Nanopore Community Meeting 2016
New York

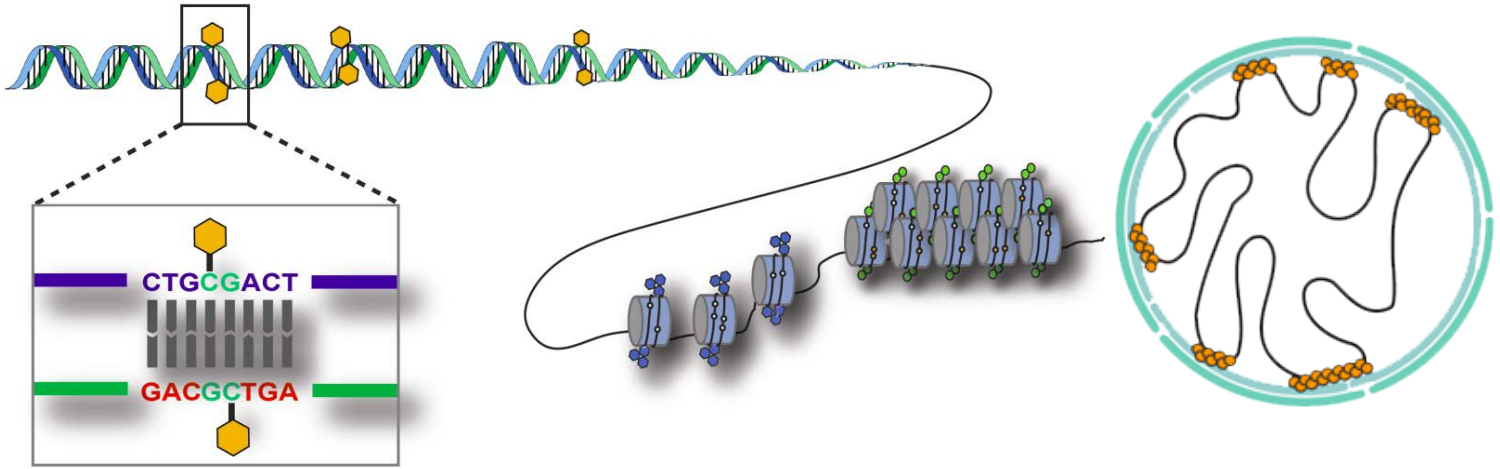
Oxford NANOPORE Technologies

JOHNS HOPKINS
BIOMEDICAL ENGINEERING

Measuring DNA Methylation with the MinION

Winston Timp
Department of Biomedical Engineering
Johns Hopkins University
12/1/16

Epigenetics: Modern

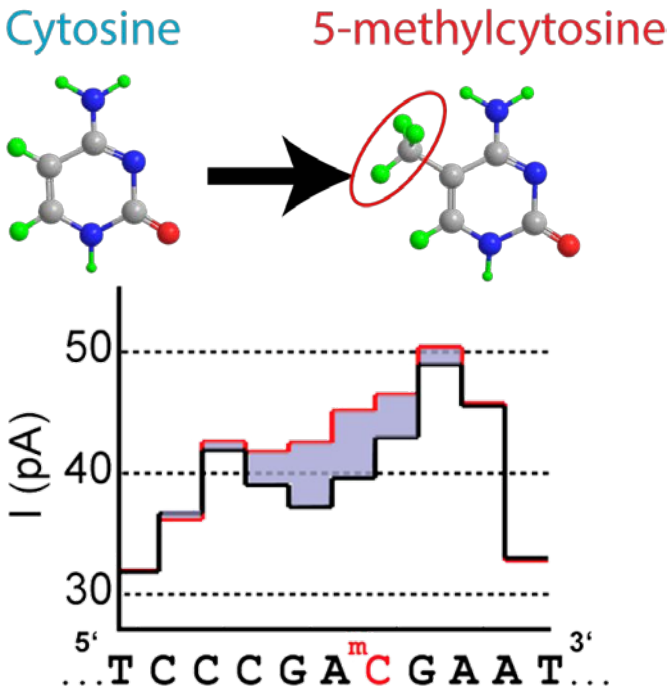


- Modern Definition of epigenetics involves heritable changes other than genetic sequence, e.g., positive feedback, high order structure, chromatin organization, histone modifications, DNA methylation.
- An analogy to a computer system:
  - DNA Sequence = Hardware
  - User input = Environment
  - Systems Biology = Running programs
  - **Epigenetics = RAM**

JOHNS HOPKINS BIOMEDICAL ENGINEERING

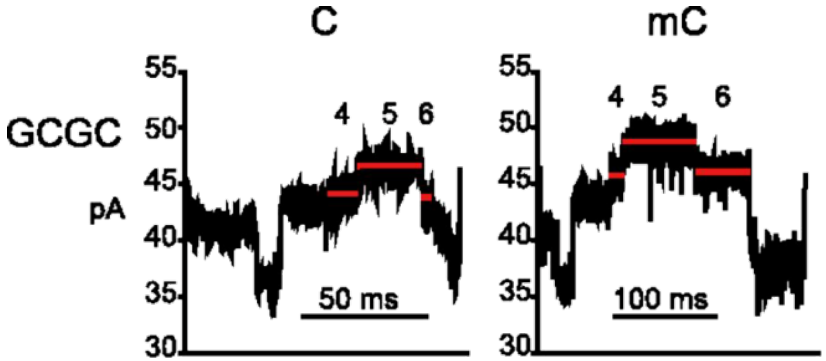
Oxford NANOPORE Technologies

Nanopore: Methylation



Laszlo, et al. *PNAS* (2013)

- Differences between methylated and unmethylated cytosine have been detected using nanopores.
- Methylation state can be called with 90% accuracy.
- We are writing a methylation detector for Oxford Nanopore for 5-methylcytosine.

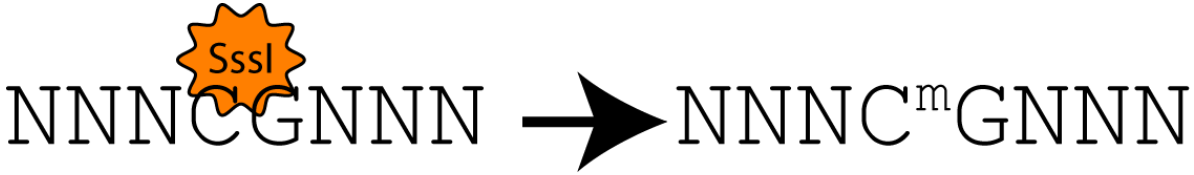


Schreiber, et al. *PNAS.* (2013)

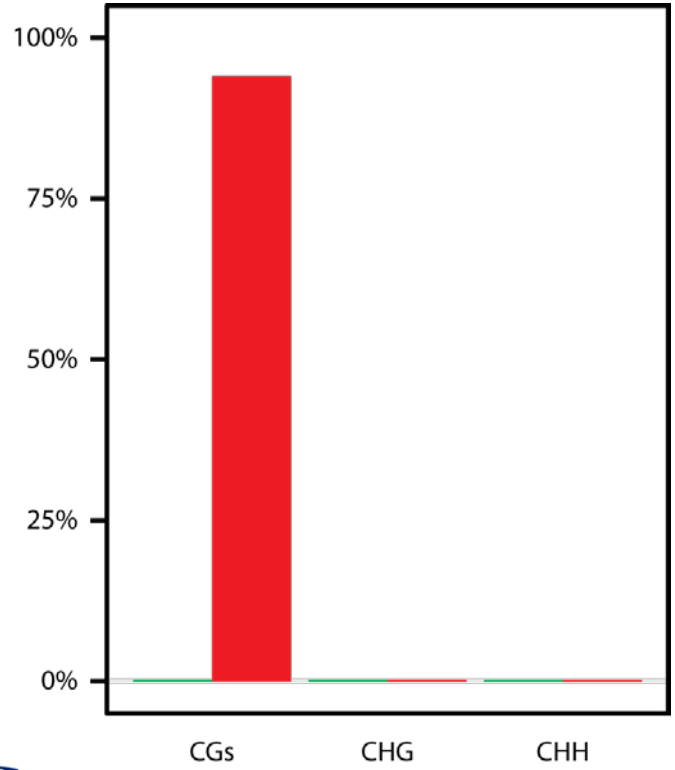
JOHNS HOPKINS BIOMEDICAL ENGINEERING

Oxford NANOPORE Technologies

Generation of methylated Samples



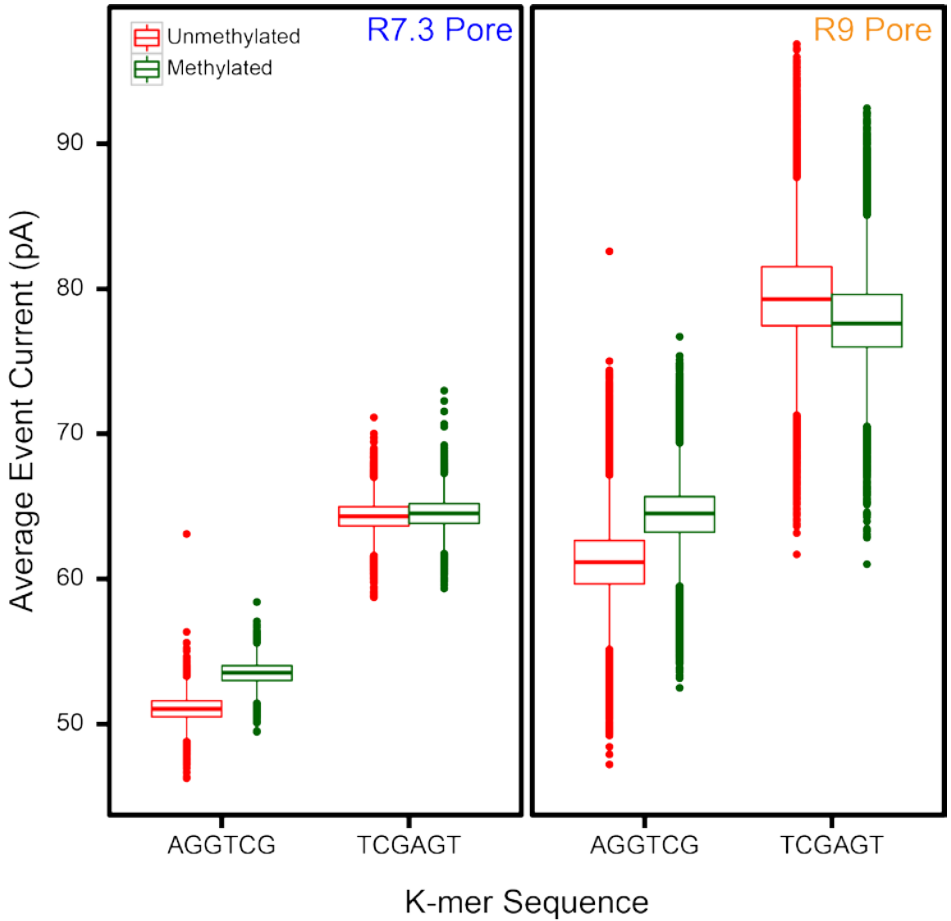
- To generated methylated samples, we treat unmethylated DNA (lambda, dam-/dcm- E. Coli, PCR product) with M. SssI methyltransferase
- We confirm the CG specific methylation using Illumina bisulfite sequencing of the sample – pictured right is methylation in different contexts E. Coli dataset treated with M. SssI (red) versus untreated (green)



JOHNS HOPKINS BIOMEDICAL ENGINEERING
Oxford NANOPORE Technologies

Emission Probabilities

- We measured distributions of current for k-mers from *E. Coli* M.SssI treated (methylated; green) and untreated (unmethylated; red) samples on both R7.3 and R9 flowcells.
- Boxplots of AGGTCG and TCGAGT k-mers which both contain CGs show significant differences in current in some cases (AGGTCG R7.3) and little to none in others (TCGAGT R7.3)
- R9 current distribution seem wider in both cases, but gives better discrimination in TCGAGT.



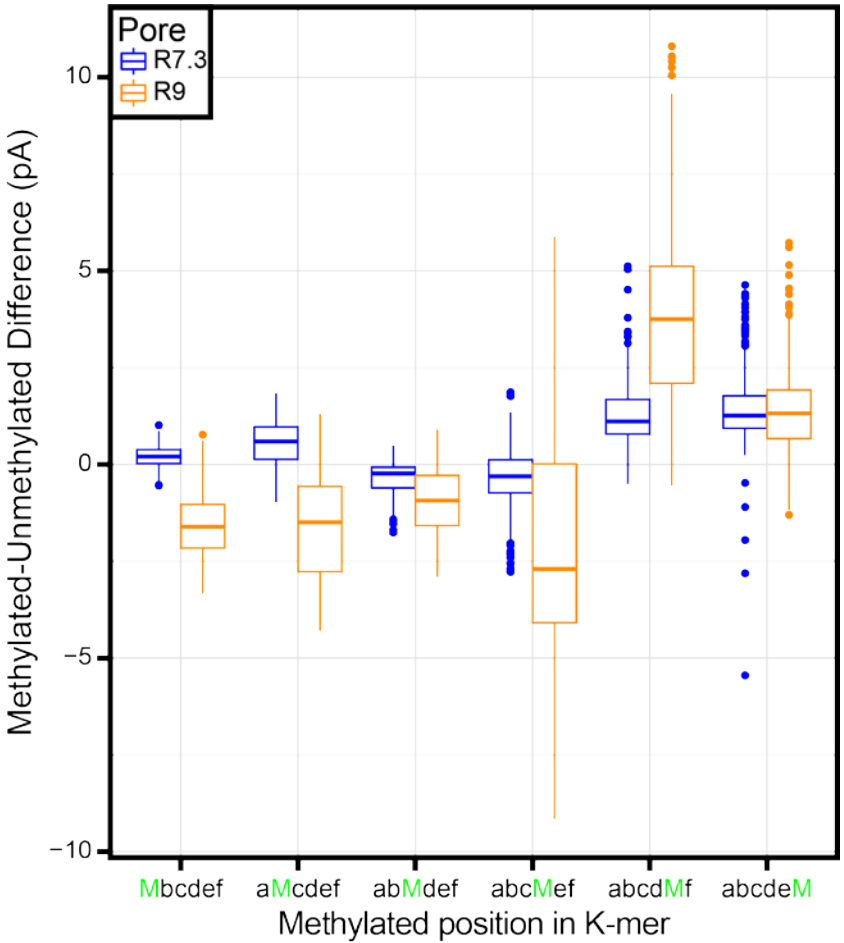
Simpson, Workman, in revision (2016)

JOHNS HOPKINS BIOMEDICAL ENGINEERING

Oxford NANOPORE Technologies

Distance of methylation effect

- We looked at the difference in current levels dependent on the position of the methylated base – plotted are the current differences for R7.3(blue) and R9 pores(orange).
- Signal seems again stronger but more variable for R9 pores than R7.3
- Methylation can either reduce current or increase it.
- Some positions are more sensitive to methylation than others.



Simpson, Workman, in revision (2016)

JOHNS HOPKINS BIOMEDICAL ENGINEERING

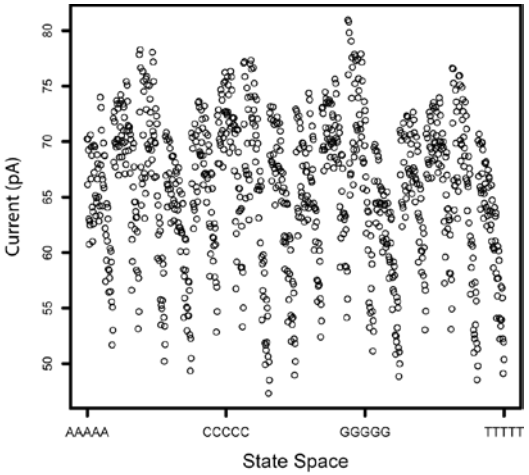
Oxford NANOPORE Technologies

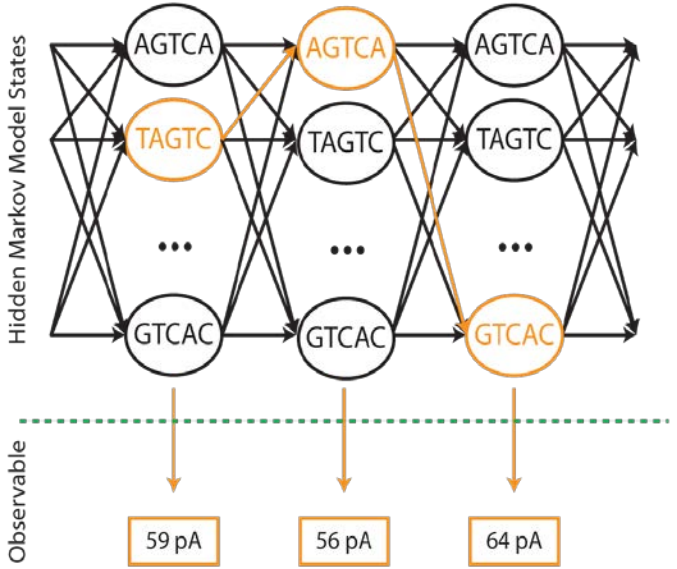
Nanopore: nanopolish methyltrain

- Multiple bases influence the current passing through the pore.
- Oxford uses 6-mers for a HMM to perform basecalling.
- Oxford basecalling does not take into account the 5th base – mC.
- With *nanopolish* we can call the probability:

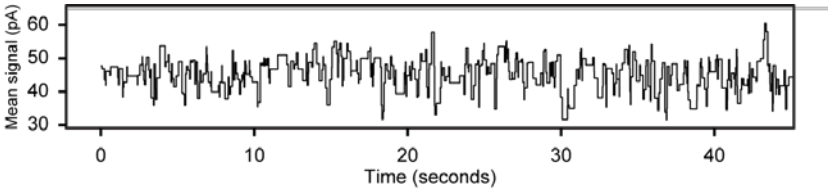
$$\frac{P(\mathcal{D}|S_m)}{P(\mathcal{D}|S_r)}$$

- Where S_m is the probability methylated for a given observable D and S_r the probability unmethylated
- We then take the log of this likelihood ratio.





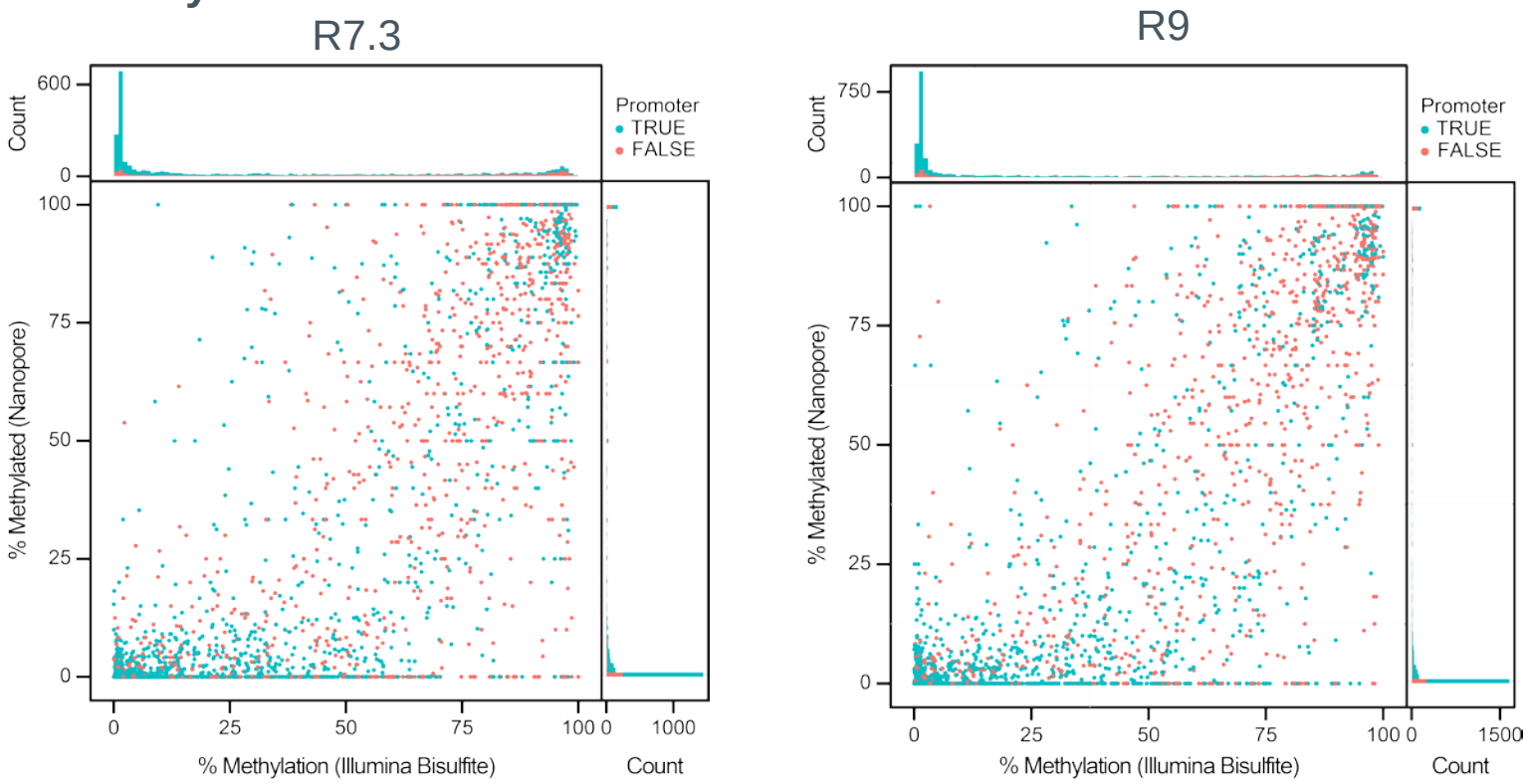
$$\delta_k \prod_t P(I(t)|k_t) \times T_{(t-1)(t)}$$



JOHNS HOPKINS BIOMEDICAL ENGINEERING

Oxford NANOPORE Technologies

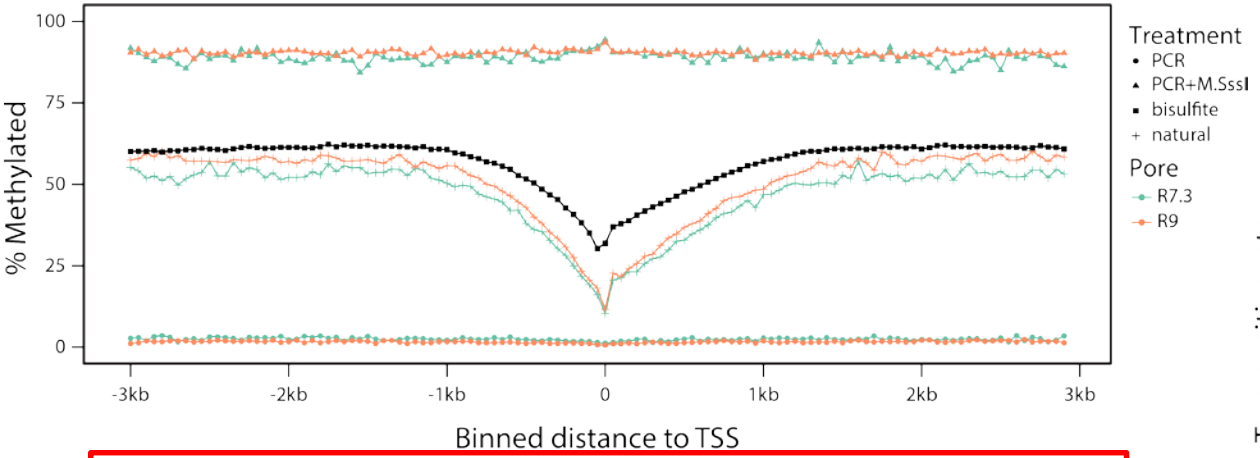
NA12878 Methylation

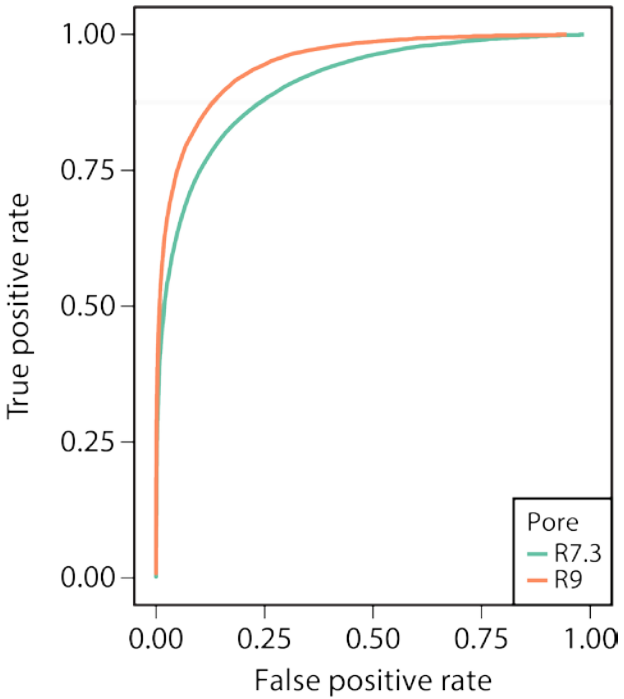


Simpson, Workman, in revision (2016)

- NA12878 (lymphoblast) gDNA: Illumina WGBS on X-axis (24X coverage) (SRA: GSM1002650) vs. R7.3 (0.02X) or R9 (0.13X) nanopore sequencing.
- Correlation of 0.83 (R7.3) and 0.84 (R9) – most gene promoters unmethylated

Binned Methylation vs. Transcription Start Sites



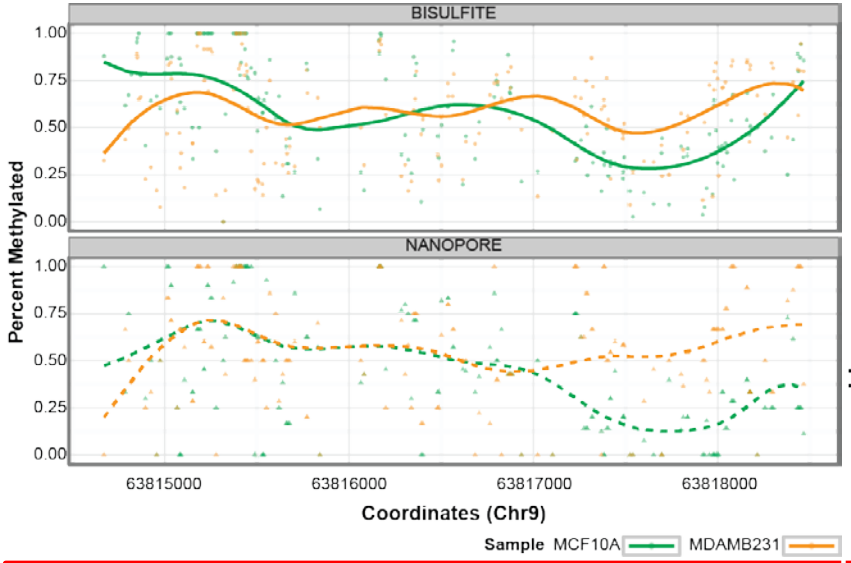


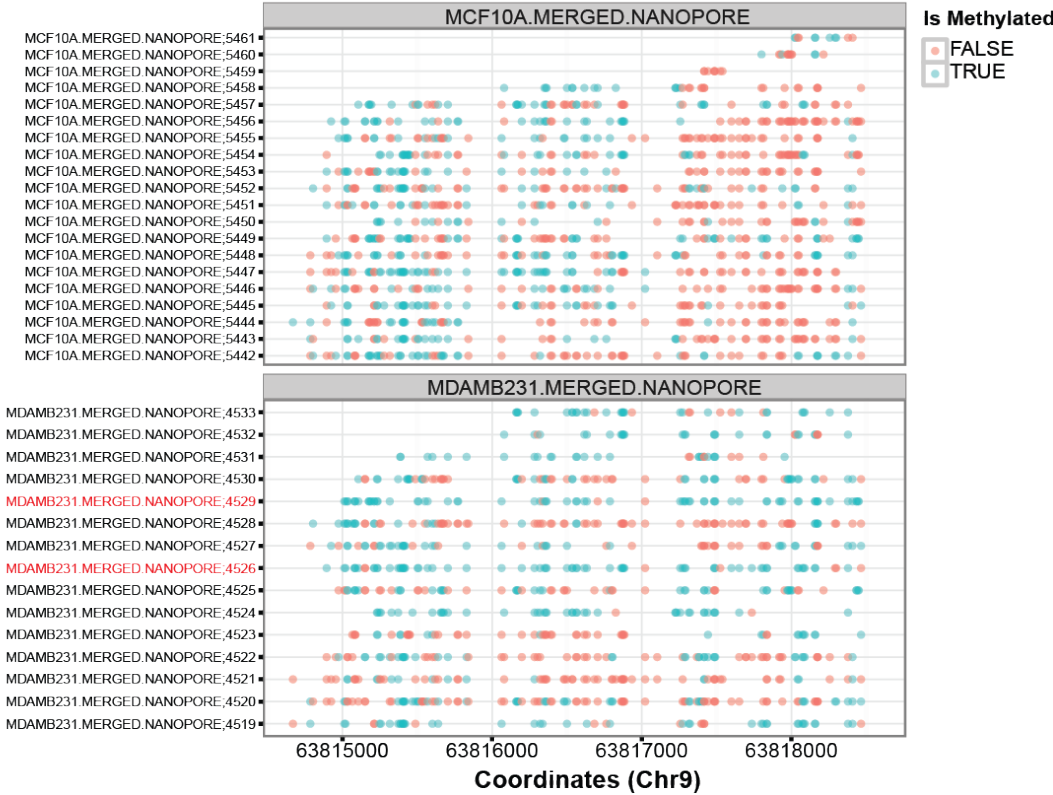
- On human genomic samples:
- Binning methylation levels vs. distance to TSS sites, compared to bisulfite data (NA12878).
- We also generated completely methylated (M.SssI treated; ~95% meth) and unmethylated – used to generate the ROC curve (right)
- R7.3 91% accurate at 68% of sites
- R9 94% accurate at 77% of sites

JOHNS HOPKINS BIOMEDICAL ENGINEERING

Oxford NANOPORE Technologies

Cancer-Normal Comparison





- Reduced representation method:12.5Mb of the genome (3.5-6kb size selection)
- We sequenced this fraction on nanopore and bisulfite Illumina seq
- Long reads measure *phased* methylation

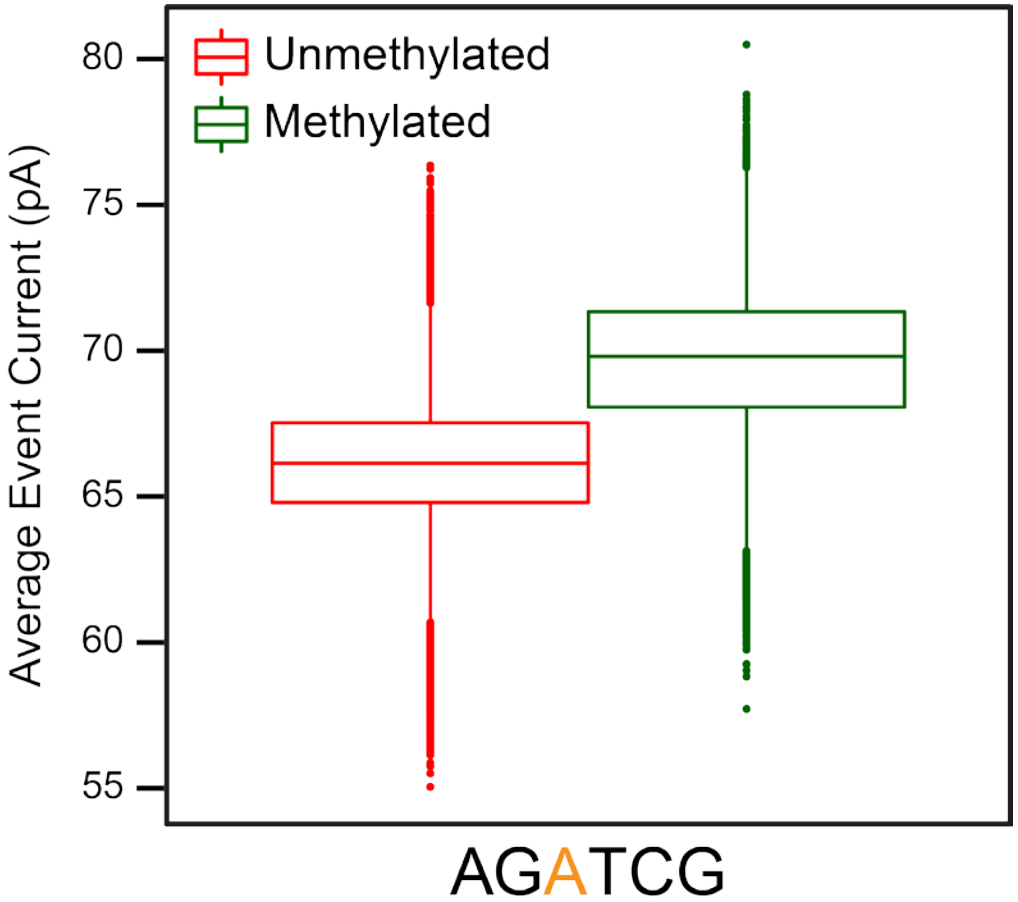
Simpson, Workman, in revision (2016)

JOHNS HOPKINS BIOMEDICAL ENGINEERING

Oxford NANOPORE Technologies

Future Work

- Expand to non-CpG methylation
- Expand to non 5-methylcytosine methylation
  - Strong signal for N6-methyladenine
- Apply to clinical samples
- Exogenous labeling of DNA and readout



JOHNS HOPKINS BIOMEDICAL ENGINEERING

Oxford NANOPORE Technologies

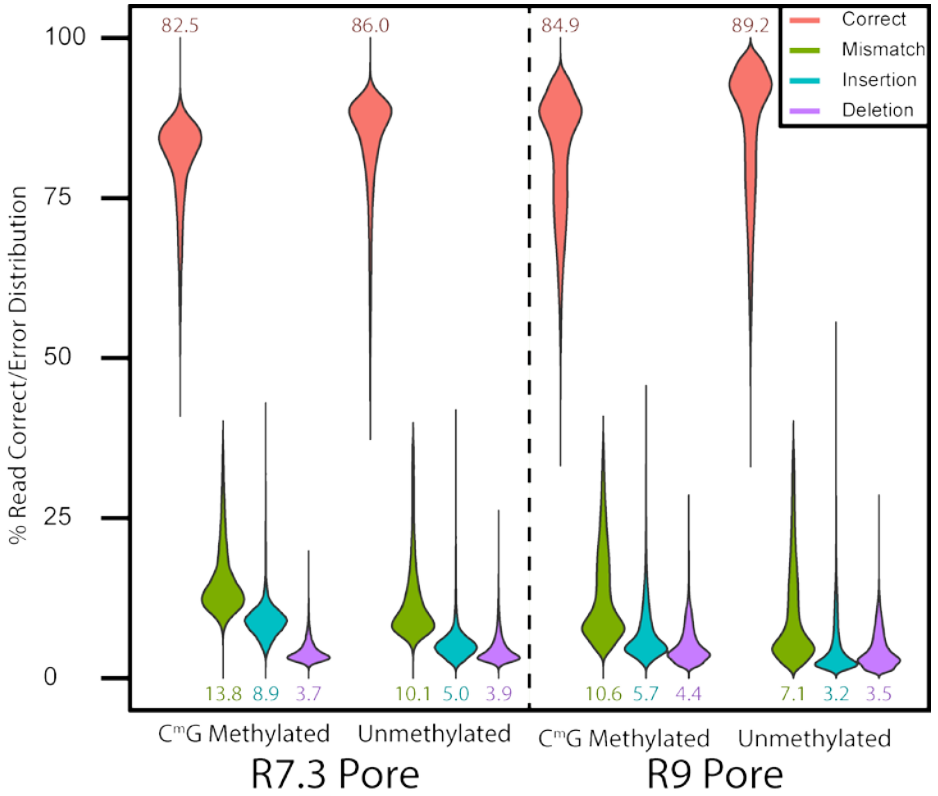
Acknowledgments

Timp Lab – Johns Hopkins University
Winston Timp, PhD
Rachael Workman, MS

Ontario Institute for Cancer Research
Jared Simpson, PhD
P.C. Zuzarte, PhD
Matei David, PhD
L. J. Dursi, PhD

Nanopore: Methylated Error

- We sequenced PCR'd *E. Coli* gDNA samples with either SssI or no treatment on v7.3 and v9 chips.
- Plotted is a distribution of the per read % correct, mismatch, insertion and deletion evens, generated with piledriver after bwa mem alignment.
- Notably, mismatch error rate and indel rate are higher on methylated samples than unmethylated – *if you aren't interested in methylation, PCR your samples.*
- R9 data has a generally higher correct rate, but still a significant change in % correct per read.



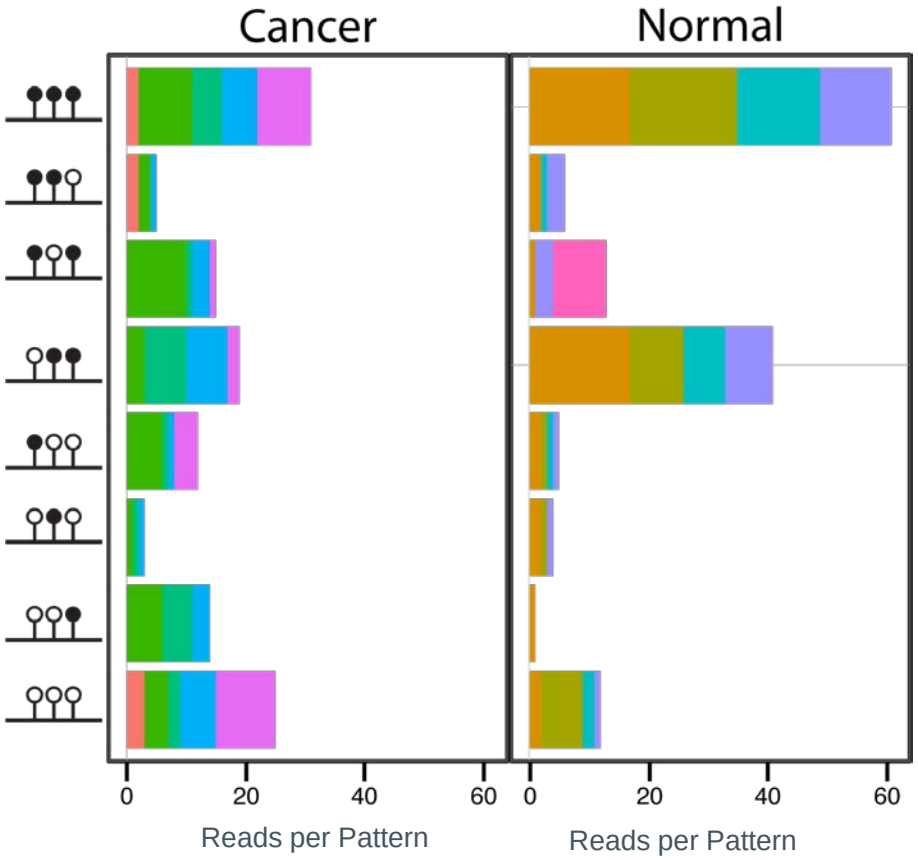
Simpson, Workman, et al. in revision (2016)

JOHNS HOPKINS BIOMEDICAL ENGINEERING

Oxford NANOPORE Technologies

Single Read Methylation: Distribution

- Using traditional short-read methods, the ability to characterize methylation pattern is limited, but intriguing.
- Distribution of methylation patterns within cancer and normal samples are shown to the right. Colors in the stacked bar graph represent different sequenced samples.
- Selected areas have significantly different methylation between normal and cancer samples



JOHNS HOPKINS BIOMEDICAL ENGINEERING

Oxford NANOPORE Technologies