Direct RNA sequencing of human transcripts using the Oxford Nanopore sequencing platform

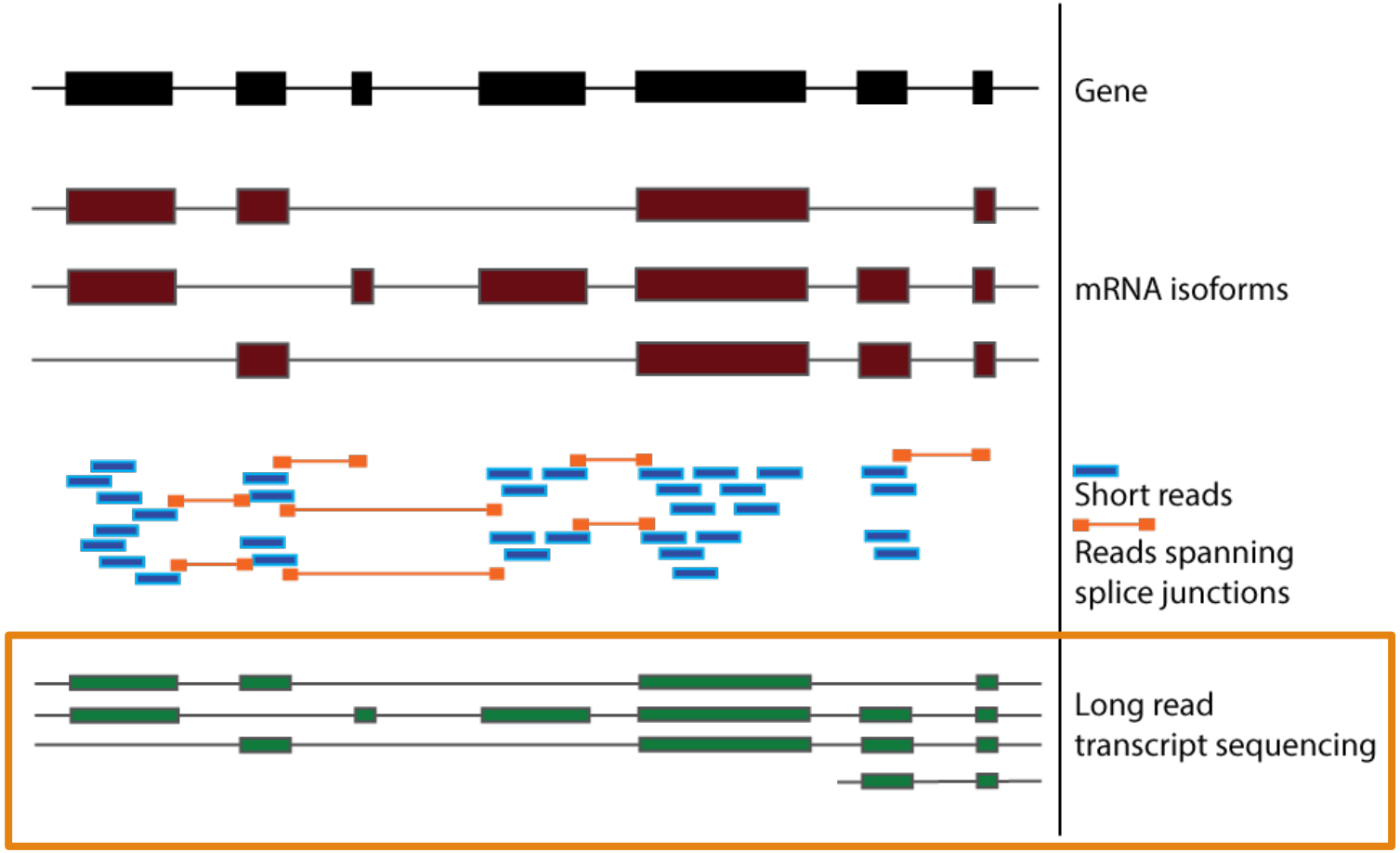
Rachael Workman

The Nanopore RNA Consortium

AGBT 2018

Direct RNA sequencing opens new frontiers for transcriptome exploration

- Analyse isoforms directly
- Poly-A length assessment
- RNA modifications
- PCR-free system



Nanopore RNA consortium

- **Six participating Universities**
  - Johns Hopkins University
  - University of Birmingham
  - University of California Santa Cruz
  - University of British Columbia
  - Ontario Institute for Cancer Research
  - University of Nottingham
- **Isolated native poly-A RNA from GM12878 CEPH cell line, sequenced direct RNA and amplified cDNA**
- **Spiked-in synthetic RNA molecules***


*SIRVs: Lexogen SIRV Set 3, ERCCs + 69 SIRV Isoforms

How much data did we generate?

- ≈13 million Direct RNA sequences, 30 flow cells
- >24 million cDNA sequences, 12 flow cells

- https://github.com/nanopore-wgs-consortium/NA12878/

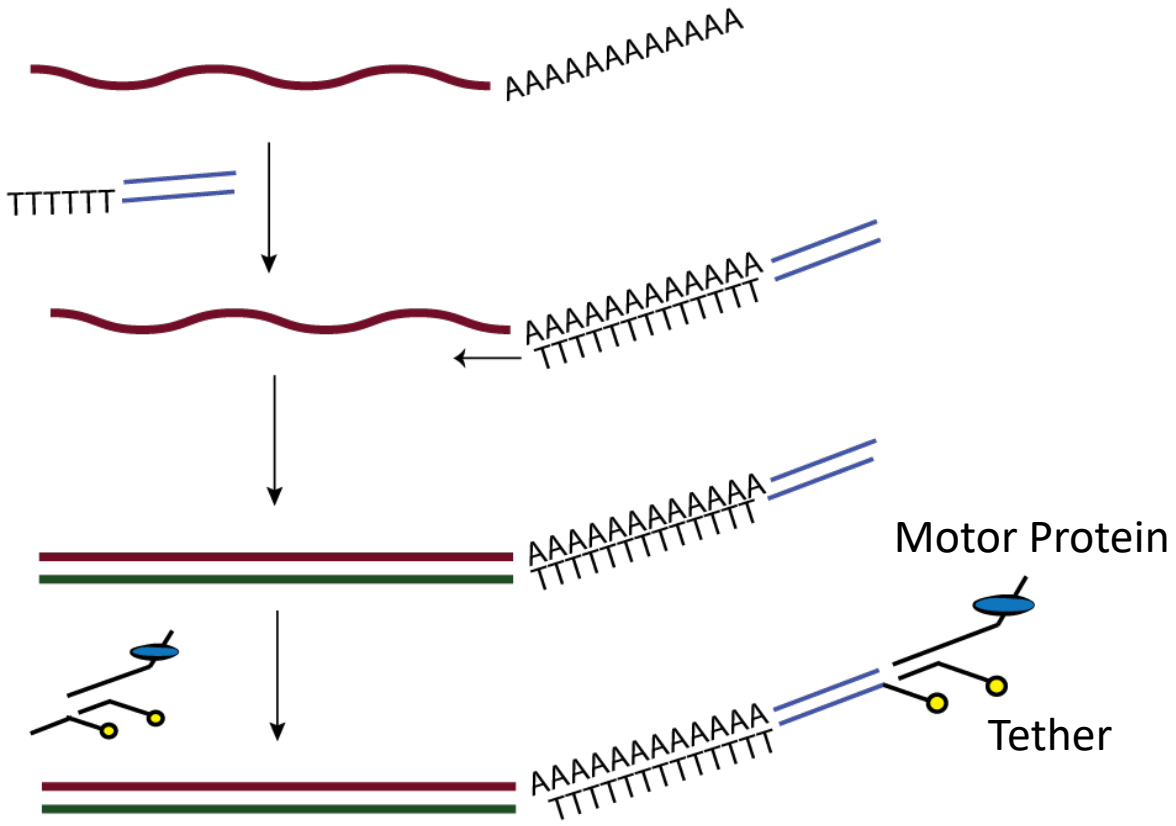
Long read RNA can now be sequenced directly with ONT

PolyA+ RNA captured

Splint poly-T adapter ligation

Reverse transcription (optional)

Sequencing adapter ligation



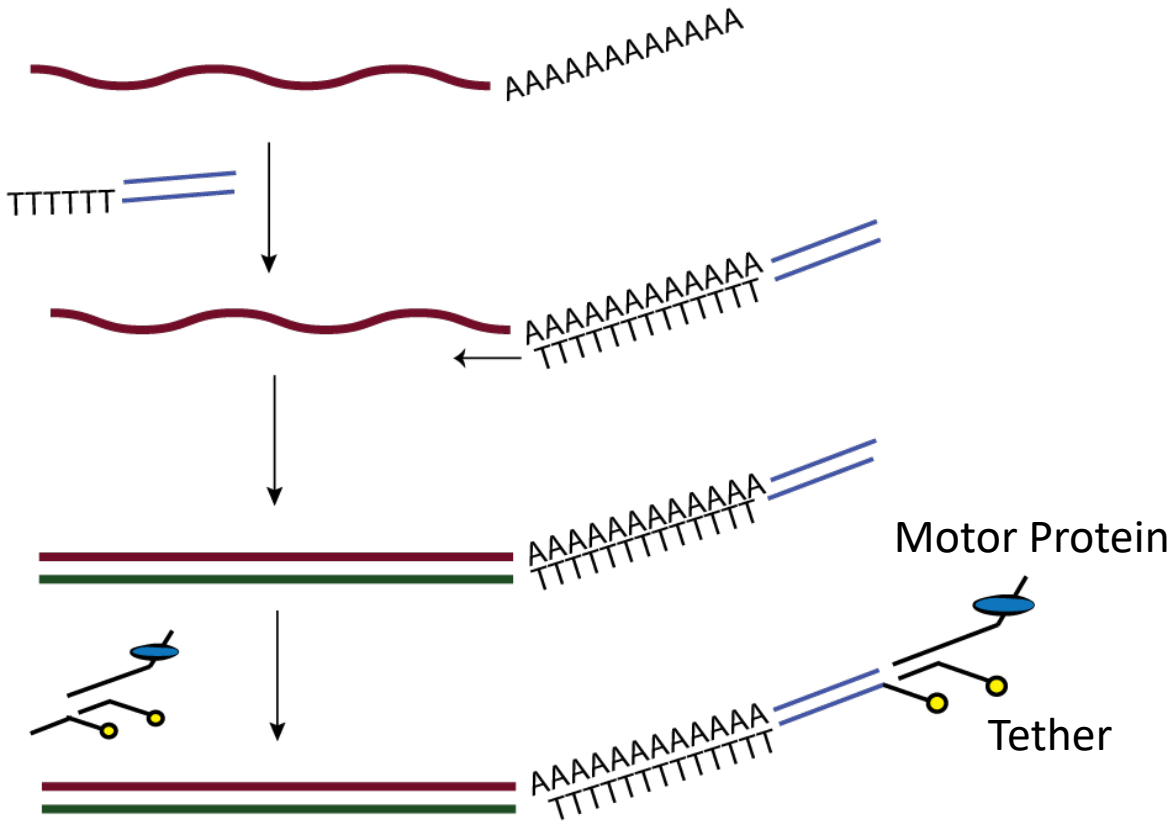
Long read RNA can now be sequenced directly with ONT

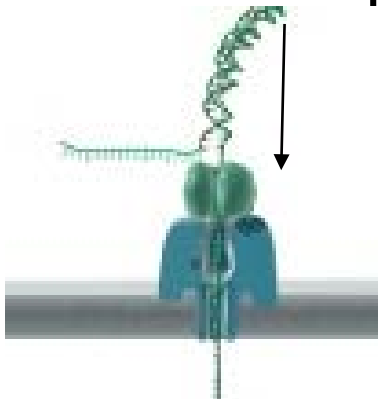
PolyA+ RNA captured

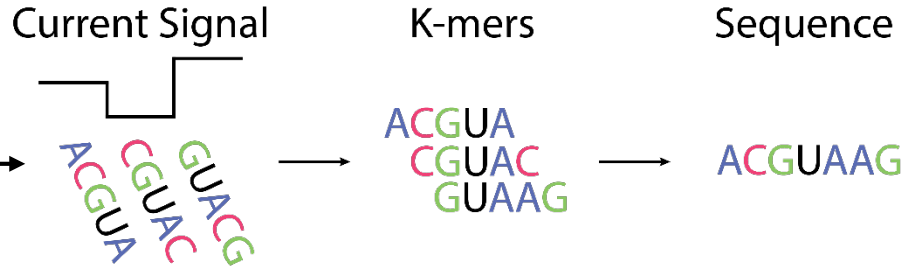
Splint poly-T adapter ligation

Reverse transcription (optional)

Sequencing adapter ligation

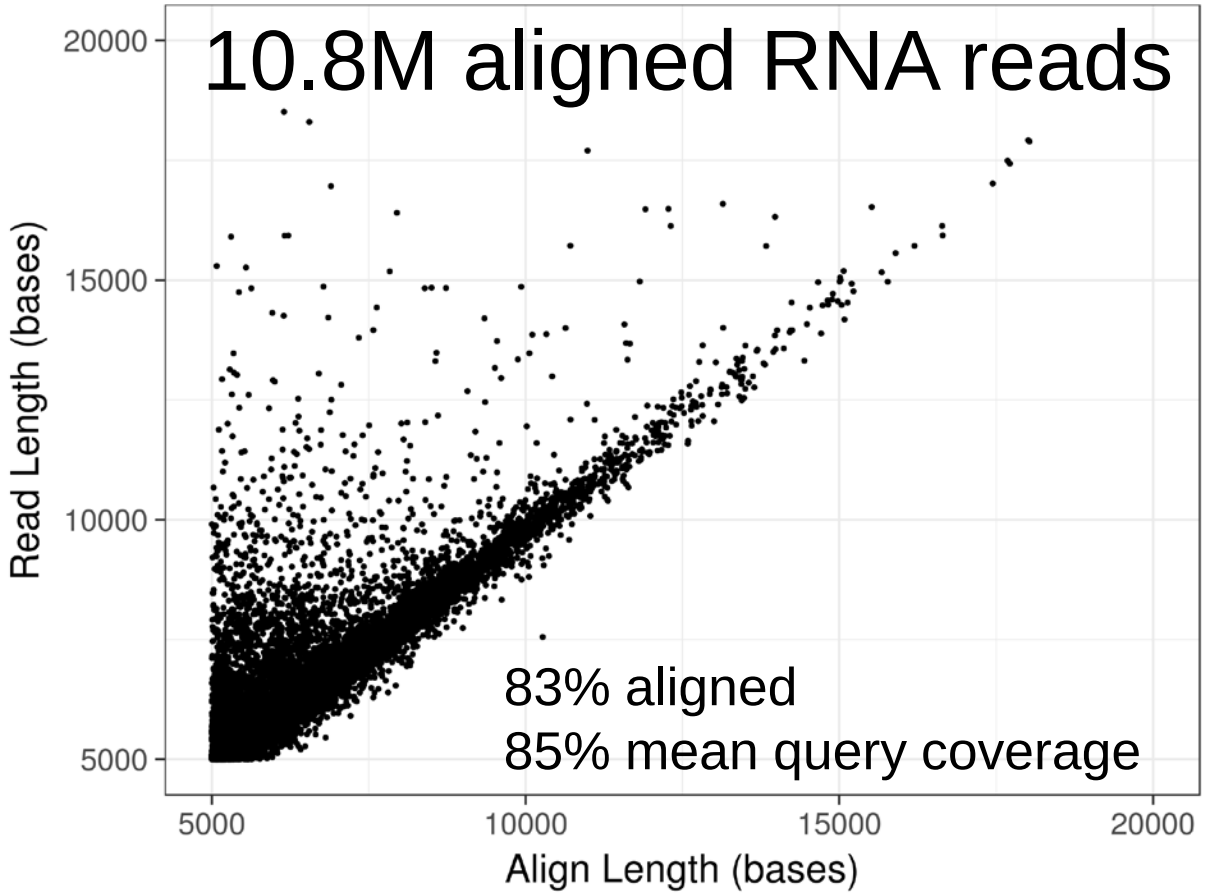


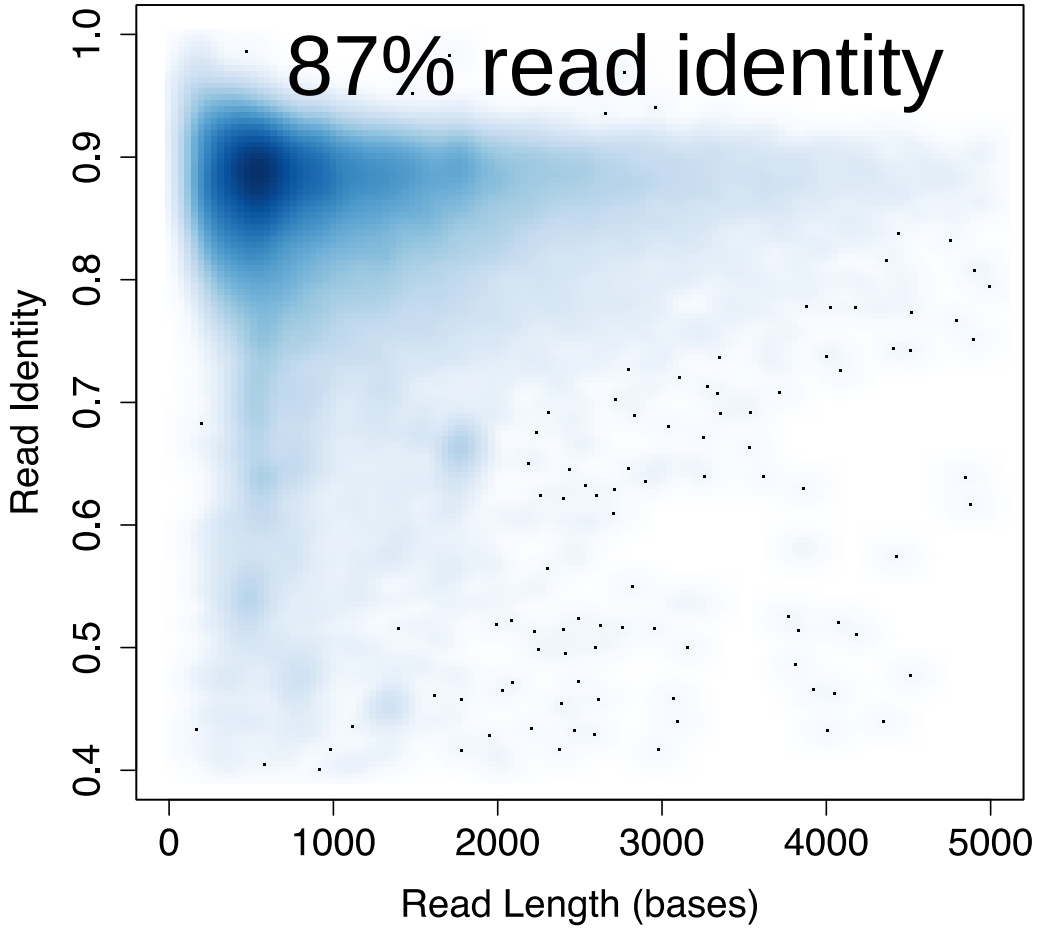




GM12878 DIRECT RNA DATA DESCRIPTION

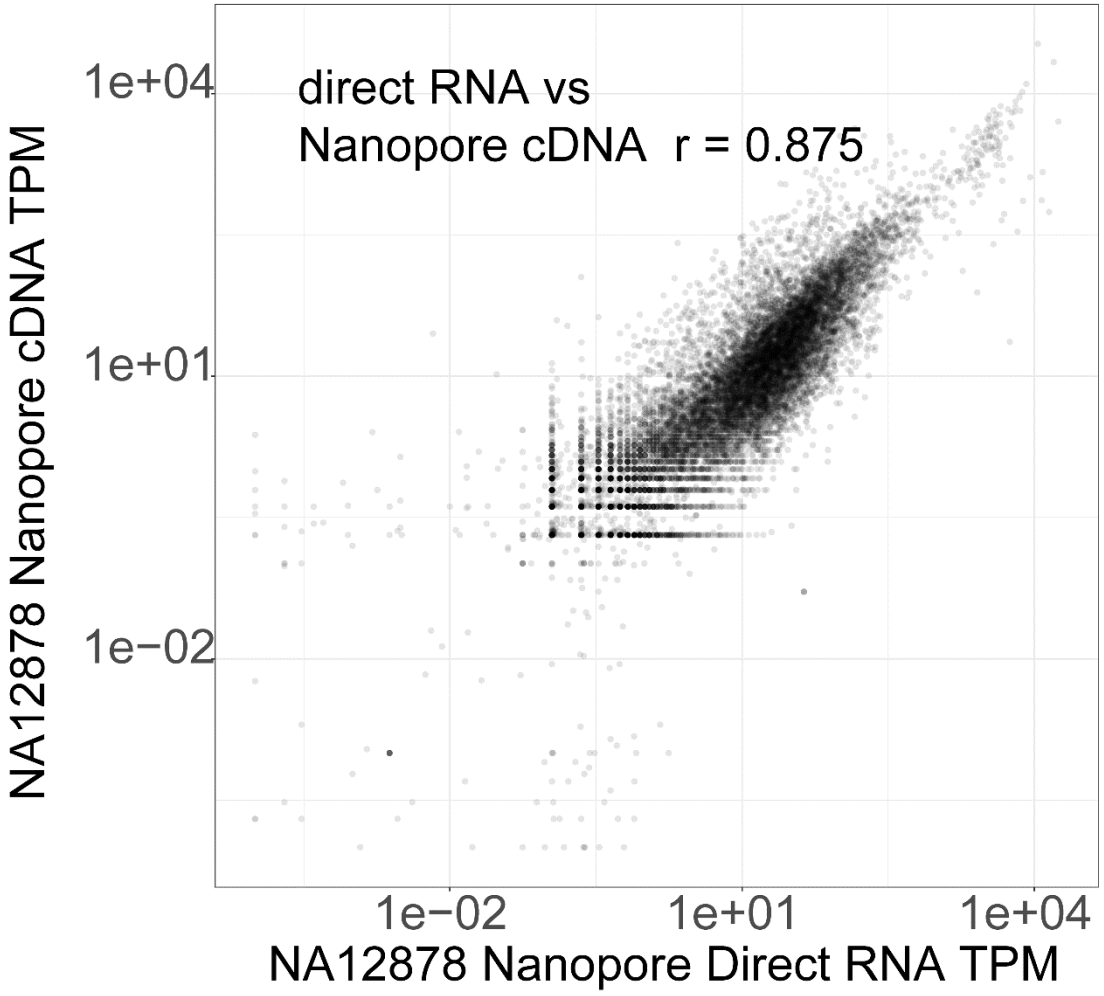
dRNA reads show good alignment to GENCODE reference

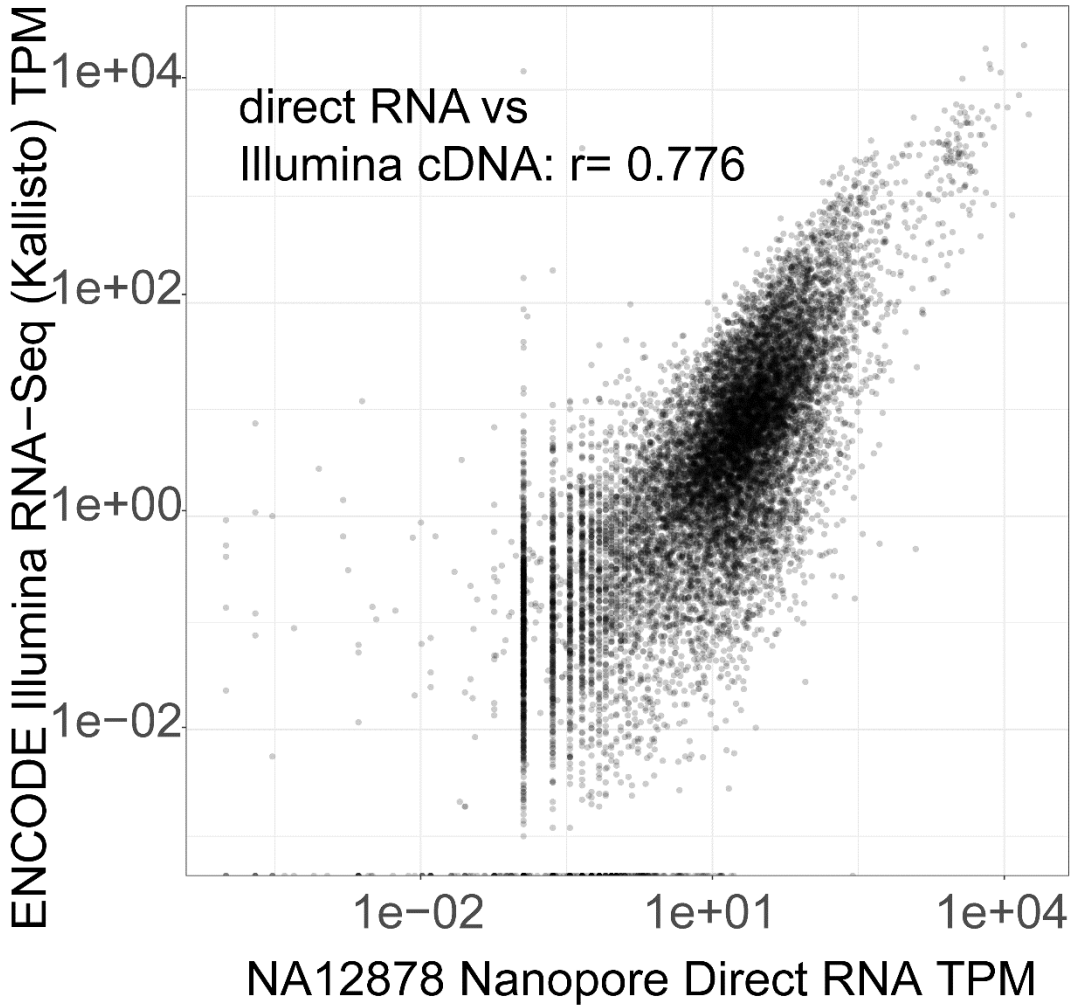




Alignment to GENCODE V27 dataset performed using Minimap2: https://github.com/lh3/minimap2

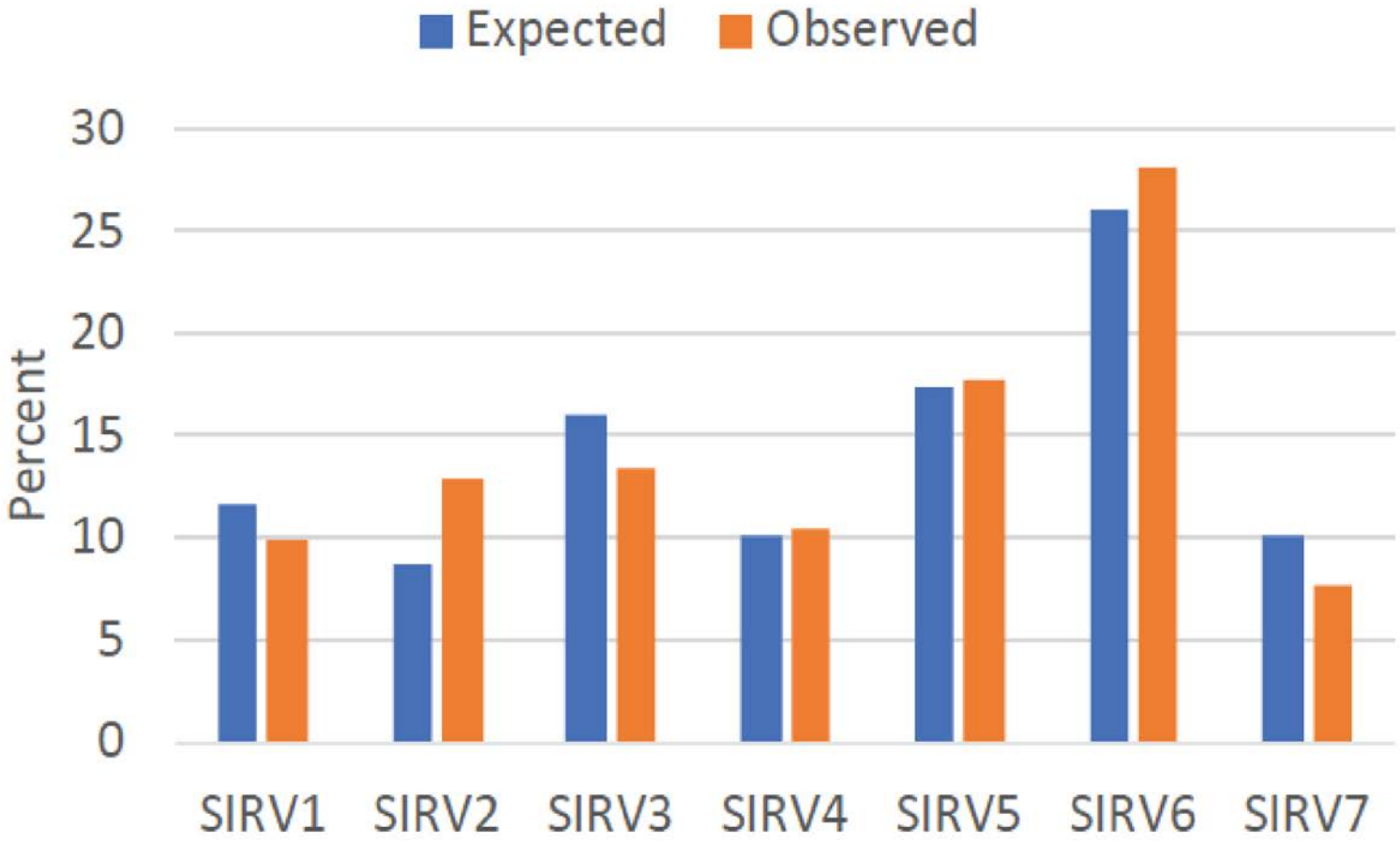
Correlation in gene-level abundance between RNA vs. ONT cDNA, Illumina



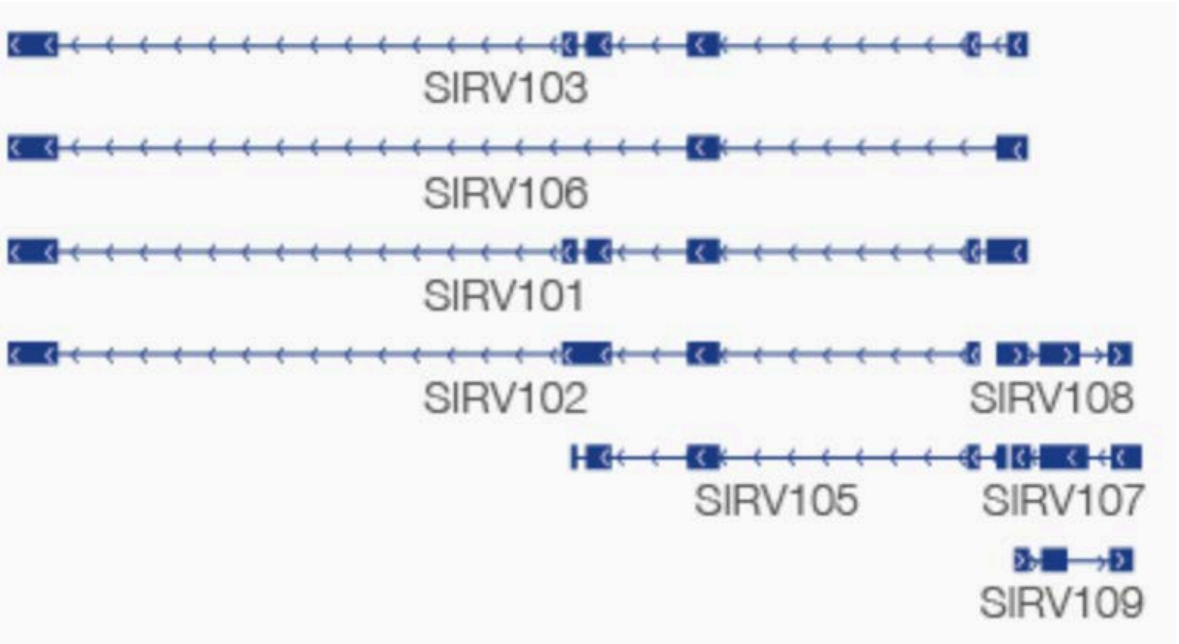


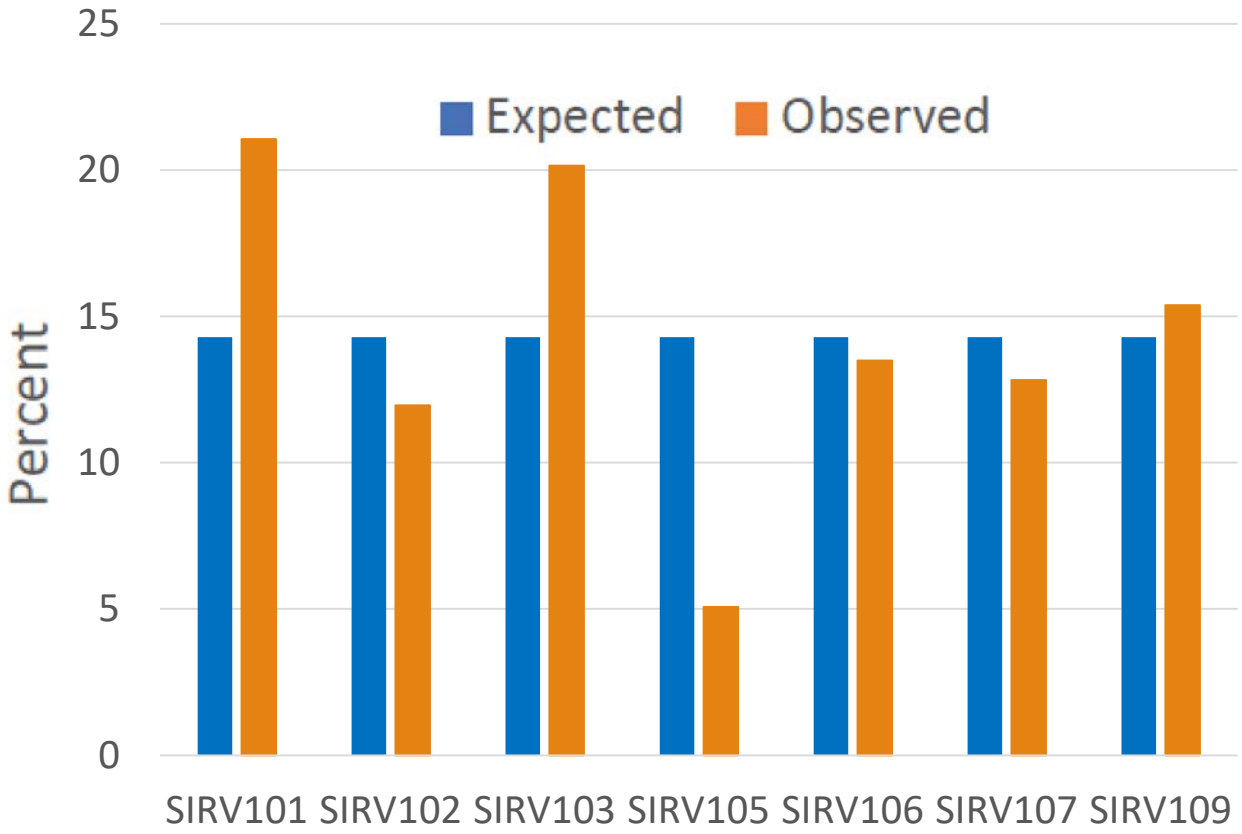
ENCODE Illumina dataset, Wold Lab: https://www.encodeproject.org/experiments/ENCSR000AEG/

SIRV gene level quantification tracks expected input well



Quantification complicated at isoform level





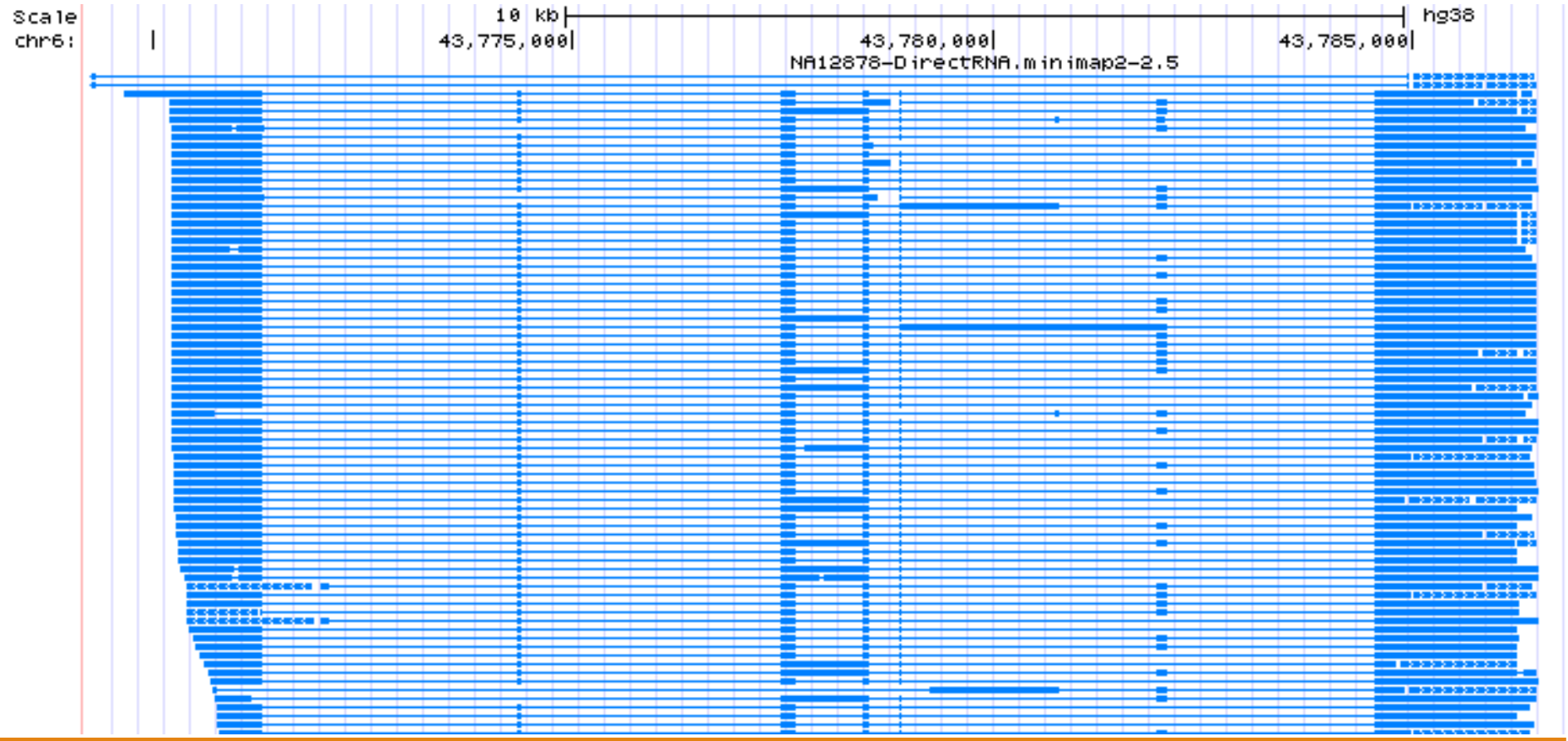
Isoform level quantification complicated by read error, multi-mapping, 5’ degradation

UNIQUE BENEFITS OF DIRECT RNA SEQUENCING

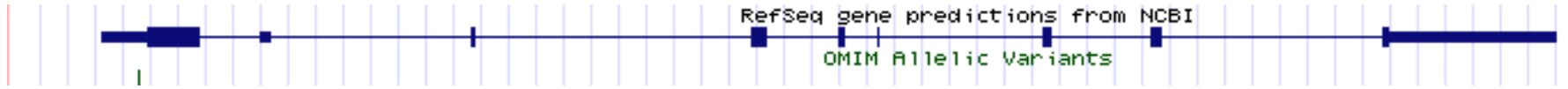
1. LONG READS PROVIDE EXON CONNECTIVITY WITHOUT PCR BIAS

Long RNA reads provide exon connectivity

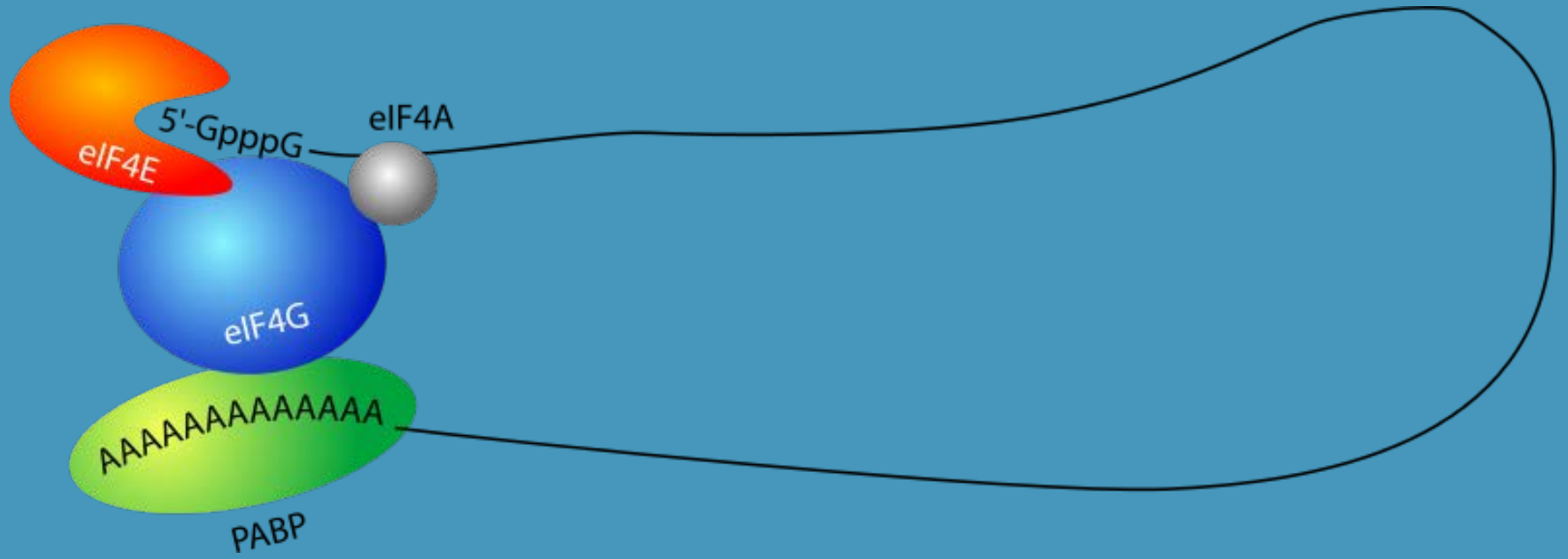
Example RNA reads capture several isoforms of VEGFA gene



Vascular endothelial growth factor A: stimulates blood vessel formation



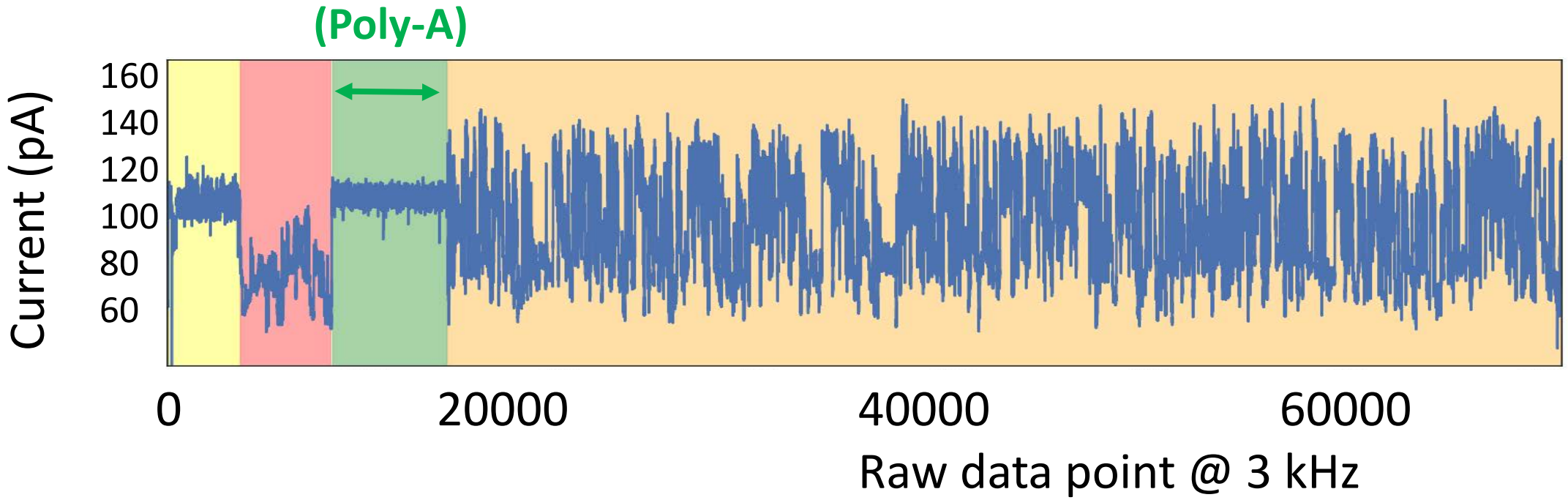
Custom track for UCSC browser, Courtesy of Heng Li: https://goo.gl/n9PFqN



2. POLYA TAIL LENGTHS CAN BE ESTIMATED

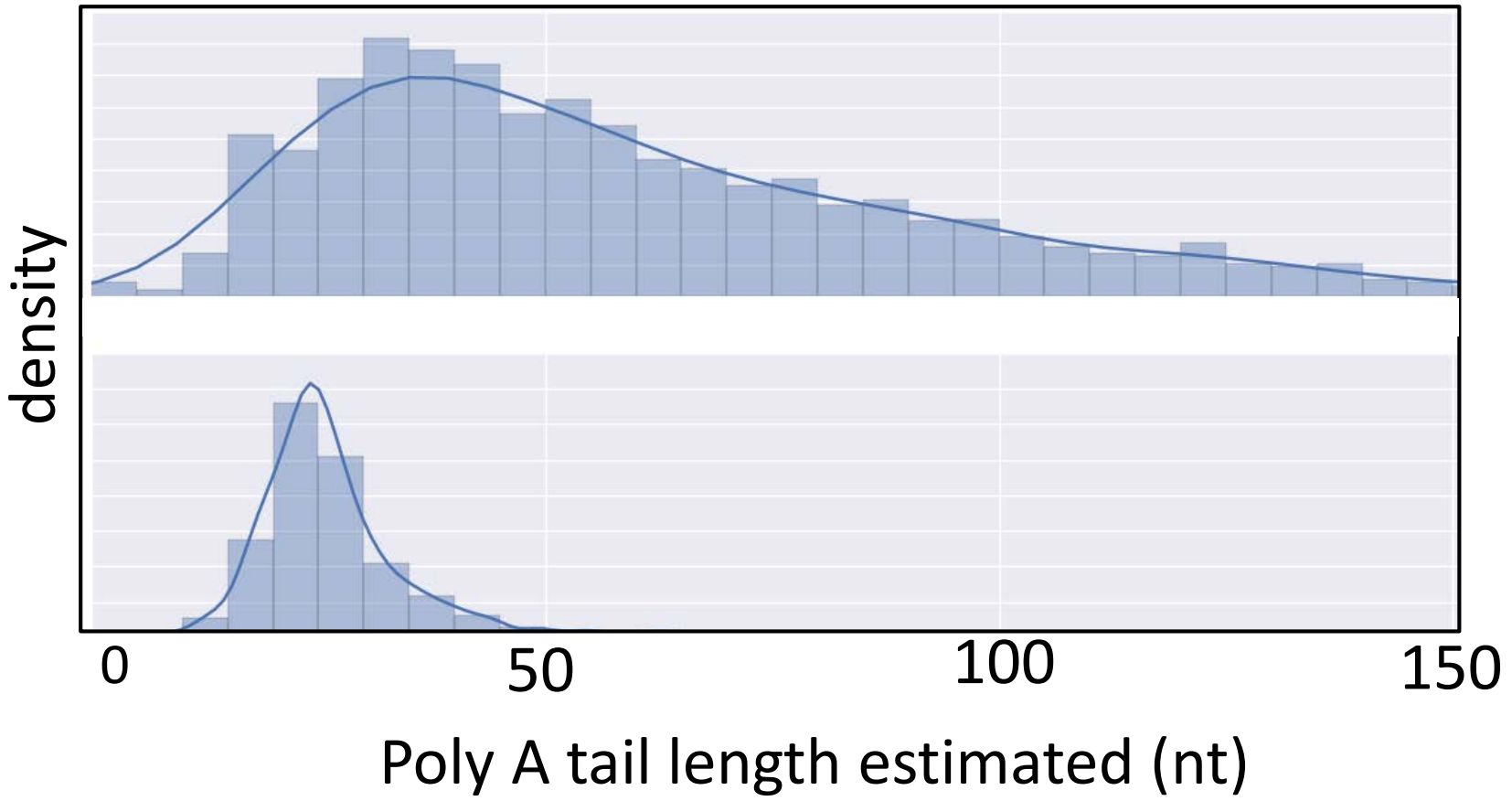
Image courtesy of: Fdardel

Ionic current dwell time can be used to estimate poly-A tail lengths



Predicting poly-A sequence length becomes tractable when consistent structural regions of dRNA reads can be identified and separated

Poly-A tail lengths for GM12878 and SIRVs consistent with expected



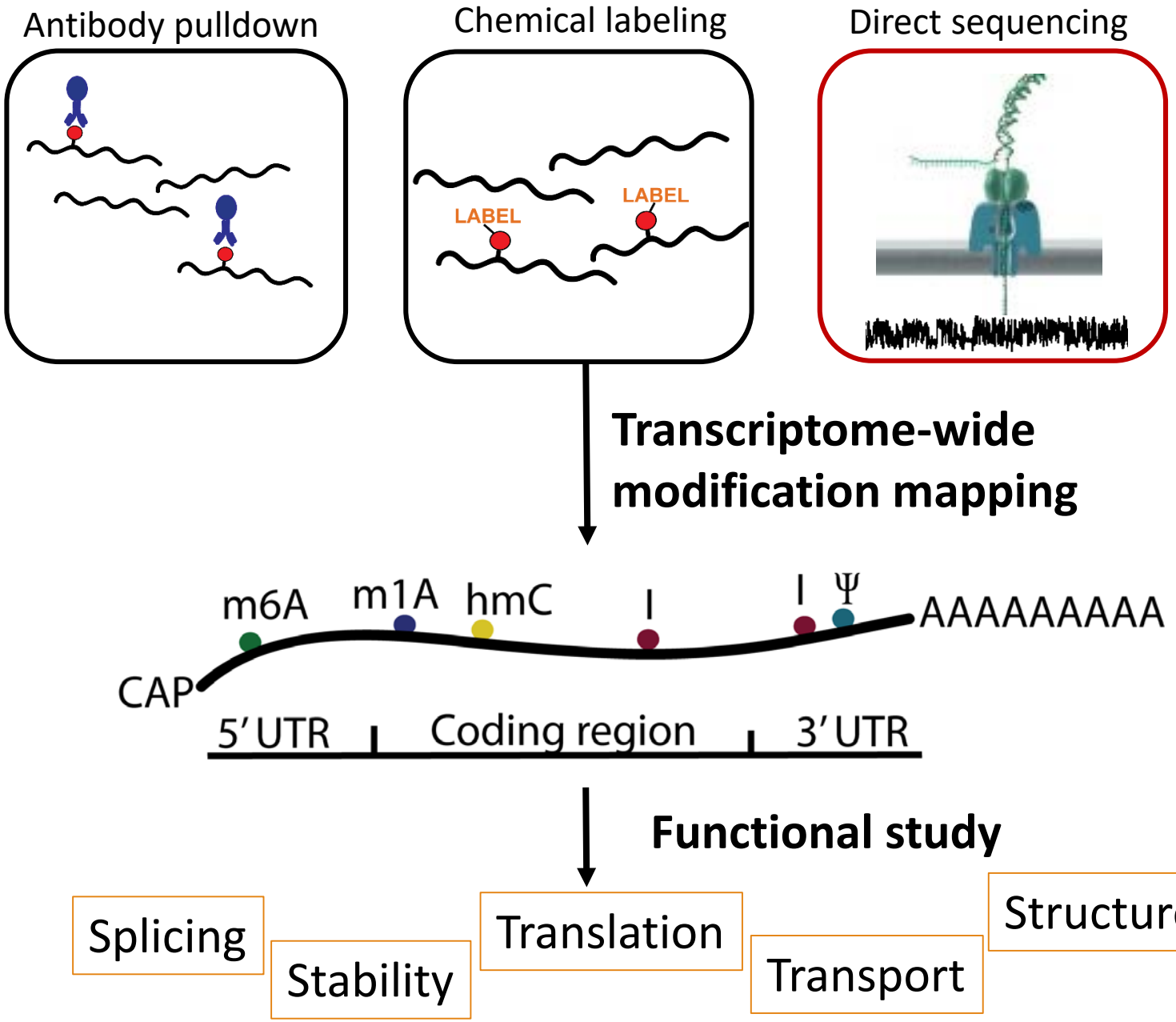
Poly-A tail lengths in humans expected 30-150+ nt

SIRV polyA tails expected 20-30nt

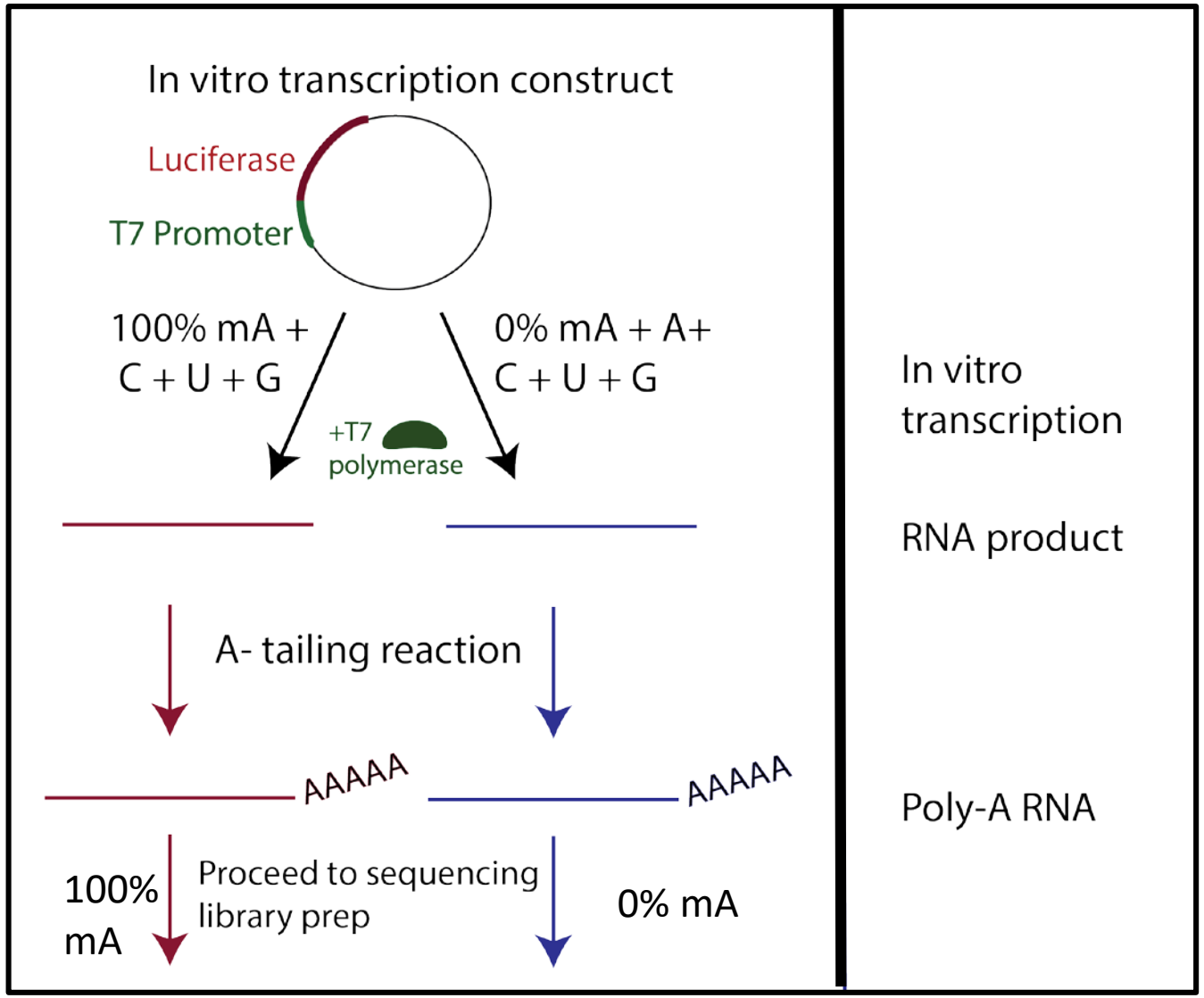
PolyA estimator (under development): https://github.com/jts/nanopolish/tree/polya_estimator

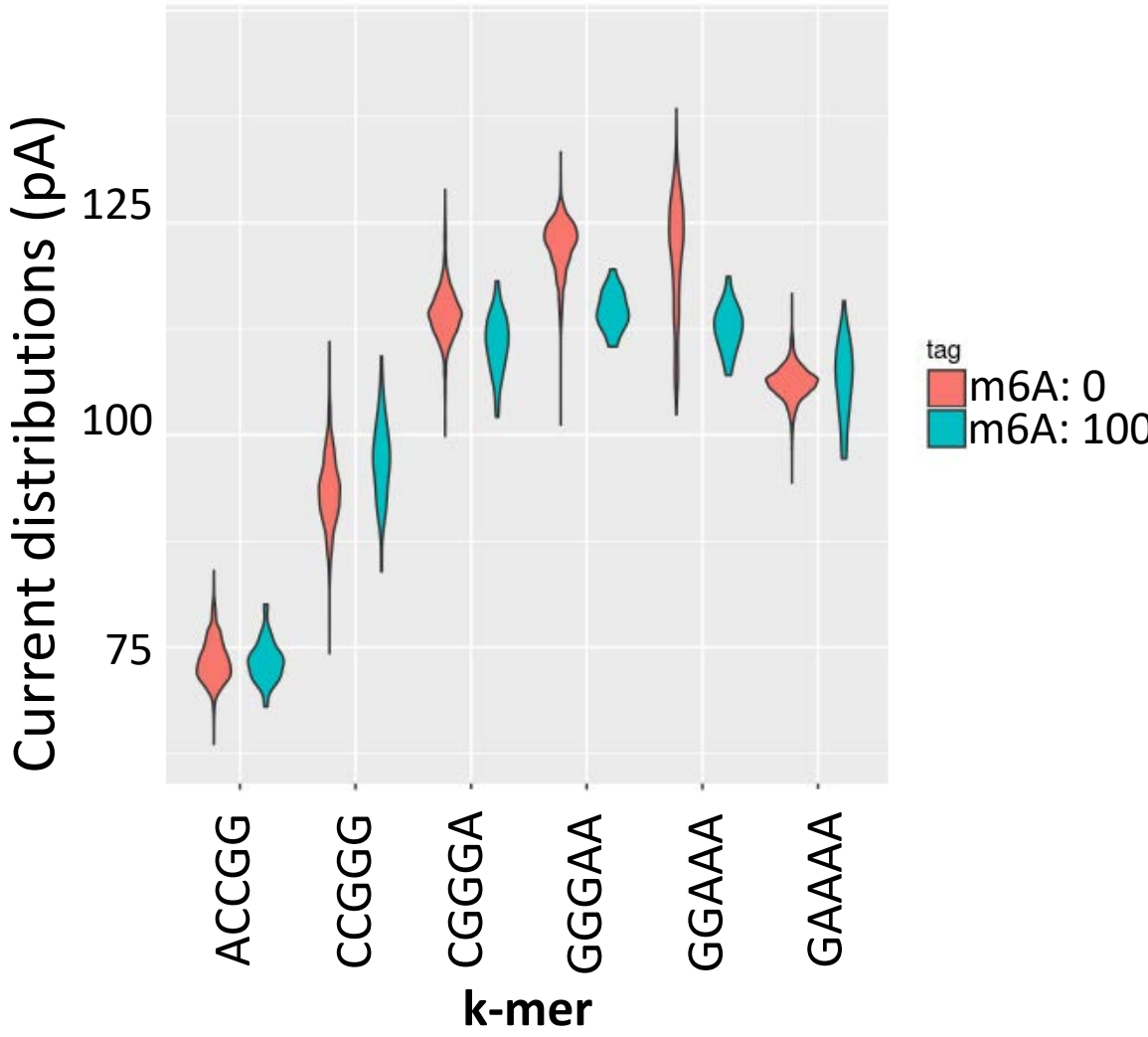
3. RNA MODIFICATIONS CAN BE DETECTED

Direct sequencing of RNA mods can supplement existing modification detection approaches



Detection of RNA modifications possible with IVT spike ins

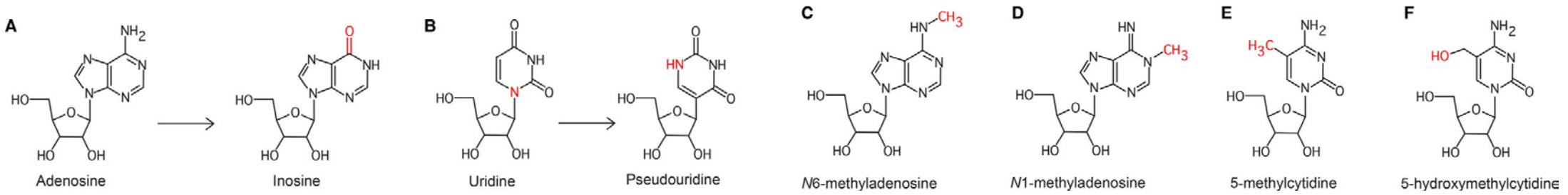




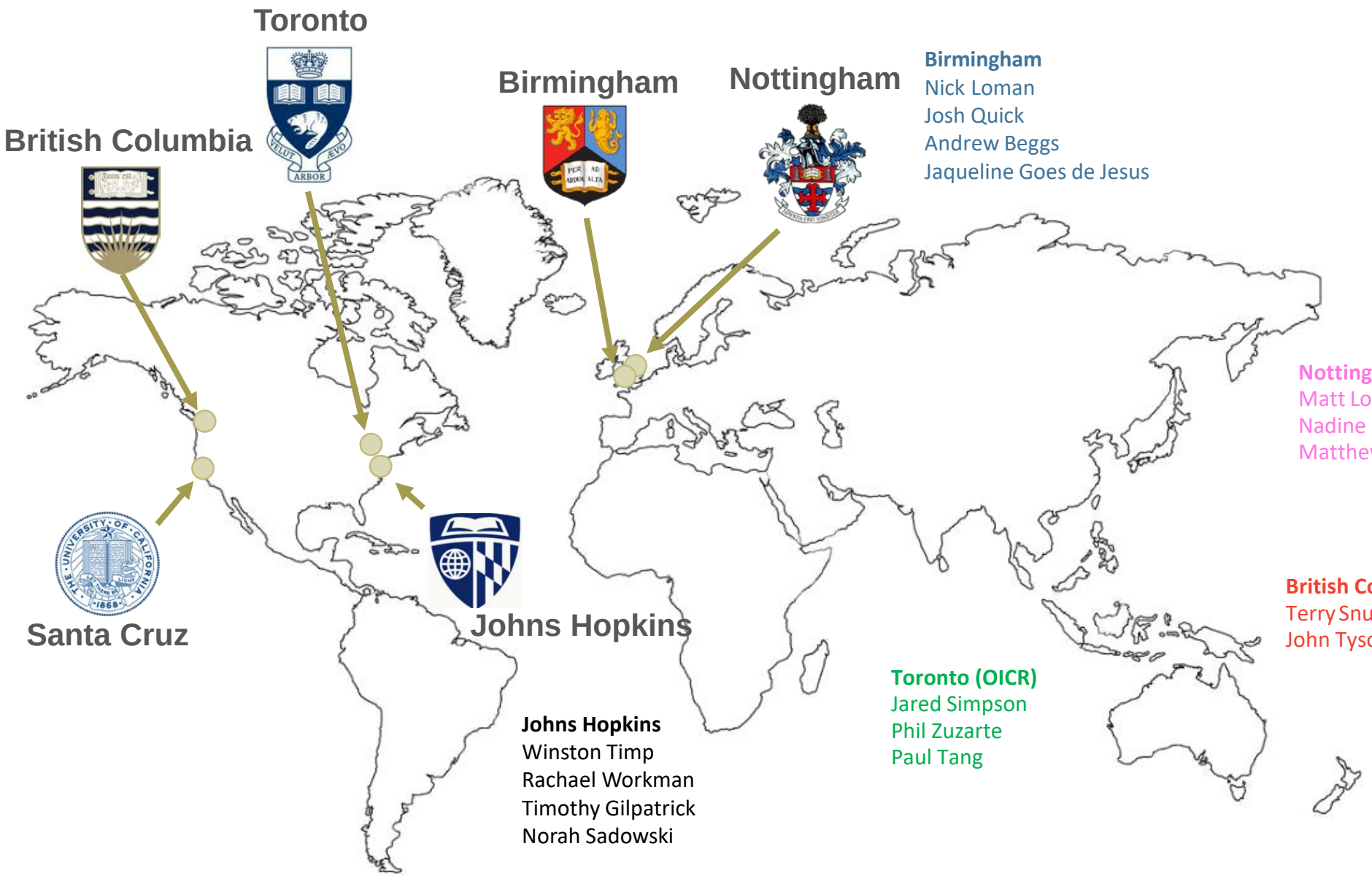
CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions and Future Directions

- Nanopore direct RNA sequencing is a promising technology for the simultaneous assessment of isoform structure and features of interest
- Unique information in this dataset can be used to improve human reference transcriptome
- RNA modification training expansion to include simultaneous detection of multiple mods



Acknowledgements



Birmingham
Nick Loman
Josh Quick
Andrew Beggs
Jaqueline Goes de Jesus

Nottingham
Matt Loose
Nadine Holmes
Matthew Carlile

Santa Cruz
Mark Akeson
Hugh Olsen
Benedict Paten
Angela Brooks
Miten Jain

British Columbia
Terry Snutch
John Tyson

Toronto (OICR)
Jared Simpson
Phil Zuzarte
Paul Tang

Johns Hopkins
Winston Timp
Rachael Workman
Timothy Gilpatrick
Norah Sadowski

ONT
Daniel Garalde
Daniel Jachimowicz
Andy Heron
Rosemary Dokos