

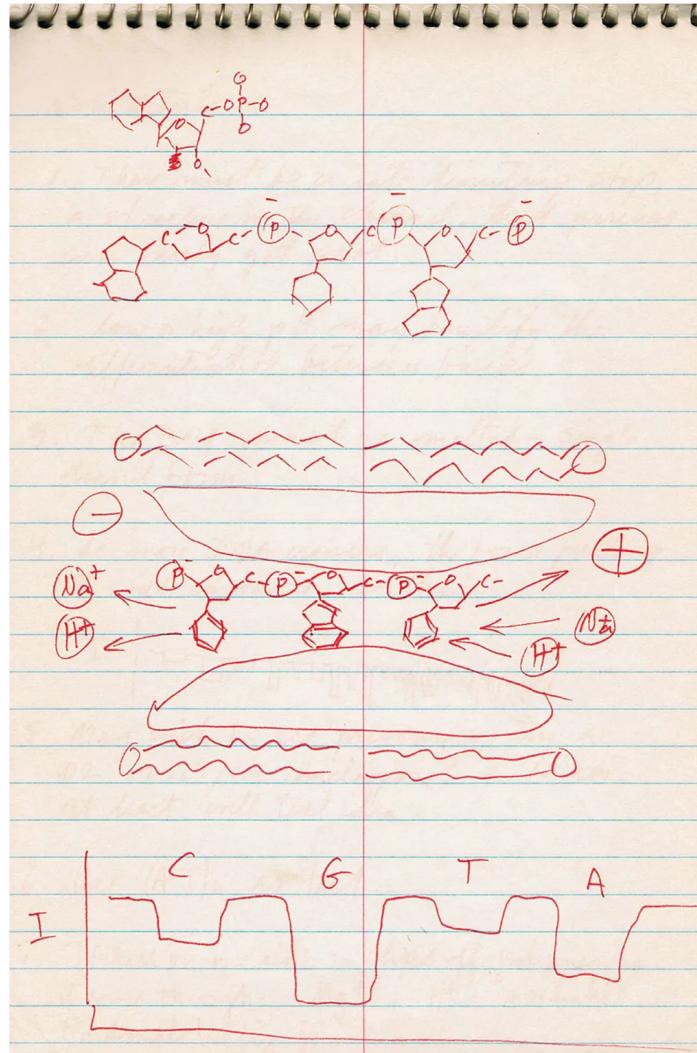
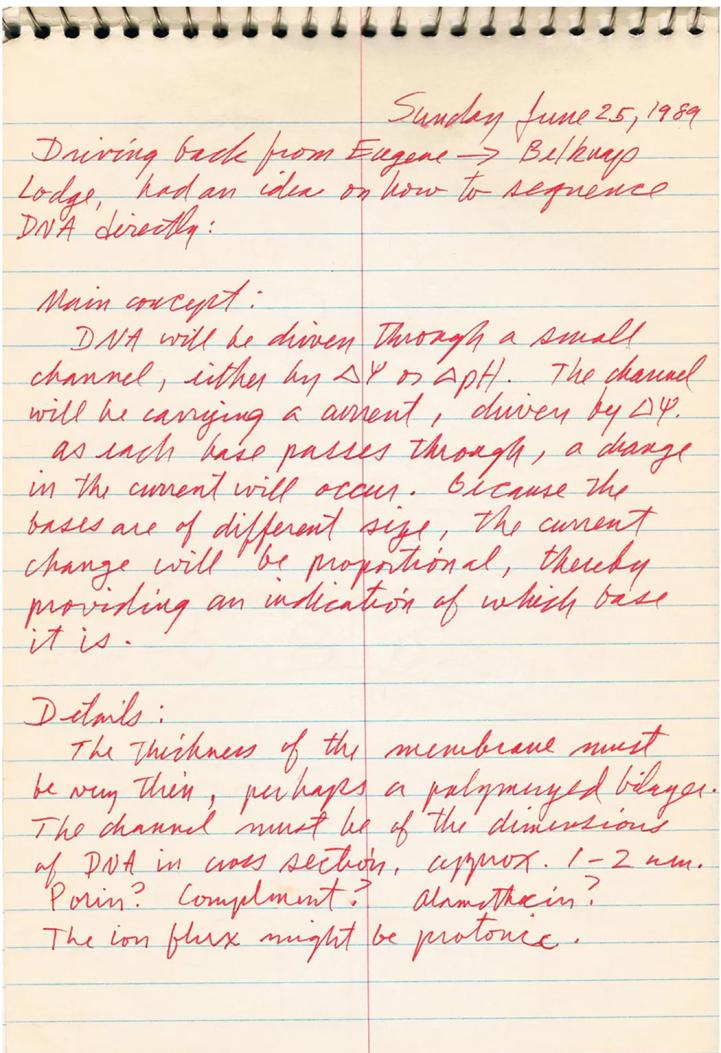
RNA modifications calling from Nanopore data

A technical overview

Tommaso Leonardi
Istituto Italiano di Tecnologia

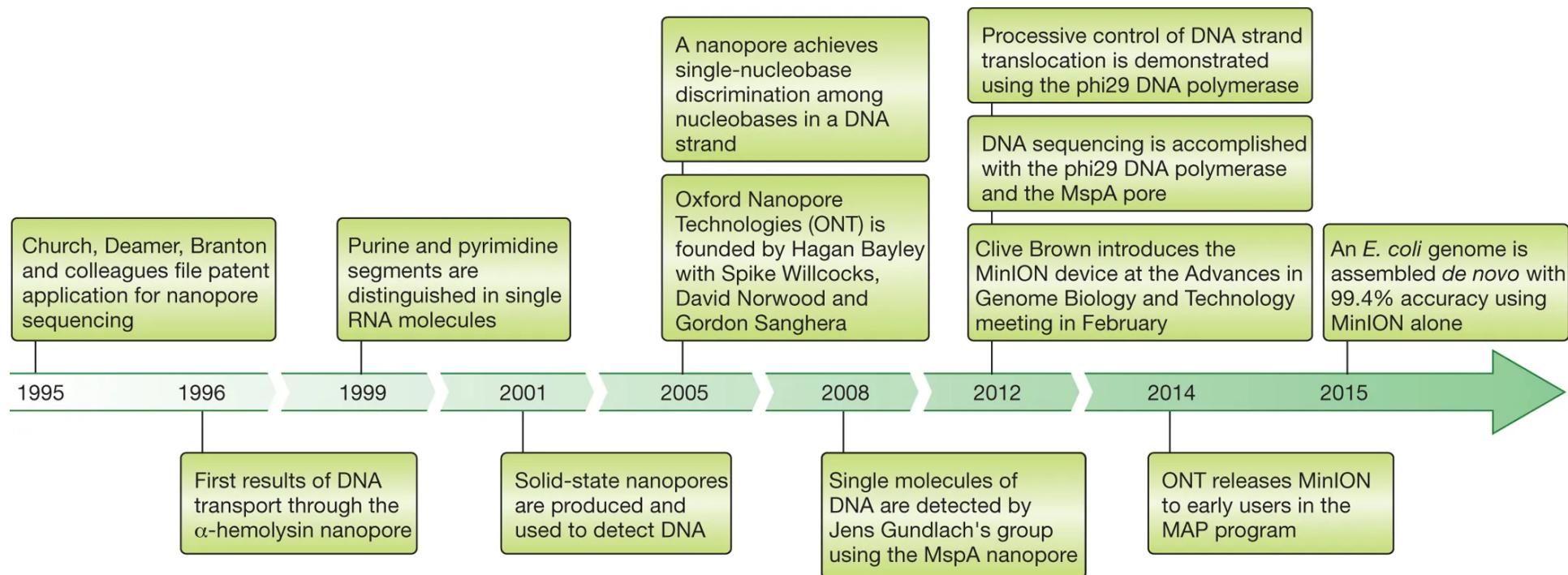
The origins of Nanopore sequencing

The origins of Nanopore sequencing



David Deamer's notebook, June 1989

ONT sequencing development



What it is, how it works



Nanopore Sequencing

Advantages

- Real-time sequencing
- Minimal sample preparation (at least in some applications)
- High portability of the sequencing platform (e.g. field sampling)



MinION: on the International Space Station, the Indian Ocean, the Arctic, Guinea (Ebola monitoring), Tanzania (frog species surveillance), Under the sea (NASA/NEEMO), Brazil (Zika surveillance)

© Copyright 2020 Oxford Nanopore Technologies

Oxford Nanopore Technologies devices are currently for research use only



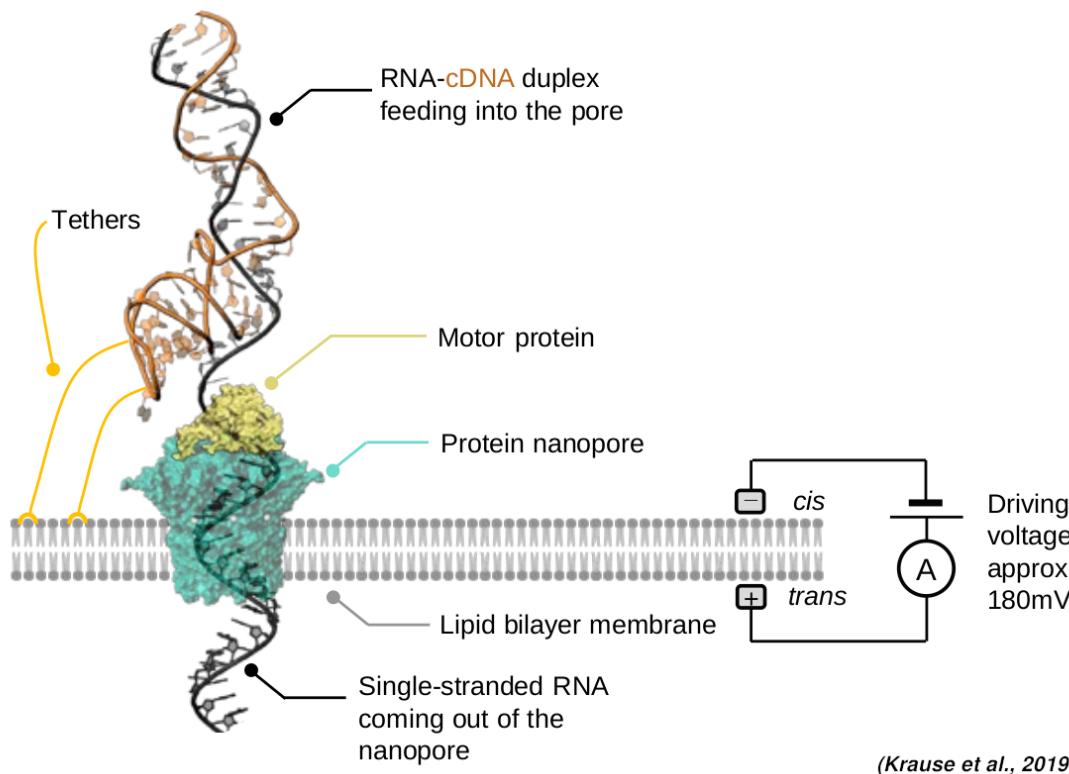
- **Ultra-long reads**
- **Direct DNA/RNA sequencing** (access to epigenetic information)

Nanopore Sequencing

Disadvantages

- Relatively low read accuracy (especially for direct RNA sequencing)
- Short reagent shelf life
- Higher sequencing costs (per coverage)
- Bioinformatic and *wet* protocols under constant development

The theory of nanopore sensing

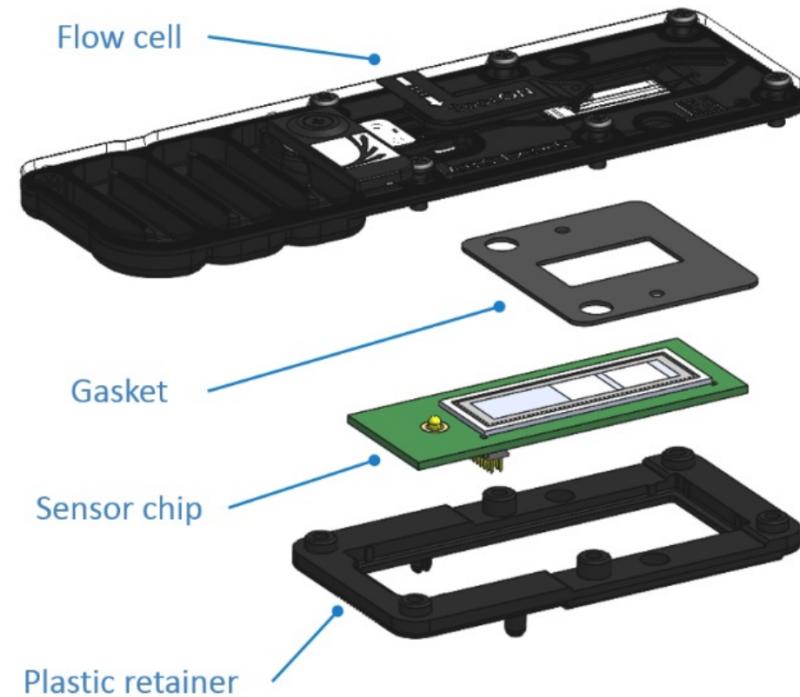
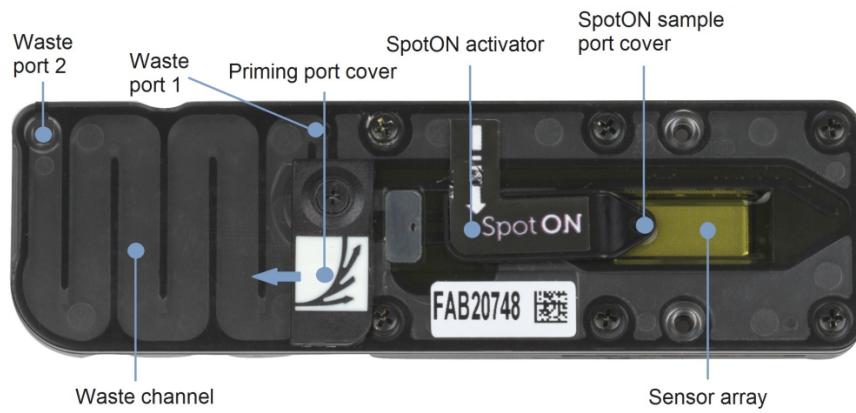


(Krause et al., 2019)

- Each base interferes with the ionic flow to a different extent
- All the bases within the pore contribute in determining the pore conductance (*Kmer* concept)
- Measurement of these current changes requires a *pico-ampere* sensitivity (10-1000) @ 2-20 kHz
- The speed of translocation is controlled by the inclusion of a *motor protein*
- A tether, incorporated during library preparation, concentrates DNA/RNA near the pore

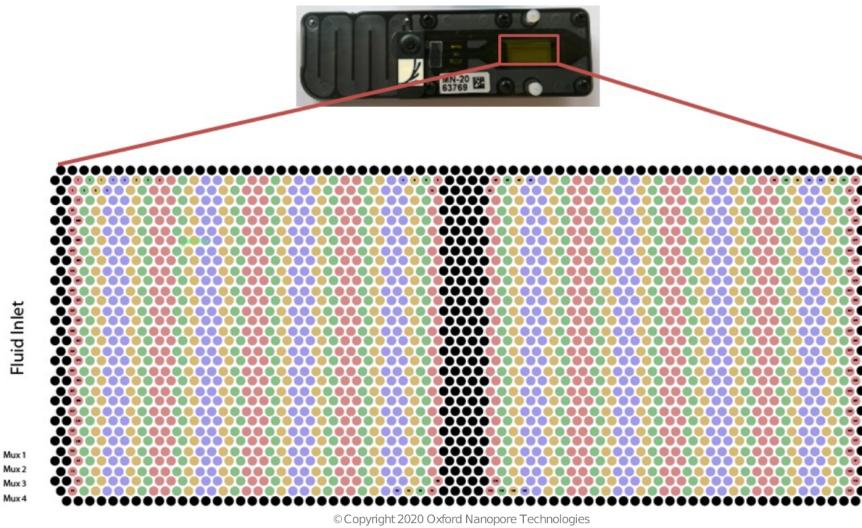
Top-down description of a MinION flow cell

The hardware assembly



© Copyright 2020 Oxford Nanopore Technologies

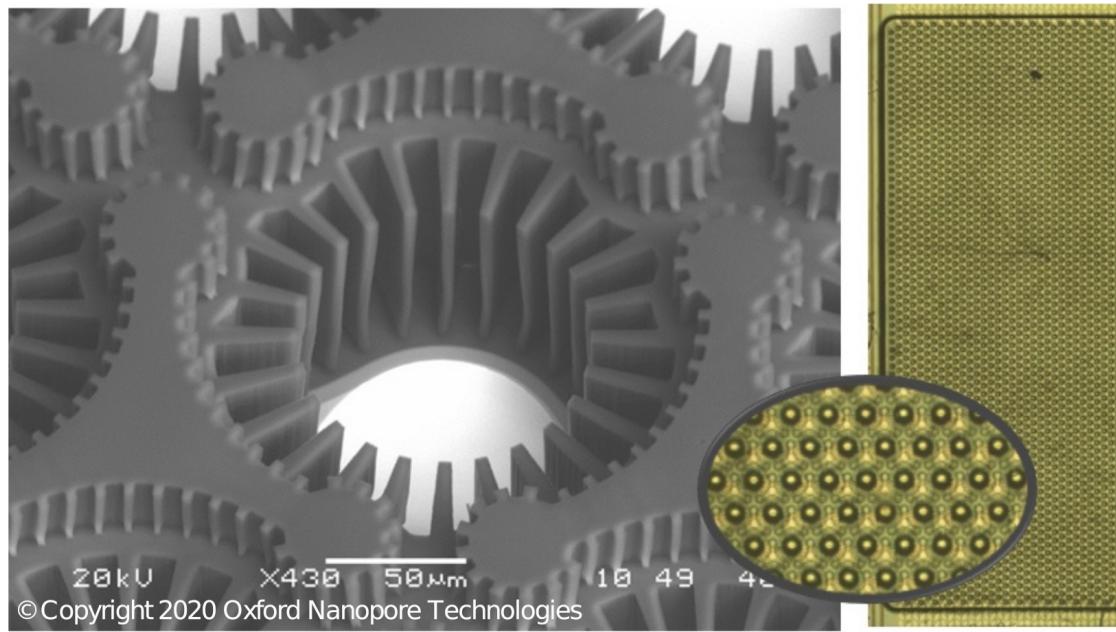
Top-down description of a MinION flow cell



Sensor Array

- 2048 active well electrodes organised hexagonally on the surface of the array
- The active wells are arranged in two blocks of 32x32
- Wells are connected to the ASIC in groups of 4 but are read one group at a time
- At most, 512 channels (pores) can be sequencing simultaneously
- The instrument periodically checks pore status and changes well if needed (mux change)

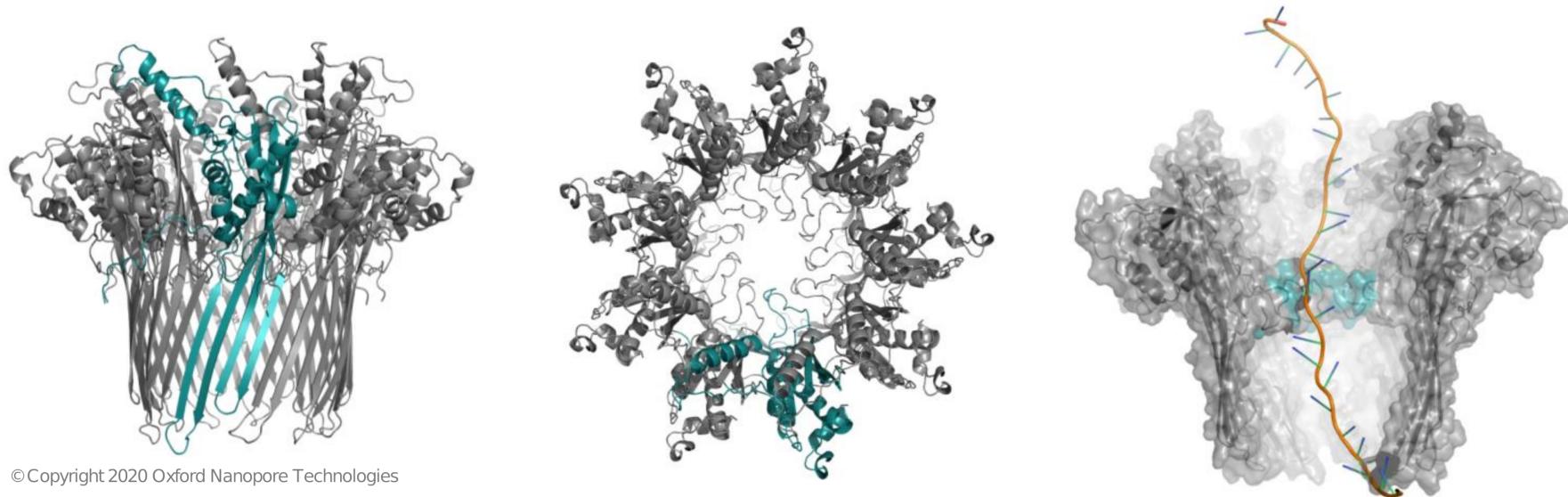
Top-down description of a MinION flow cell



Active well

- 90 μm deep micro-patterned structure with a platinum electrode at the base
- An insulating synthetic polymer membrane is formed over the array
- Each well is supposed to embed a single pore protein

Top-down description of a MinION flow cell



Channel ("pore") protein

- The current pore (version R9) is a mutant of the *CsgG* lipoprotein from *E. coli*
- The *CsgG* lipoprotein is made of 9 identical subunits which form a 36-stranded beta-barrel
- The pore has been engineered to translocate DNA instead of peptides

Applications

- Genome: DNA sequencing with ultra long reads
- **Transcriptome:** cDNA and direct RNA sequencing of full-length transcripts
- Proteins: feasibility recently demonstrated

Nanopore long-read RNA sequencing

3 different flavours:

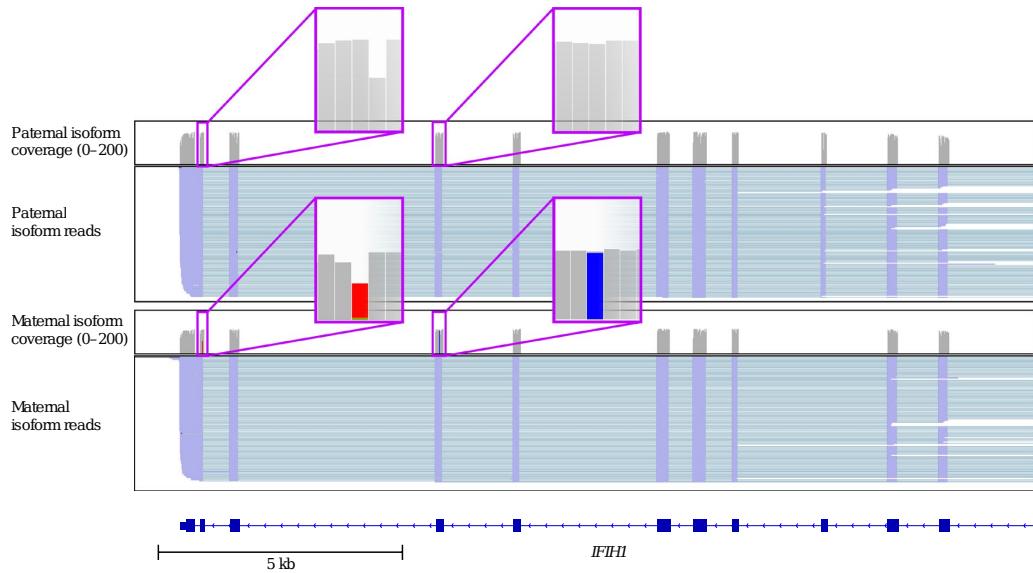
- Direct RNA Sequencing
- Direct cDNA Sequencing
- cDNA-PCR Sequencing

Advantages of *Direct* RNA sequencing

- No PCR amplification-derived biases (e.g. GC-content, length)
- No RT biases (e.g. certain transcripts may be difficult to reverse-transcribe)
- RNA modification detection

dRNA Sequencing applications

- Differential Gene Expression (DGE)
- Differential Transcript Usage (DTU)
- Transcriptome assembly
- Genome annotation
- Structural analyses (fusion transcripts)
- Epitranscriptomics
- Viral Genome Sequencing



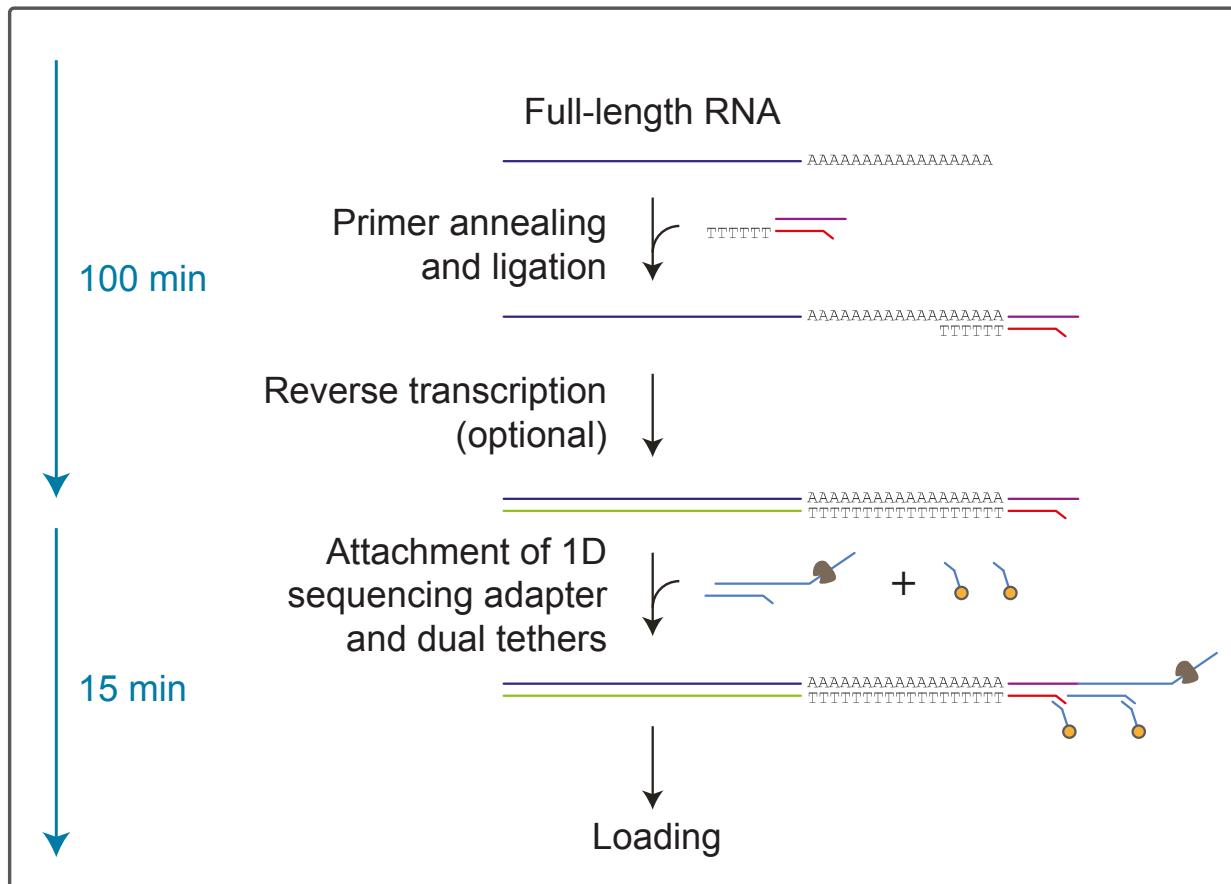
Workman et al., Nat Meth 2019

"Much of what we know about RNA biology was seen through a cDNA lens"

Matthew Keller, CDC

RNA-seq Library Prep Strategies

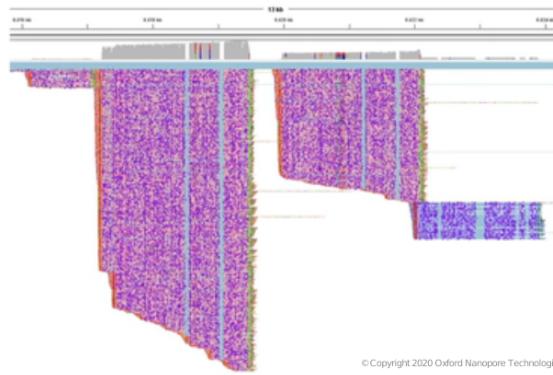
Direct RNA Sequencing



Input Samples for dRNA-Seq

Recommended input RNA:

- Full length (RIN>7)
- 500ng
- Size >200nt
- poly(A)+ (oligo-dT selected)

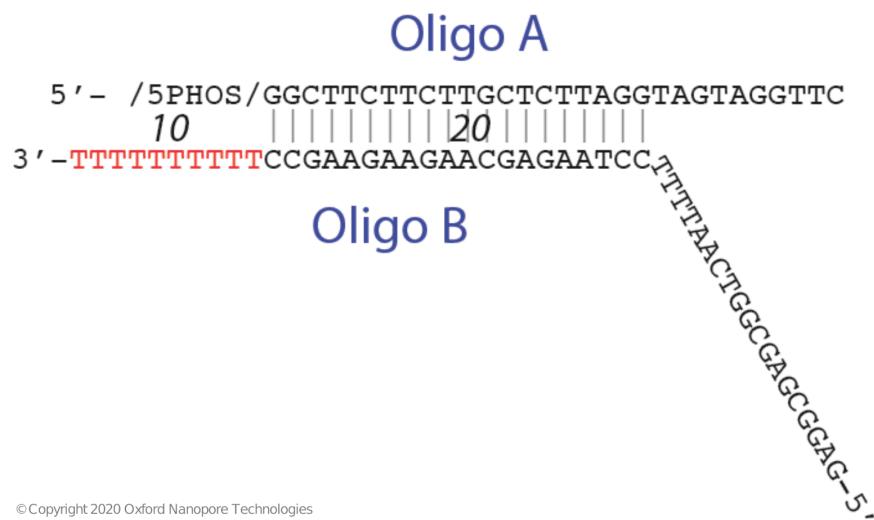


© Copyright 2020 Oxford Nanopore Technologies

Alternative approaches for non-poly(A) species:

- Enzymatic poly-adenylation (*E. coli* poly(A) polymerase)
- Targeted sequencing

Sequence-specific (targeted) Sequencing



© Copyright 2020 Oxford Nanopore Technologies

Targeted Direct RNA Sequencing:

- The Reverse Transcription Adapter (*RTA*) is modular
 - Users can design their own RTA to target a specific RNA sequence (e.g. *16S rRNA*)
 - Target molecule needs to terminate precisely at its 3' end to allow ligation
 - Not suited for poly(A)+ RNA targets

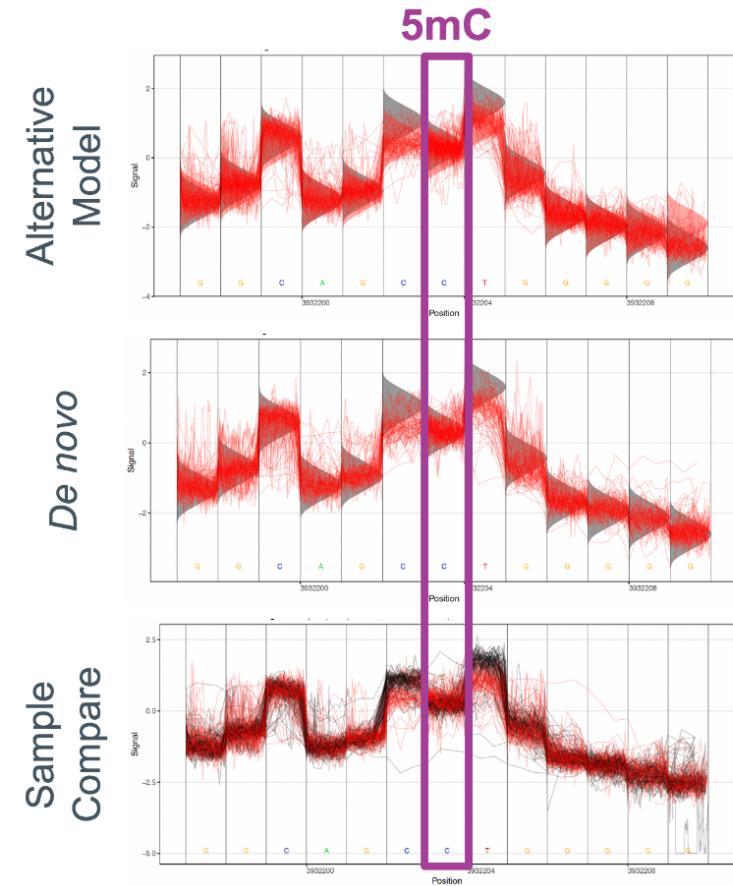
RNA modification detection from dRNA-Seq data: Tombo

nanoporetech/tombo (doi:10.1101/094672)

MODIFIED BASE DETECTION

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

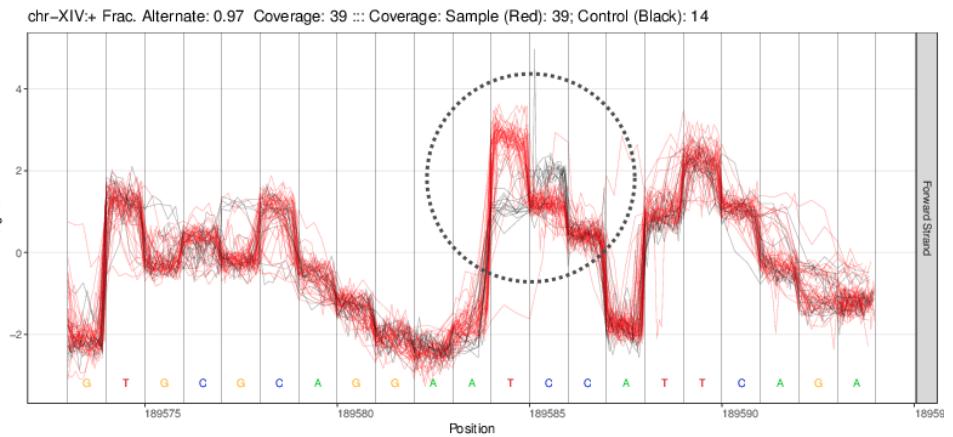
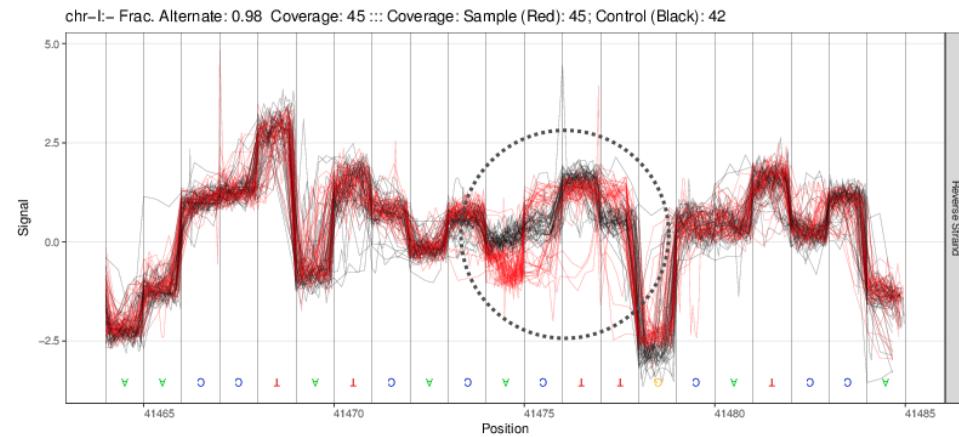
	Advantages	Disadvantages
Alternative Model	<ul style="list-style-type: none">Known alt. baseExact alt. loc.Good accuracy	<ul style="list-style-type: none">Requires alt. model estimation
<i>De novo</i>	<ul style="list-style-type: none">Apply to any sample	<ul style="list-style-type: none">High error rateInexact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



RNA modification detection from dRNA-Seq data

- Sequence based methods
 - Differential rate of errors (e.g. Eligos2, Epinano)
 - Model based (e.g. Epinano)
- Signal based methods
 - Differential signal (e.g. Nanocompore)
 - Model based (e.g. Tombo)

Comparative identification of RNA modifications



- Compare a sample to a matching "unmodified" control
- Without any modifications: Synthesized or in vitro transcribed RNA
- Specific RNA modification(s) targeting: RNA mod writer KD or KO
- Proportion of reads modified and efficiency of the KD or KO
- SNPs also induce significant signal shift -> No SNPs between the 2 conditions.

Nanocompore

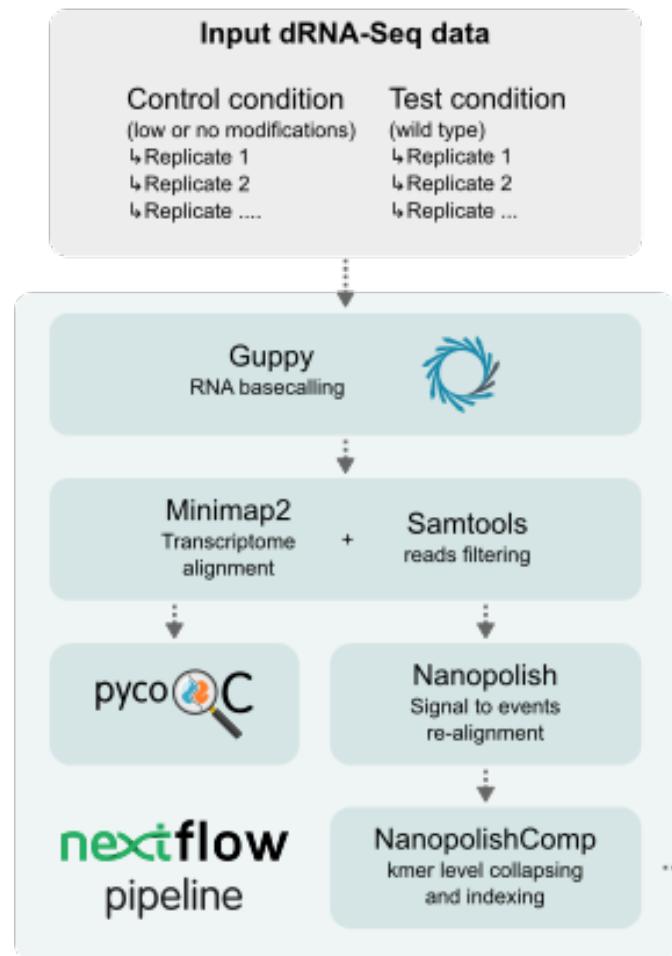
[tleonardi/nanocompore \(doi:10.1101/843136\)](#)



- 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.)

How to go from the raw data to modification calls

- Basecall the fast5 files
- Map to the transcriptome
- Filter alignments
- Resquiggle with Nanopolish
- Collapse and index results with eventalign_collapse
- Analyse with Nanocompare



Leger et al., bioRxiv 2019

Questions?

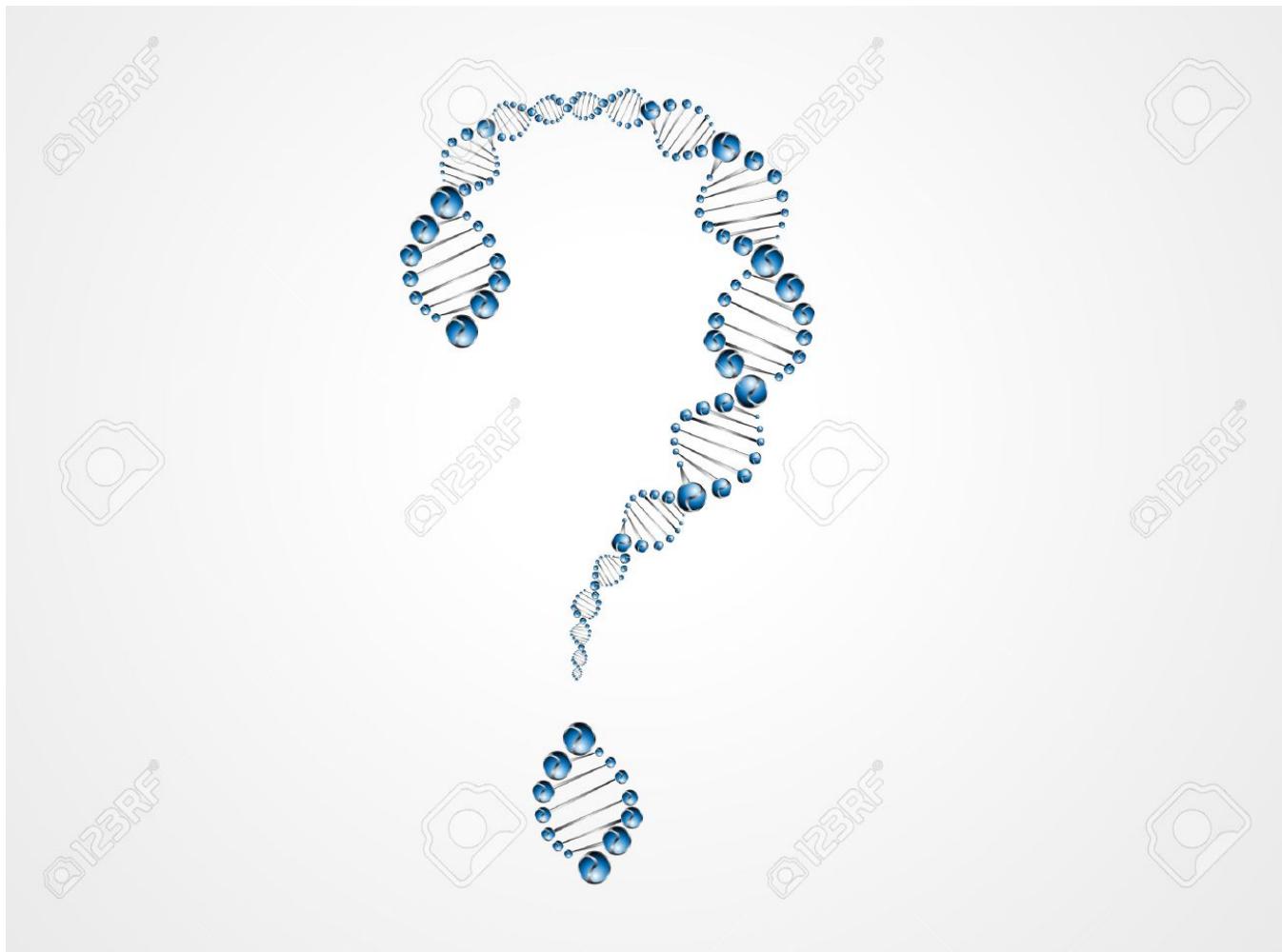


Image credit: 123rf.com