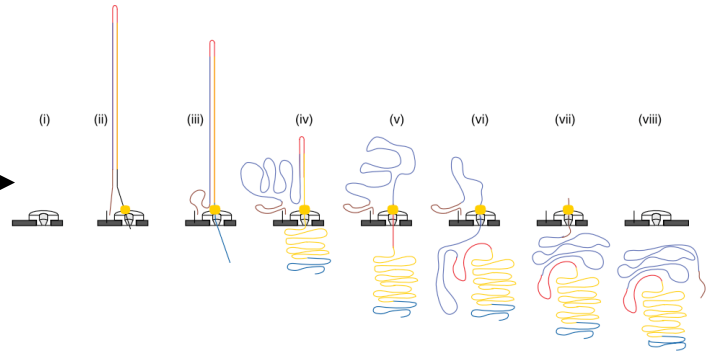
Bioinformatics Lab - Computational Methods for 4th generation Sequencing

Ivan Gesteira Costa
IZKF Research Group Bioinformatics

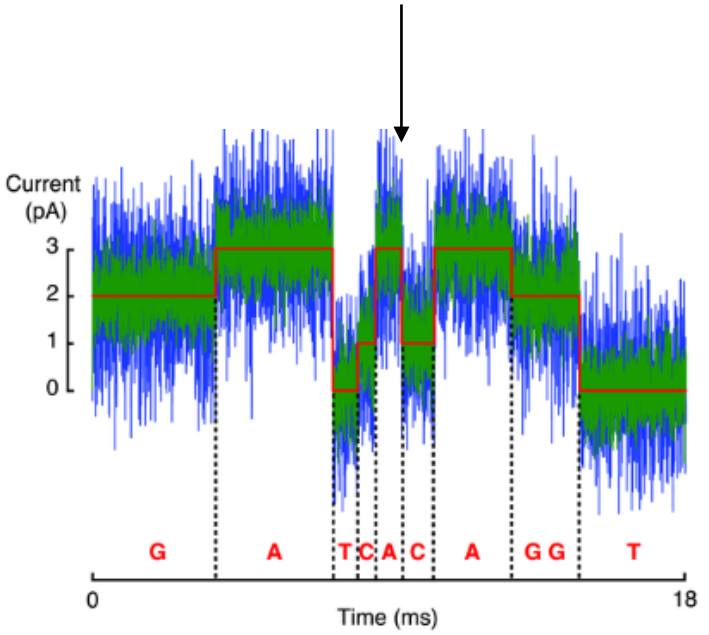
IZKF Interdisziplinäres Zentrum für Klinische Forschung | RWTH AACHEN UNIVERSITY

Oxford Nanopore Minion - Overview



Characteristics:

- cheap equipment and portable
- low throughput (1GB)
- long reads (100 kb) but high error rate (13%)
- applications: infection agents in point-of-care; target sequencing, novel genomes. DNA methylation …



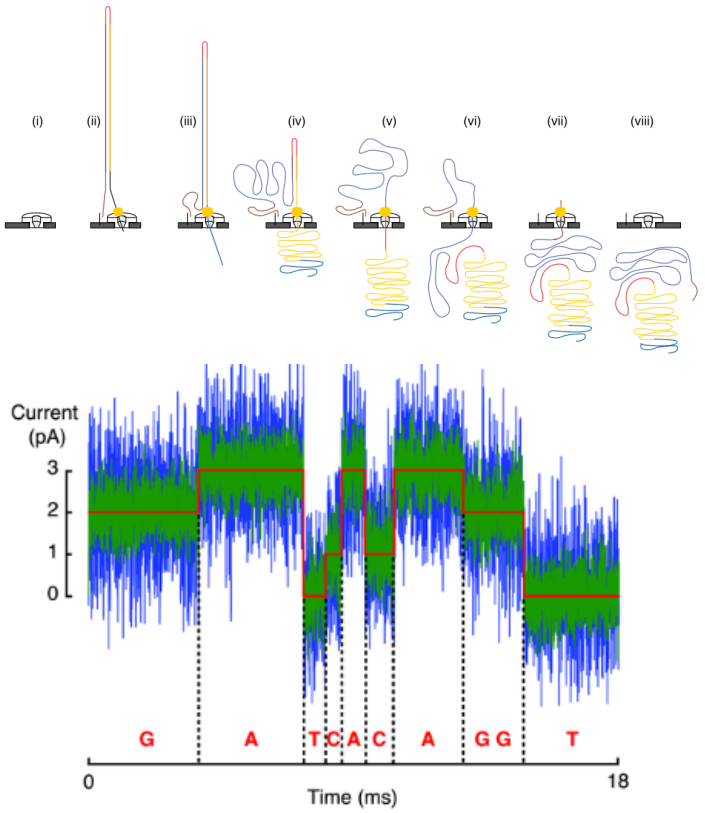
Oxford Nanopore Minion - Base calling (1)

Base calling:

1. the pA value determines the base.
2. duration of a base passing through the pore varies
3. Two run modes:
   - 1D (one strands)
   - 2D (two strands)

Tools available:

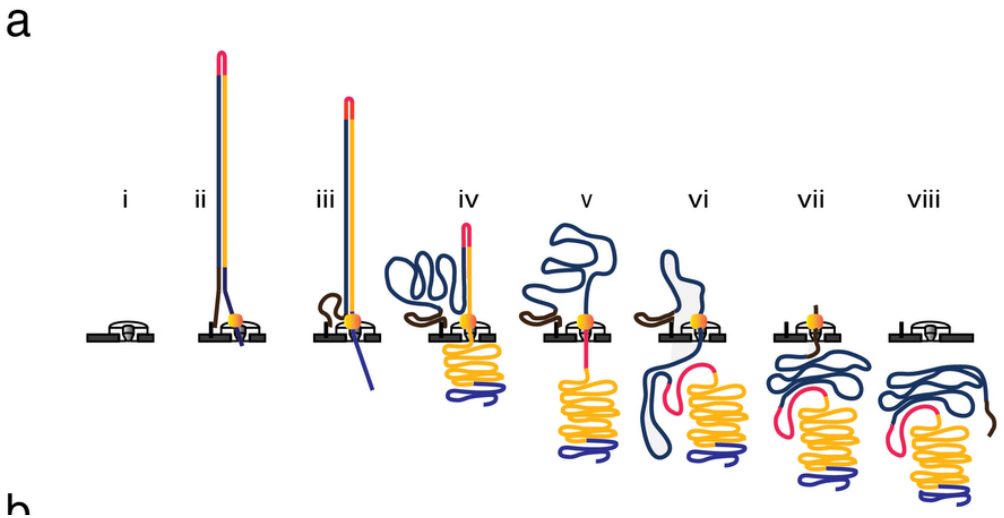
- MinKNOW (1D only - official)
- Metrichor (2D - official/cloud based)
- Nanocall (HMM based)
- DeepNano (Deep learning)



David M, Nanocall: An Open Source Basecaller for Oxford Nanopore Sequencing Data. bioRxiv 2016;33. 046086.

Boža V, Brejová B, Vinař T. DeepNano: Deep Recurrent Neural Networks for Base Calling in MinION Nanopore Reads. arXiv 2016:1-12.

Oxford Nanopore Minion - Base calling (2)



Read Structure: 2D mode

- lead adaptor (blue)
- template strand (gold)
- hairpin adaptor (red)
- complement strand (dark blue)
- trailing adaptor (brown)

Important points

1. adaptors sequence is given
2. complement strand should agree with template strand

Base calling - Nanocall

Hidden Markov model

- each state for each 6 mer

- transitions only allowed to "valid" shifts

i.e. CGACAT -> GACATA, GACATC, …

shifts of 2,3 are also possible

(with smaller probability)

- emissions emits mean(μ)/variance (σ)

of a pA for a particular base

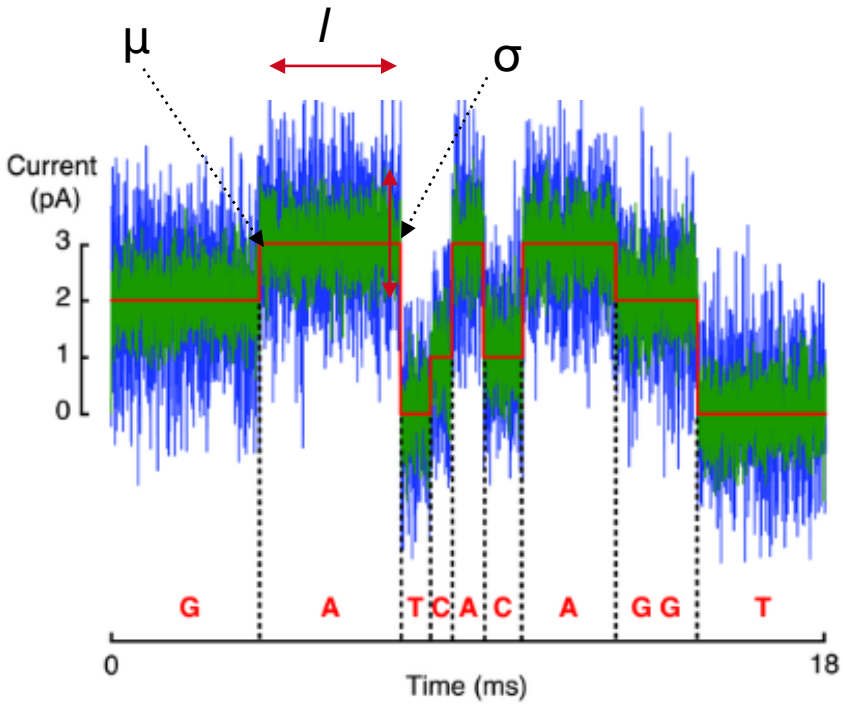
- μ is modelled by a gaussian distribution

- σ by a negative gaussian distribution

- length *l* is not used (stochastic ???)

- global pA values are-scaled/

normalized previous to analysis



Base calling - Nanocall

Caveats:

- based on Viterbi path

- no shred scores are provided

- only 1D calls are possible

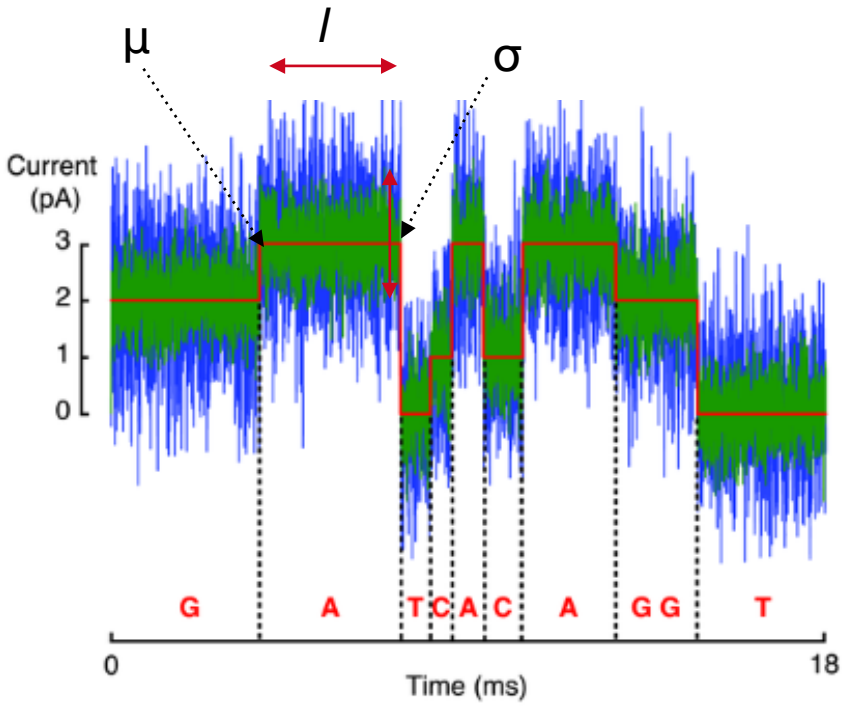
Ideas:

- use posterior decoding

- evaluate use of other k-mers

- use length data

- combine posterior with alignment to do 2D calling



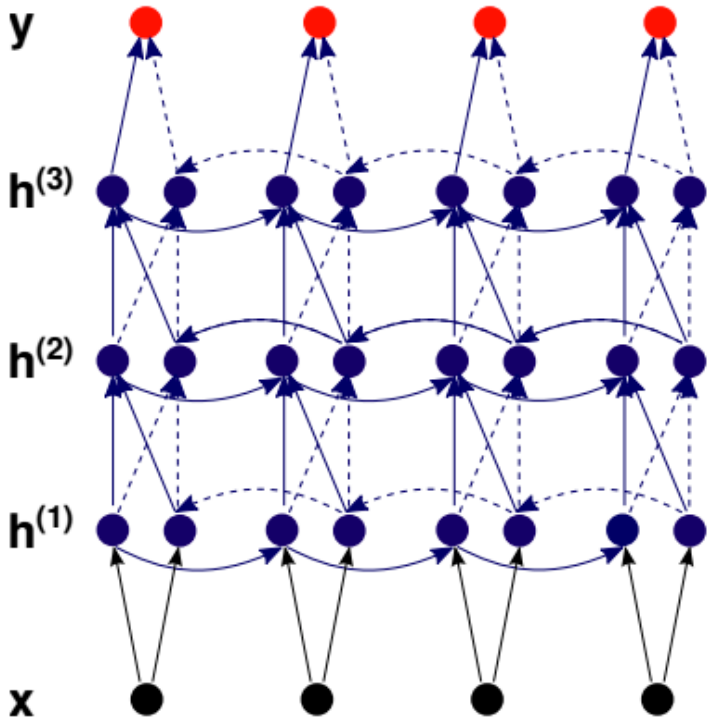
Base Calling - Deep Nano

Model: Bi-directional Recurrent Neural Network

Training: stochastic gradient descent (SGD) combined with Nesterov momentum

Labels: genomic sequences of alignments

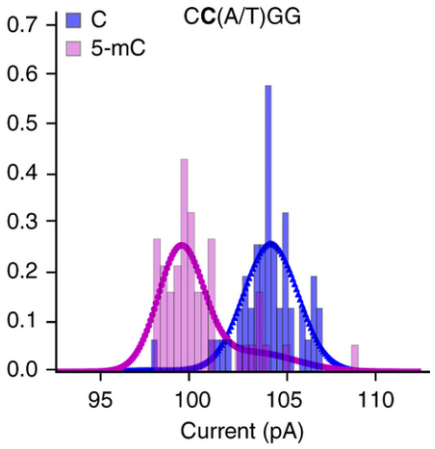
Caveats: only works as a post-processing step after me tricolor



Oxford Nanopore Minion - DNA methylation

Detection of DNA methylation:

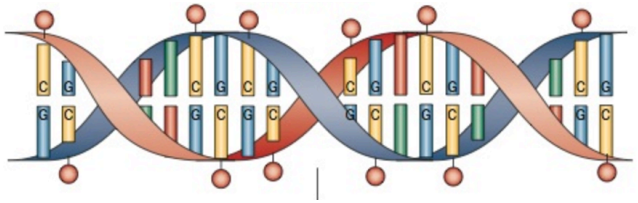
- minION allow direct detection of DNA modifications
- changes in pA values



Tools available:

Nanopolish - HMM based/expands nanocall

SignalAlign - variable order HMM & Bayesian estimation



Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W. Detecting DNA cytosine methylation using nanopore sequencing. Nat Meth 2017;advance on(April 2016):1-7.

Rand AC, Jain M, Eizenga JM, et al. Mapping DNA methylation with high-throughput nanopore sequencing. Nat. Methods 2017;(C).

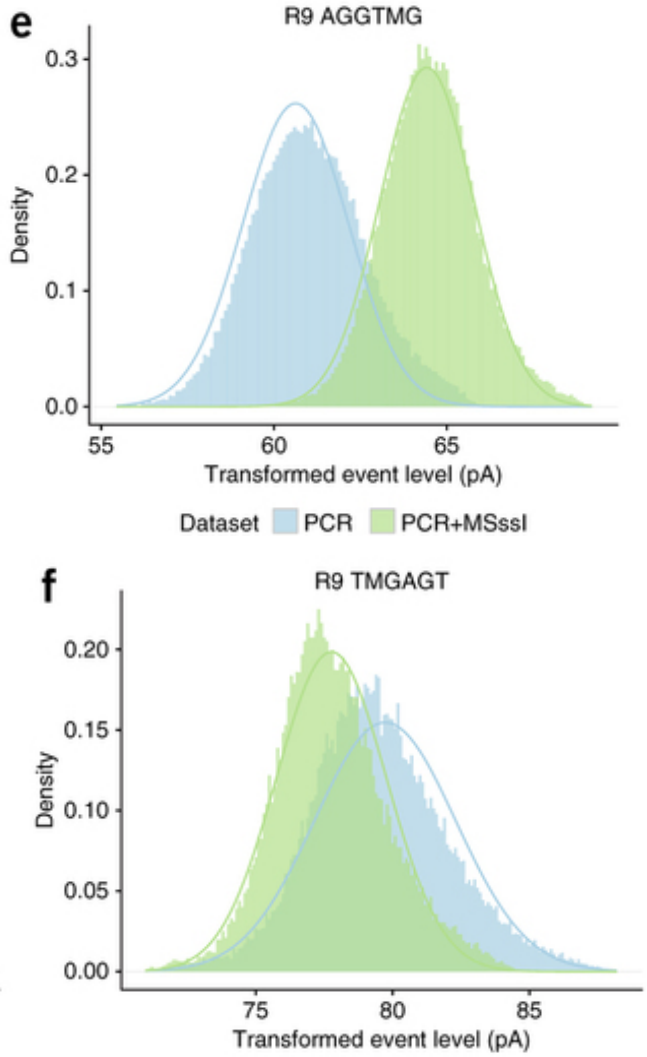
DNA methylation - Nanopolish

Nanopolish - expanded the 6-mers with met. Cs (M).

Data - *Escherichia coli* DNA before after methyltransferase M.SssI treatment

Caveats:

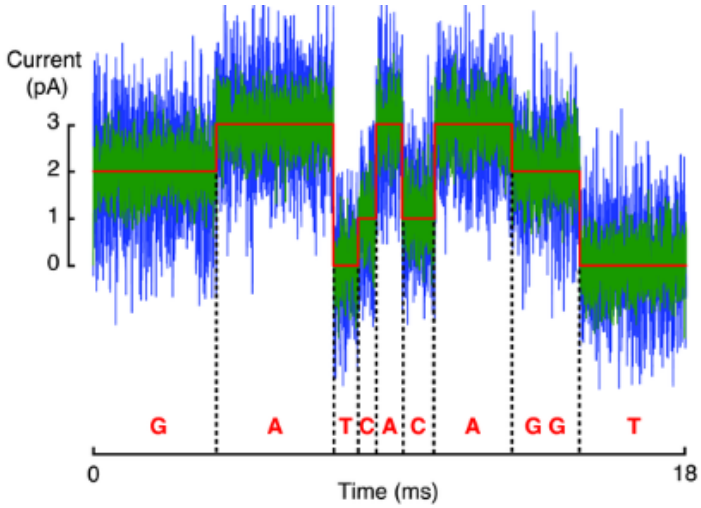
data with M.sssI treatment was hypermethylated / high local error (which C was methylated) - misses other modifications (5-hmC[8], 5-formylcytosine 5-fC) …

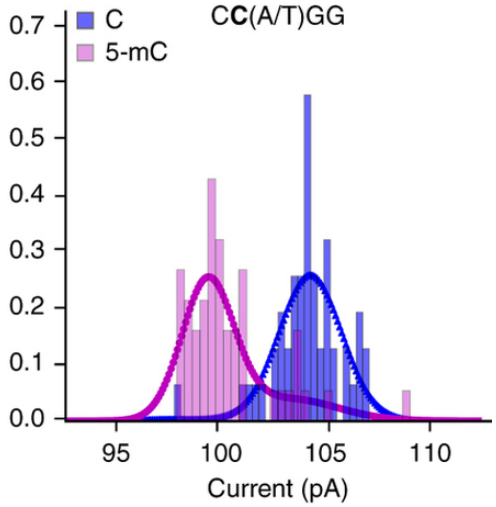


Oxford Nanopore Minion - Problems

Detection of DNA methylation:

- use machine learning (Neural networks)
- think about strategies to train data





Project - To dos

1. Methods to improve error detection
   - i.e. now phred scores (error rate) are not provided
2. Detection of DNA methylation
   - use of machine learning (instead of generative models) for detection of DNA methylation
3. Methods for sequence alignment …

Thank you!