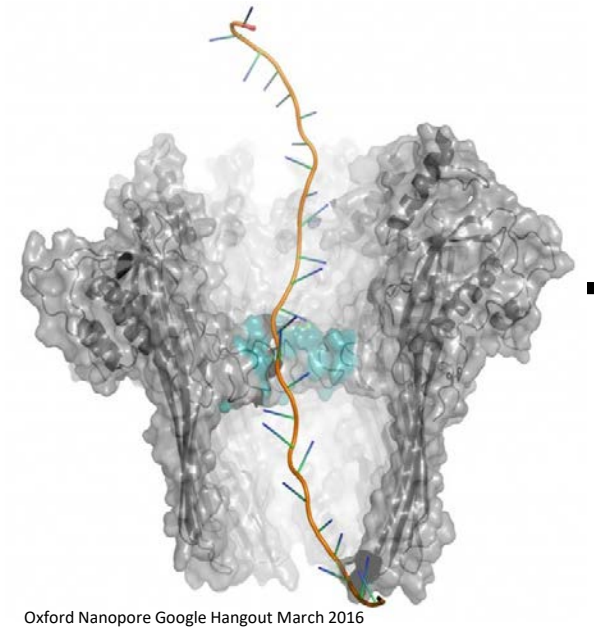
JOHNS HOPKINS
BIOMEDICAL ENGINEERING

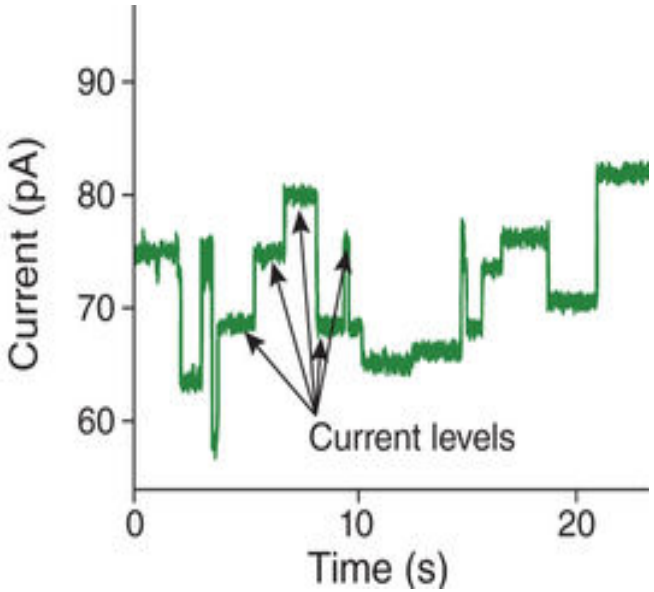
Direct RNA sequencing of human transcripts using the Oxford Nanopore sequencing platform

For the Nanopore RNA Consortia:
Winston Timp
Department of Biomedical Engineering
Johns Hopkins University

Revolutions in Genomics: Single Molecule Sequencing



Oxford Nanopore Google Hangout March 2016



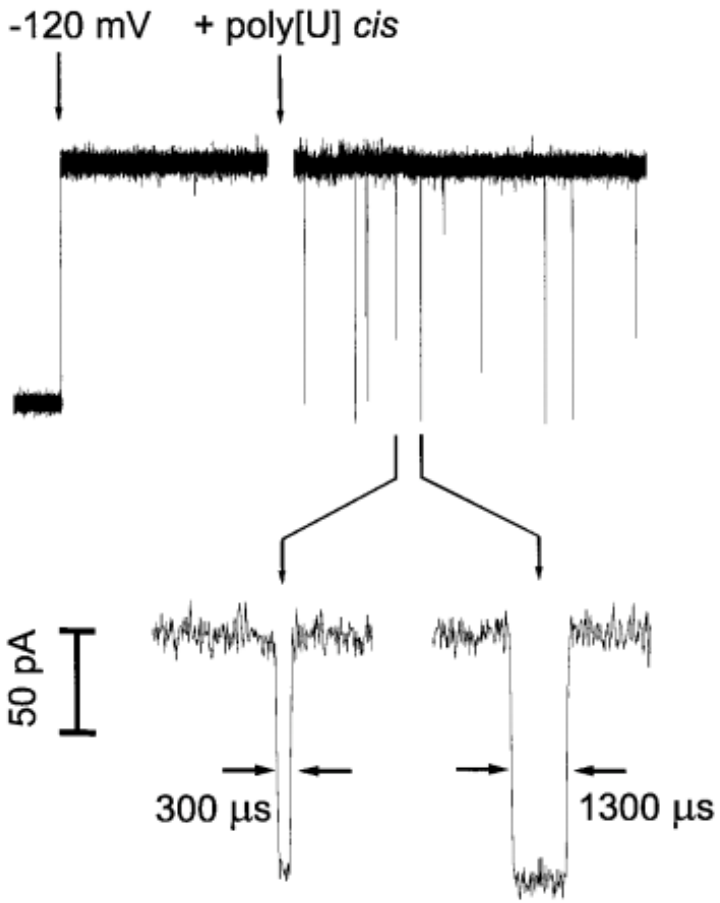
Deamer et al 2016, Nature Biotech

ATCGATCGATAGTA
TTAGATACGACTAG
CGATCAG

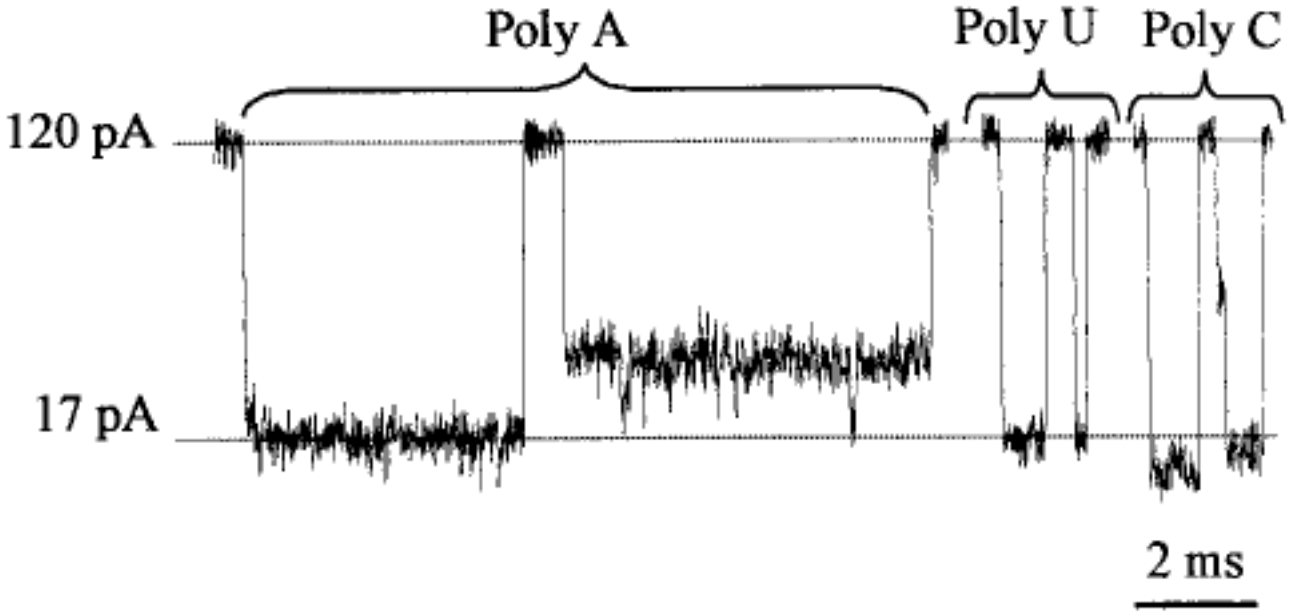
- First patented back in 1995, commercialized in 2014
- No theoretical upper limit to sequencing read length, practical limit only in delivering DNA to the pore intact
- Palm sized sequencer
- Sequencing output 5-10Gb

Disclosure: Timp has two patents (US 2011/0226623 A1; US2012/0040343 A1) licensed to ONT

Earliest nanopore experiments analyzed RNA



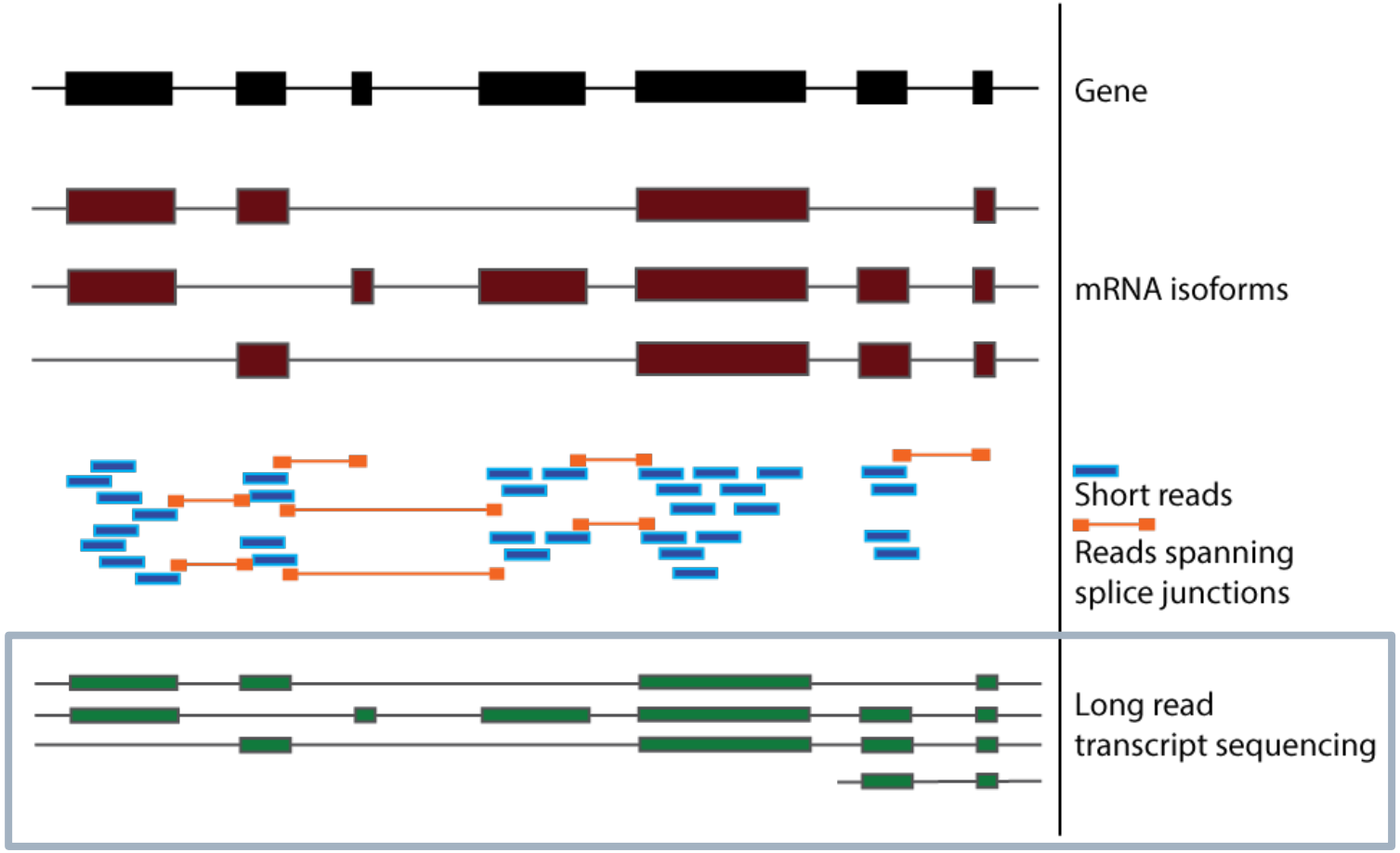
Kasianowicz, Brandin, Branton, & Deamer
PNAS 1996



Akeson, Branton, Brandin & Deamer
Biophysical J. 1999

Building Transcriptomes: Native RNA sequencing

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



Nanopore RNA consortium

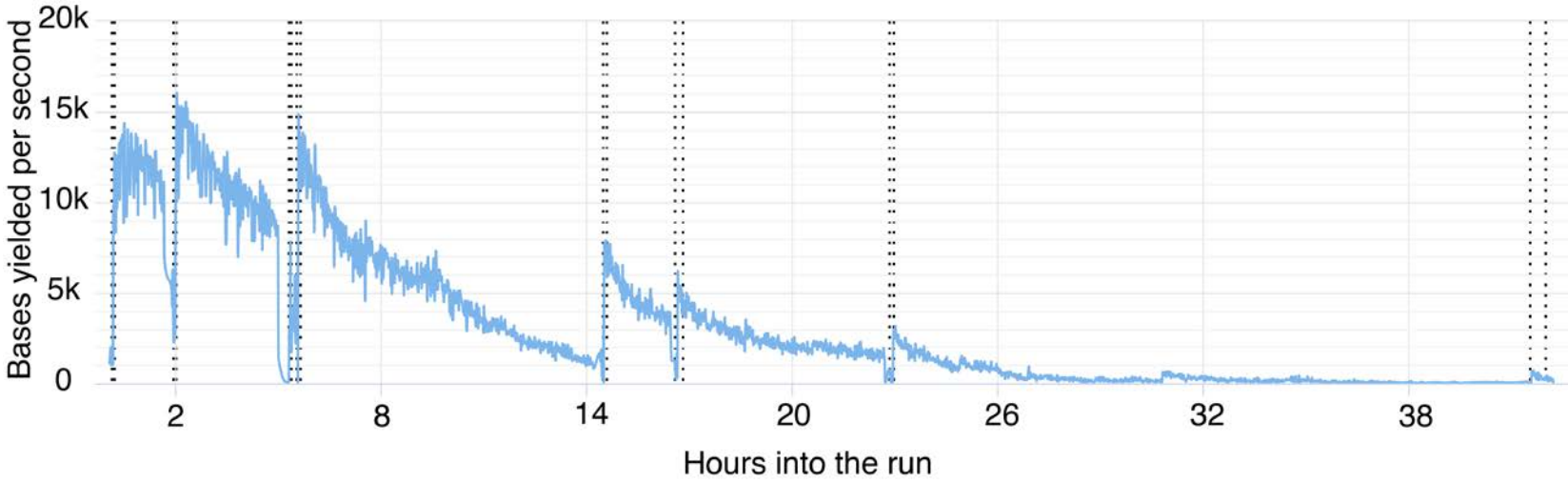
- **Six participating Universities**
  - Johns Hopkins University
  - University of Birmingham
  - University of California Santa Cruz
  - University of British Columbia
  - Ontario Institute for Cancer Research
  - University of Nottingham
- Isolated poly-A RNA from GM12878 CEPH cell line, sequenced direct RNA and amplified cDNA
- Spiked-in synthetic RNA molecules
  - Lexogen SIRV Set 3 (ERCC + 69 SIRV Isoforms)

How much data did we generate?

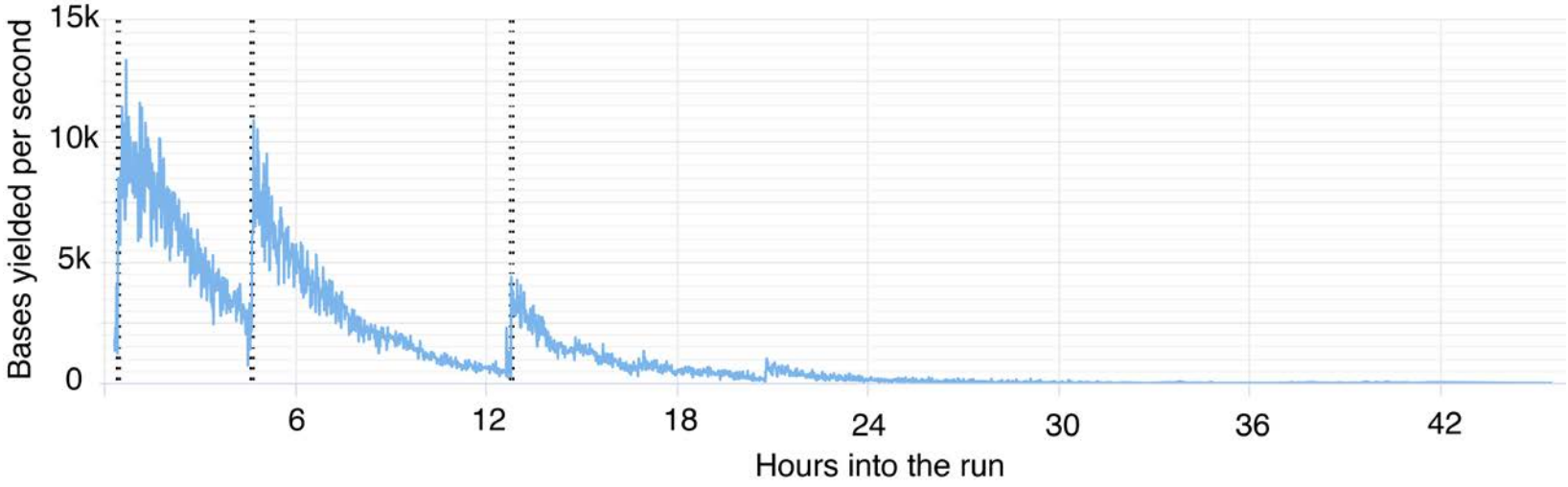
- ≈13 million Direct RNA sequences, 30 flow cells
- >24 million cDNA sequences, 12 flow cells
- https://github.com/nanopore-wgs-consortium/NA12878/

Voltage tuning improves throughput

With Tuning



Without Tuning



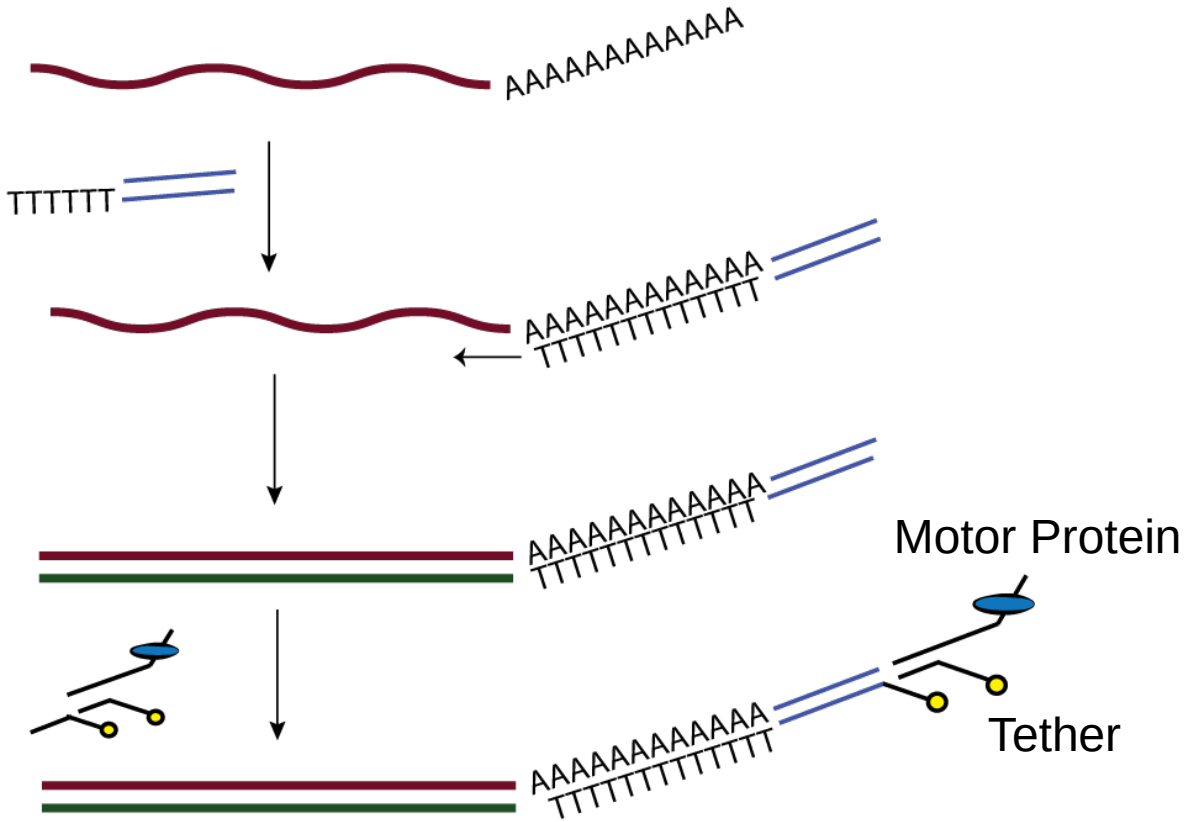
Long read RNA can now be sequenced directly with ONT

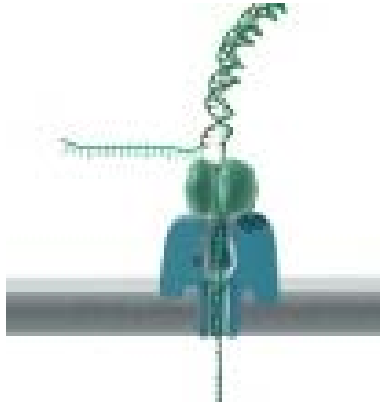
PolyA+ RNA captured

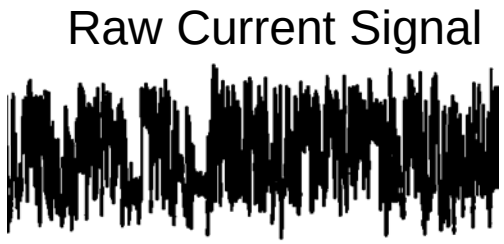
Splint poly-T adapter ligation

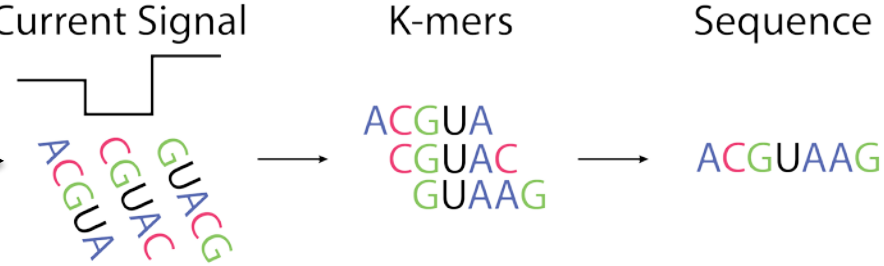
Reverse transcription (optional)

Sequencing adapter ligation



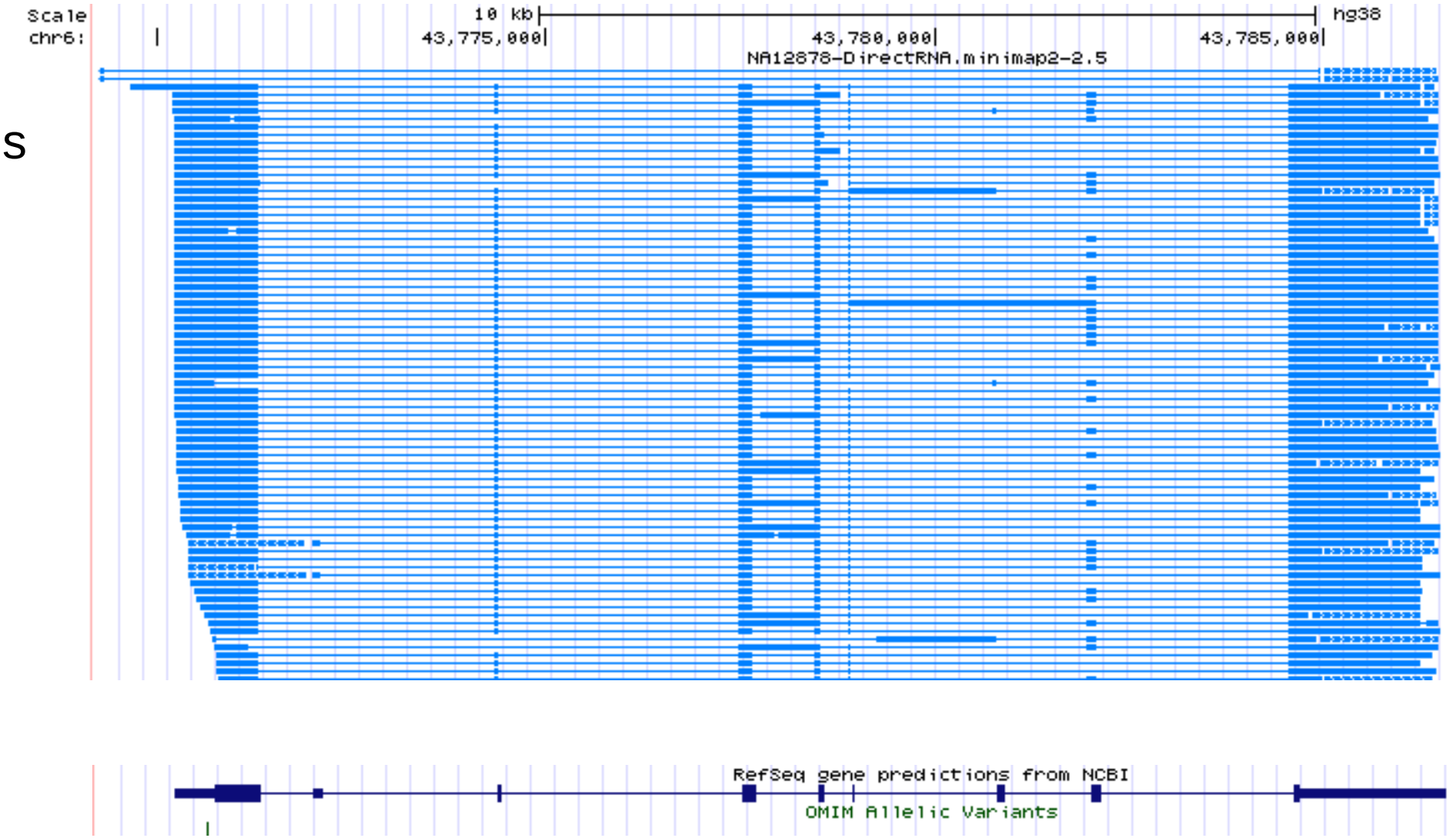




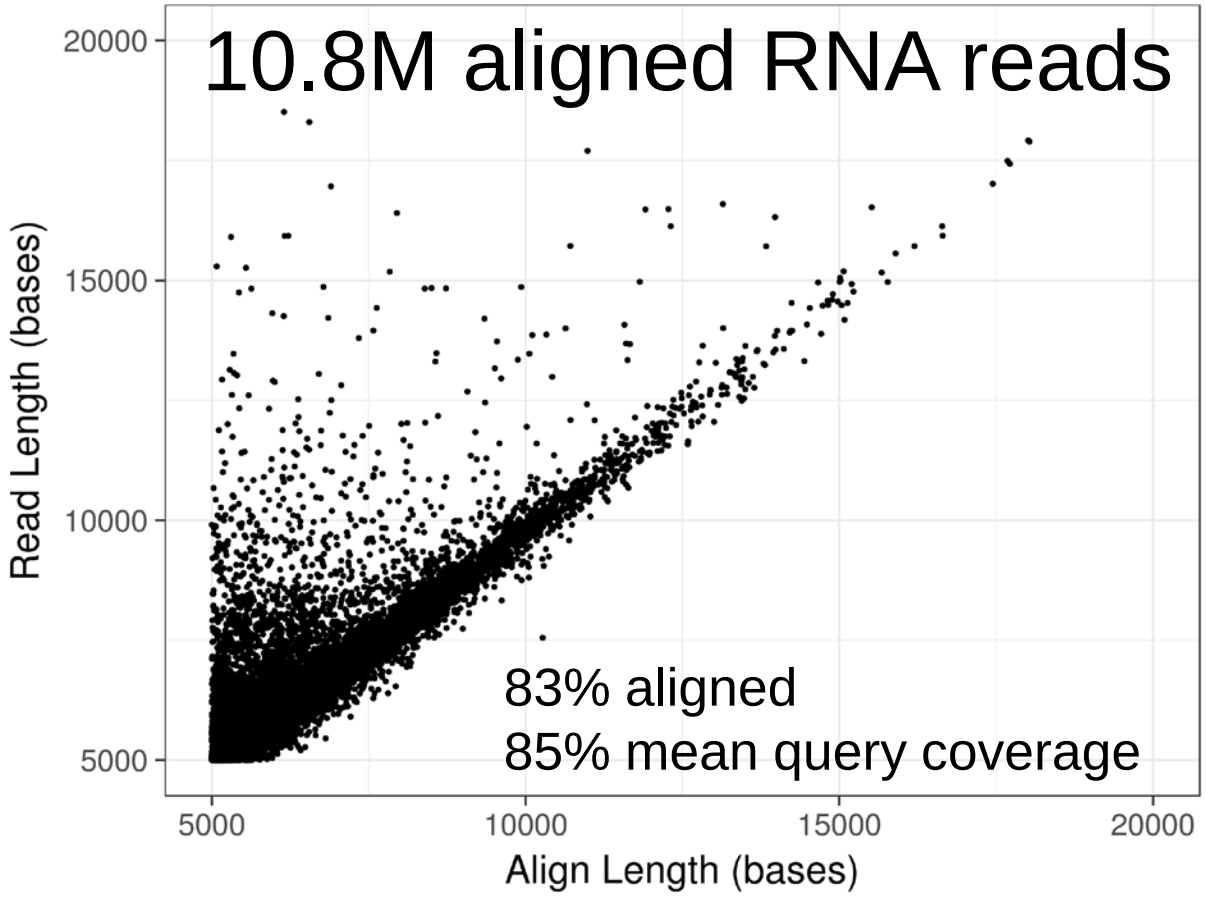


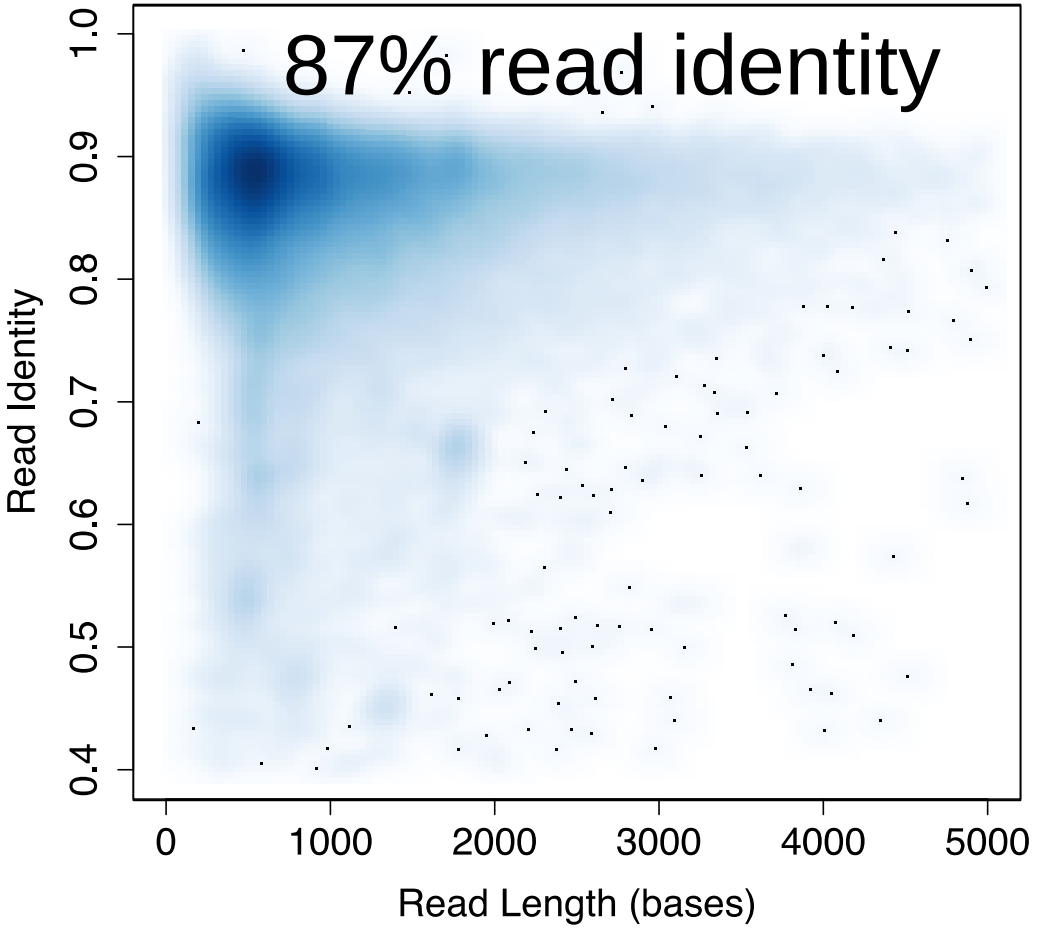
Building Transcriptomes: Direct RNA sequencing

- Long reads allow us to see the entire VEGF gene
- Different isoforms clearly visible
- We can even see transcripts in the process of splicing (introns kept in)



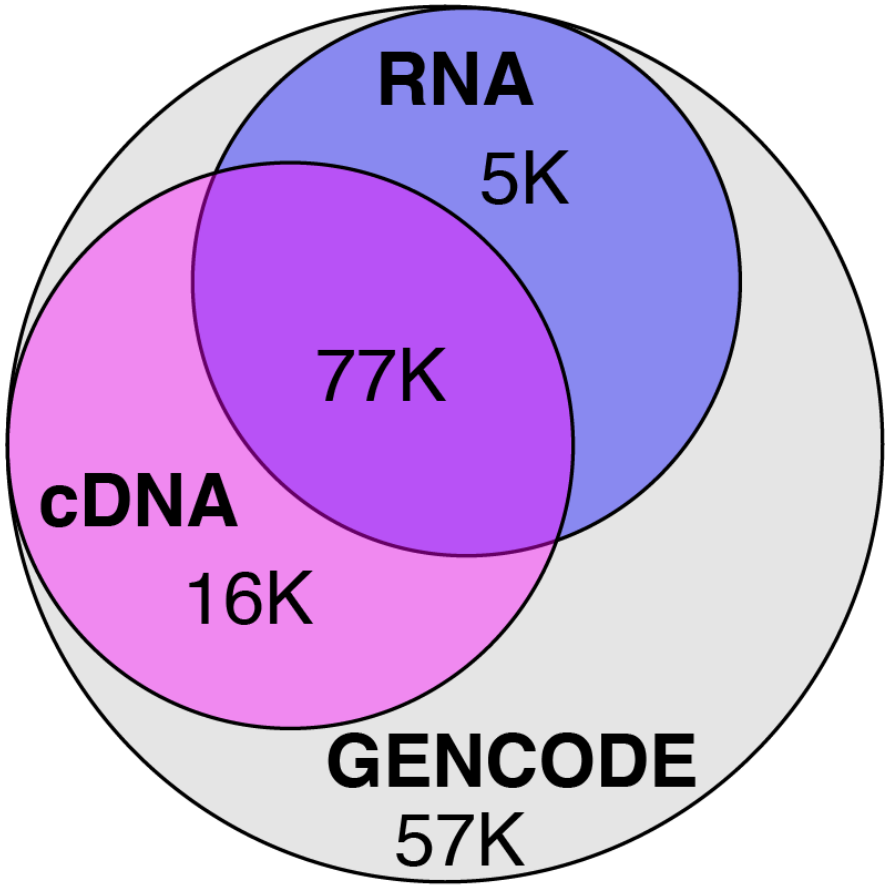
dRNA reads show good alignment to GENCODE reference

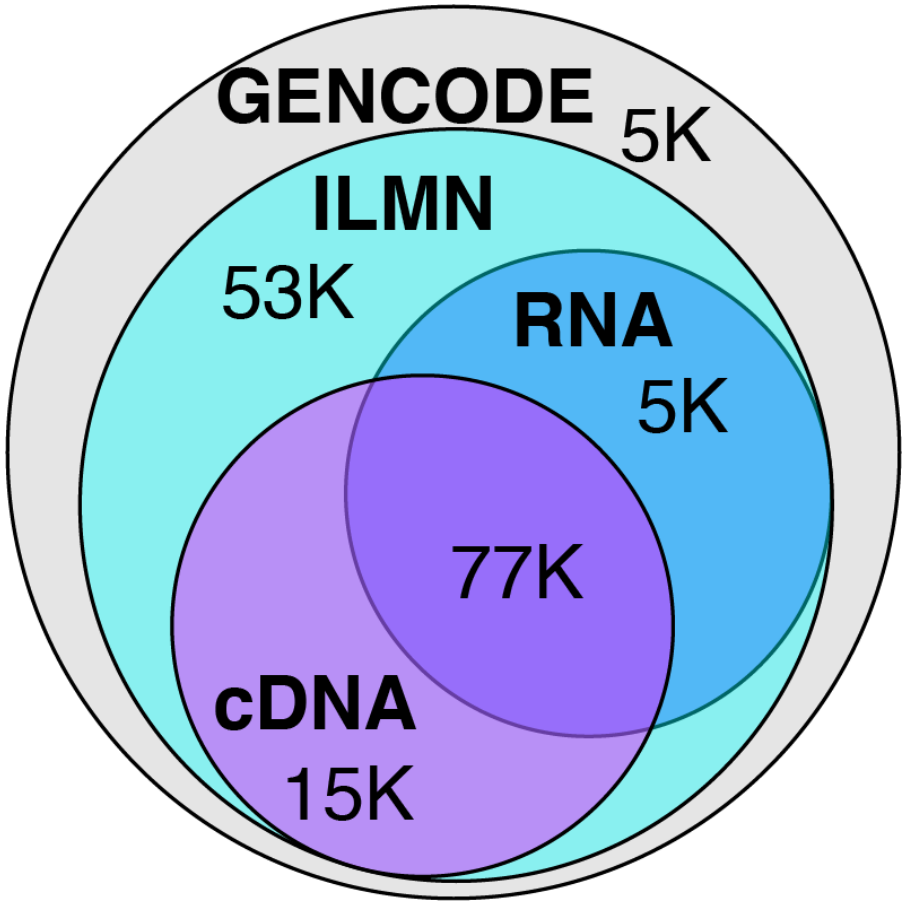




Alignment to GENCODE V27 dataset performed using Minimap2: https://github.com/lh3/minimap2

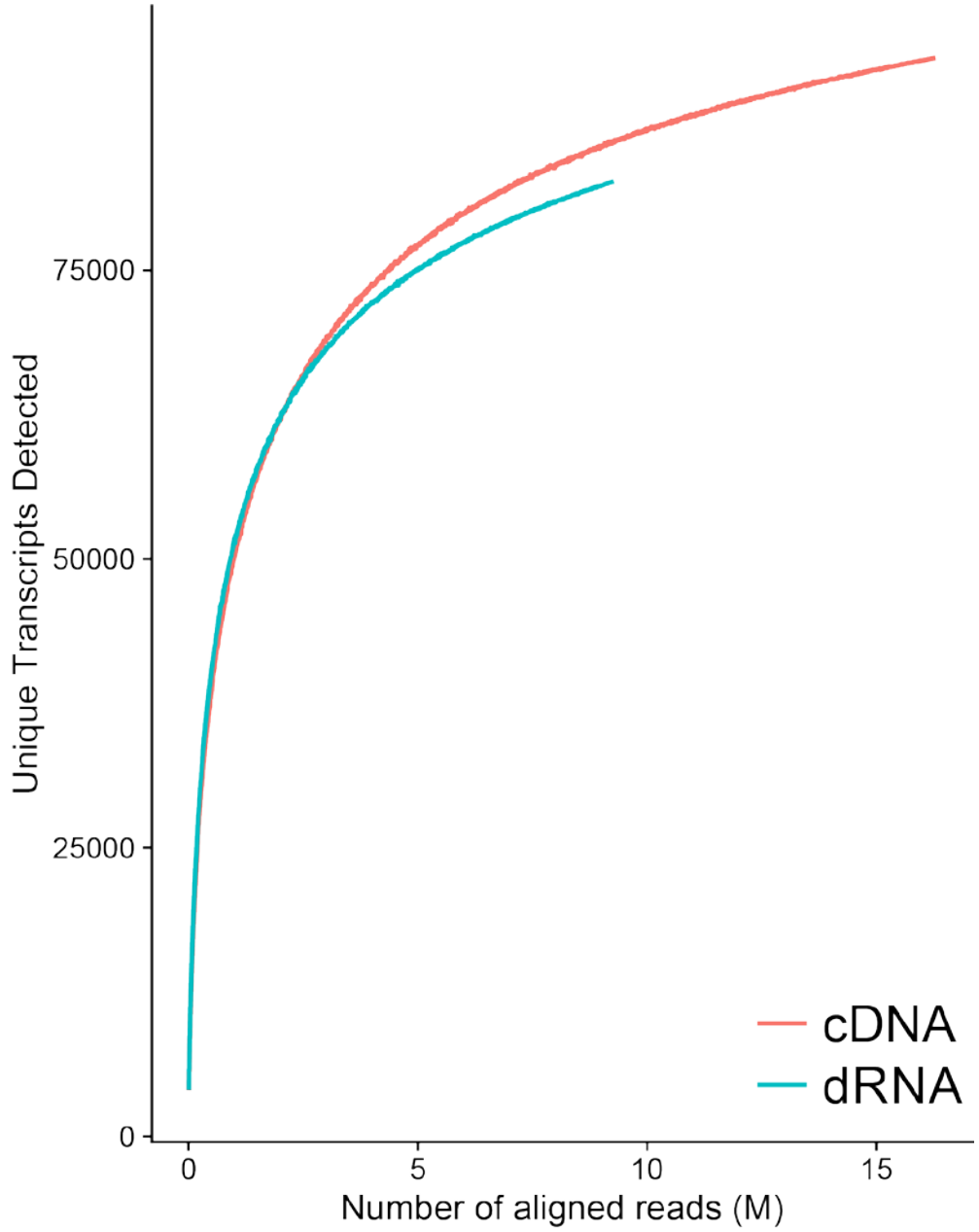
Large number of annotated genes identified for dRNA



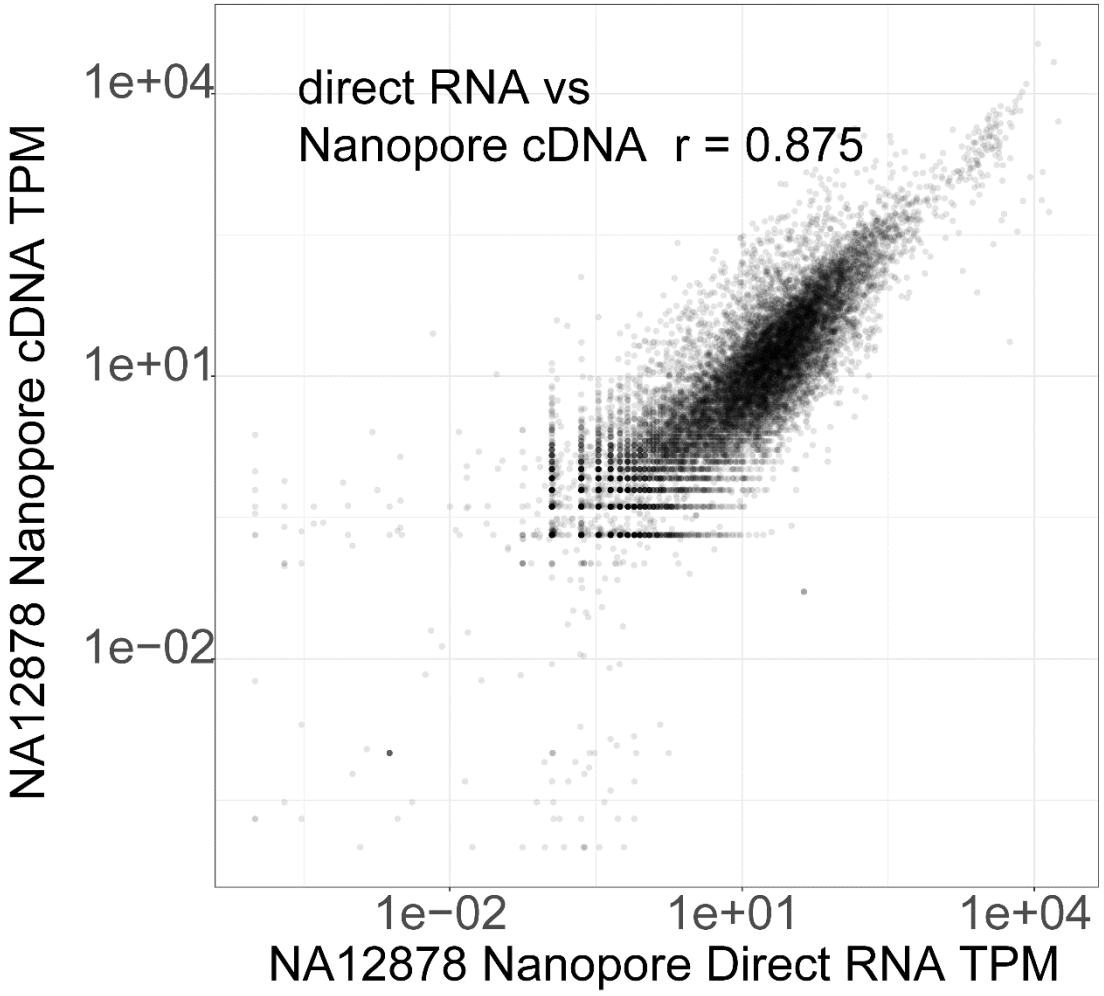


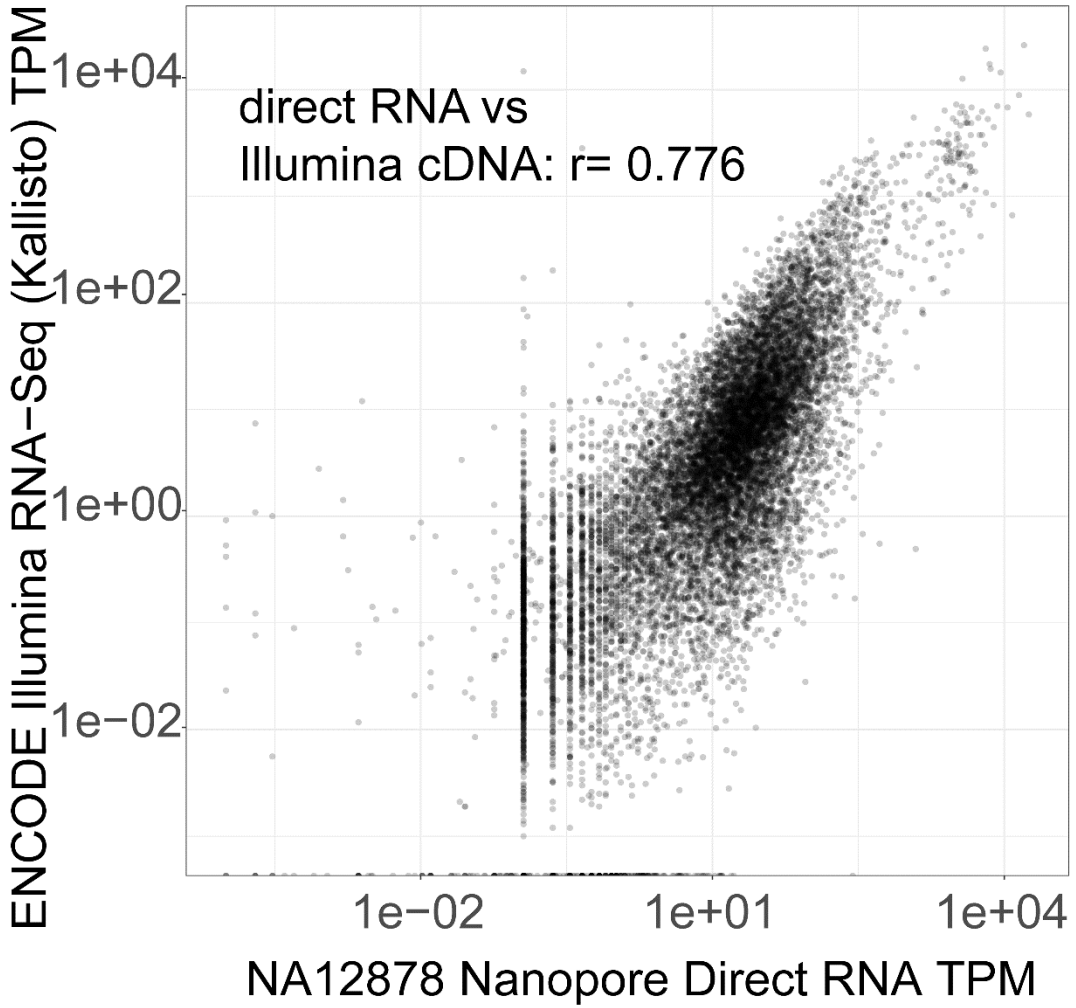
Unique transcripts detected approaches “saturation”

- How important is capturing “every transcript”?
- After 5M reads we already start to see diminishing returns



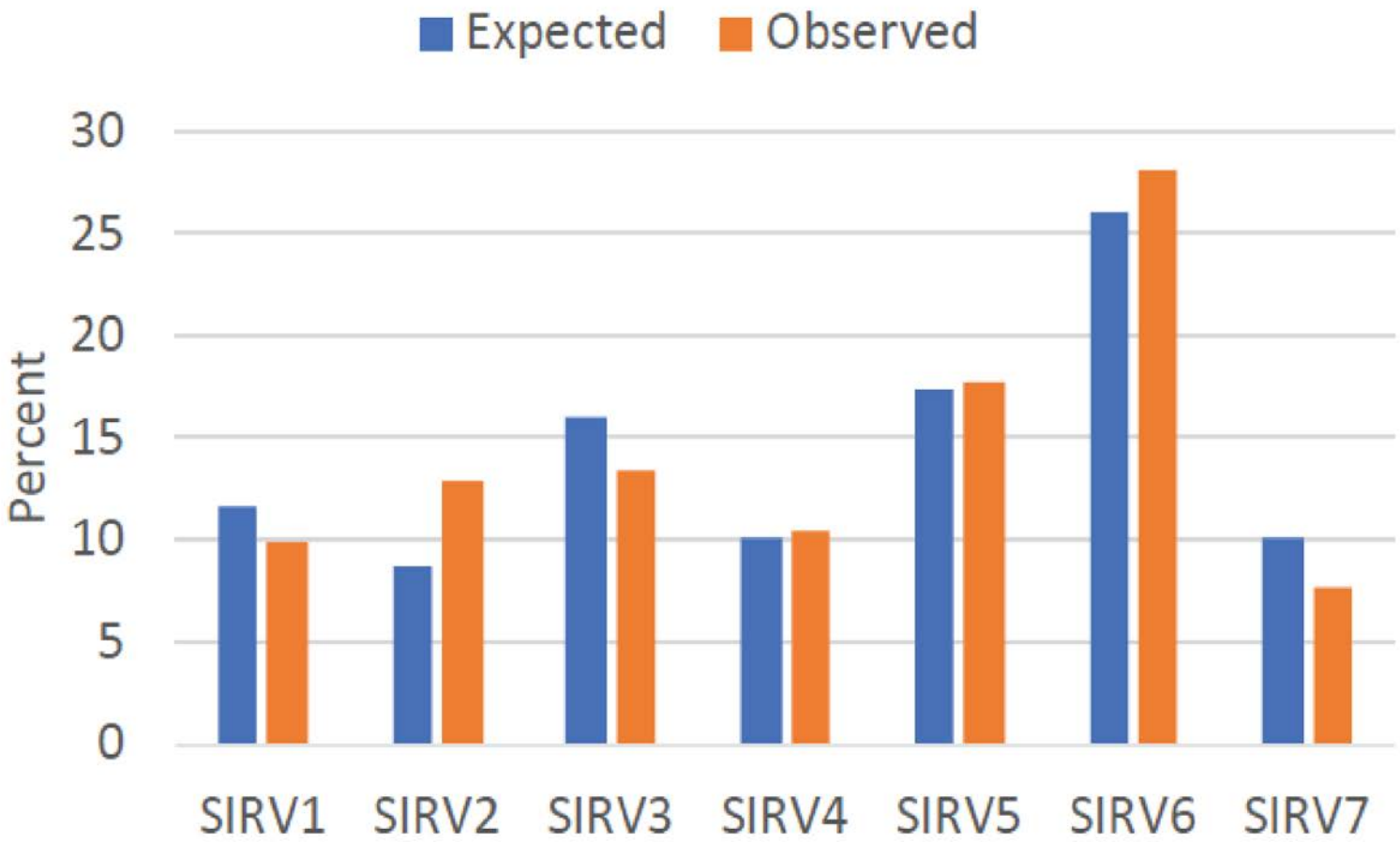
Correlation in gene-level abundance between RNA vs. ONT cDNA, Illumina



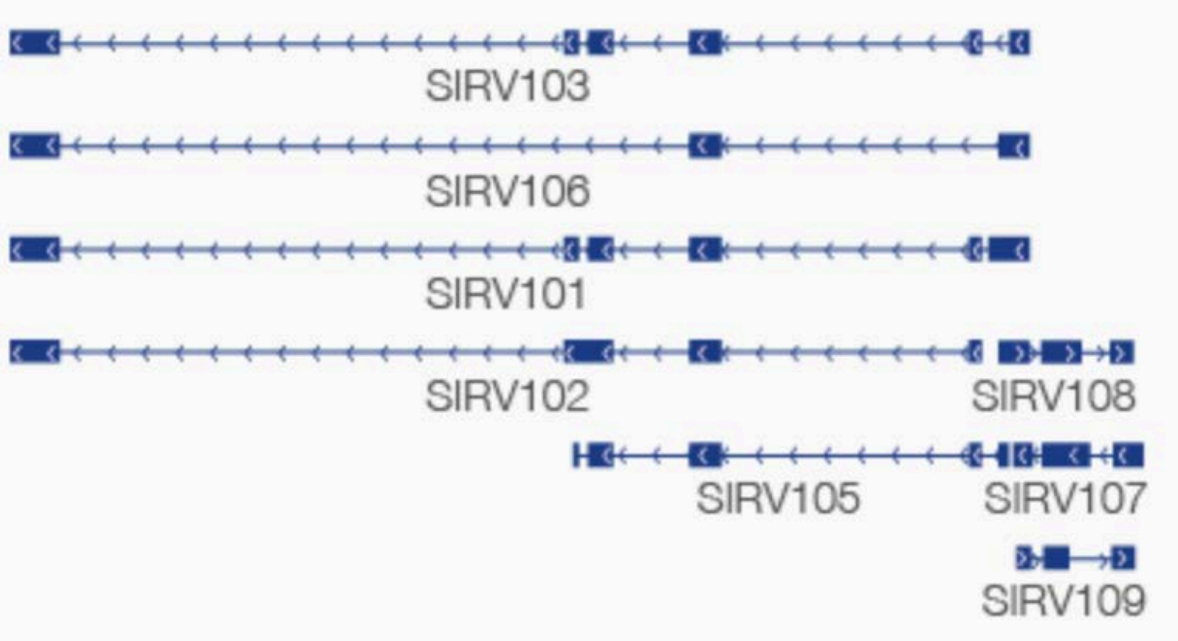


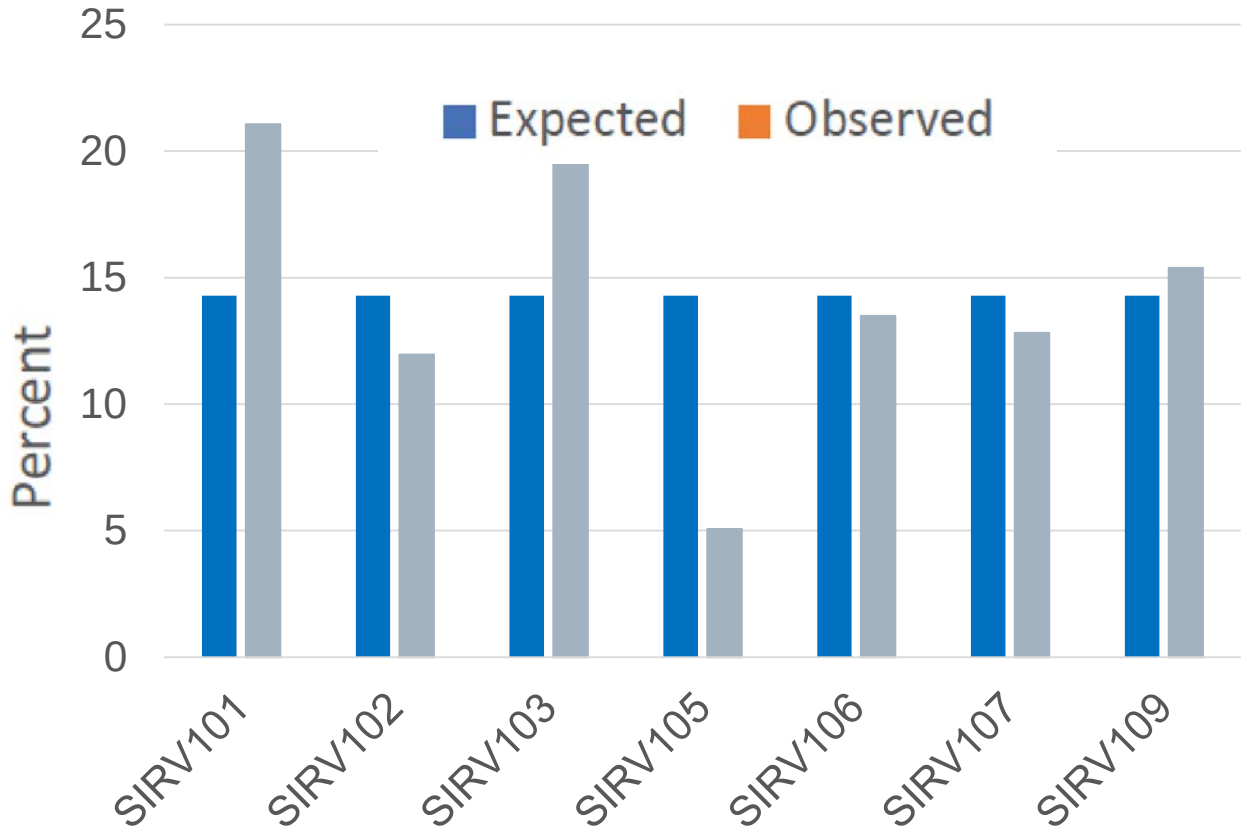
ENCODE Illumina dataset, Wold Lab: https://www.encodeproject.org/experiments/ENCSR000AEG/

SIRV gene level quantification tracks expected input well



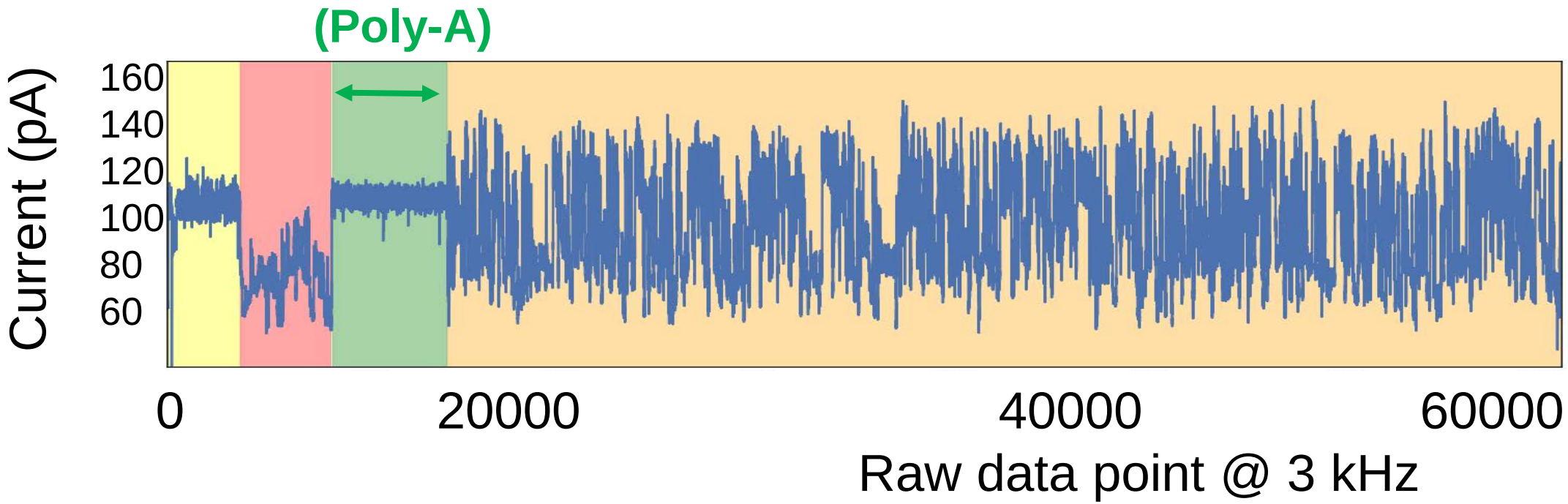
Quantification more complicated at isoform level





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

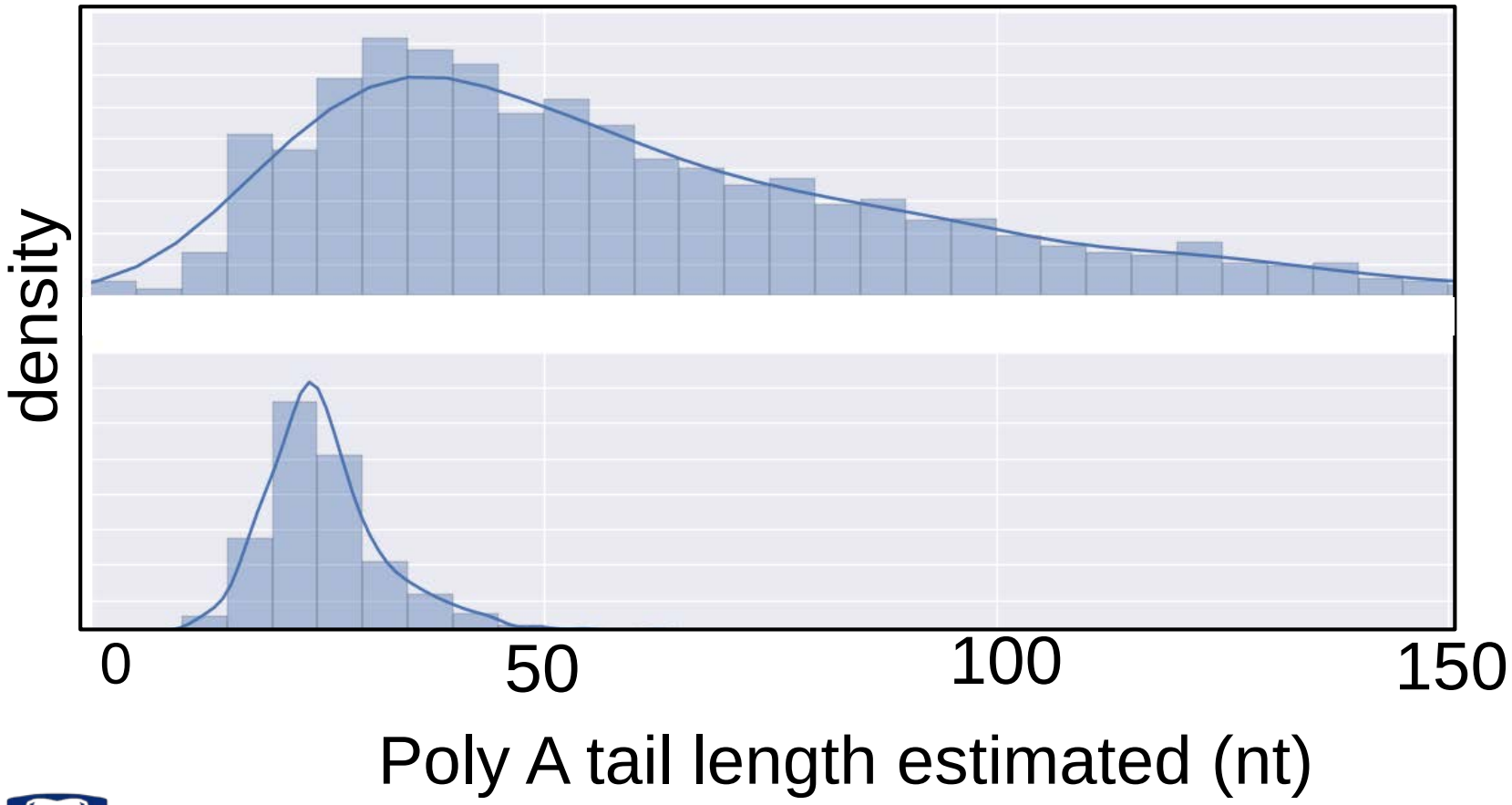
Ionic current dwell time can be used to estimate poly-A tail lengths



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

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

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

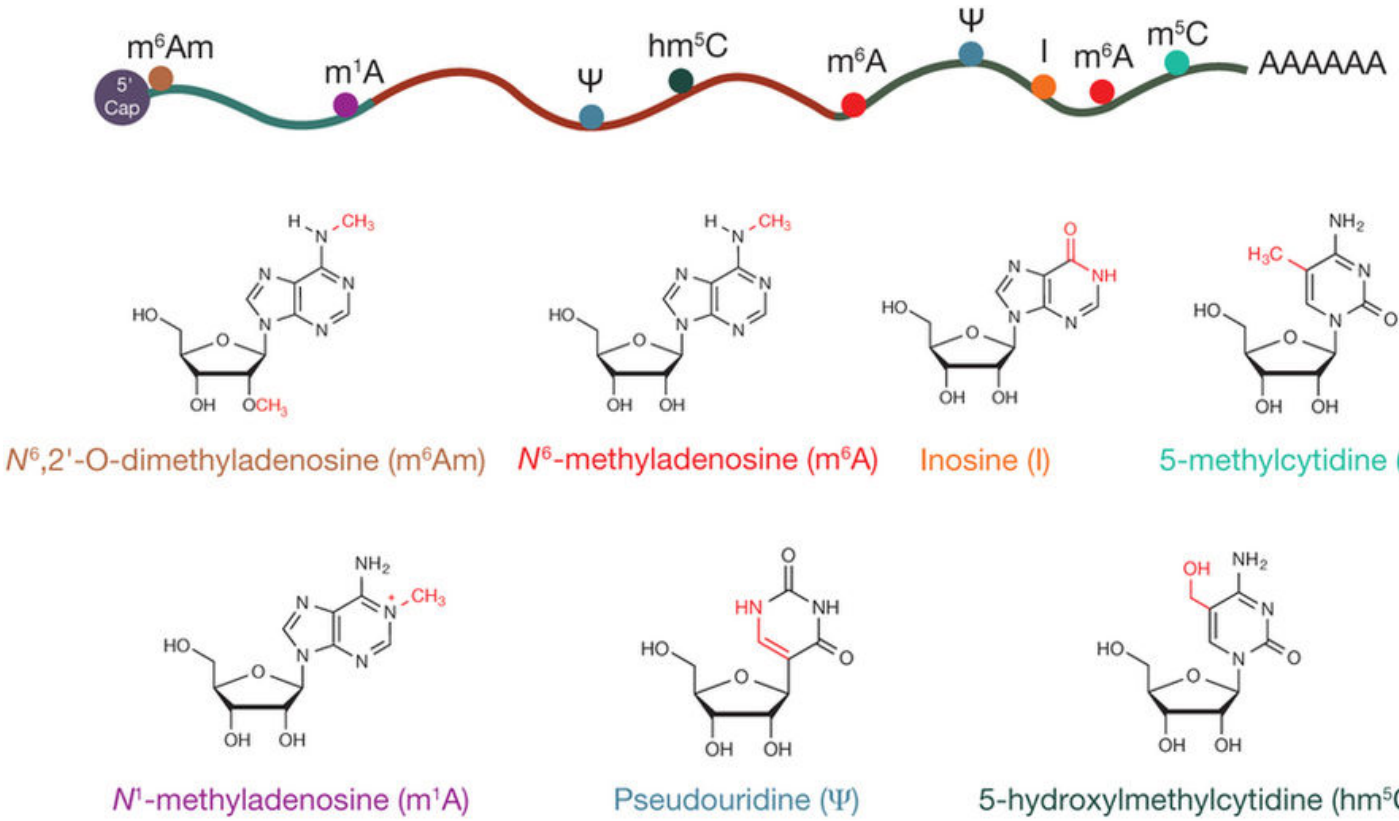


Poly-A tail lengths in humans expected 30-150+ nt

SIRV polyA tails expected 20-30nt

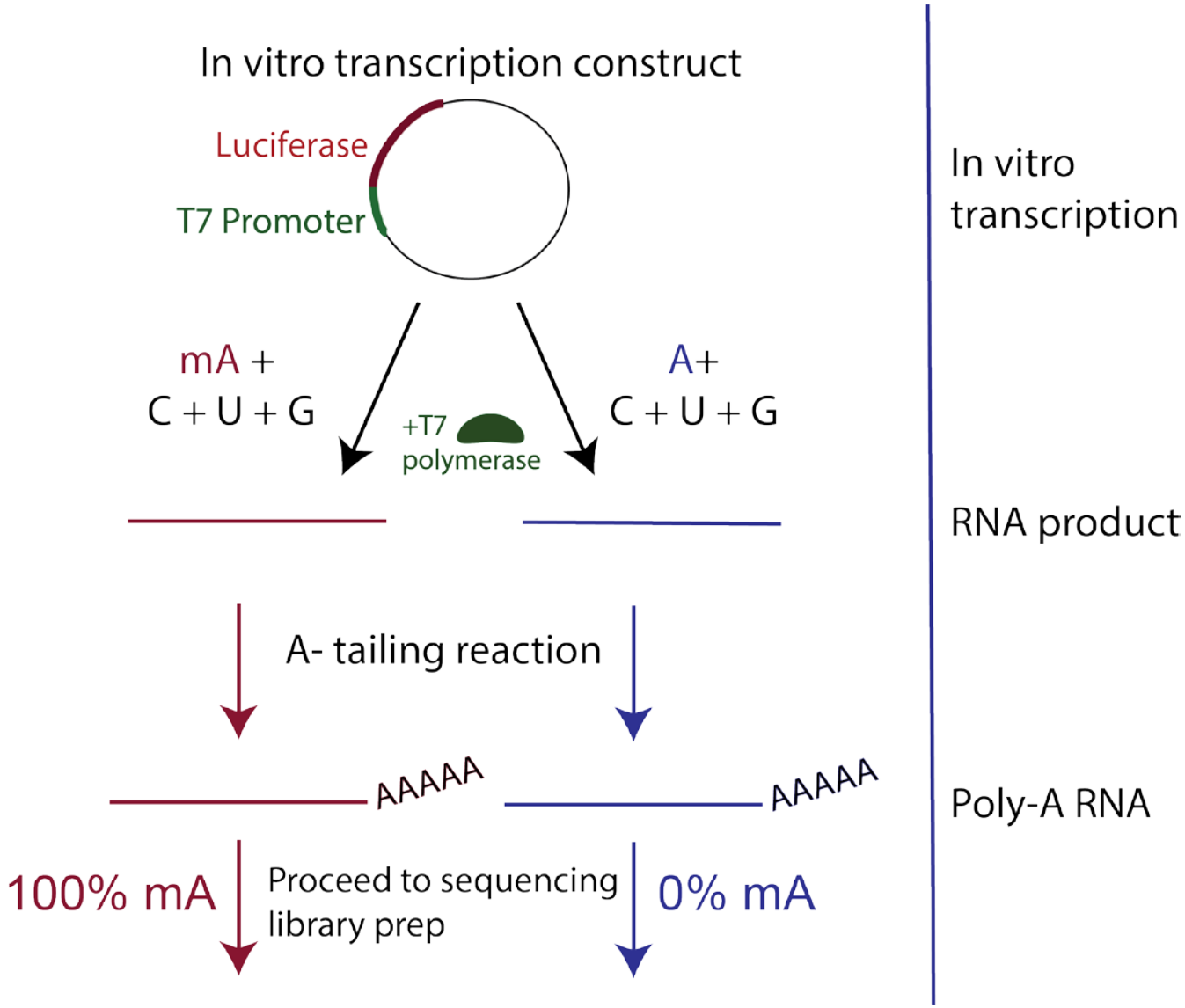
Building Transcriptomes: Direct RNA Sequencing

- We can use this to understand RNA modifications – the **epitranscriptome**
- Other methods are challenging – either inefficient, or lack resolution, and always only one modification at a time



Detection of RNA modifications with modIVT

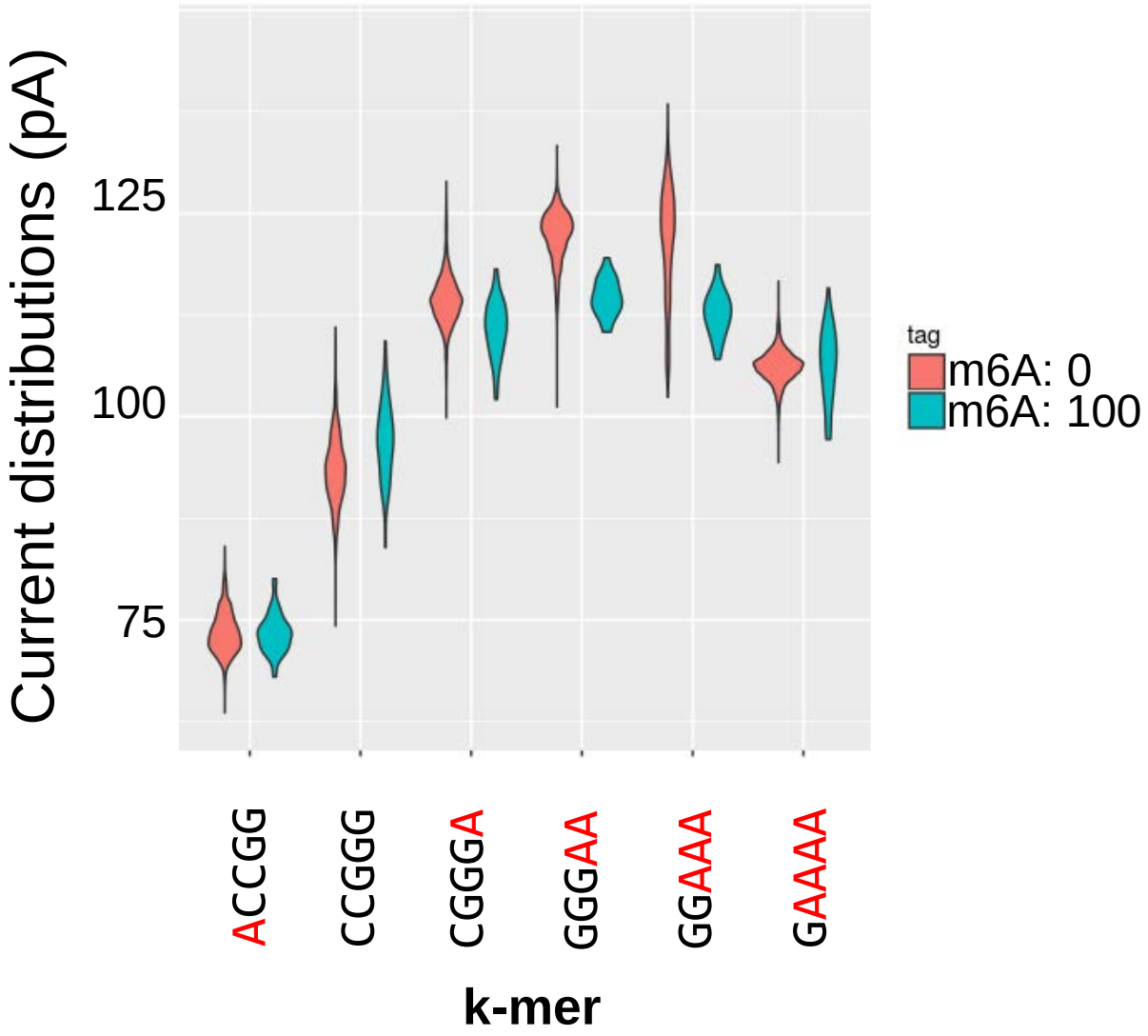
- IVT based RNA synthesis allows incoproation of labeled nucleotides
- All or none reaction right now, T7 has a strong preference for the unmodified nulceotides, making mixtures hard



NEB HiScribe™ T7 High Yield RNA Synthesis kit

Detection of RNA modifications with modIVT

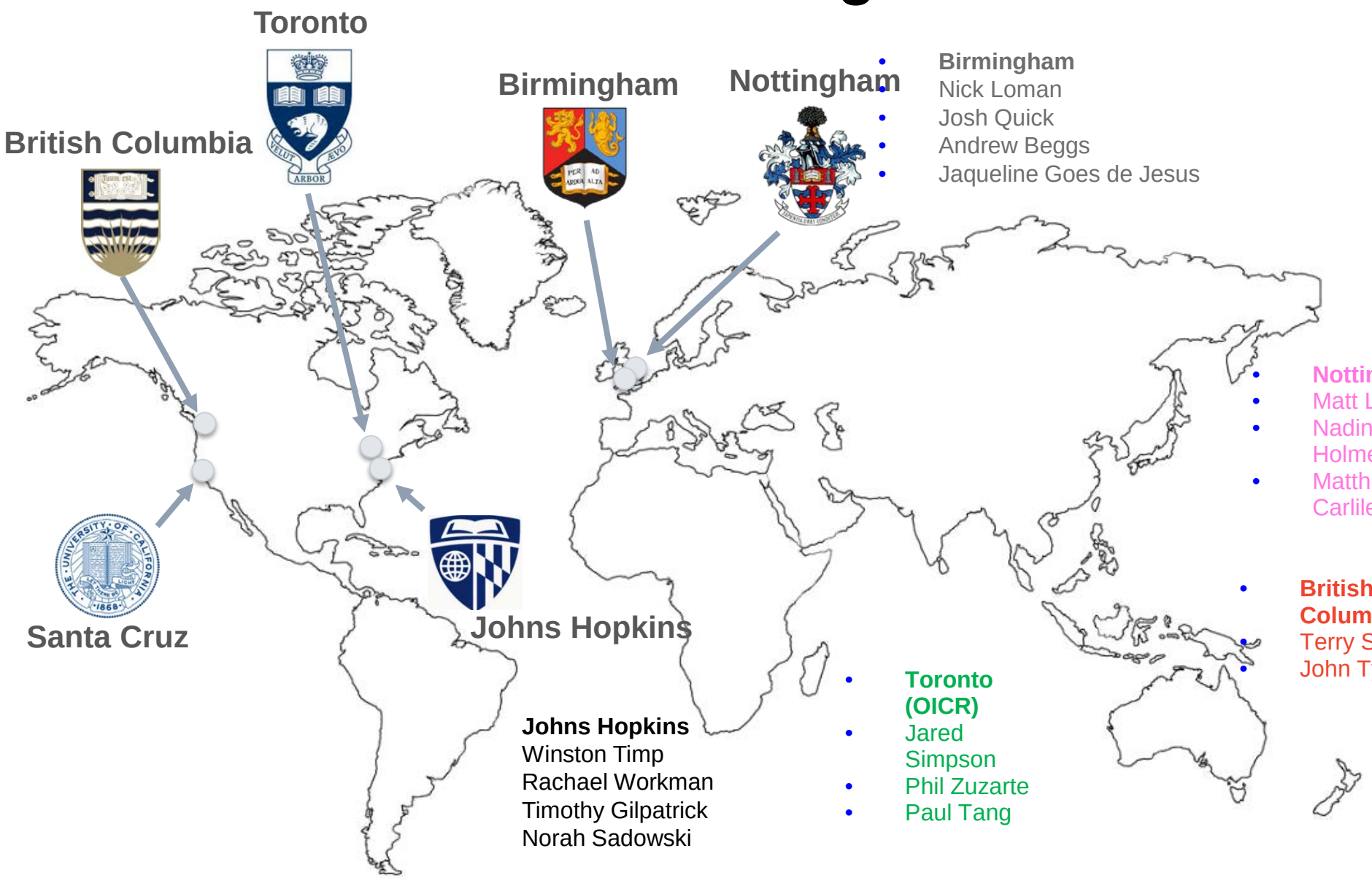
- From Luciferase we can already see strong signal depending on context
- Using nanopolish eventalign, we can extract the distribution of current values along the RNA strand



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
- Poly-A tail length measurement may reveal new insights into its function
- 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

Johns Hopkins
- Winston Timp
- Rachael Workman
- Timothy Gilpatrick
- Norah Sadowski

Toronto (OICR)
- Jared Simpson
- Phil Zuzarte
- Paul Tang

ONT
- Daniel Garalde
- Daniel Jachimowicz
- Andy Heron
- Rosemary Dokos