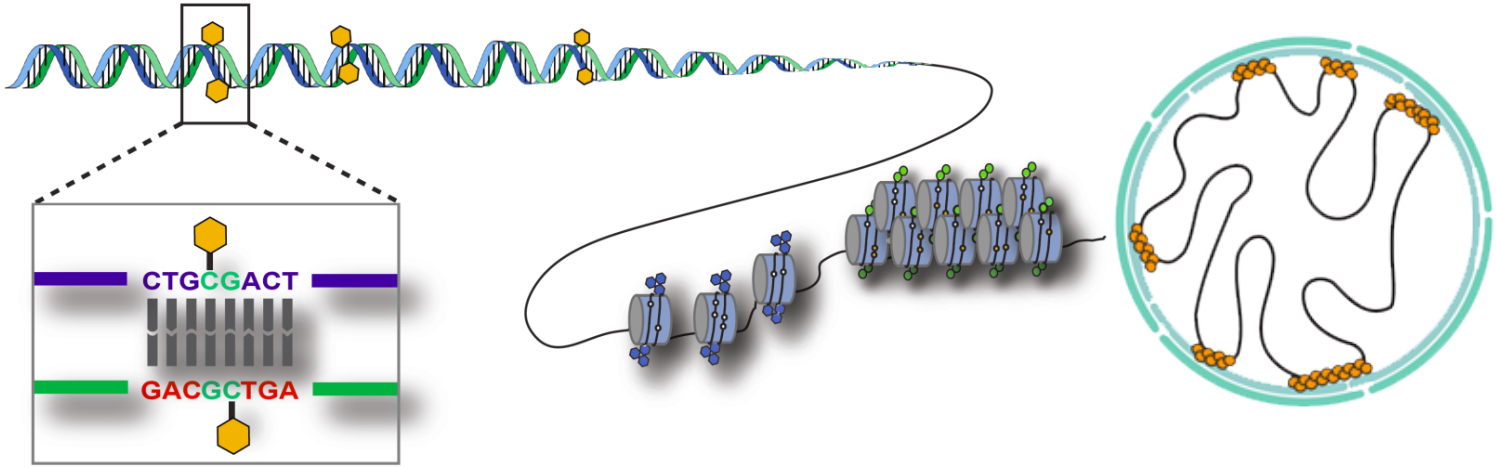
2017 LONDON CALLING

Measuring DNA Methylation with the MinION

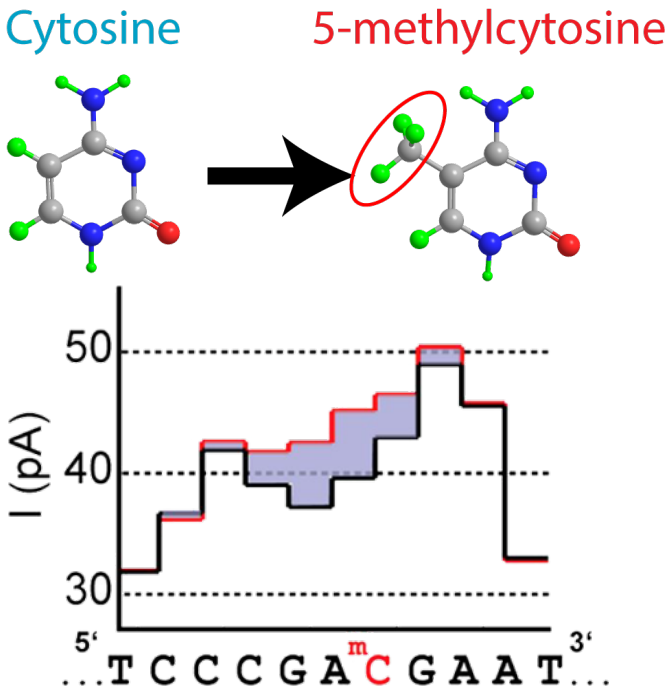
Winston Timp
Department of Biomedical Engineering
Johns Hopkins University

Epigenetics: Modern



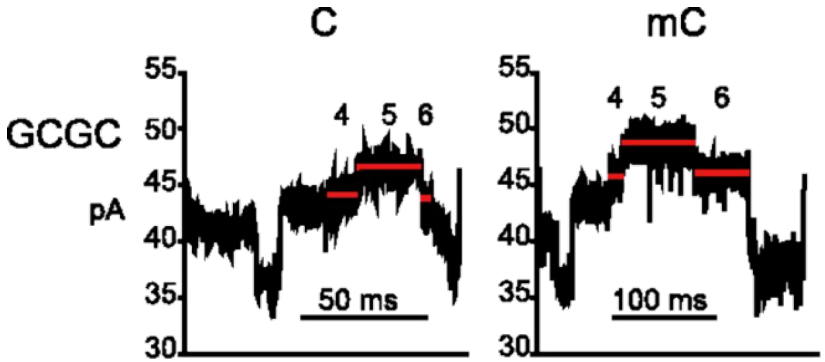
- Modern Definition of epigenetics involves heritable changes other than genetic sequence, e.g., positive feedback, high order structure, chromatin organization, histone modifications, DNA methylation.
- An analogy to a computer system:
  - DNA Sequence = Hardware
  - User input = Environment
  - Systems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

Nanopore: Methylation



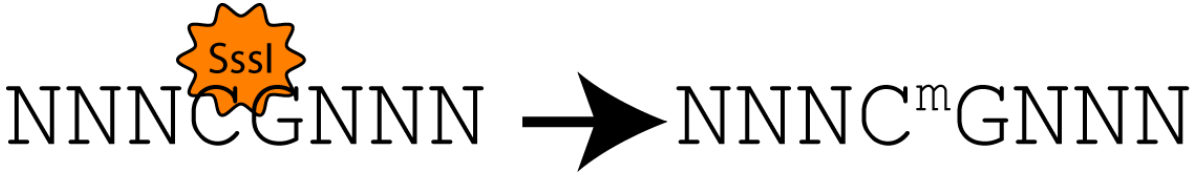
Laszlo, et al. *PNAS* (2013)

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

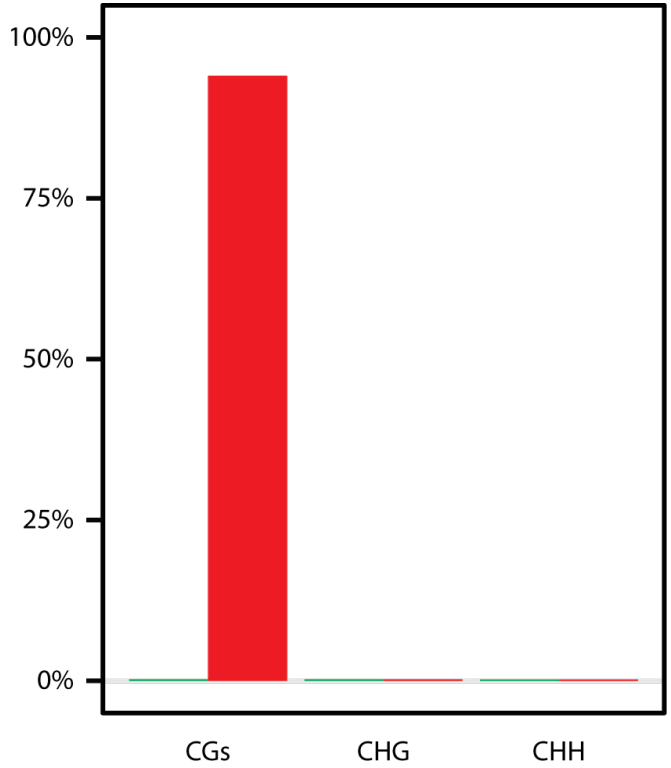


Schreiber, et al. *PNAS.* (2013)

Generation of methylated Samples

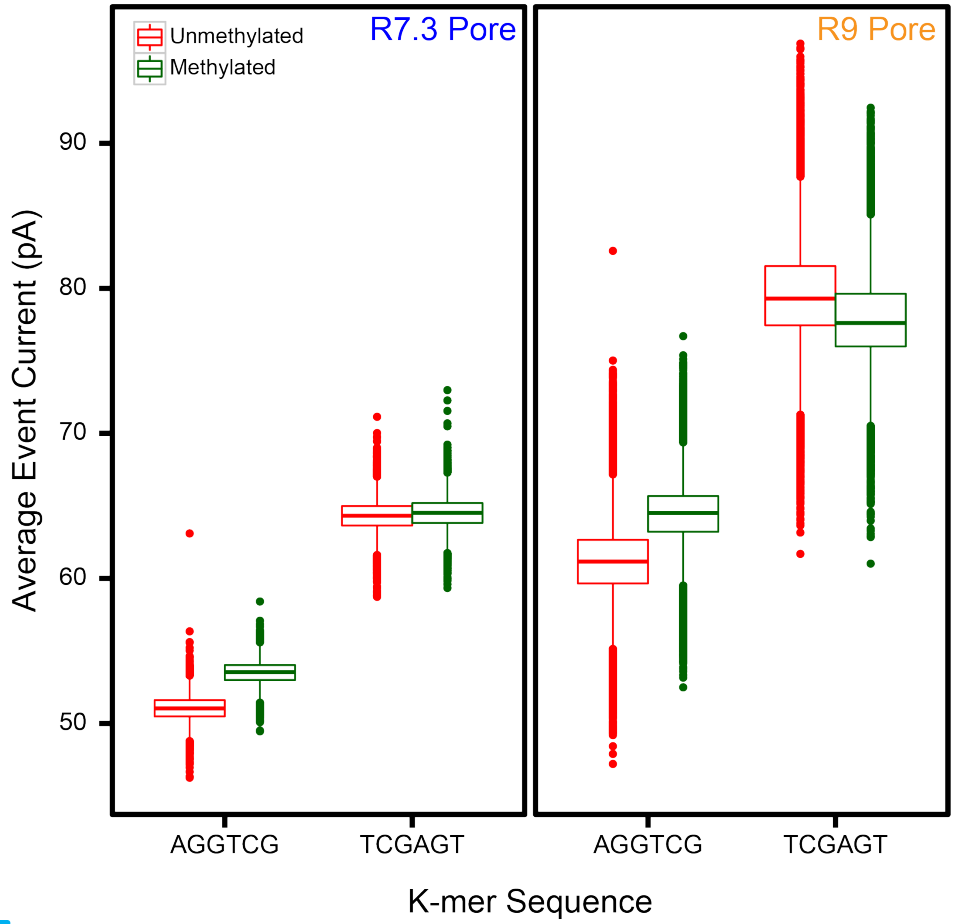


- To generated methylated samples, we treat unmethylated DNA (lambda, dam-/dcm- E. Coli, PCR product) with M. SssI methyltransferase
- We confirm the CG specific methylation using Illumina bisulfite sequencing of the sample – pictured right is methylation in different contexts E. Coli dataset treated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



Emission Probabilities

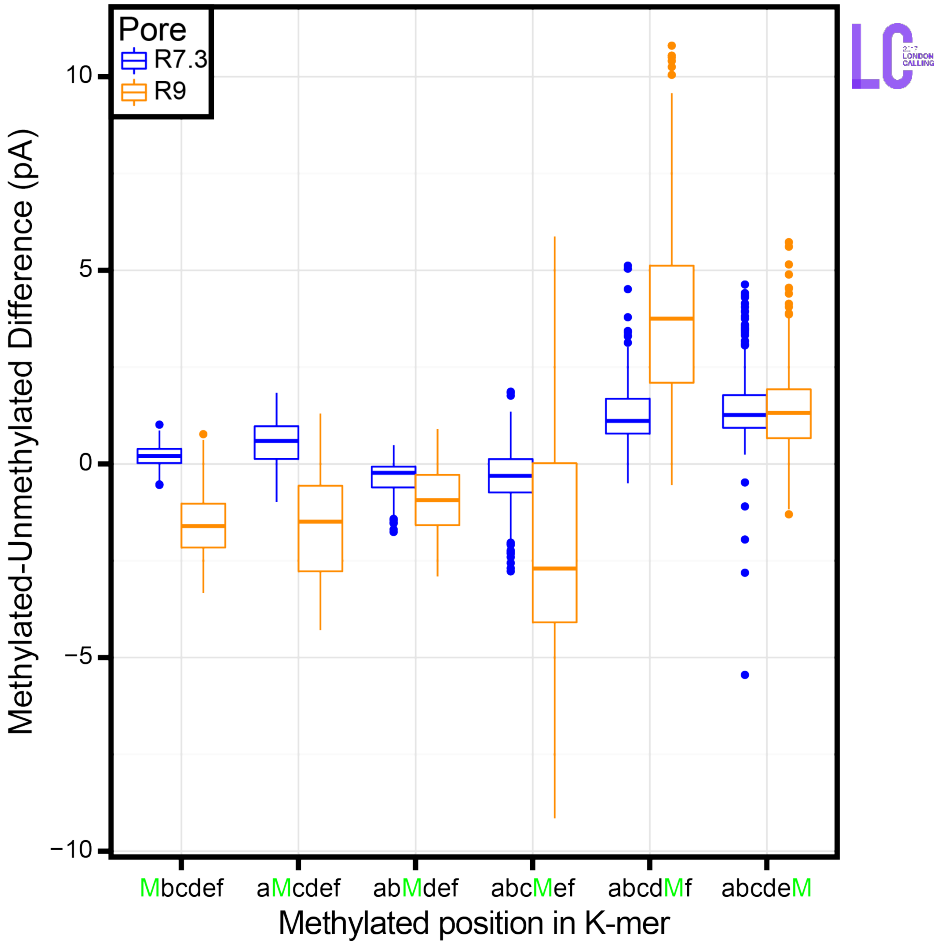
- We measured distributions of current for k-mers from *E. Coli* M.SssI treated (methylated; green) and untreated (unmethylated; red) samples on both R7.3 and R9 flowcells.
- Boxplots of AGGTCG and TCGAGT k-mers which both contain CGs show significant differences in current in some cases (AGGTCG R7.3) 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, *Nature Methods* (2017)

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.



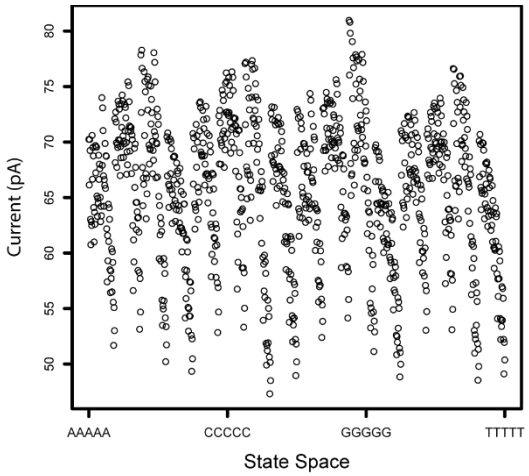
Simpson, Workman, *Nature Methods* (2017)

Nanopore: nanopolish methyltrain

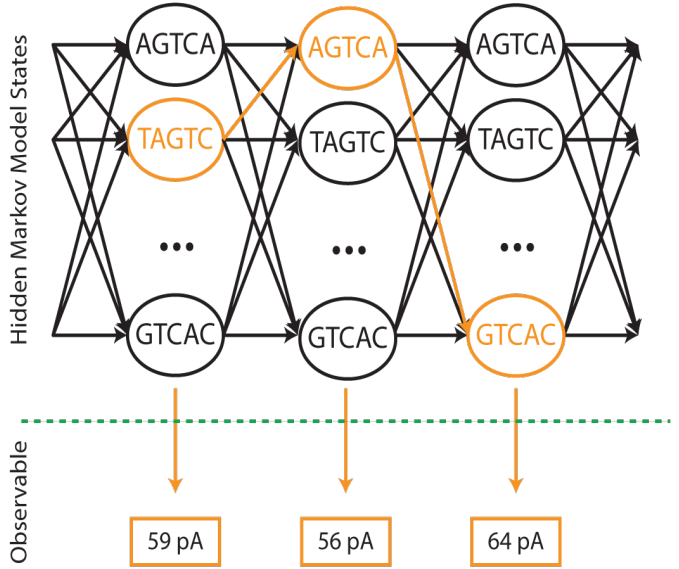
- Multiple bases influence the current passing through the pore.
- Current basecallers use a neural-network based methodology to call bases.
- We currently use a HMM based classifier to call methylation
- With *nanopolish* we can call the probability:

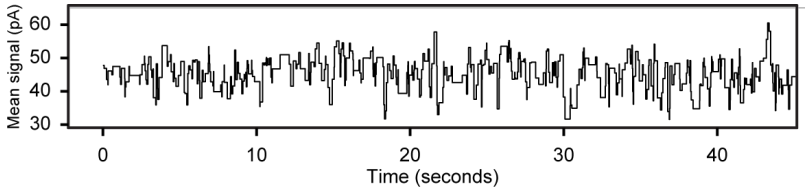
$$\frac{P(\mathcal{D}|S_m)}{P(\mathcal{D}|S_r)}$$

- 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.

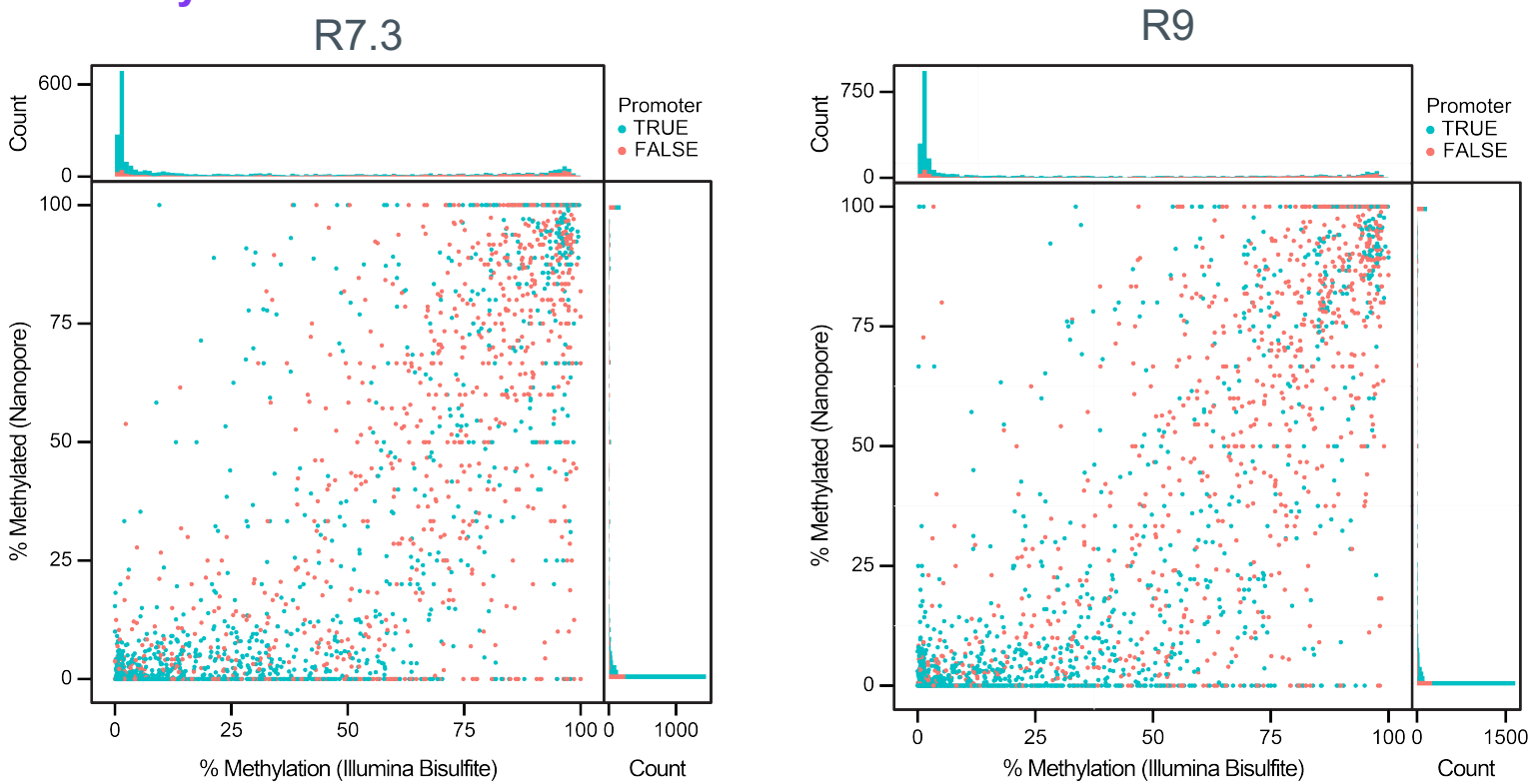


$$\delta_k \prod_t P(I(t)\,|\,k_t) \times T_{(t-1)(t)}$$





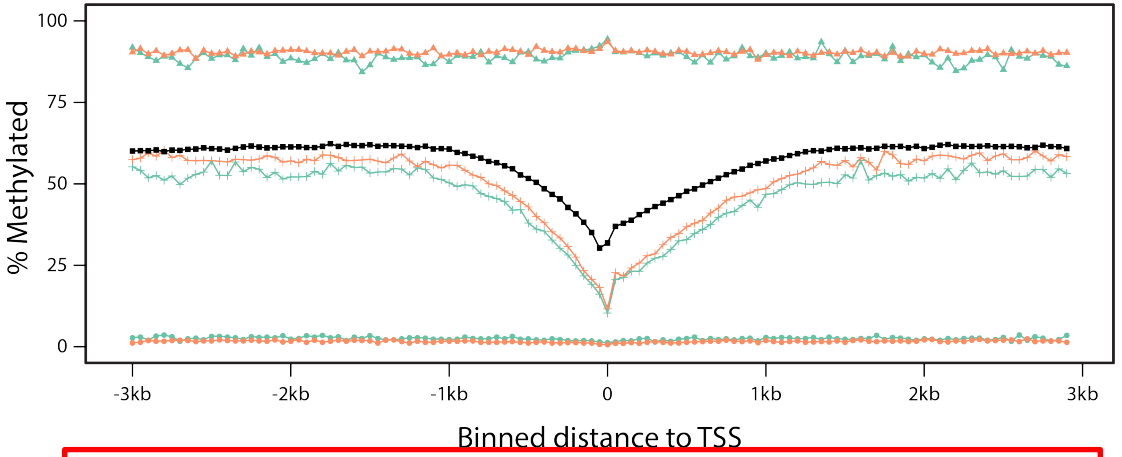
NA12878 Methylation

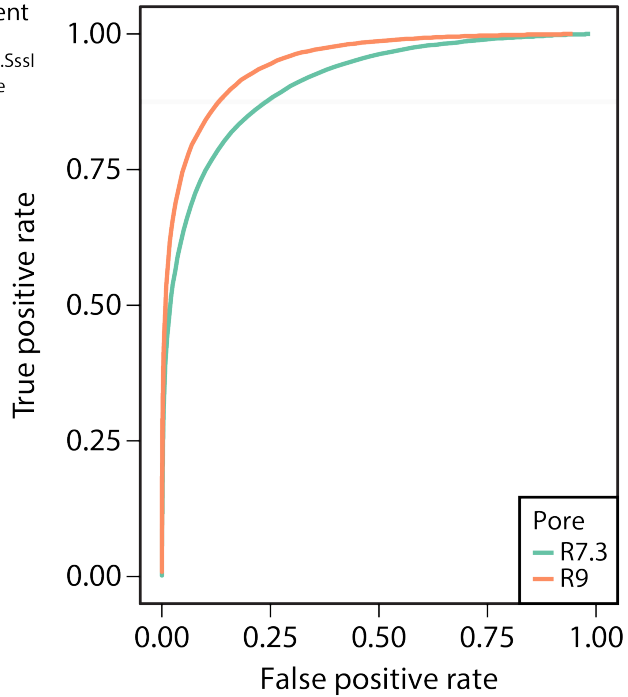


Simpson, Workman, *Nature Methods* (2017)

- 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

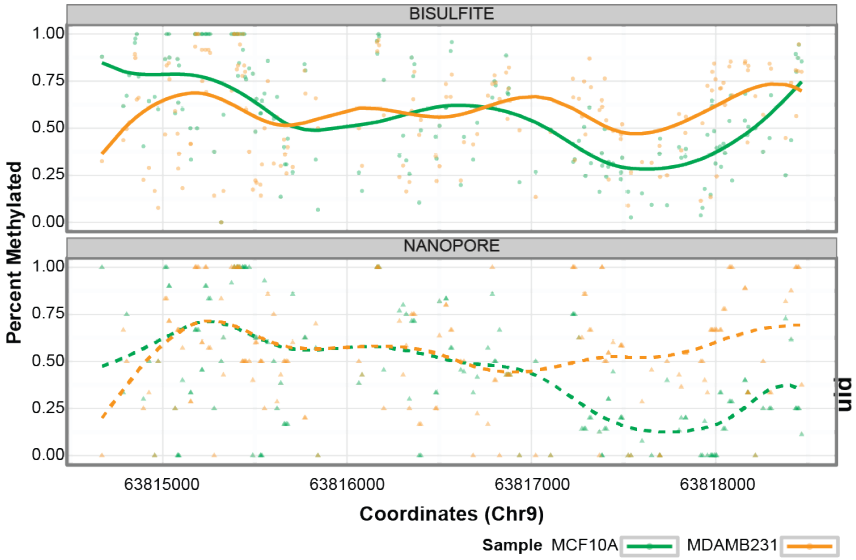
Binned Methylation vs. Transcription Start Sites

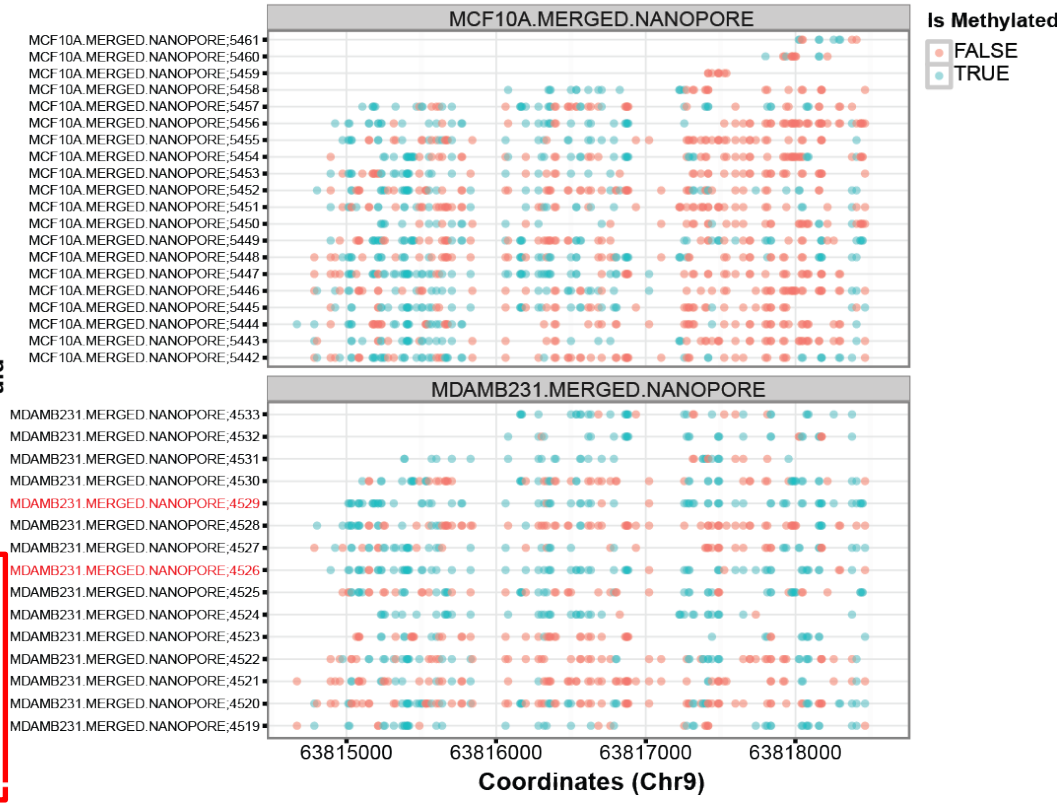




- On human genomic samples:
- Binning methylation levels vs. distance to TSS sites, compared to bisulfite data (NA12878).
- We also generated completely methylated (M.SssI treated; ~95% meth) and unmethylated – used to generate the ROC curve (right)
- R7.3 91% accurate at 68% of sites
- R9 94% accurate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

Cancer-Normal Comparison



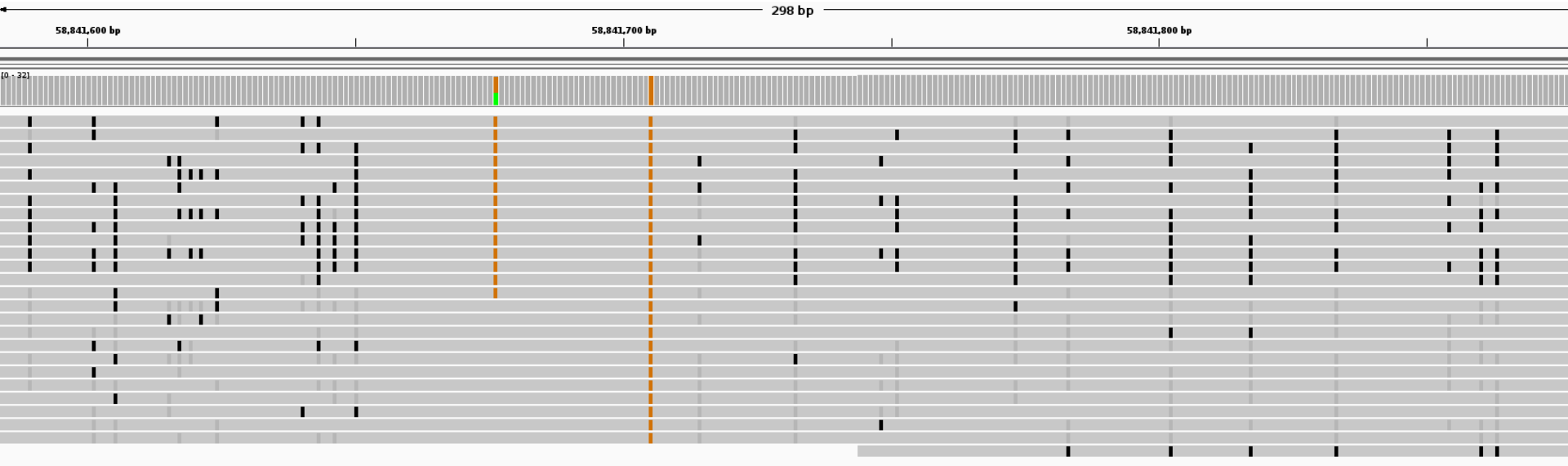


- Reduced representation method:12.5Mb of the genome (3.5-6kb size selection)
- We sequenced this fraction on nanopore and bisulfite Illumina seq
- Long reads measure *phased* methylation

Simpson, Workman, in revision (2016)

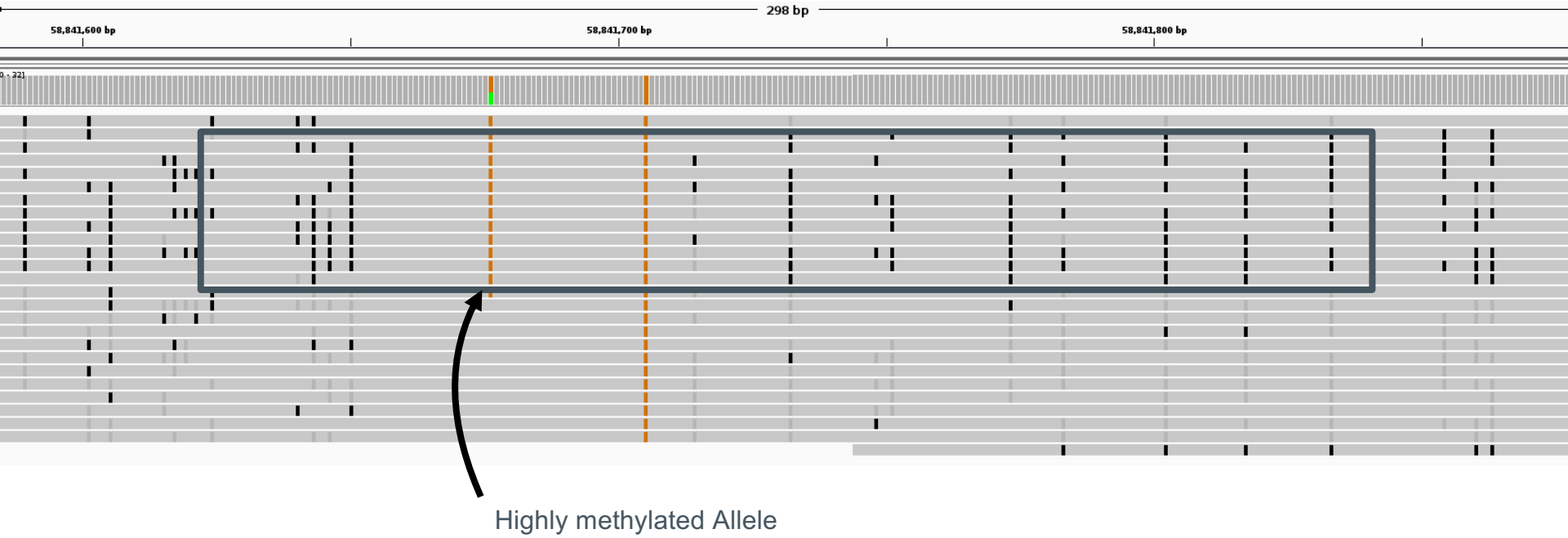
Haplotype-Phased Methylation

nanopolish has experimental support for phasing methylation patterns



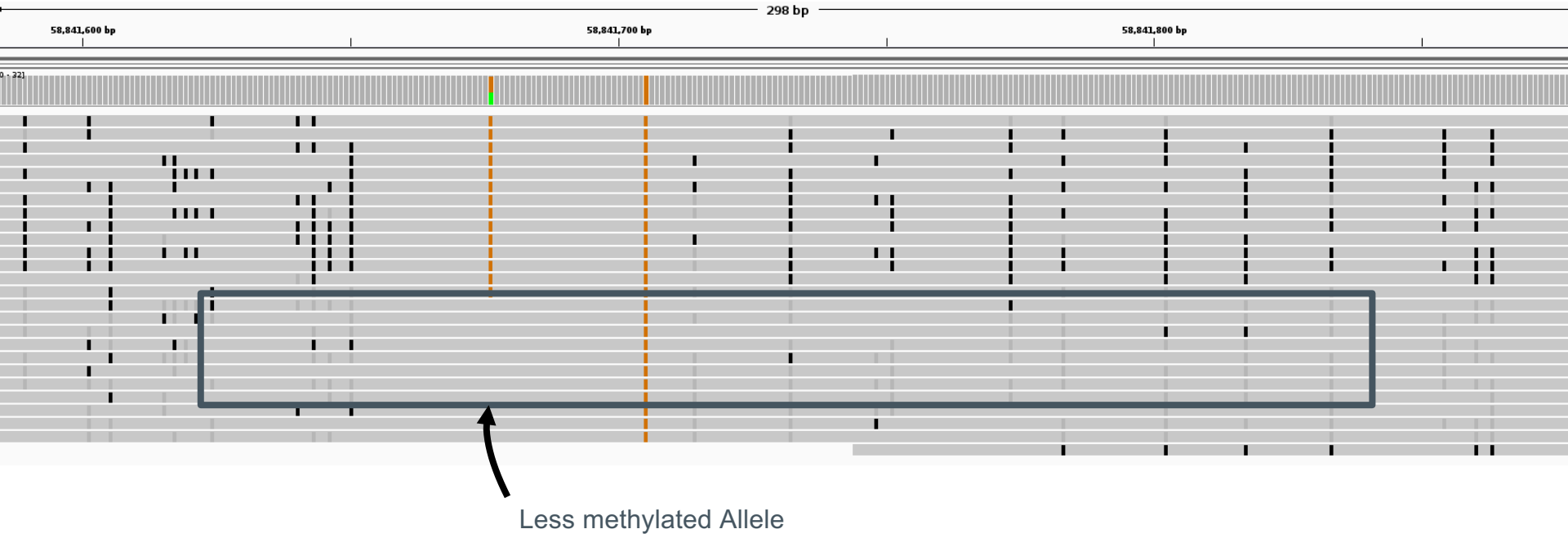
Haplotype-Phased Methylation

Nanopolish has experimental support for phasing methylation patterns. (From NA12878 data)

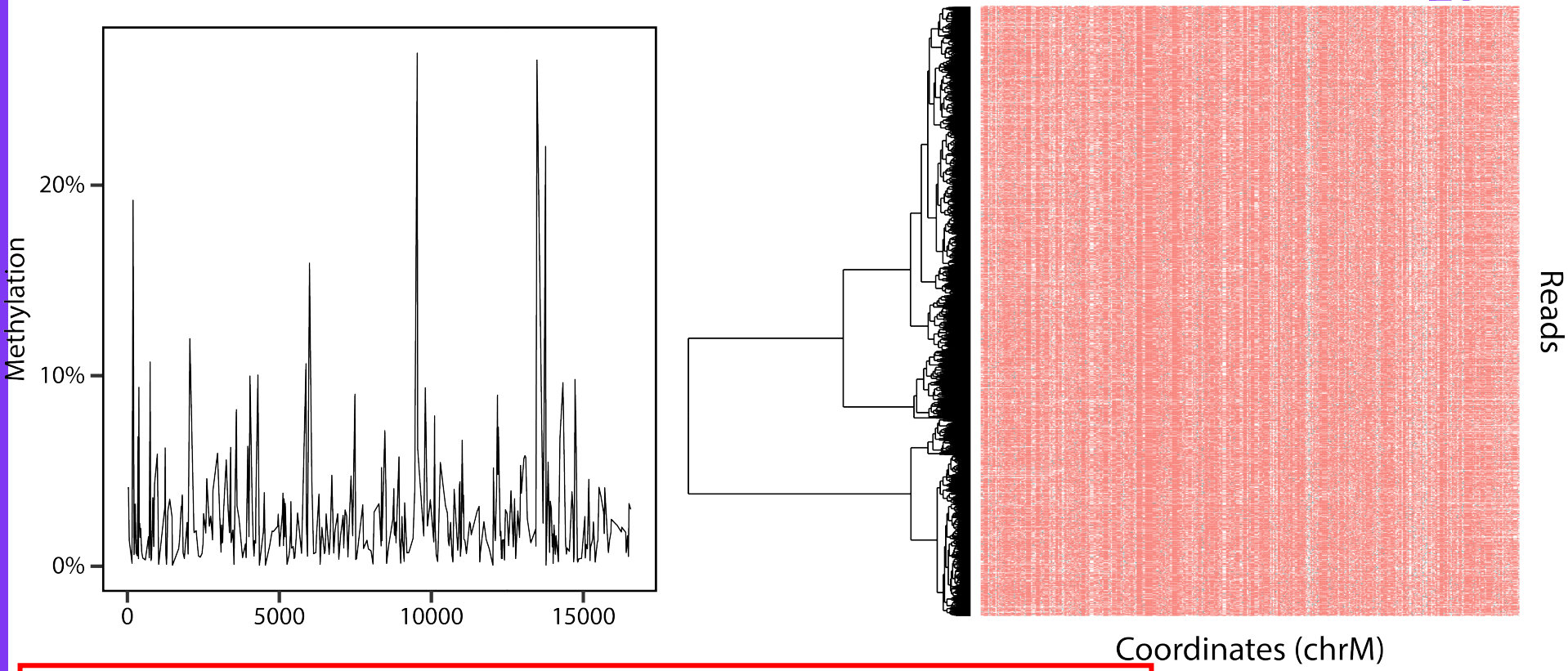


Haplotype-Phased Methylation

Nanopolish has experimental support for phasing methylation patterns. (From NA12878 data)



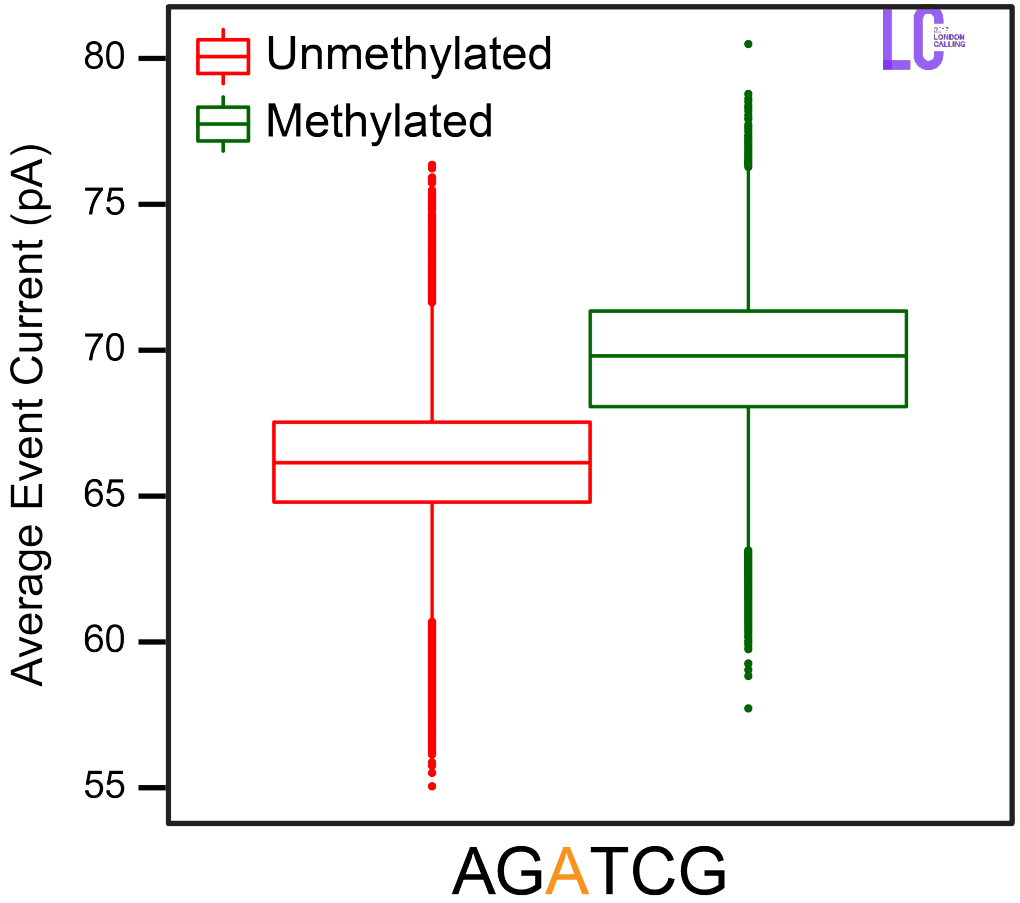
Mitochondrial methylation/clustering



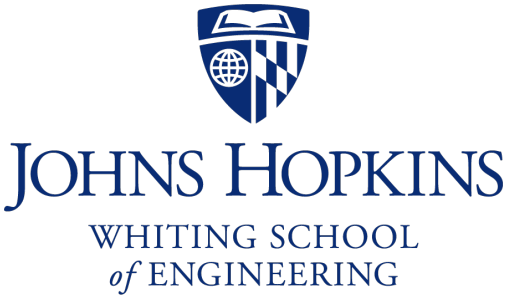
Preliminary data clustering mitochondrial CG methylation from MCF10A

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



Acknowledgments



Timp Lab – Johns Hopkins University
Winston Timp, PhD
Rachael Workman, MS
Yunfan Fan
Amy Vandiver, MD, PhD

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

1

LC 2017 LONDON CALLING

ORGANISED BY NANOPORE