RNA modifications: traces in the ONT signal

39th TBI Winterseminar in Bled

Jannes Spangenberg
Dr. Christian Höner zu Siederdissen
Dr. Manja Marz

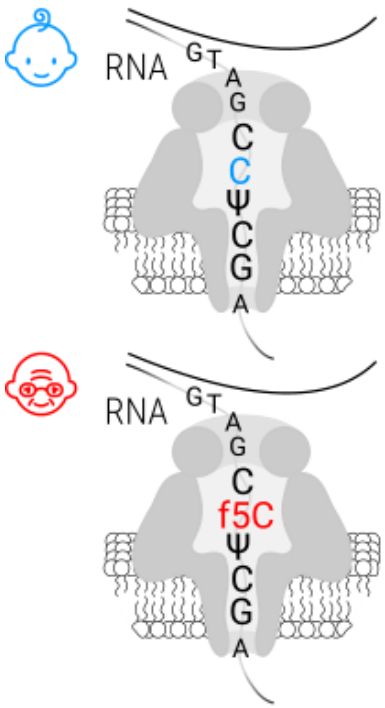
Outline

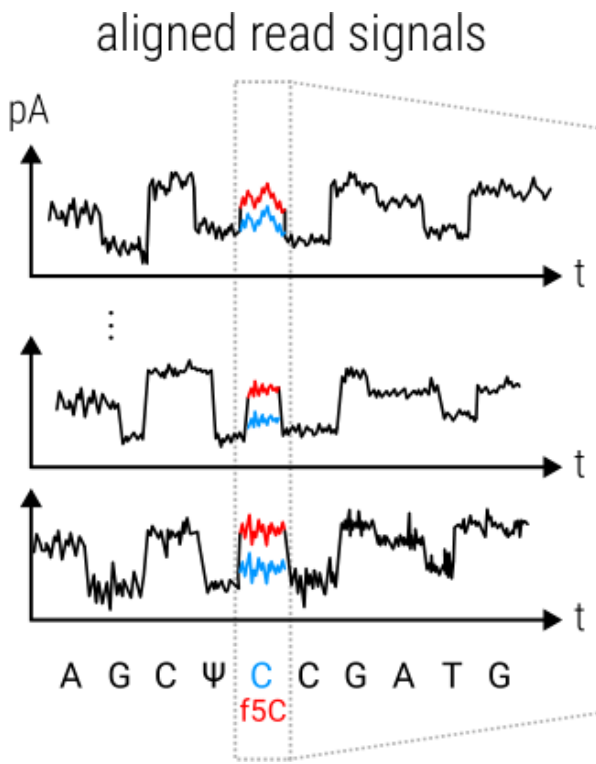
- Differential signal analysis with Magnipore
- IVT experiments with modifications

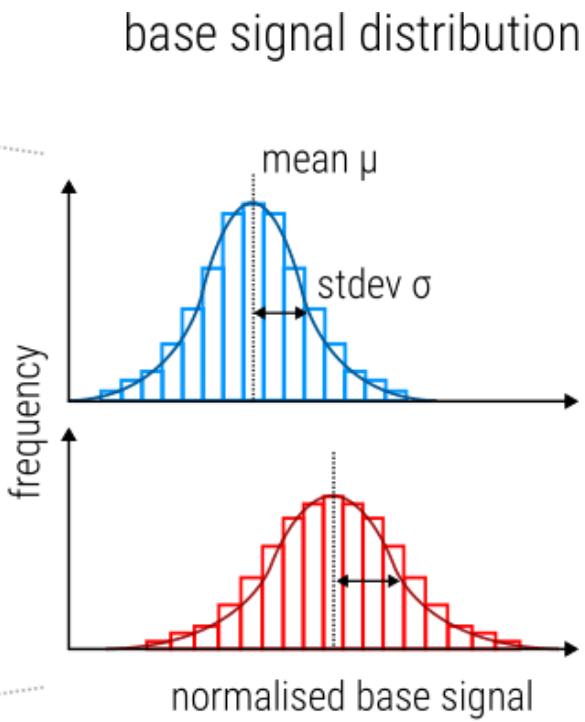
Nanopore

Nanopore

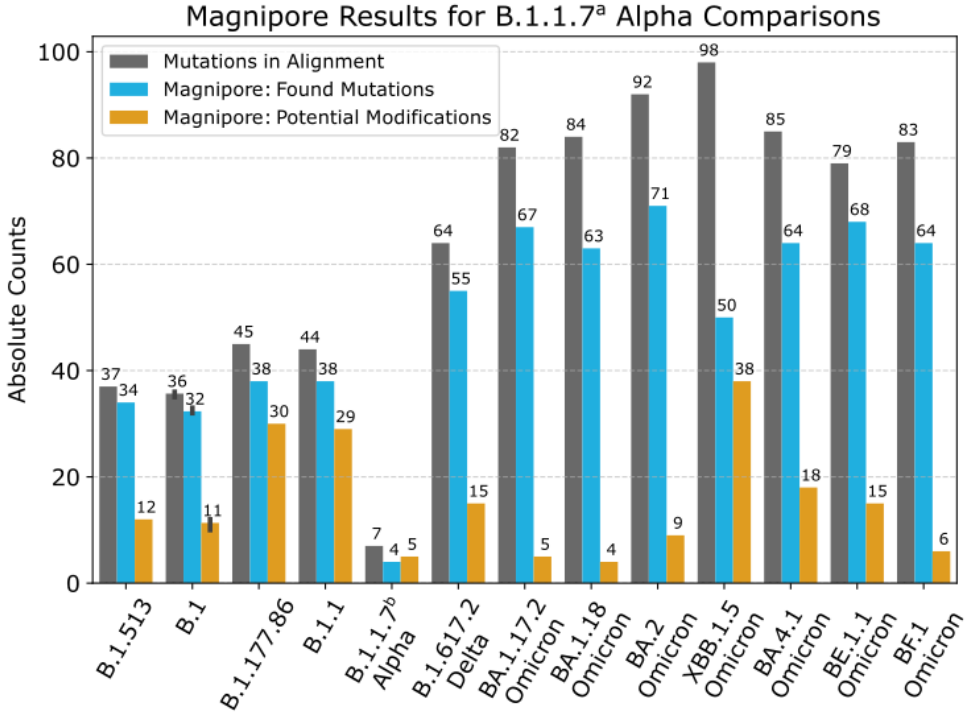
Mognipore pipeline







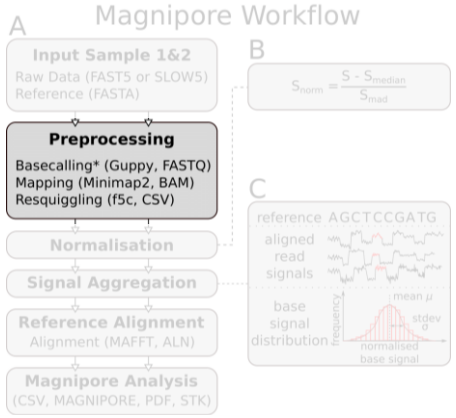
Can Magnipore detect all mutations?

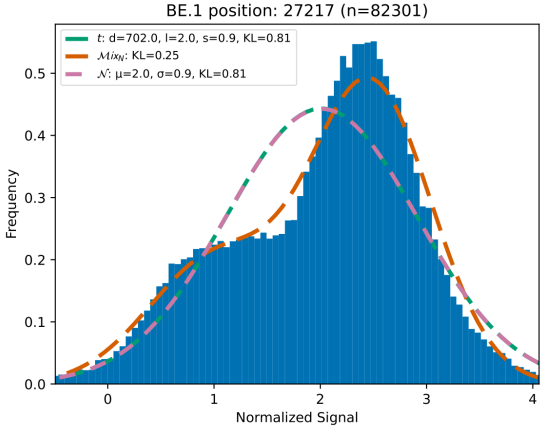


- 136 pairwise comparisons of 16 samples
  - 13928 mutations
  - 12359 mutations found (88.7%)
- Undetected mutations are especially substitutions between
  - majorly C : U (Pyrimidine)
  - A : G (Purine)

Magnipore Roadmap

- Currently: evaluating Magnipore with a ground-truth dataset
- Improve signal segmentation
  (and resquiggling/basecalling-error-correction)
- Use other distribution models (mixture, t-distribution)
- Extent model (add parameters like segment length distribution)
- Allow pod5 format as input (possible if f5c updates to allow pod5 or segmentation algorithm is replaced)



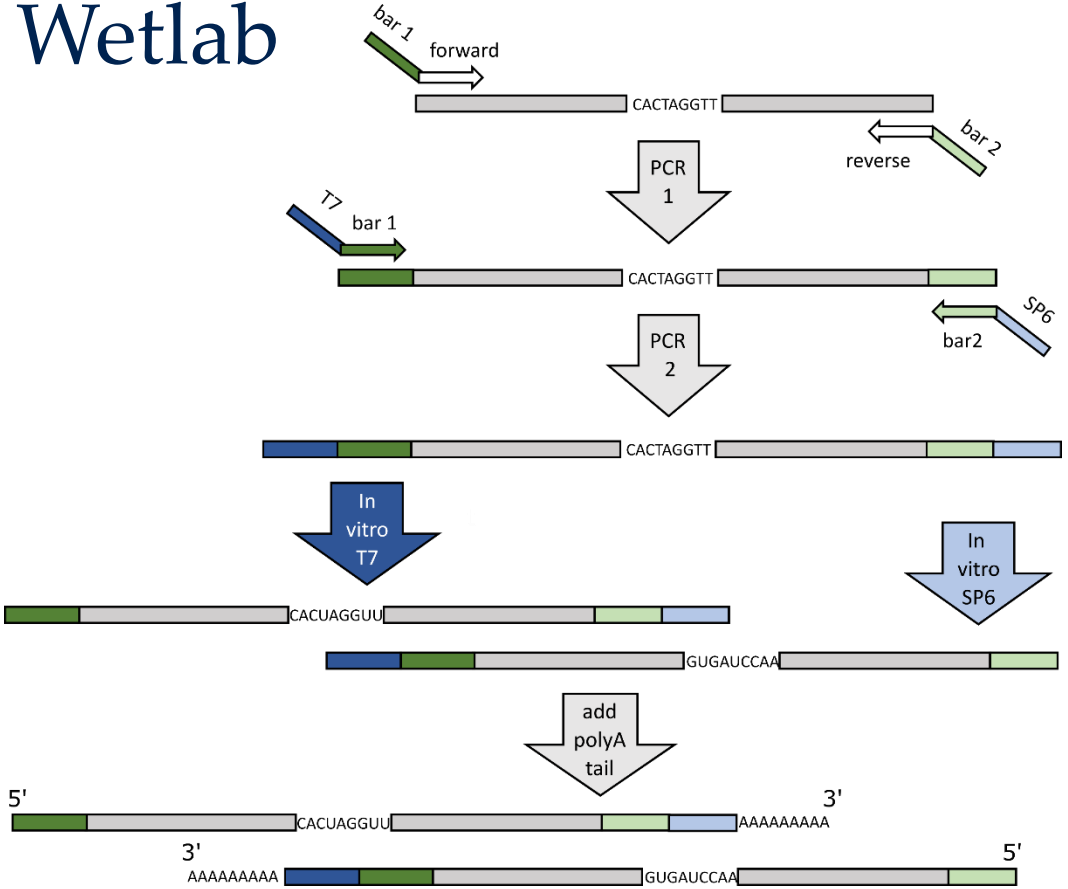


https://github.com/JannesSP/magnipore

https://anaconda.org/jannessp/magnipore

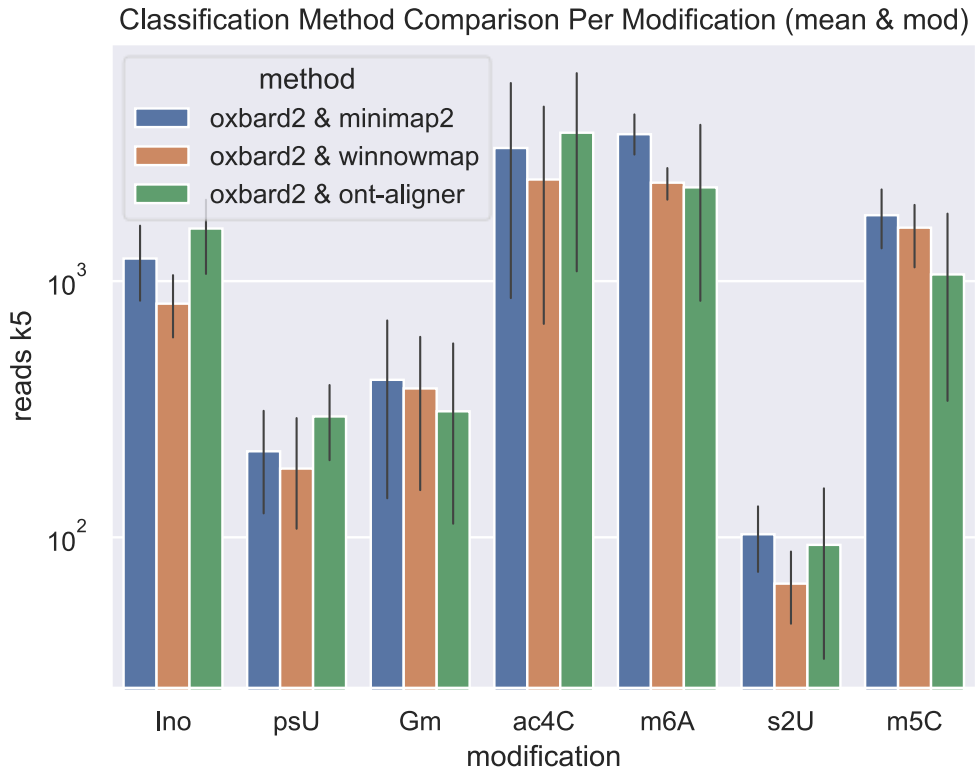
IVT experiments with modifications

Wetlab



1. Order designed DNA templates with desired sequence context / motifs (9 templates * 2k bases)
2. Add 2 barcode sets
   a. Transcription with modifications
   b. Transcription without modification
3. Add T7 Promotor and SP6 Promotor for strand specific transcription
4. Transcribe desired strand with desired nucleotides and modification
5. Prepare everything for ONT sequencing

Drylab



- Reads for 7 different modifications
  - Inosine
  - Pseudouridine
  - 2'-O-methylguanosine
  - N4-acetylcytidine
  - N6-methyladenosine
  - 2-thiouridine
  - 5-methylcytidine

- The mentioned problems reduce the number of classifyable reads immensely

e.g. Pool1:
Total: 308890 reads
Classified as mod: 19182 (~6%)
Classified as can: 89004 (~29%)
Unclassified: 200704 (~65%)

Drylab

1. Sequence modified and unmodified sequences

2. Identify the barcode set -> modified or unmodified?
    a. Currently align barcodes to barcode regions of each read

Problems:

- Did we order the templates correctly? (mostly yes)
- Did we design good primers? Did the PCR work? (meh, some dimerization, misannealing)
- Modifications can introduce more or less basecalling errors
- The basecaller eats away bases from 5‘ and 3‘ end, where the barcodes are
- …

Take Home Messages

1. ONT direct RNA sequencing is not that easy

2. Switch to R10, new chemistry, pod5 and Dorado (basecaller server) ASAP

3. A lot of improvements need to be done regarding ONT signal processing

Thanks to:
The RNA Bioinformatics/High-Throughput Analysis Team of the Friedrich Schiller University Jena
Website: www.rna.uni-jena.de

Thanks for your attention!

RNA BIOINFORMATICS & HIGH-THROUGHPUT ANALYSIS

Funding (in part):

- German DFG Collaborative Research Centre AquaDiva (CRC 1076 AquaDiva)
- The German state of Thuringia via the Thüringer Aufbaubank (2021 FGI 0009)
- Carl-Zeiss-Stiftung within the program Scientific Breakthroughs in Artificial Intelligence (project “Interactive Inference”)
- Ministry for Economics, Sciences and Digital Society of Thuringia (Landesprogramm ProDigital, DigLeben-5575/10-9)
- Durchbrüche "Werkstatt" FKZ0563-2.8/738/2

Supported by Oxford Nanopore Technologies

Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation

AquaDiva

Thüringer Zentrum für
Lernende Systeme und Robotik