LC 2019 LONDON CALLING

Signal level RNA modifications detection in eukaryotic ncRNAs

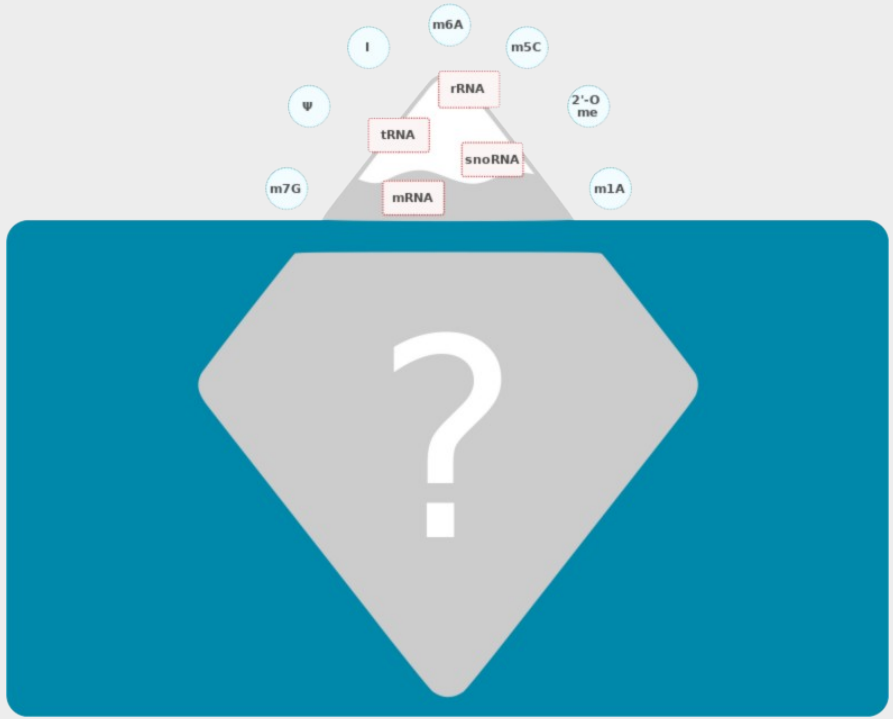
Tommaso Leonardi

Center for Genomic Science
Italian Institute of Technology

Organised by Oxford NANOPORE Technologies

The zoo of RNA post-transcriptional modifications

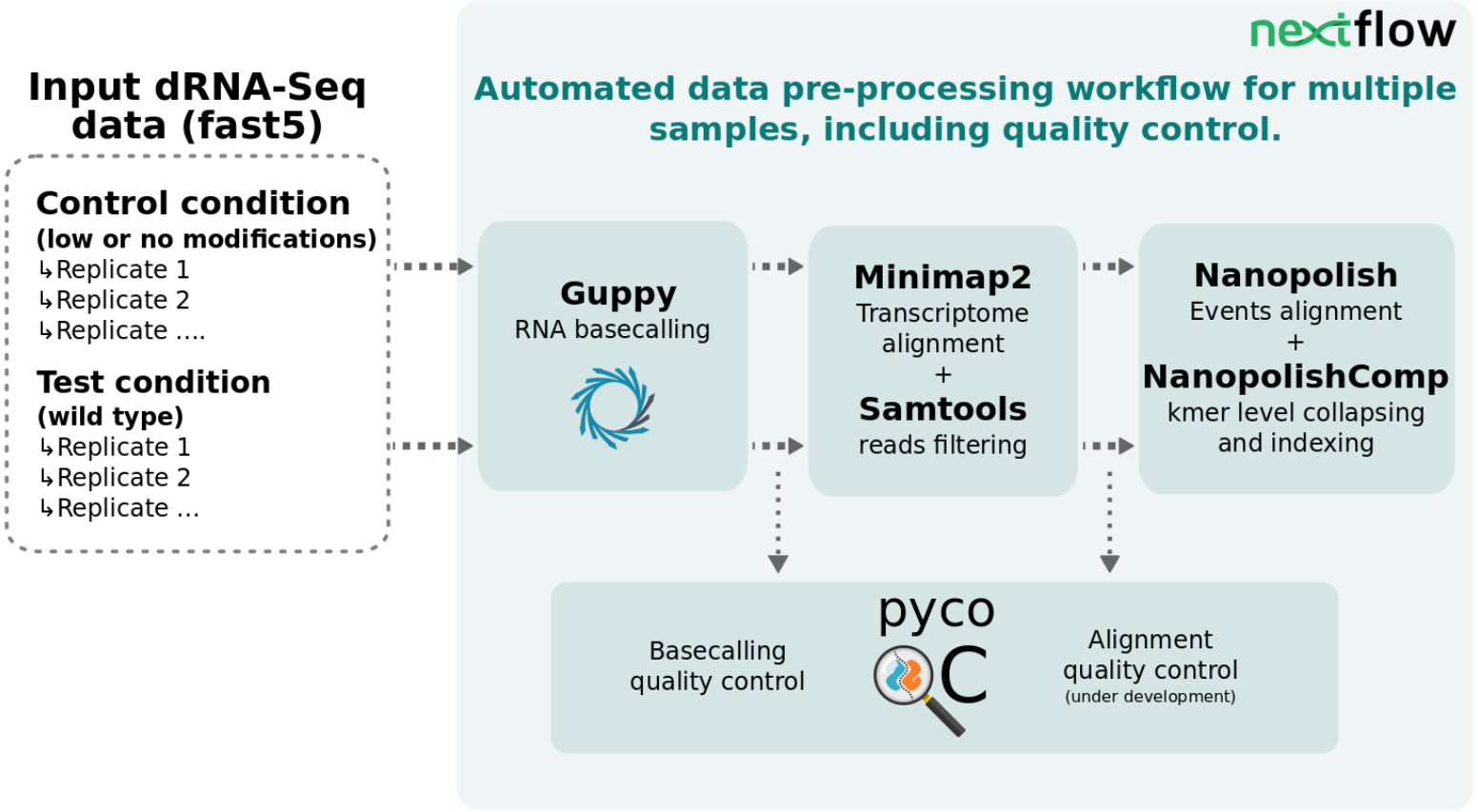
- Highly conserved feature found in archaea, bacteria and eukarya
- Impact on RNA structure and interaction properties
- More than 150 known RNA modifications



Meet 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.)

The Nanocompore analysis workflow



The Nanocompore analysis workflow

Compares 2 ONT dRNA sequencing datasets from different experimental conditions expected to have a significant impact on the epitranscriptome. Replicates aware framework with continuous integration testing.

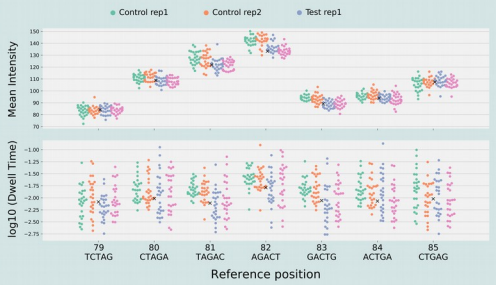
Whitelist

Select reads mapped on transcripts with sufficient coverage

Position-wise data aggregation

Intensity and dwell time values are collected at transcript position level. The replicate structure is conserved

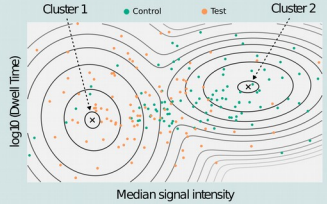
Kmer phased signal intensity and dwell time



Statistical pairwise comparison

- Mann-Whitney, Kolmogorov-Smirnov or T test
- Gaussian mixture model clustering + Anova/logit
- Complex statistical design (multi-factor, batch effects ...)

1 2D GMM clustering



2 Compare GMM clusters against condition labels

- ANOVA if replicates
  OR
- Logistic regression

3 Multiple tests correction

- Benjamini/Hochberg

4 Results saving and analysis

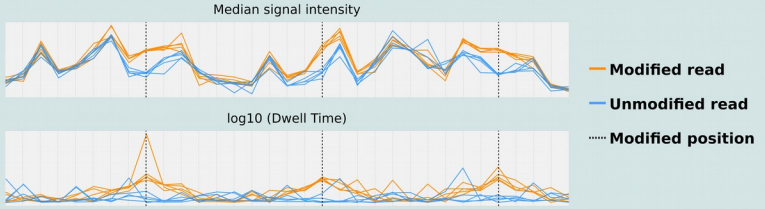
- Save data in lightweight GDBM database
- Extensive plotting API to explore data

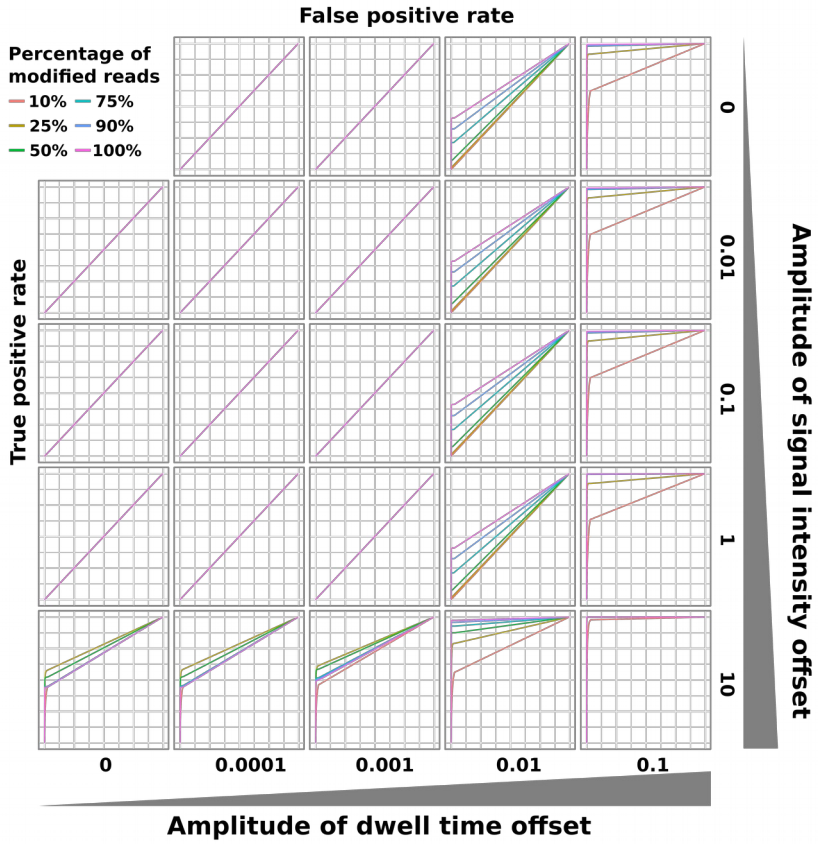
Benchmarks on *in silico* generated data

A **kmers model**
From an in vitro generated non-modified RNA dataset, find the best distribution to fit each kmers for both median intensities (*Logistic*) and dwell time (*Wald*)

B **Artificial reference**
Set of 2000 semi-random references 500 bases long. Covers all 5-mers (median: 970 times per kmer) and 99.67% of the possible 9-mers (median: 4 times per kmer)

C ***In silico* modifications simulation**
100 reads per reference using the probability density functions of the kmer models. Simulation of RNA modification effect on signal by offseting the model density function.

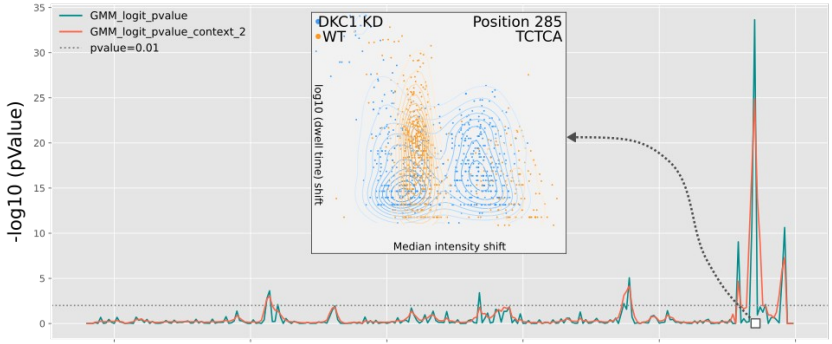




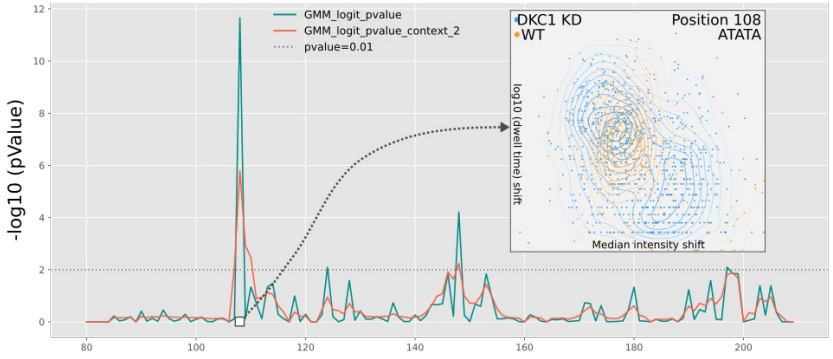
Use-case 1: RNA modifications in targeted human ncRNAs

Pseudouridine

RMRP (ENST00000602361) / MOLM-13 cells WT vs DKC1 KD (ψ)

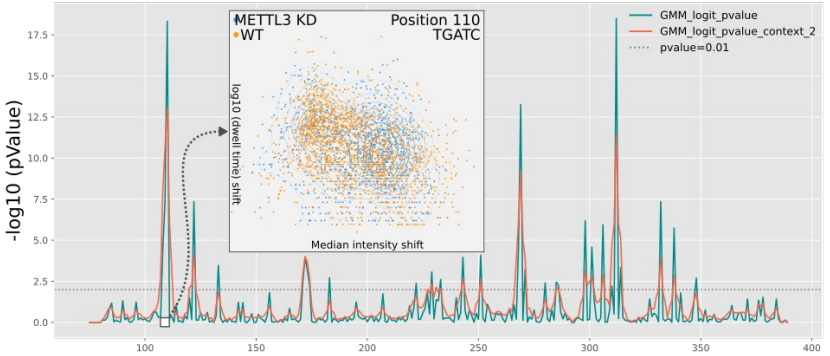


RNU2-1 (ENST00000618664) / MOLM-13 cells WT vs DKC1 KD (ψ)

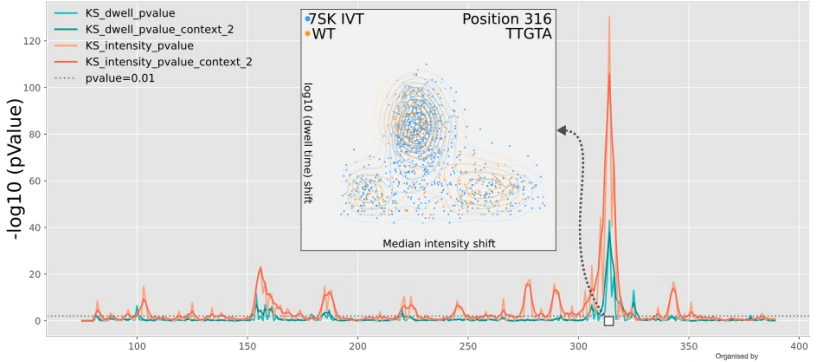


N6-methyl-adenosine

RN7SK (ENST00000636484) / MOLM-13 cells WT vs METTL3 KD (m6A)

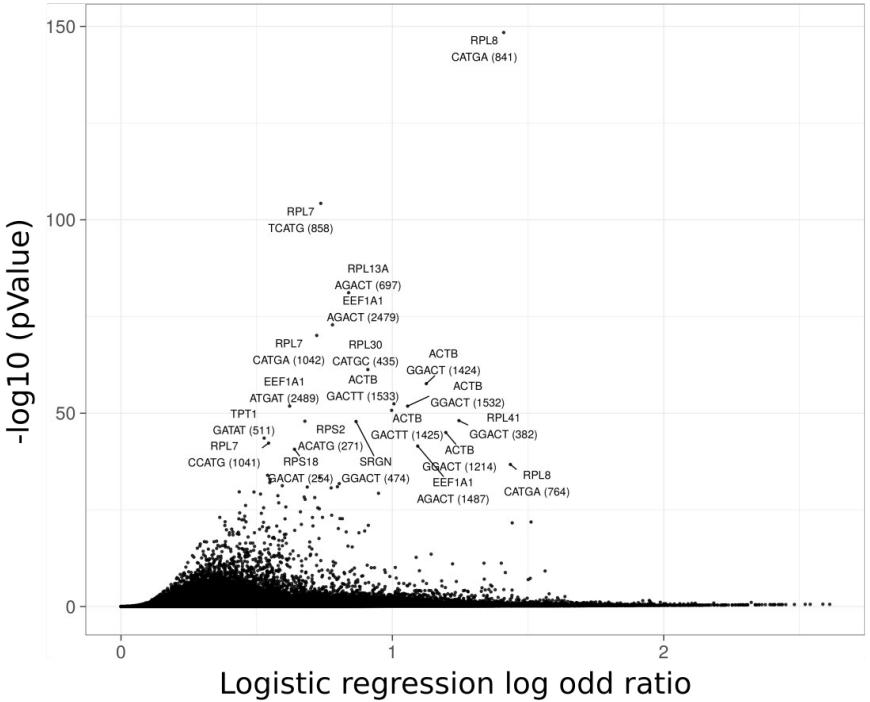


RN7SK (ENST00000636484) / MOLM-13 cells WT vs IVT 7SK (all mods)



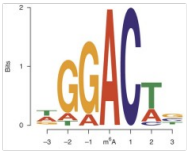
Use-case 2: m6A in a human polyA transcriptome

Top candidate kmer sites found by Nanocompore



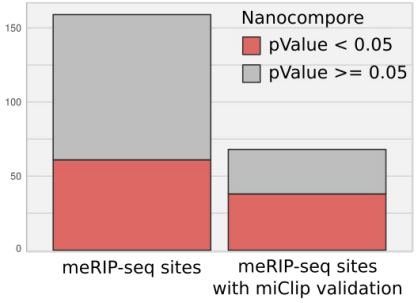
De novo motif enrichment analysis in top 100 hits

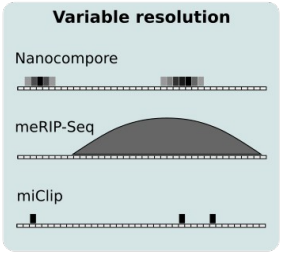
Rank	Motif	P-value	log P-pvalue	% of Targets
1		1e-2	-5.971e+00	22.45%
2		1e0	-1.172e-01	6.12%



Dominissini et al.
Nat Prot vol 8, (2013)

Orthogonal METTL3-dependent m6A mapping methods





Acknowledgements

Adrien Leger
EMBL-EBI

Paulo Amaral
STORM Therapeutics Ltd.

Birney lab, EMBL-EBI
Adrien Leger
Tomas Fitzgerald

Kouzarides lab, University of Cambridge
Paulo Amaral
Luca Pandolfini
Harvey Che
Helena Santos Rosa

Enright Lab, University of Cambridge