

TRAINING COURSE IN

Computational Methods

for Epitranscriptomics

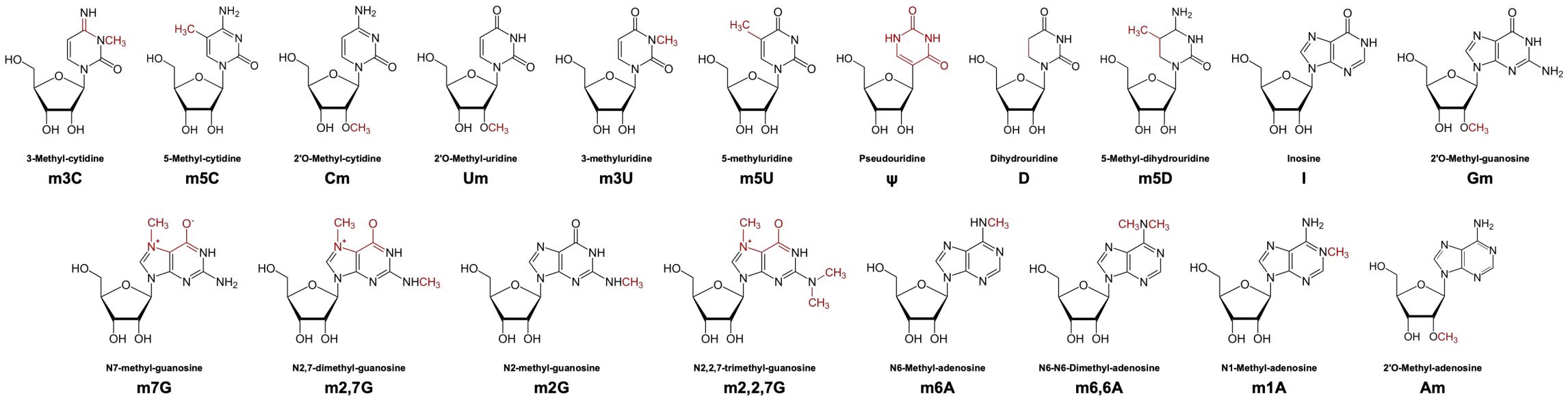
Using Nanopore sequencing to detect RNA modifications

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

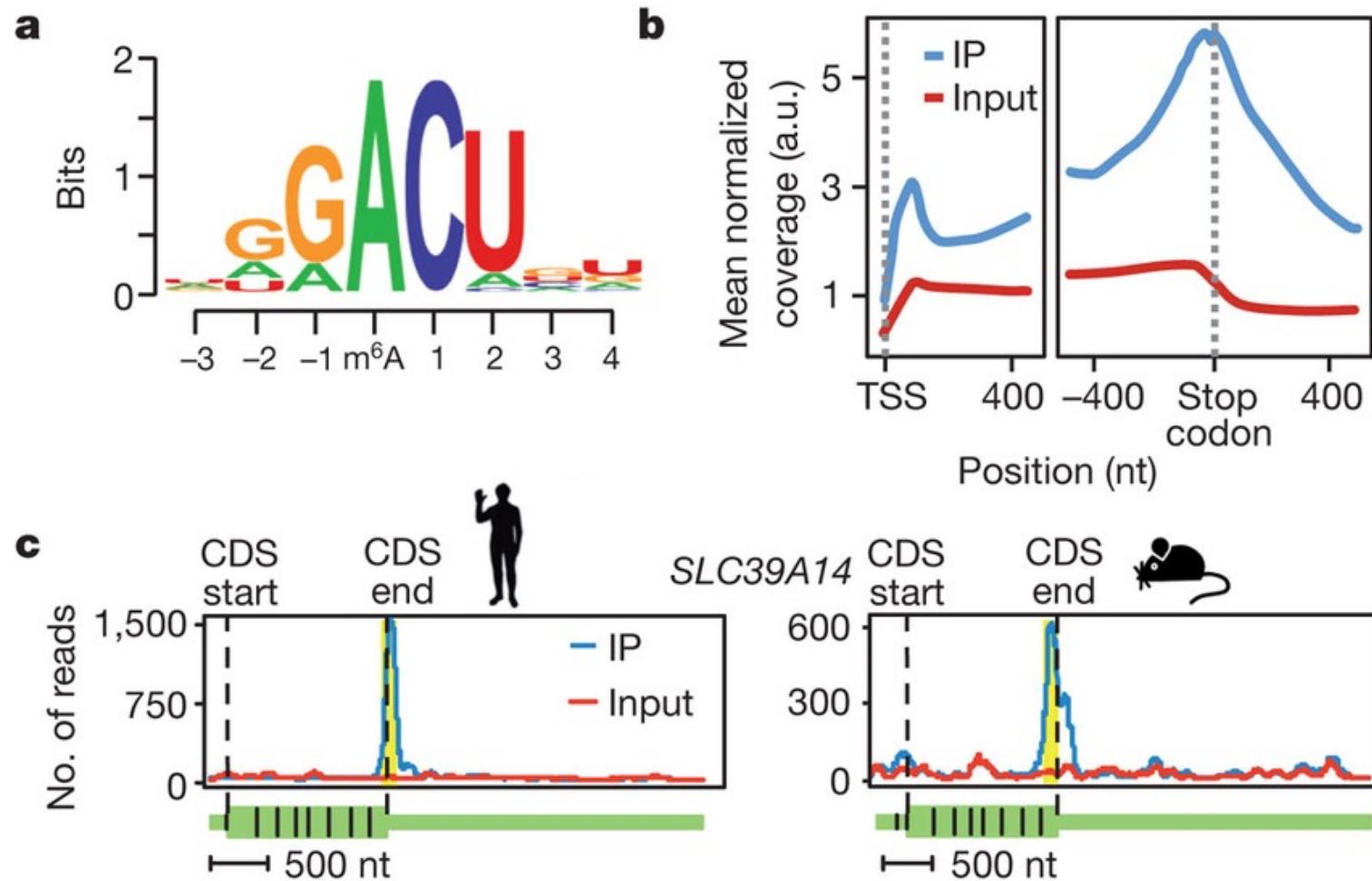


The world of RNA mods

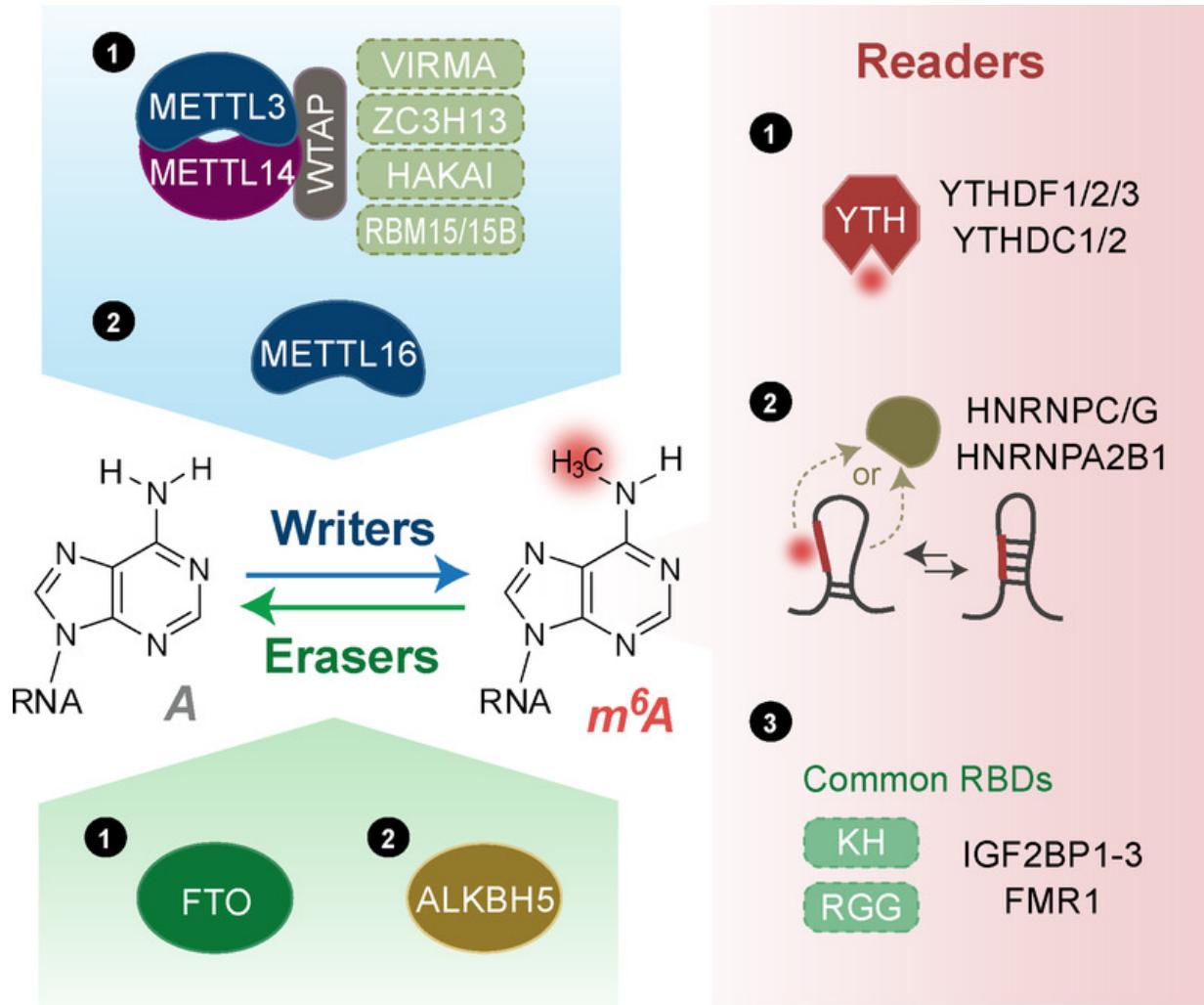
- Highly conserved features found in archaea, bacteria, and eukarya
- Impact on RNA structure and interaction properties



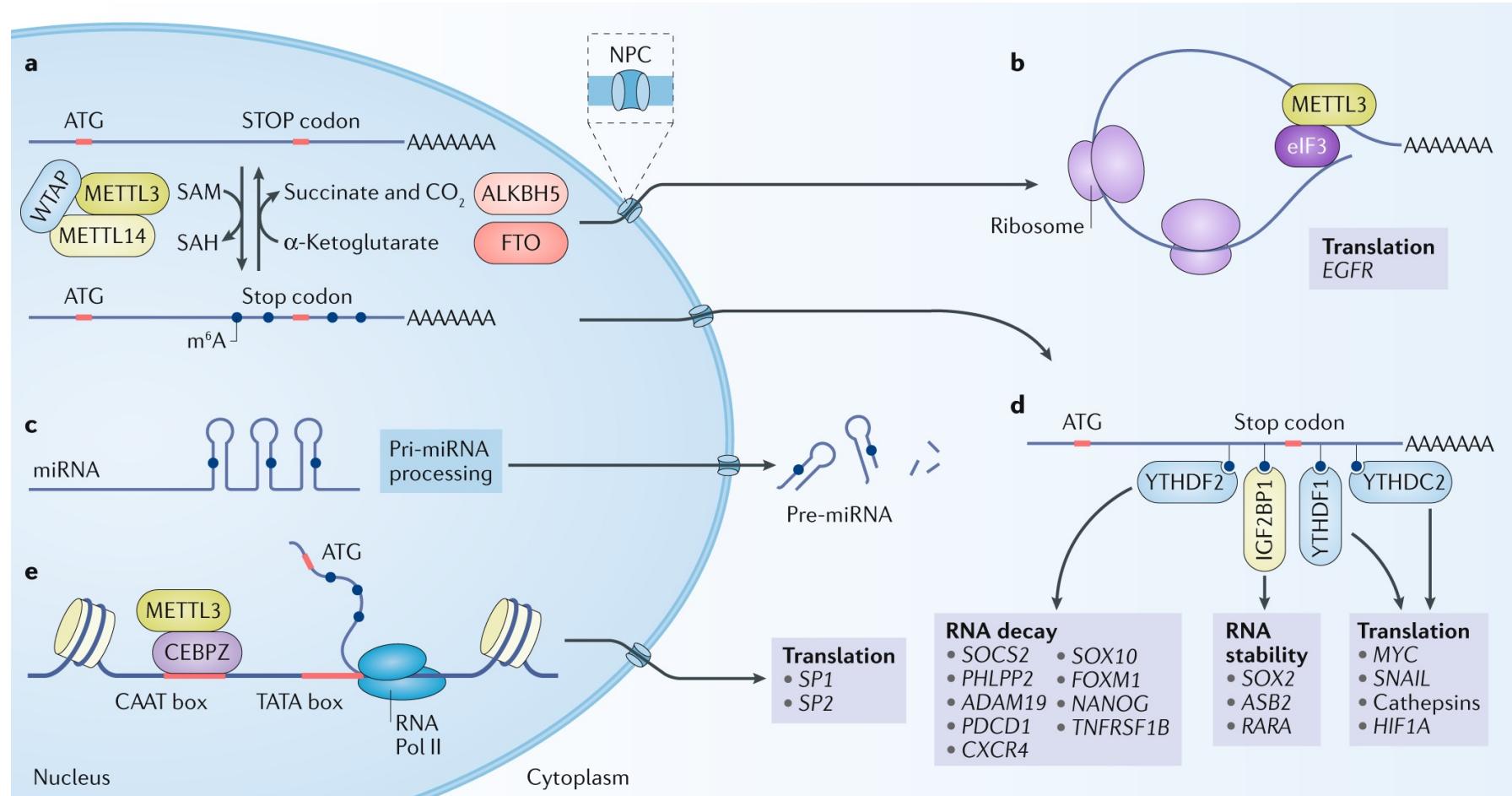
Modifications are widespread in mRNAs



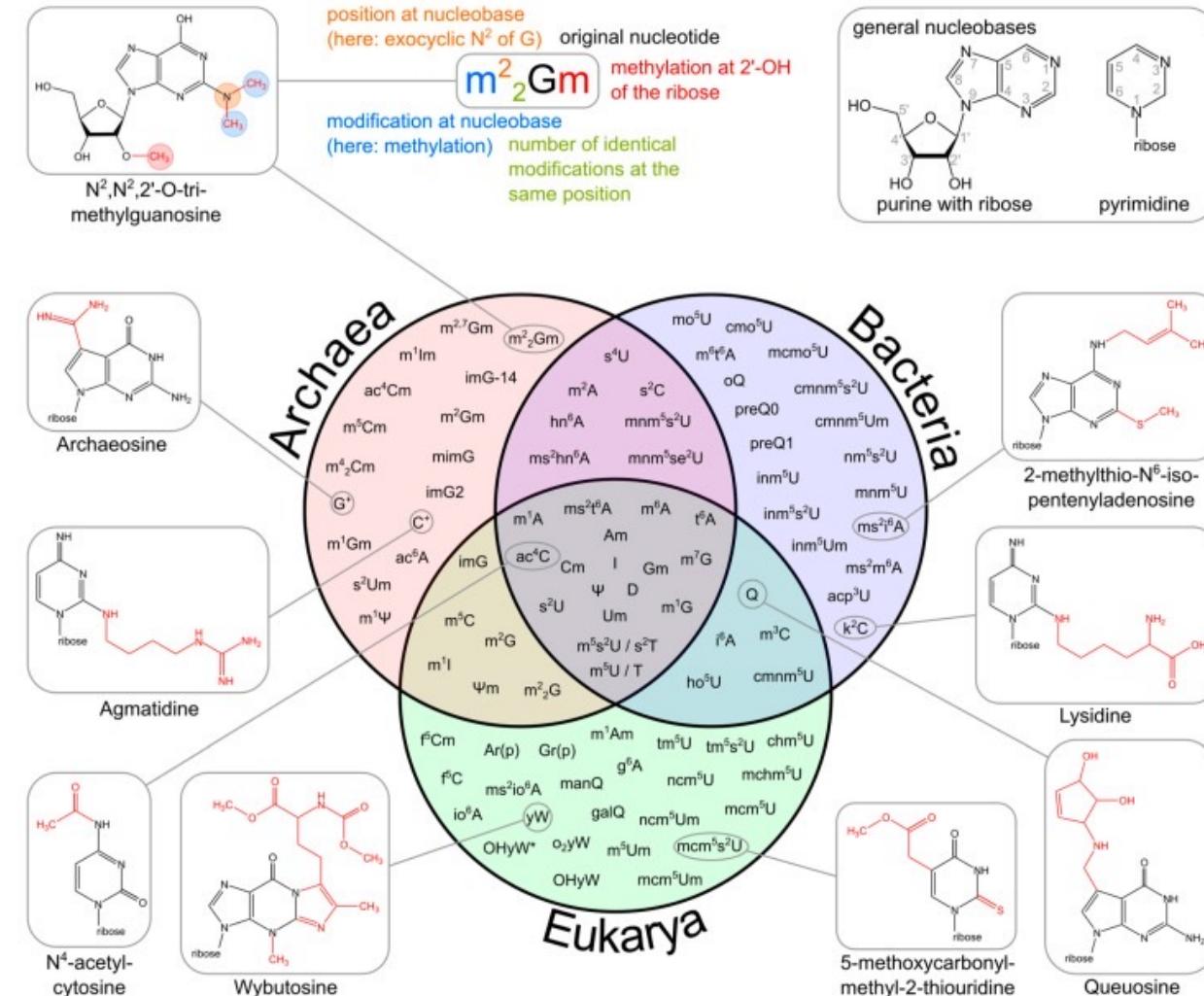
RNA modifications are dynamic



RNA modifications have been implicated in diseases, such as cancer



There are ~170 known naturally occurring RNA modifications



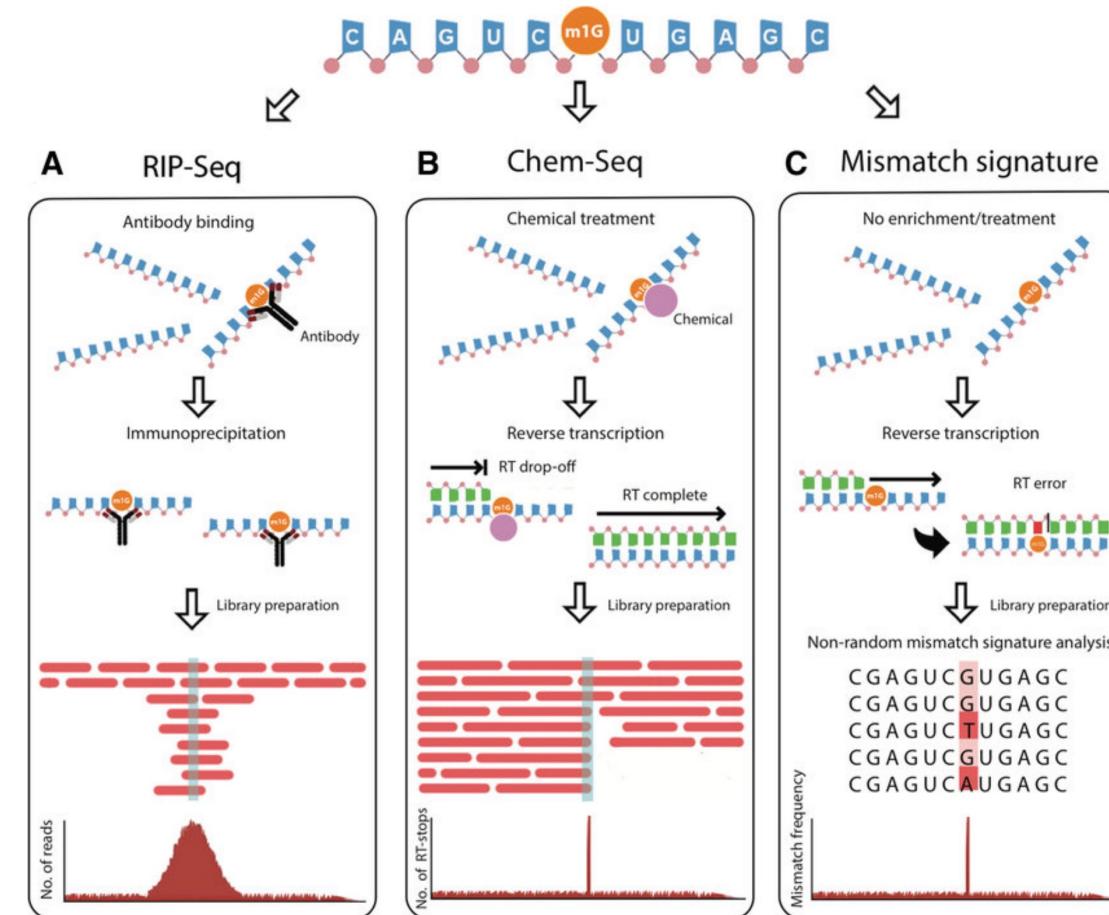
Established techniques for detecting RNA modifications

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



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

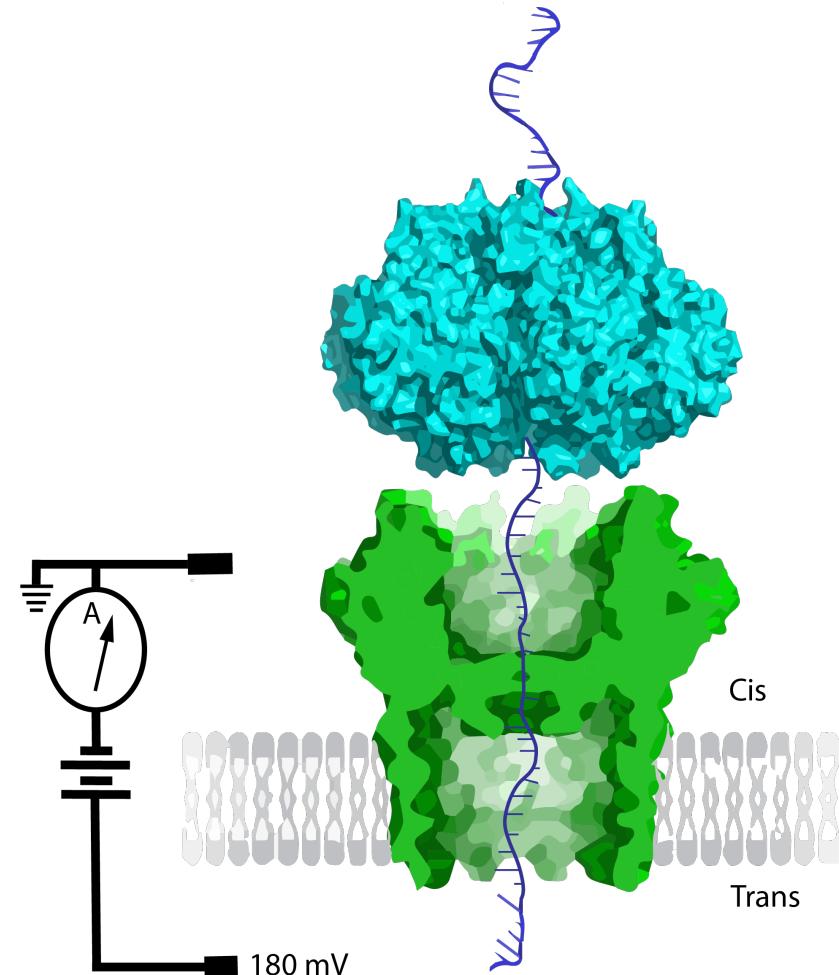
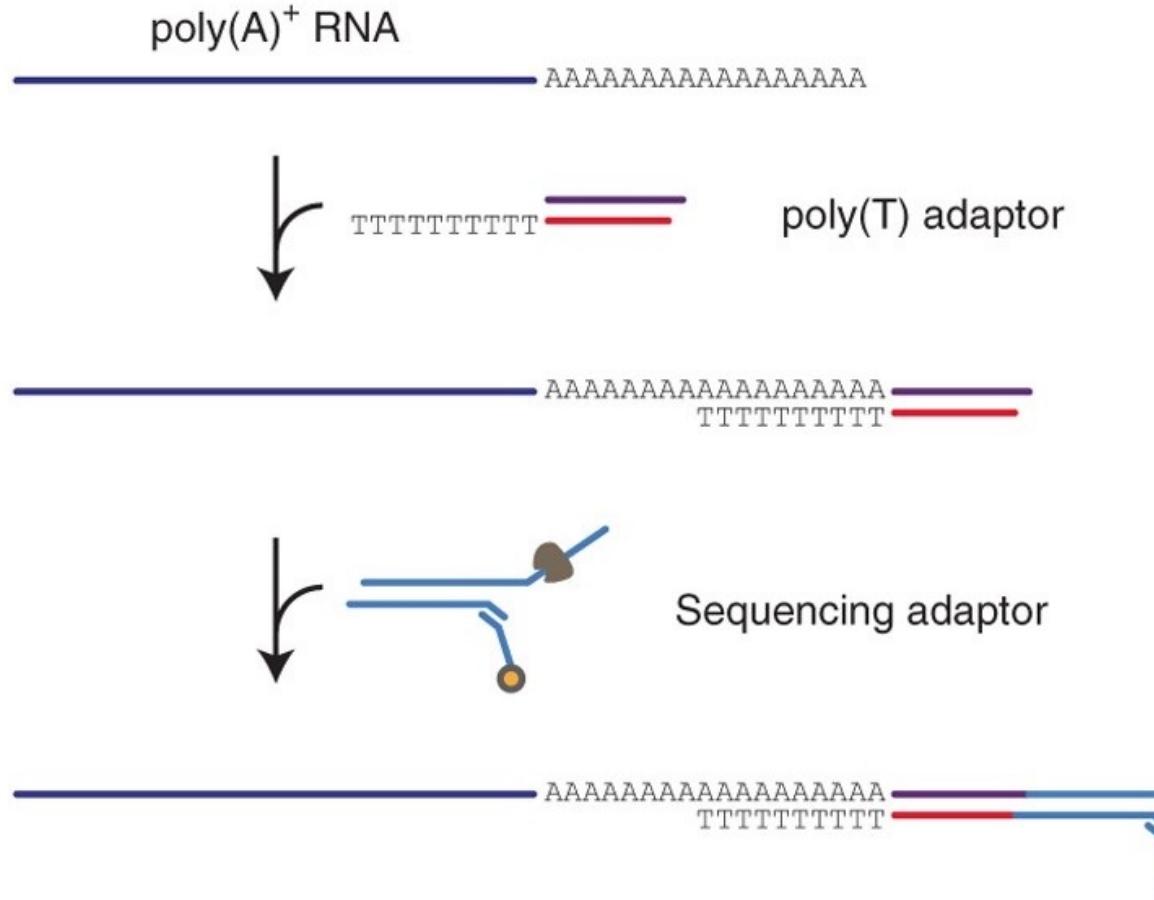


Science 2016 Epitranscriptomics

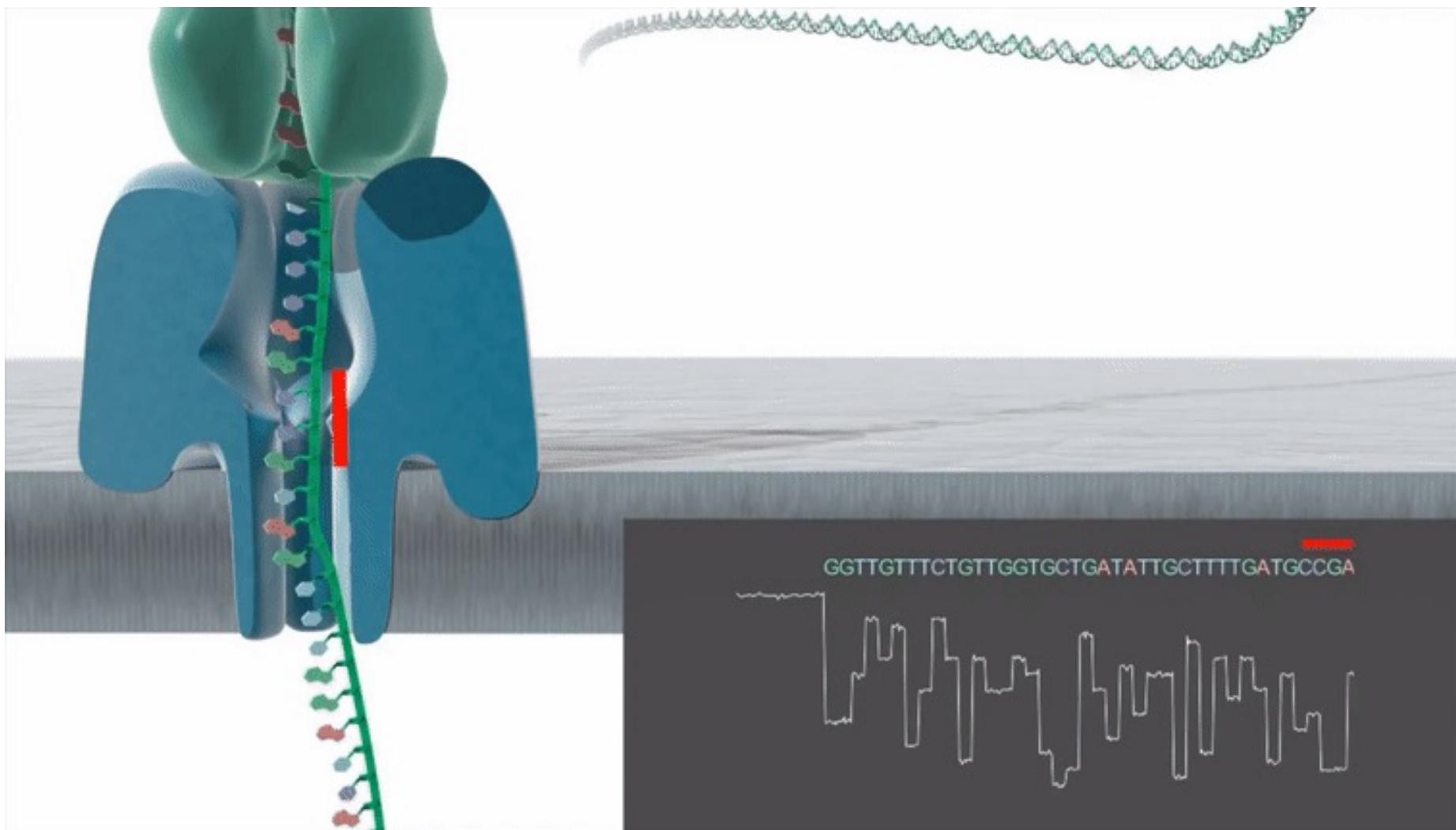


Nat. Med. 2018 Direct RNA Sequencing

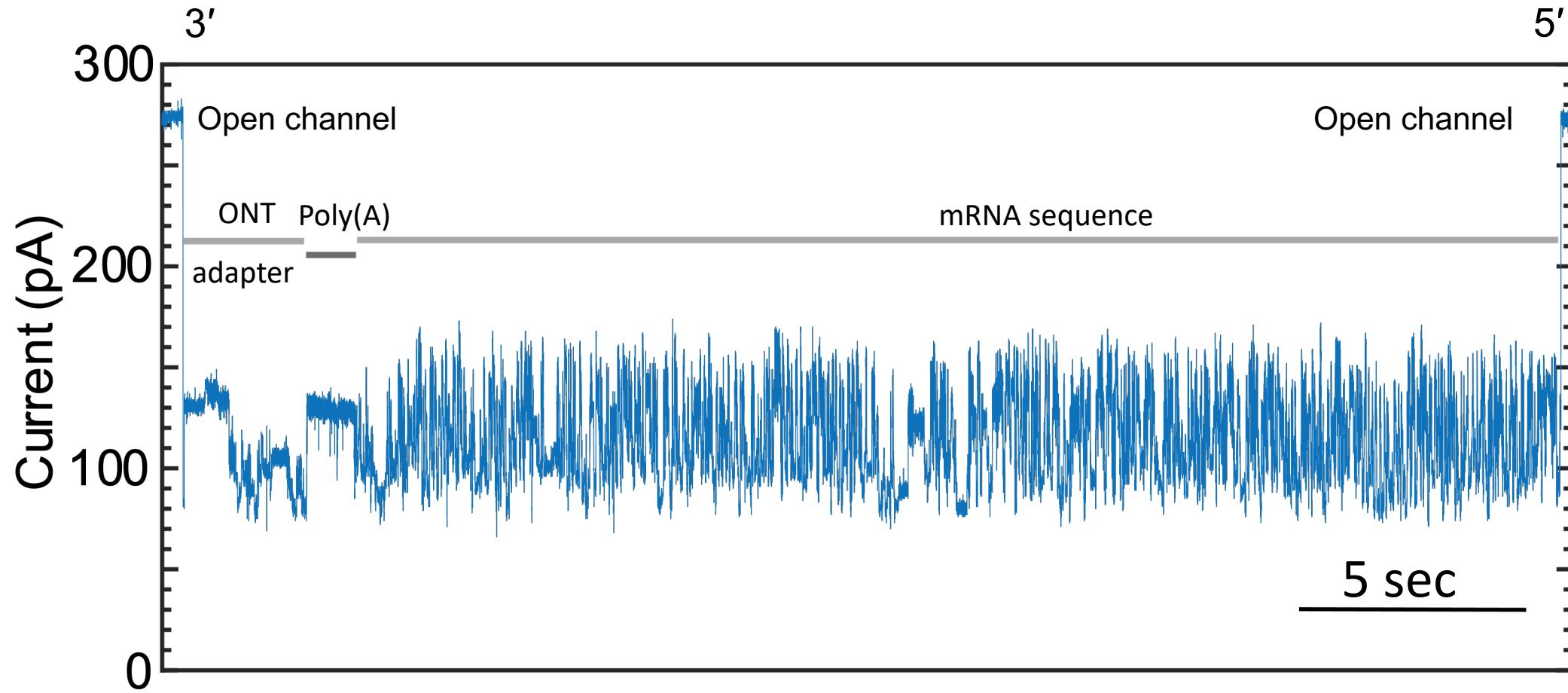
Direct RNA nanopore sequencing



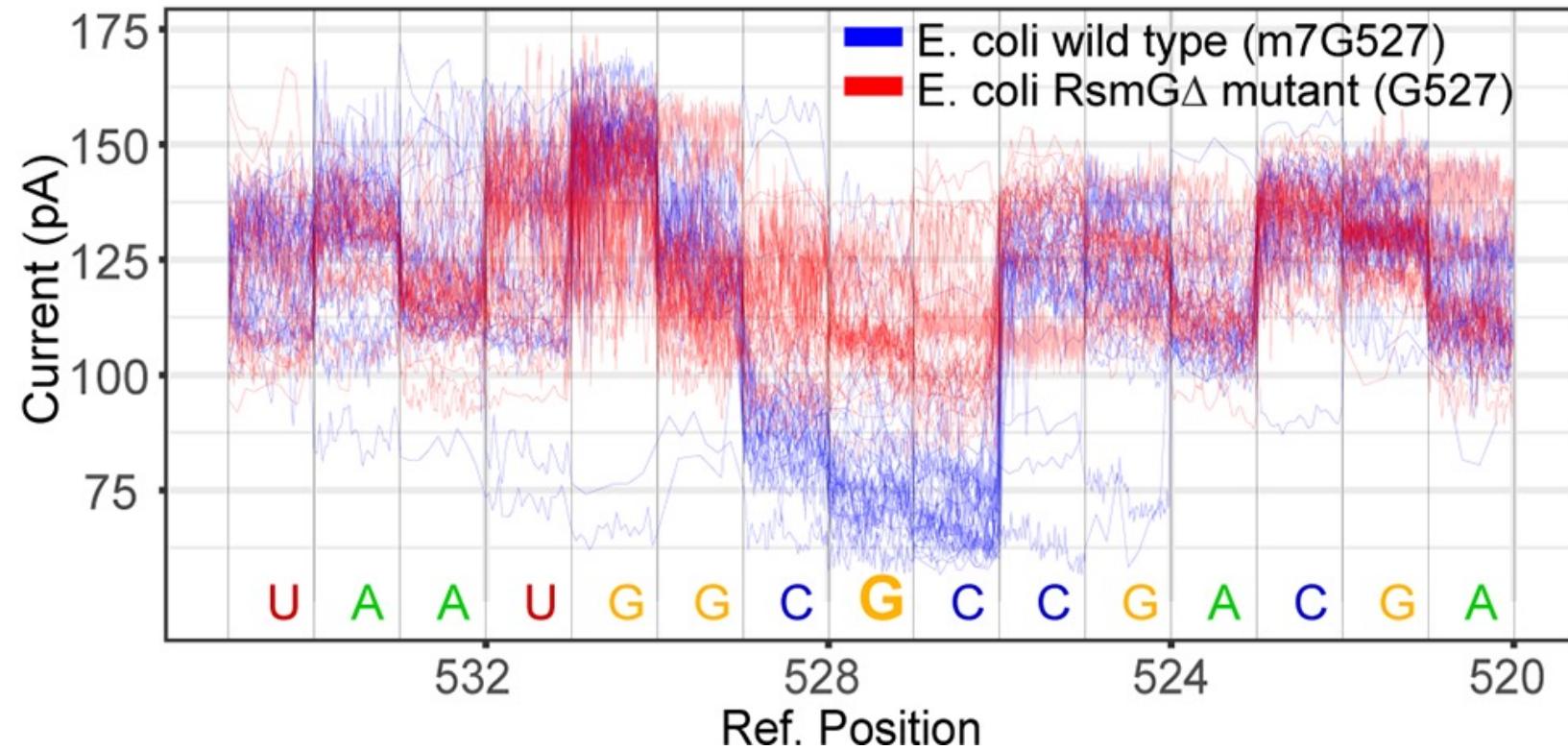
The nanopore sensor directly interacts with the nucleotides



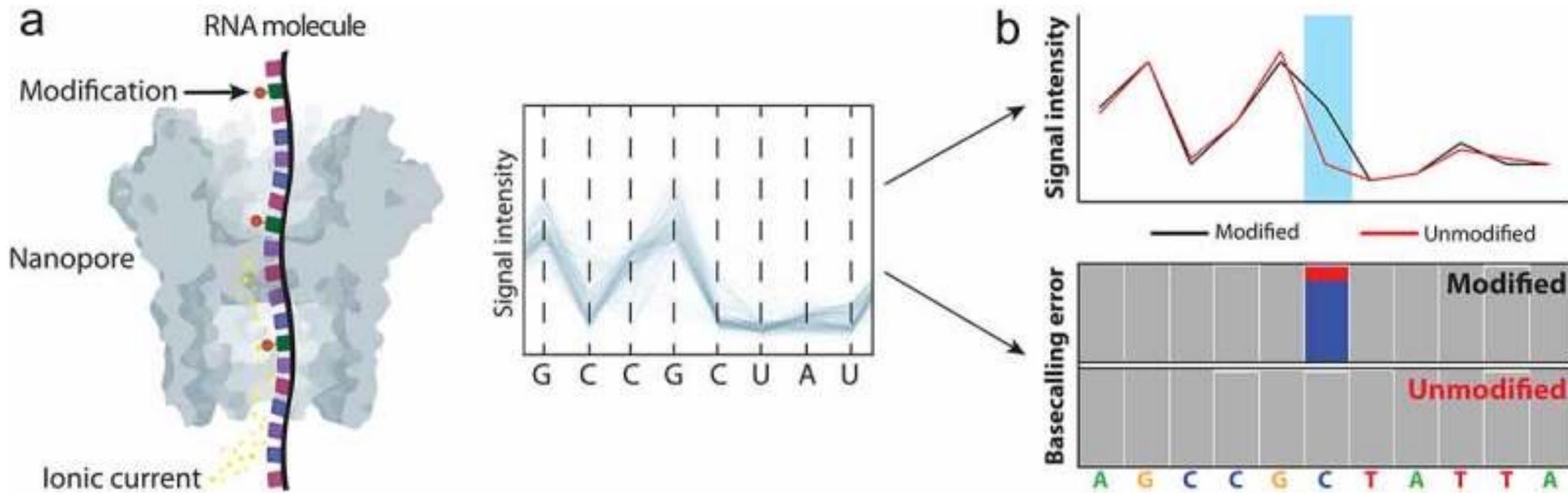
Typical direct RNA nanopore ionic current trace



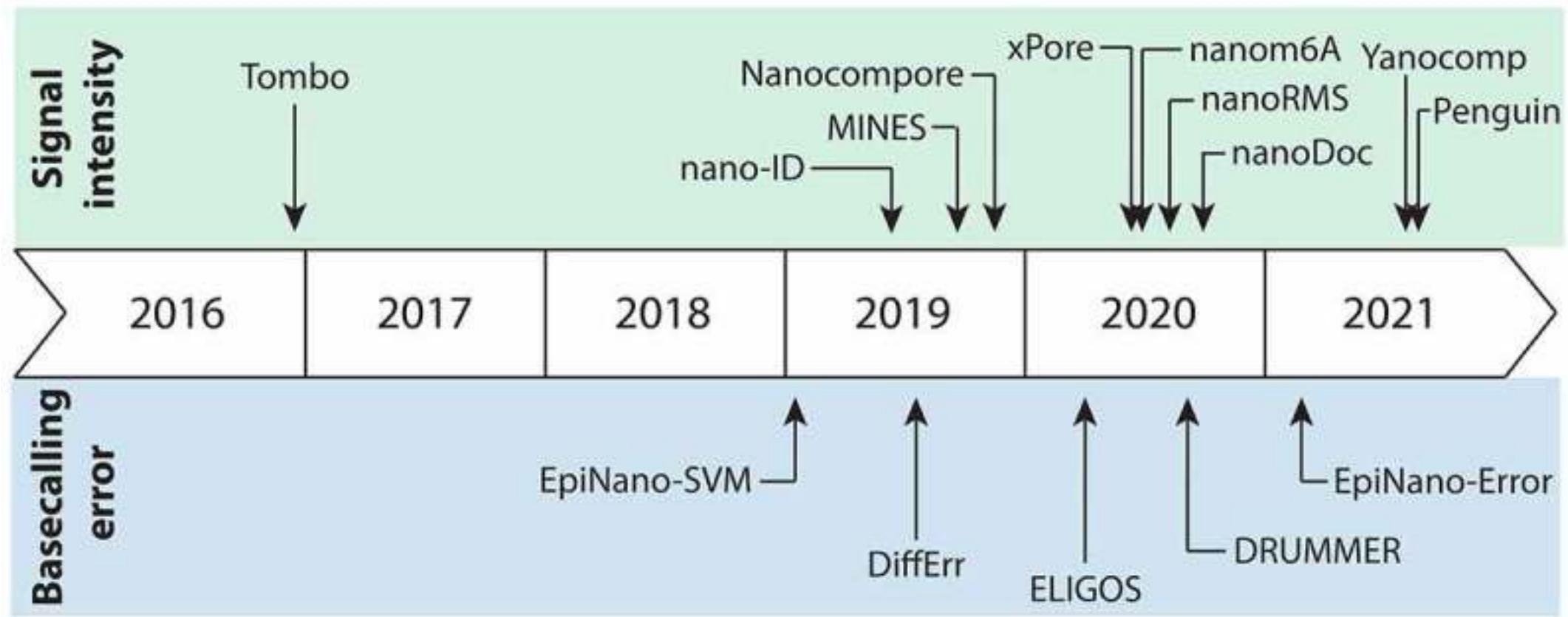
Modified nucleotides alter the ionic current in nanopore traces relative to canonical nucleotides



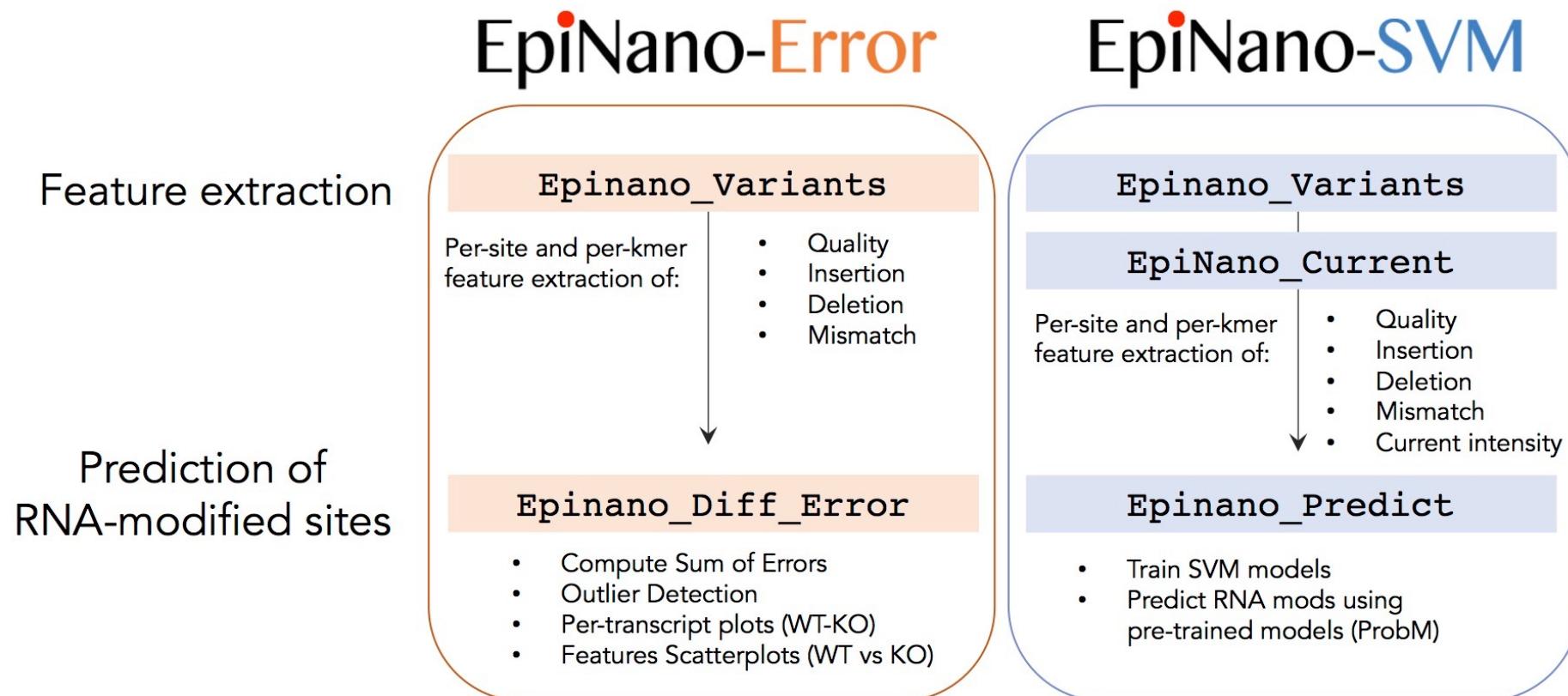
Two general strategies to detect RNA modifications from direct RNA nanopore data



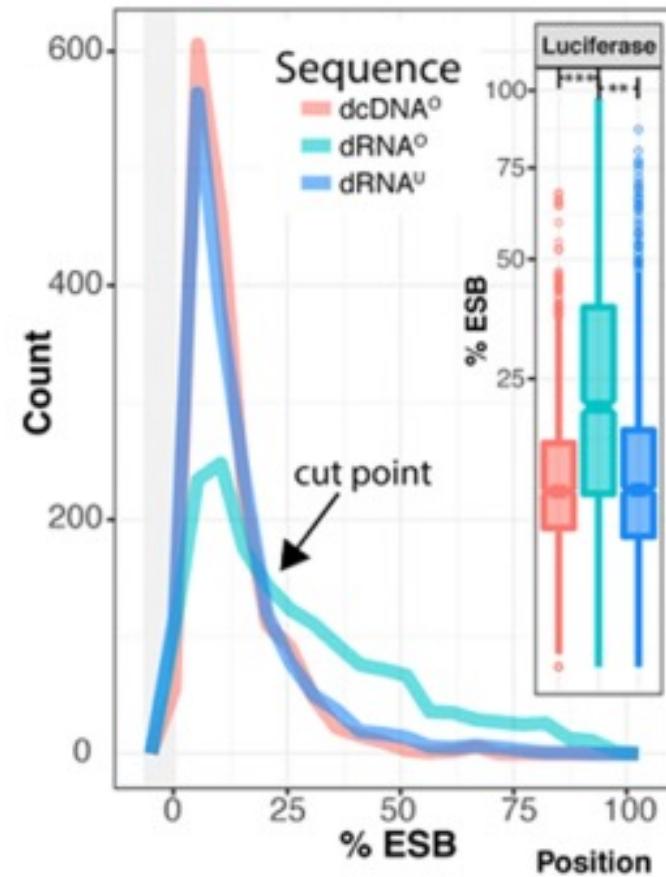
More than 15 software tools exist to detect RNA modifications from direct RNA nanopore data



EpiNano



ELIGOS

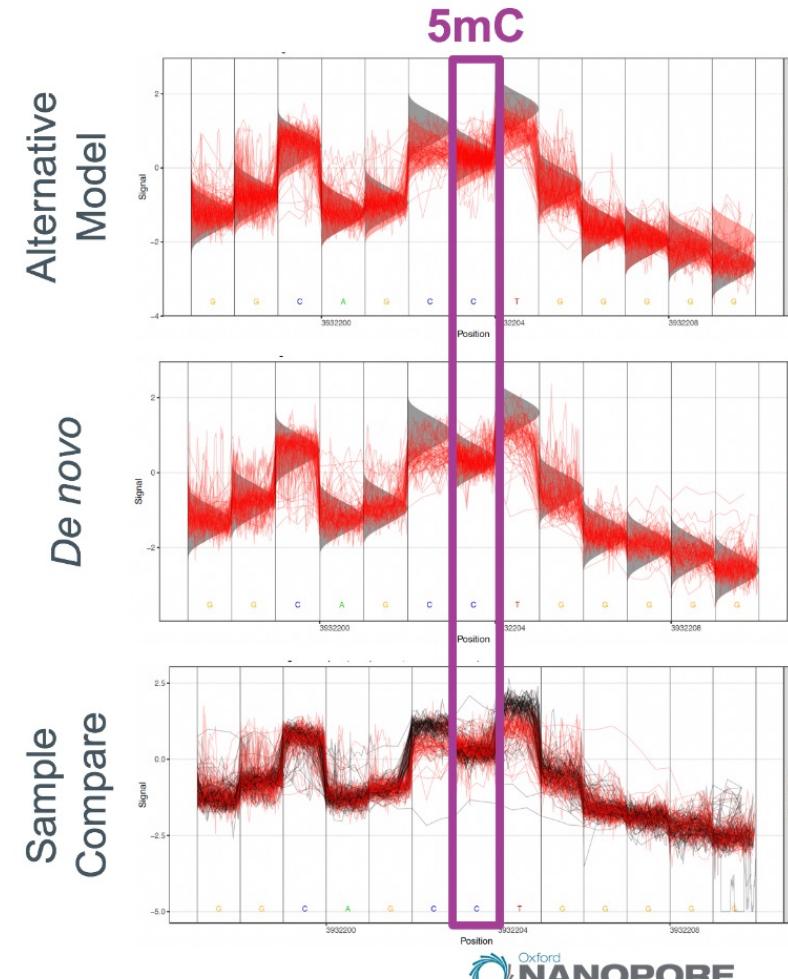


Tombo

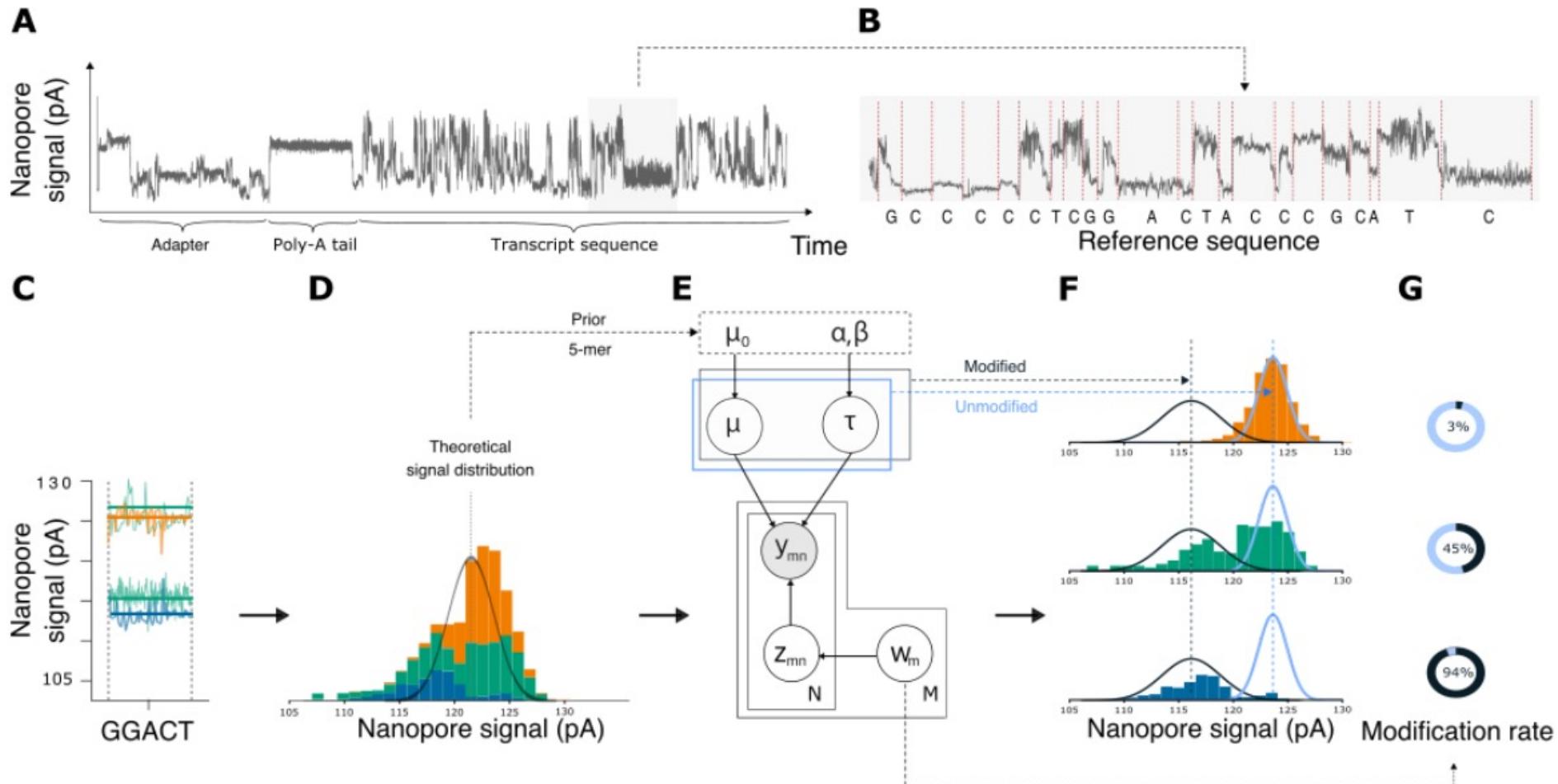
MODIFIED BASE DETECTION

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

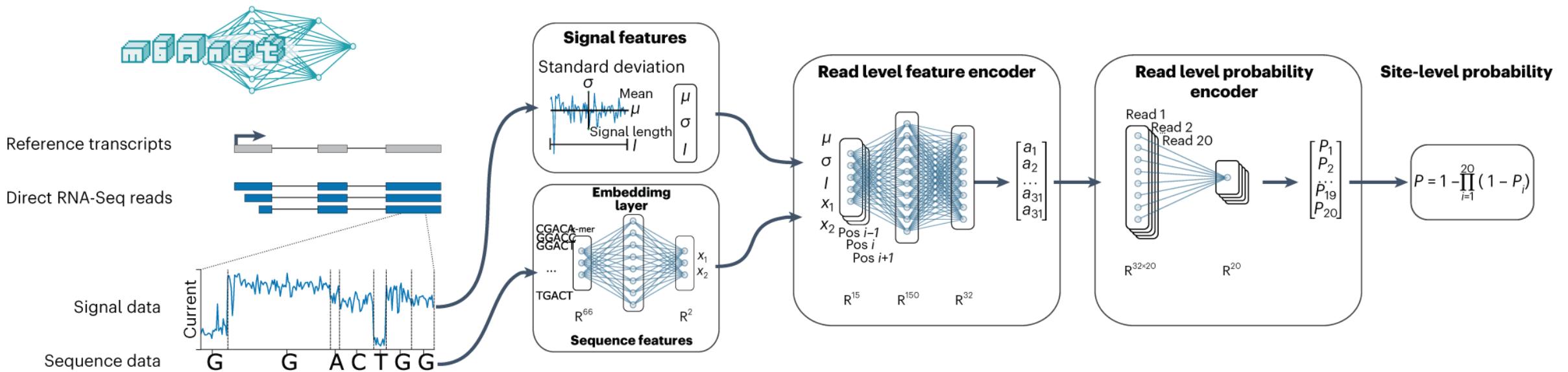
	Advantages	Disadvantages
Alternative Model	<ul style="list-style-type: none">Known alt. baseExact alt. loc.Good accuracy	<ul style="list-style-type: none">Requires alt. model estimation
<i>De novo</i>	<ul style="list-style-type: none">Apply to any sample	<ul style="list-style-type: none">High error rateInexact locationAlt. base unknown
Sample Compare	<ul style="list-style-type: none">Best AUCMost robust	<ul style="list-style-type: none">Inexact locationAlternative base unknown



Xpore



m6Anet

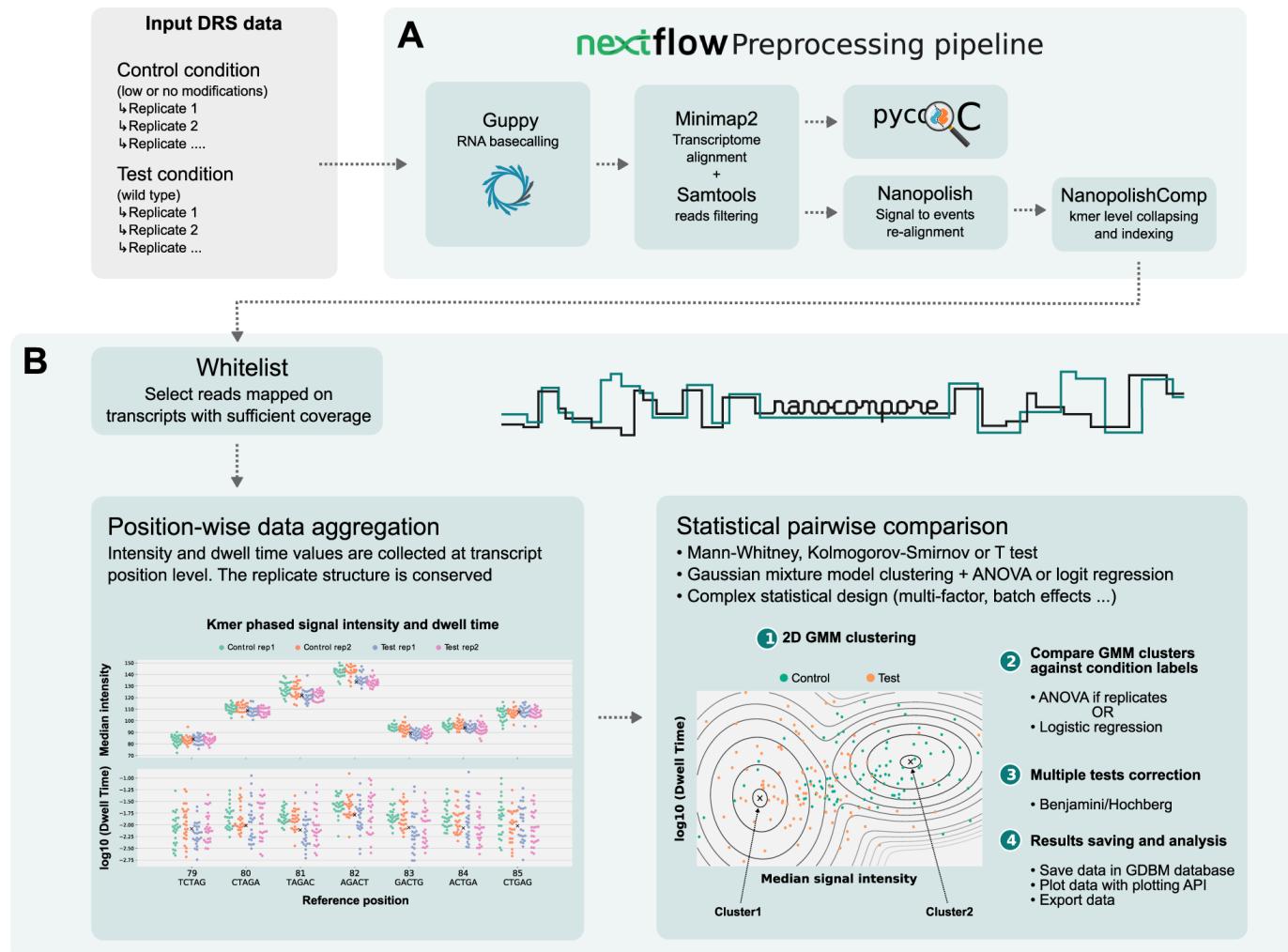


Nanocompare

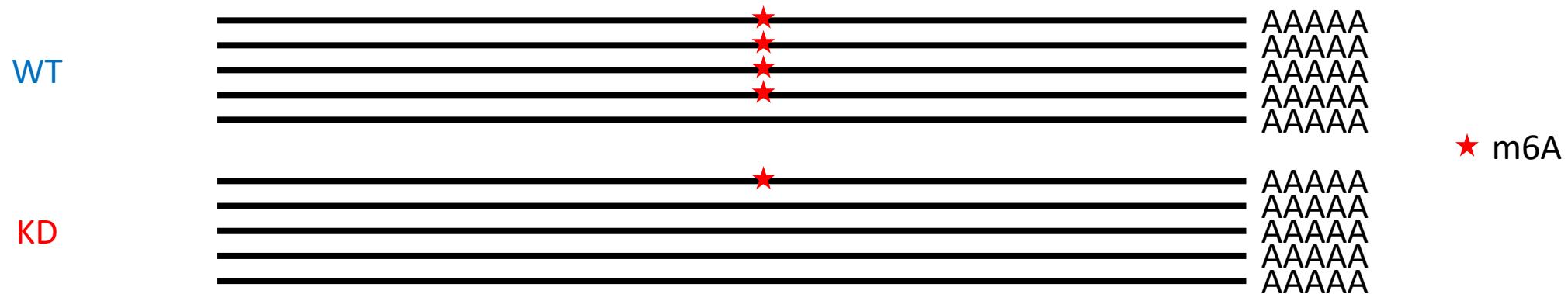


- Identifies signal-level differences between conditions (e.g. WT vs IVT/KD/KO)
- Based on Nanopolish resquiggling
- Robust and flexible statistical framework
- Takes into account biological variability
- Allows for complex statistical designs (e.g. multi-factor designs, batch effects, etc.)

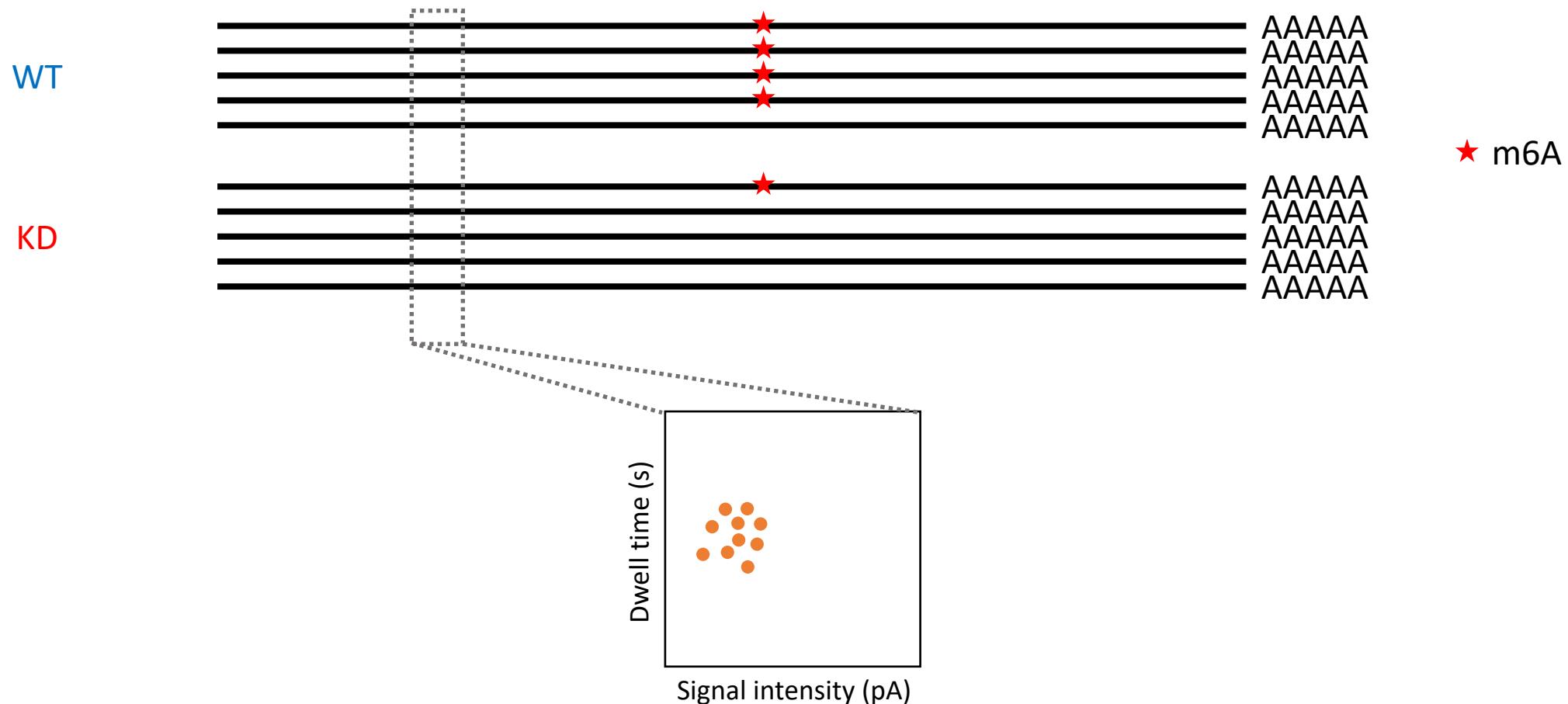
Nanocompare analysis pipeline



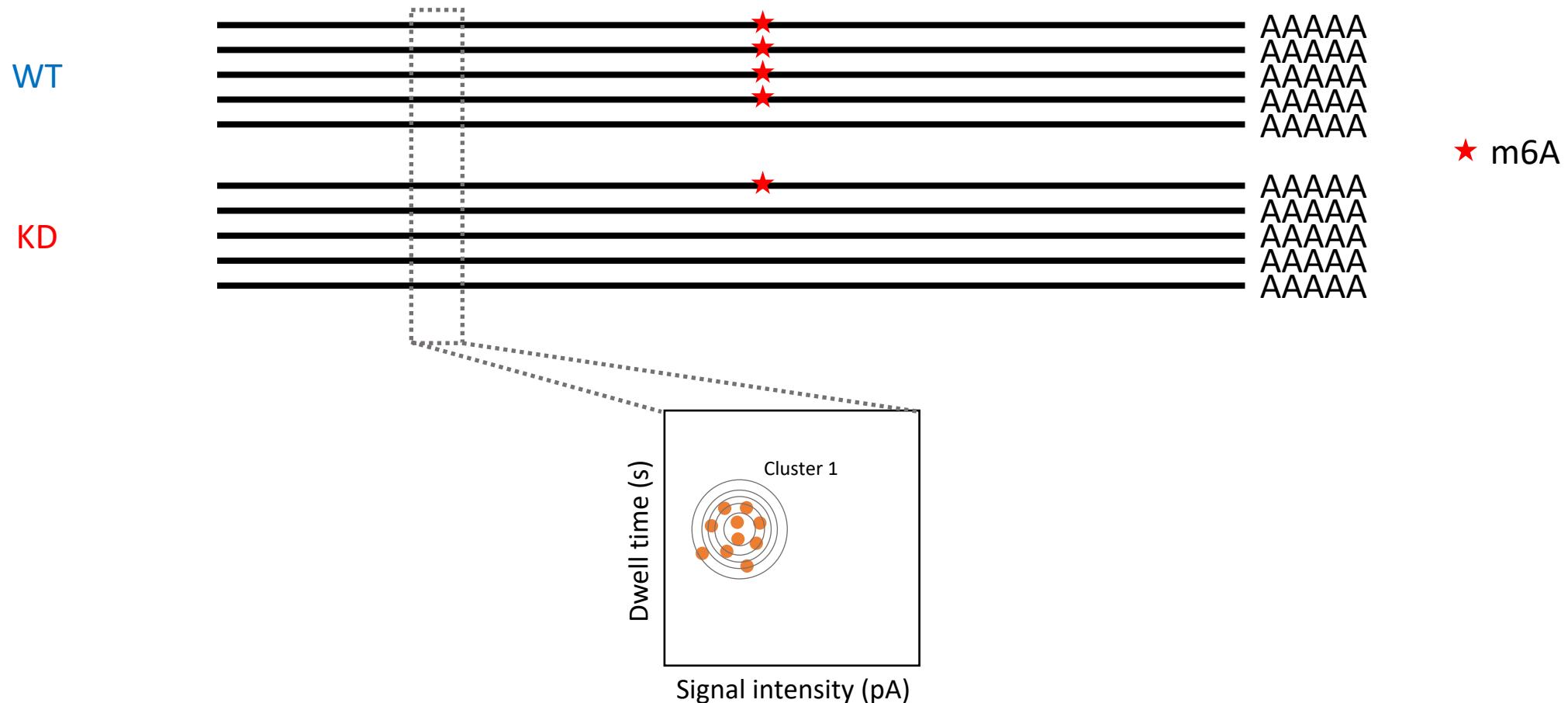
GMM for modification detection



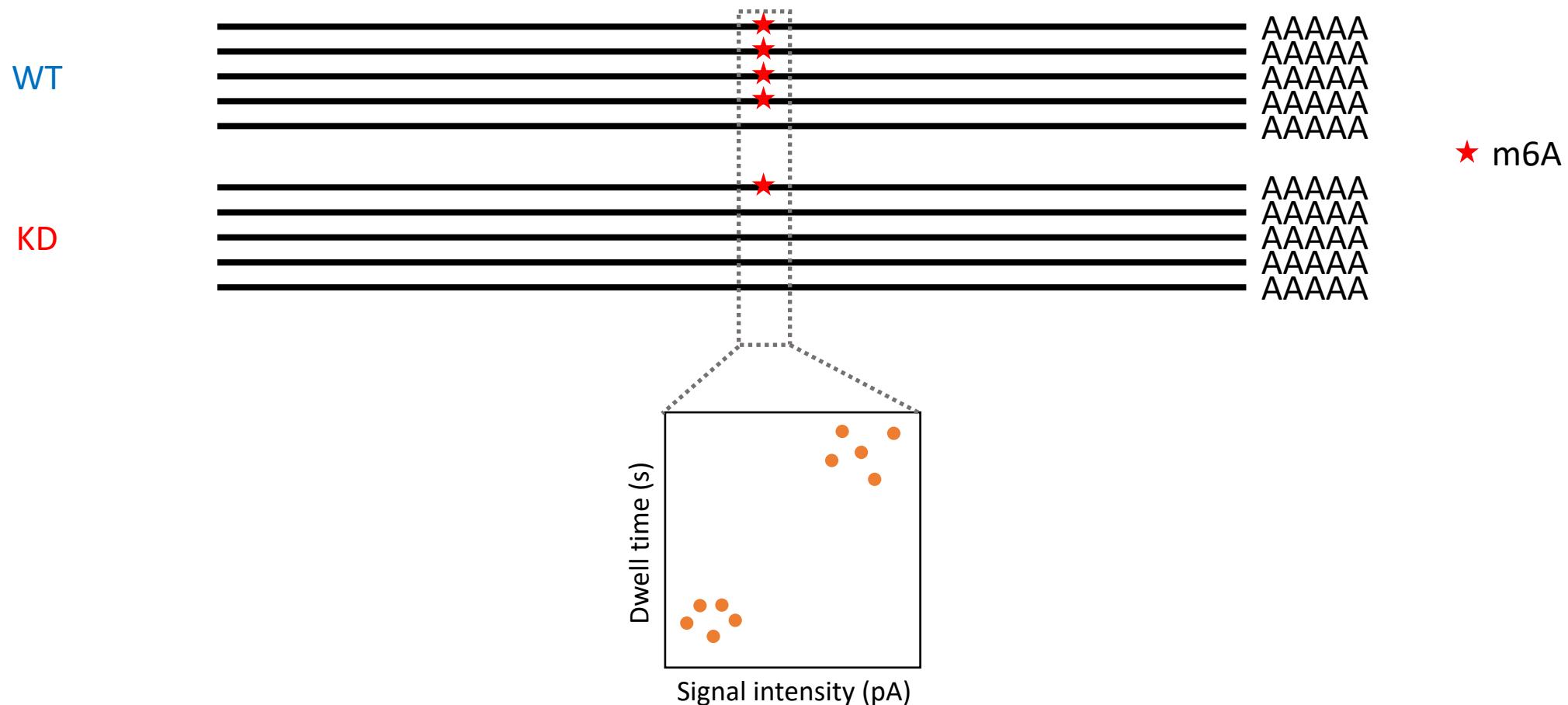
GMM for modification detection



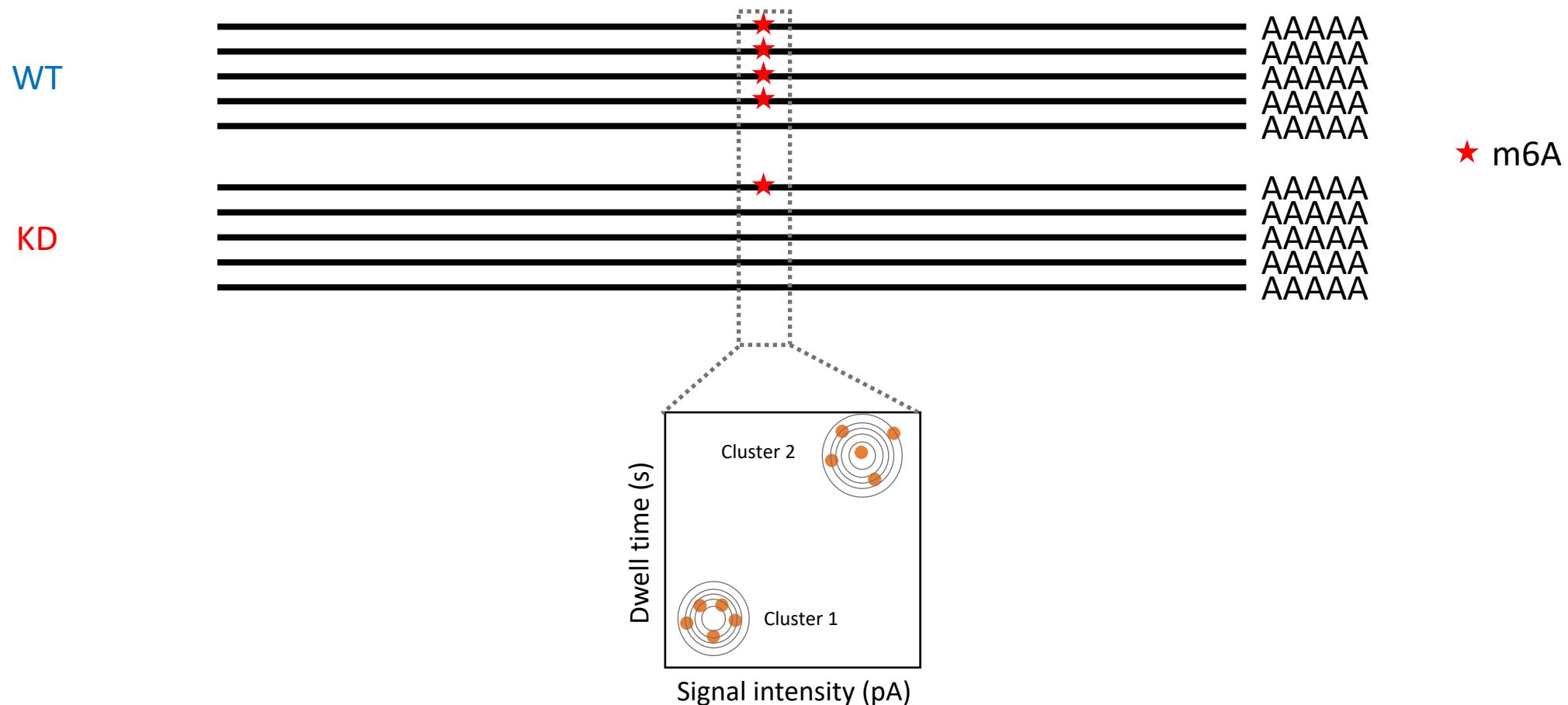
GMM for modification detection



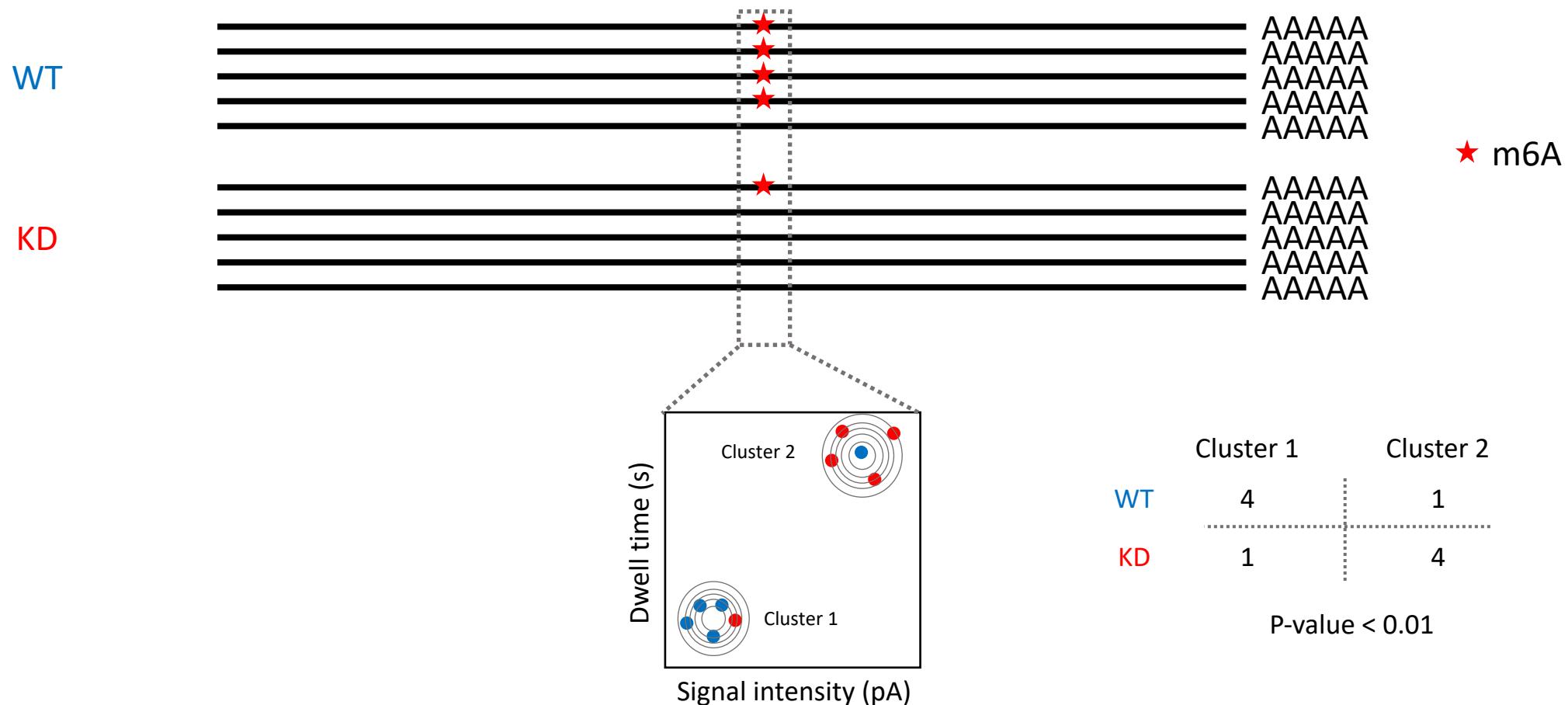
GMM for modification detection



GMM for modification detection

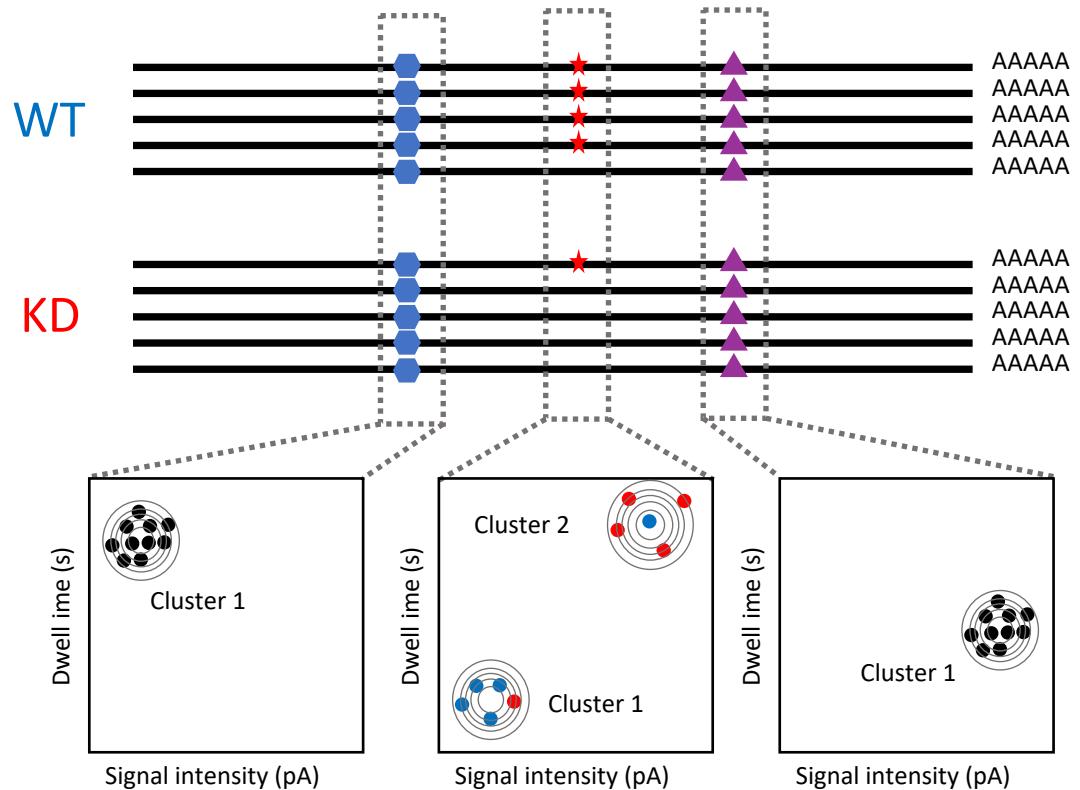


GMM for modification detection

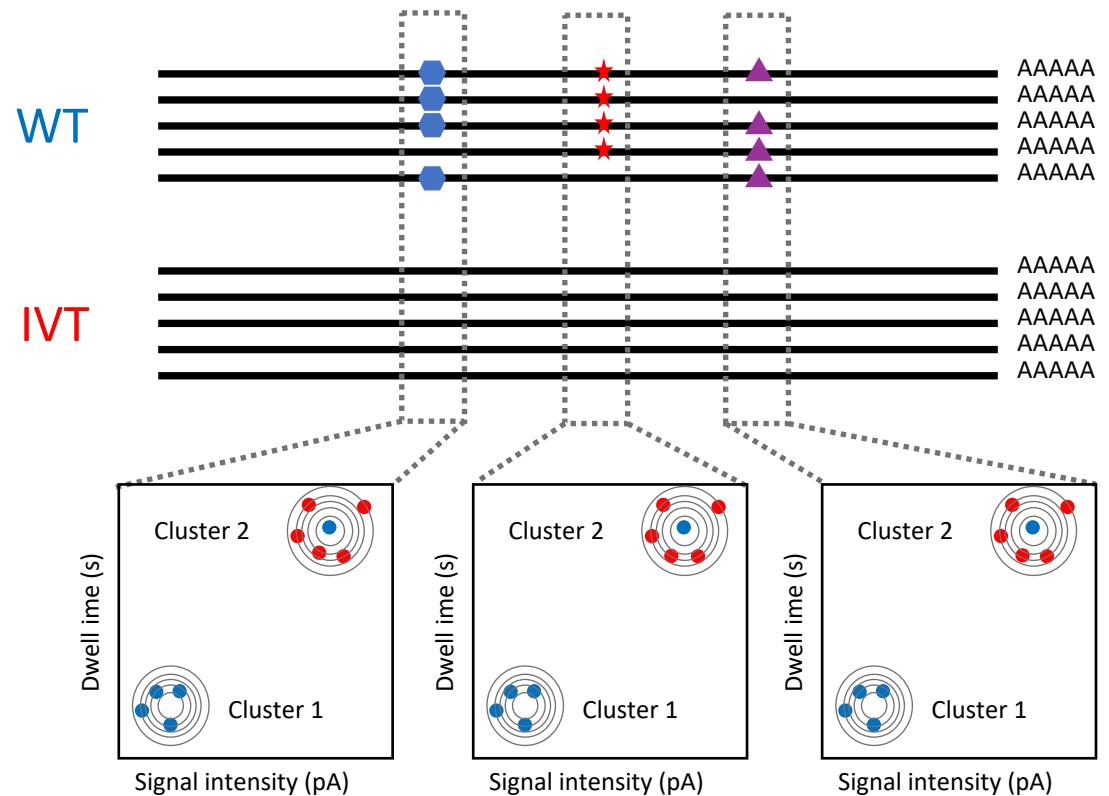


The effect of different reference samples

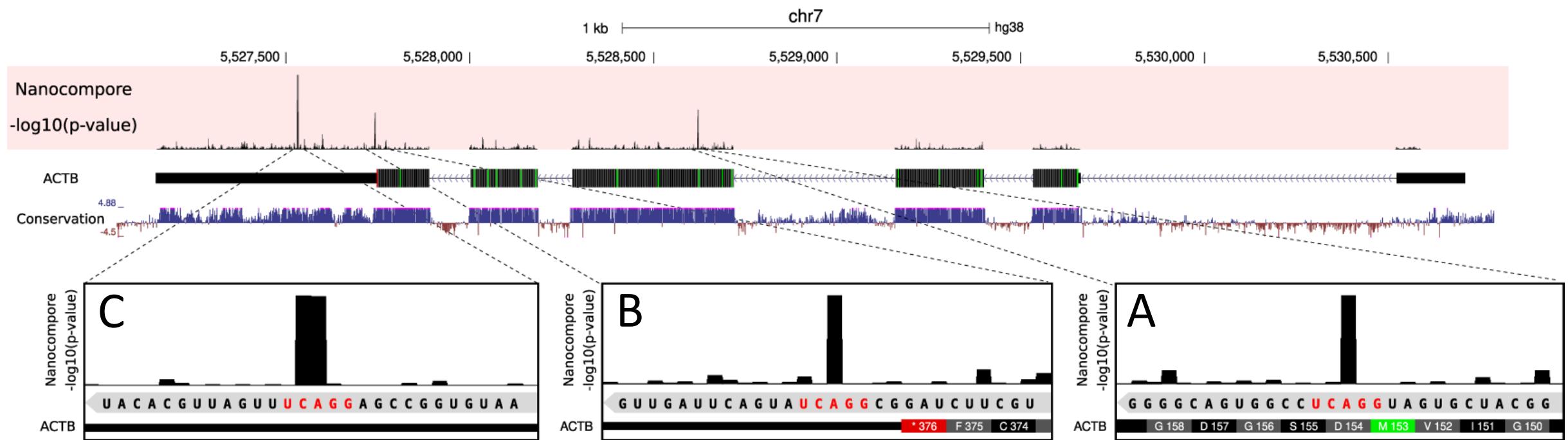
Manipulated RNA modifying enzymes



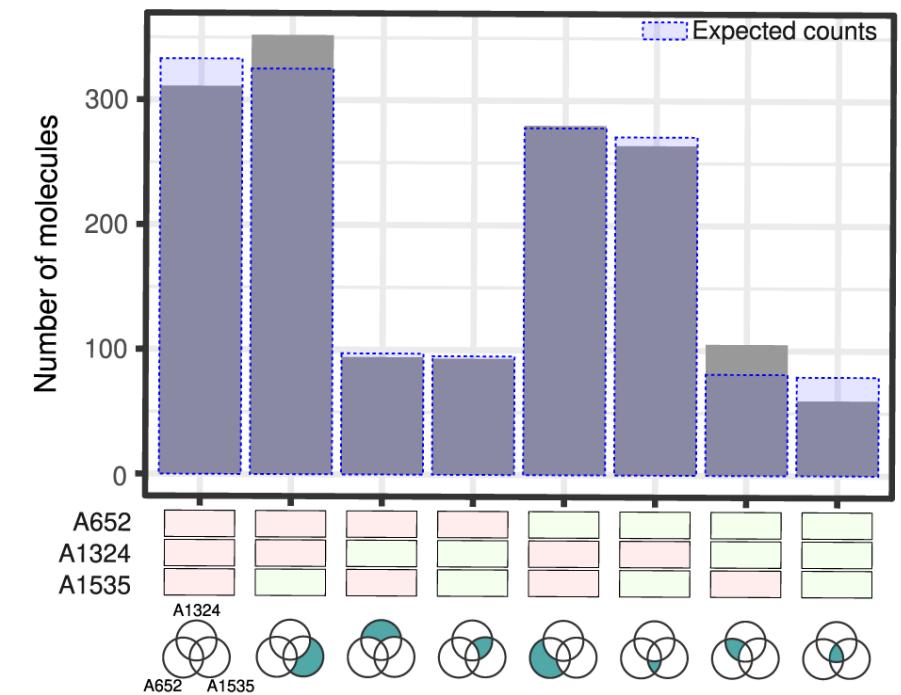
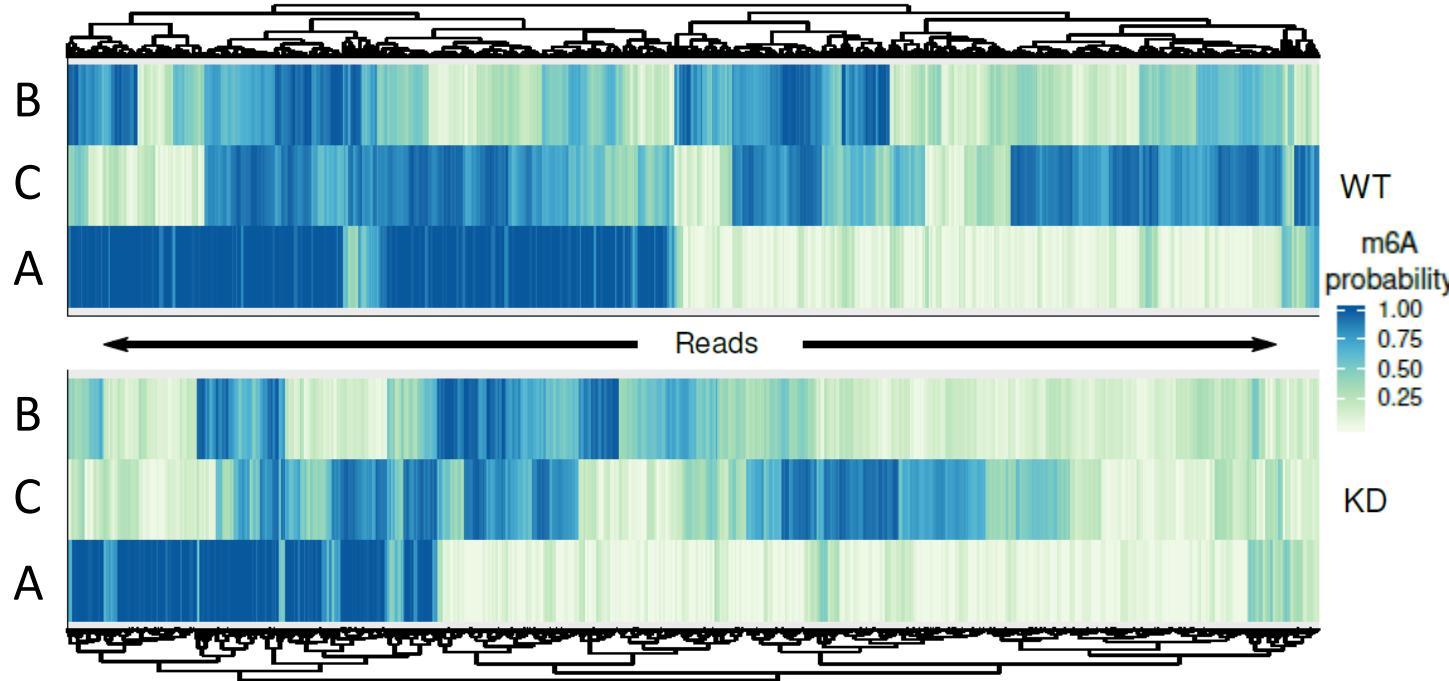
In Vitro Transcribed RNA



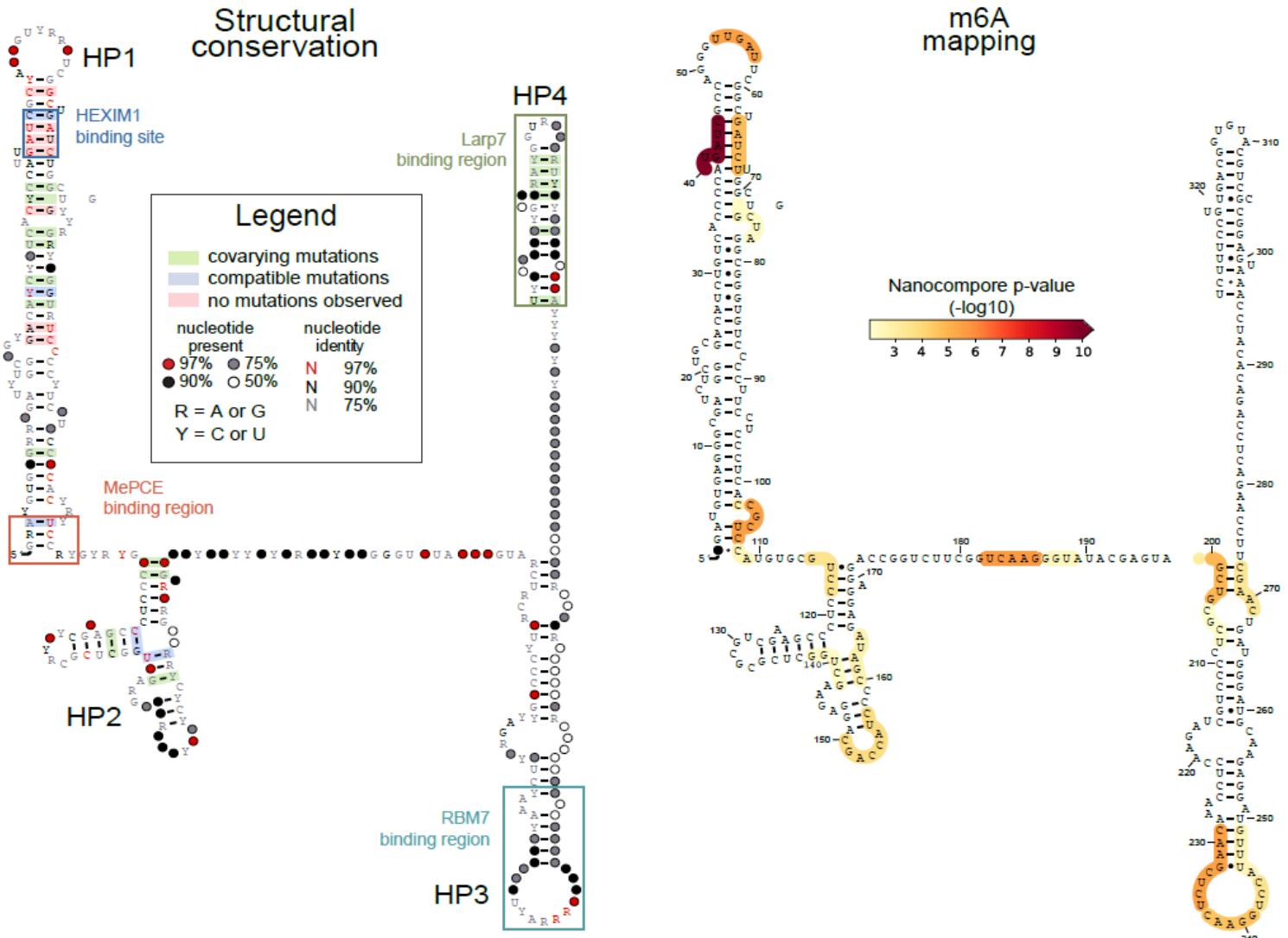
Detecting m⁶A on ACTB transcripts



Detecting m⁶A on ACTB transcripts



m6A in 7sk snRNA protein binding sites



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?

