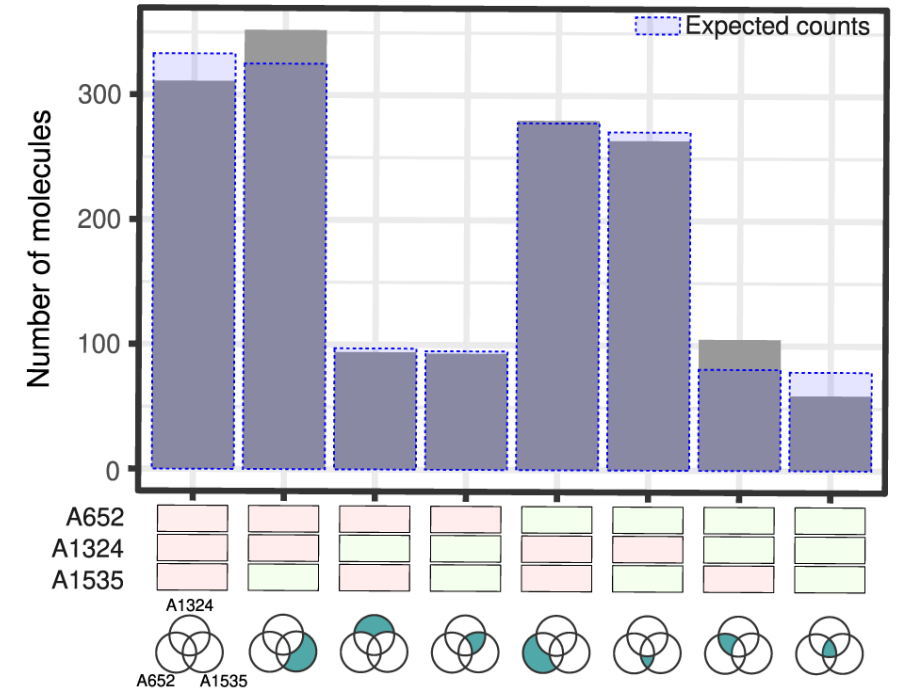
m6A in 7sk snRNA protein binding sites

Leger, et al. *Nature Comm* (2021)

Present Black boxes and *caveats*

- How the modified base interacts with the pore:
  - Kmer-size
  - Dwell Time
  - Base Stacking and RNA secondary structure (unfolding)

- Lessons from rRNA and tRNA:
  - RNA mods often appear in *clusters*
  - RNA mods often *depend* on each other

Future perspectives

- Development of layered multiple-tool pipelines

- Multi-step approaches may be beneficial
    - Pre-computed *priors*
    - *Ad hoc* trained model for deep learning
    - Comparison with a negative set

Questions?

Oxford Nanopore Technologies animation