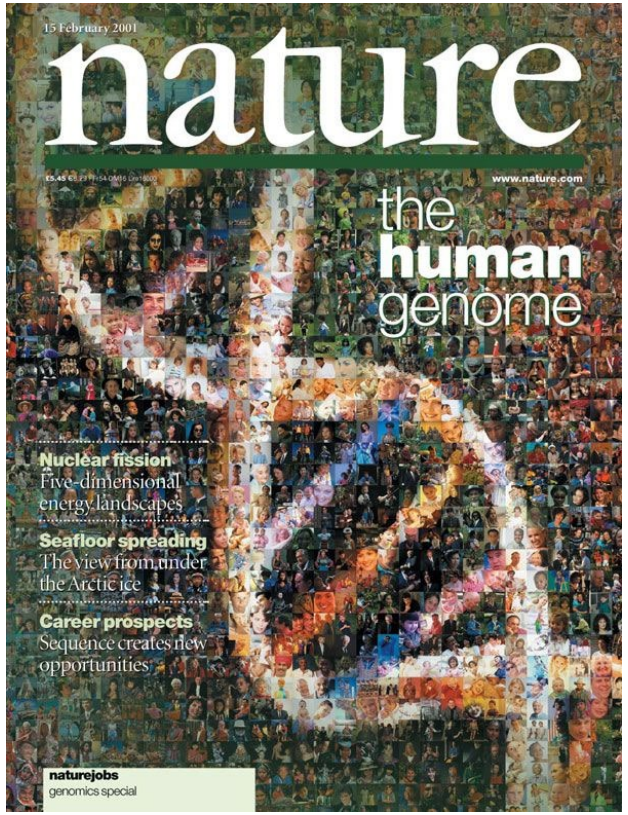
Workshop on Long-Read Sequencing
Farmington, CT
September 2019

JOHNS HOPKINS
BIOMEDICAL ENGINEERING

Using nanopore sequencing to interrogate the genome, epigenome and transcriptome

Winston Timp
Department of Biomedical Engineering
Johns Hopkins University

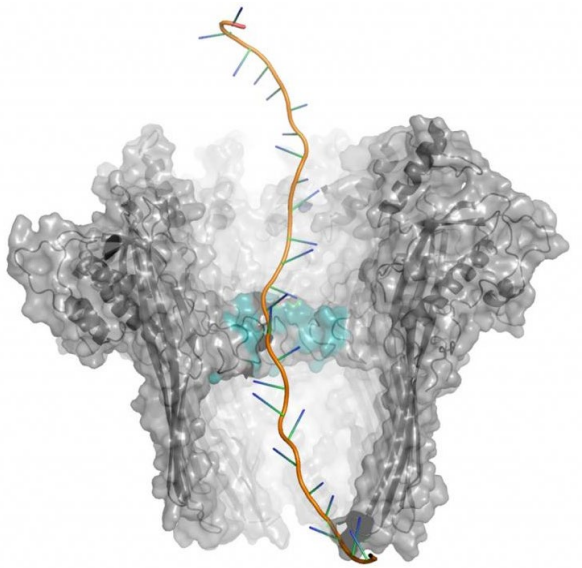
Revolutions in Science: Genomics



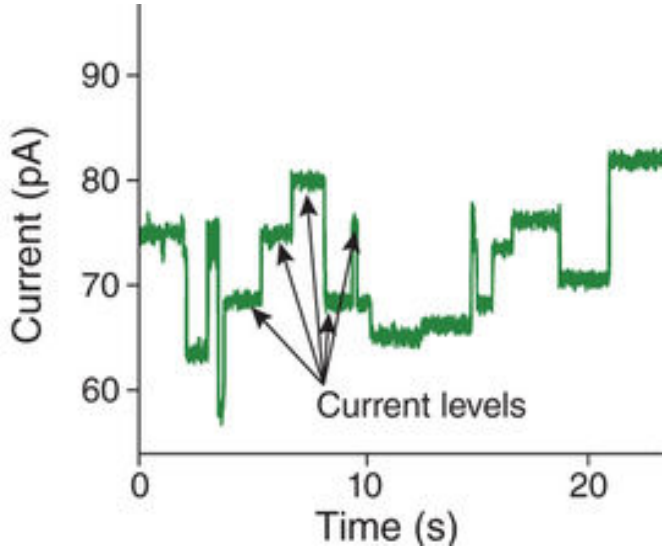
- Draft of the human genome was completed in 2001
- ~3 billion bases in size
- Think about this like the first transistor (1947) – the watershed after which genomic and epigenomic engineering has exploded

Nanopore: Single Molecule Sequencing

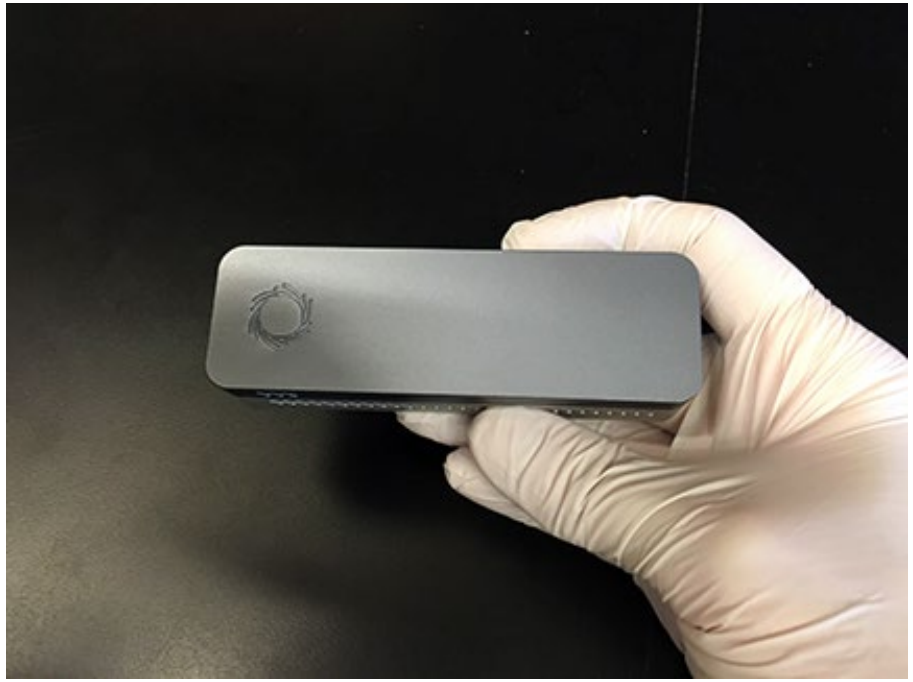
- Oxford Nanopore Technologies, CsgG biological pore
- No theoretical upper limit to sequencing read length, practical limit only in delivering DNA to the pore intact
- Palm sized sequencer
- Sequencing output 5-15Gb



Oxford Nanopore Google Hangout March 2016



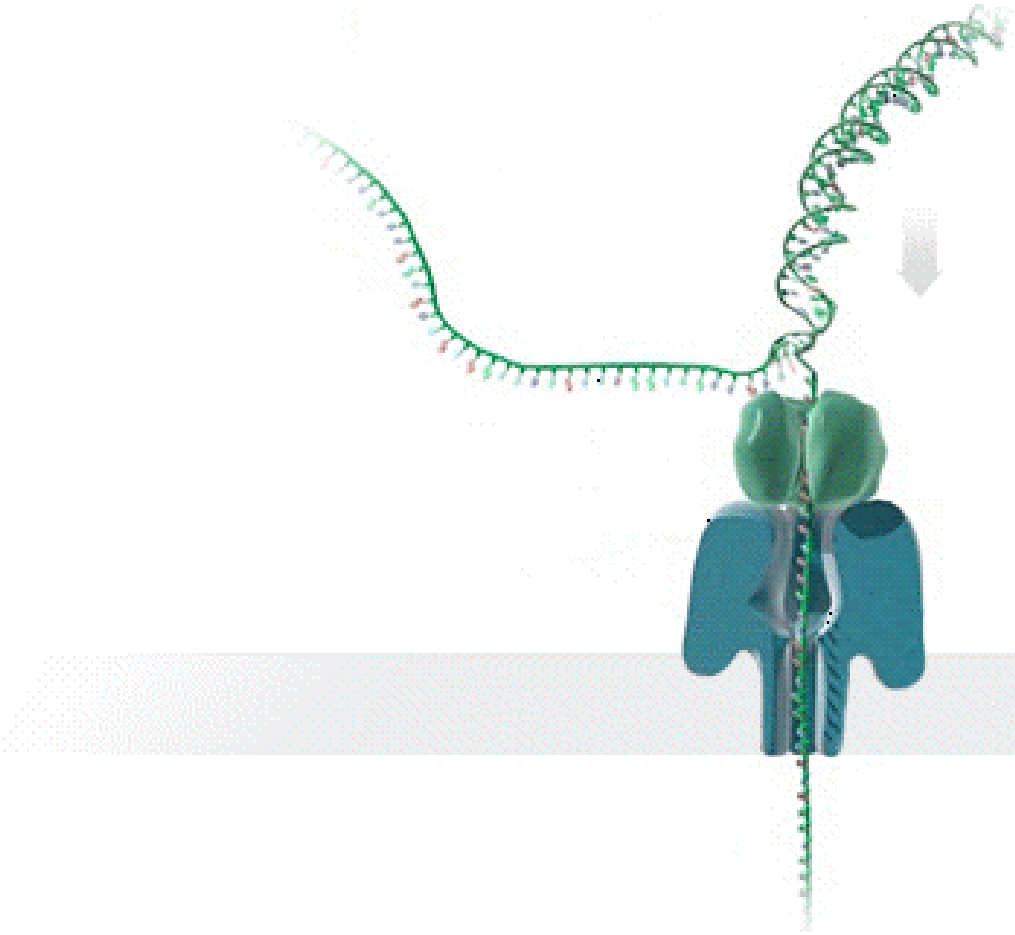
Deamer et al 2016, Nature Biotech

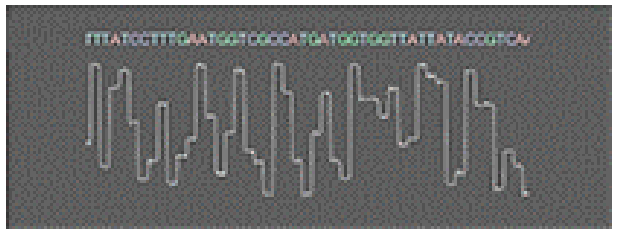


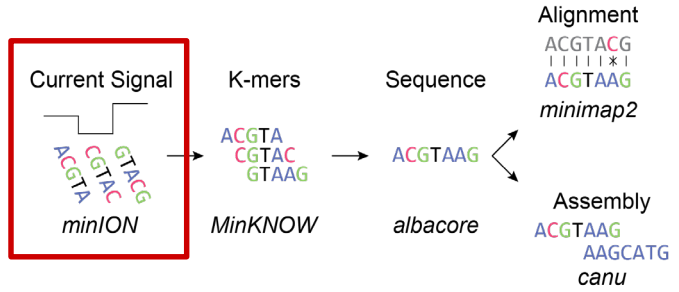
ATCGATCGATAGTAT
TAGATACGACTAGC
GATCAG

Disclosure: Timp has two patents (US Patent 8,748,091; US Patent 8,394,584) licensed to ONT

Sequencing Operation



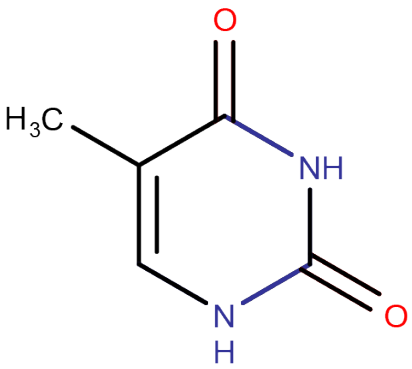




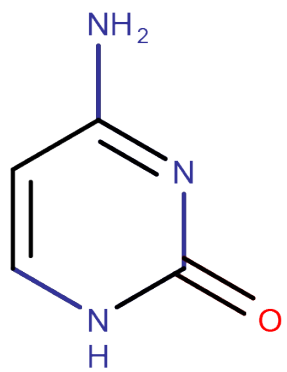
Oxford Nanopore Technologies

- Protein nanopores on a synthetic polymer
- Multiple base-pairs at a time (“k-mers”)
- Characteristic current signature is converted to nucleotide sequences

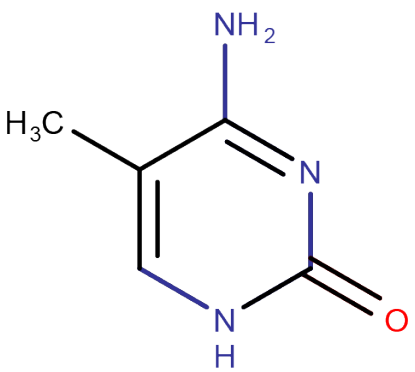
Nanopore Sequencing of Modifications



Thymine

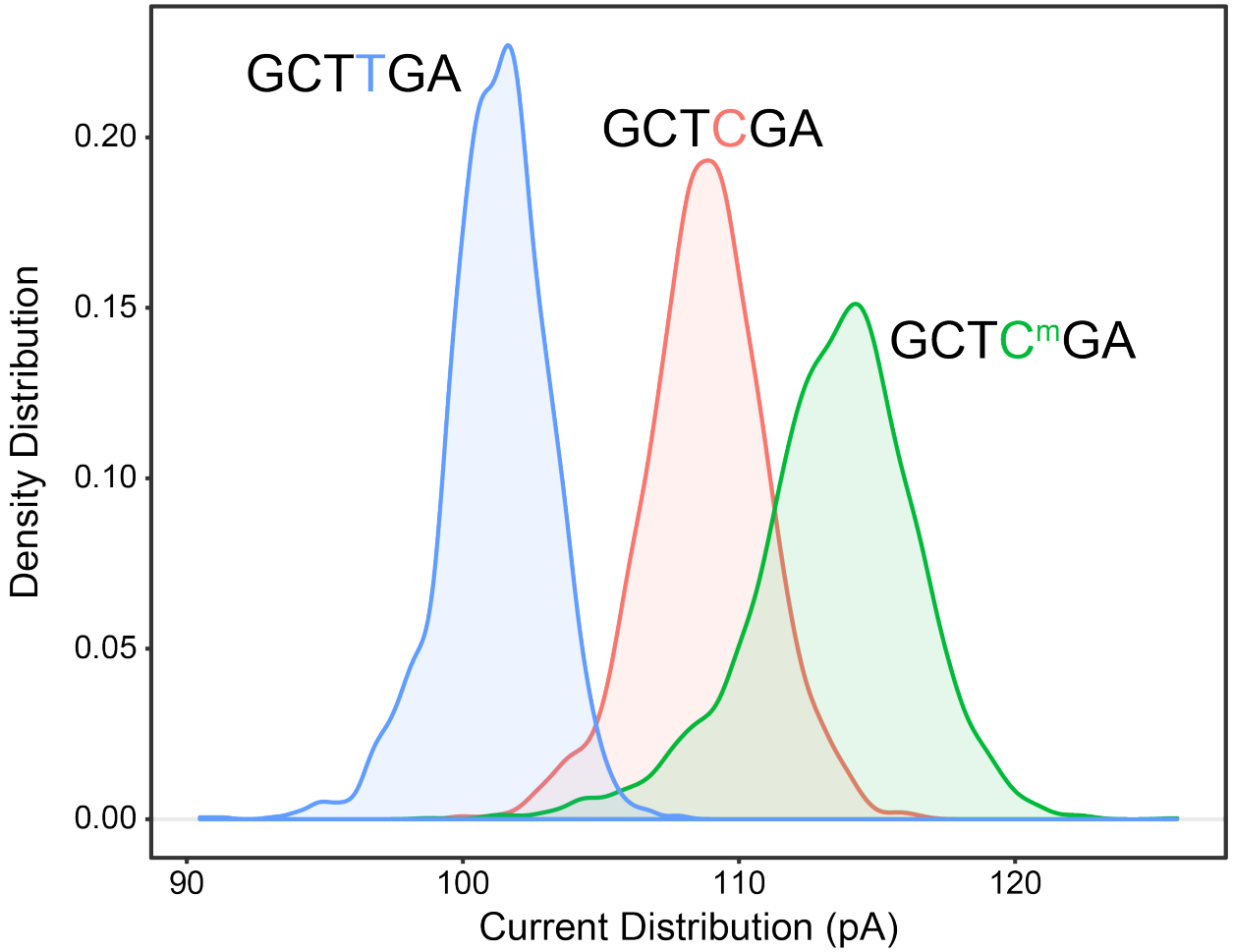


Cytosine



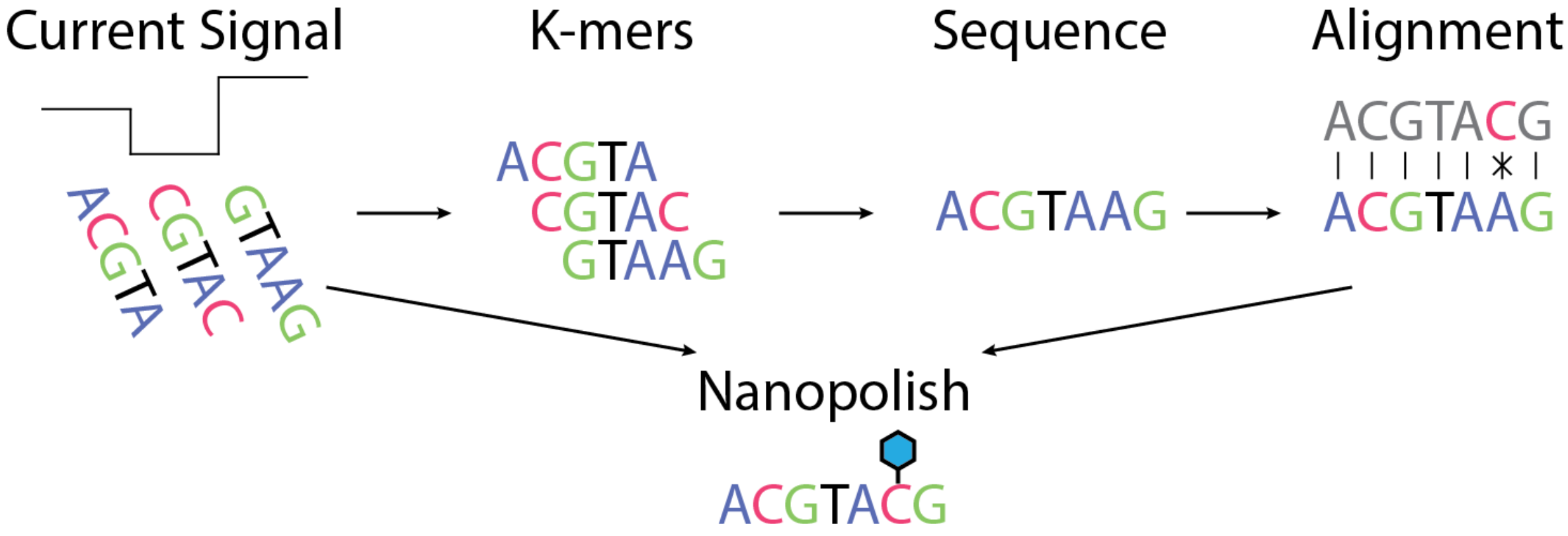
5-methylcytosine

- To generate methylated samples, we treat unmethylated DNA (PCR amplified E. Coli gDNA) with M. SssI methyltransferase
- Distributions of observed current for GCT[T/C/mC]GA demonstrate the type of signal between methylated and unmethylated k-mers



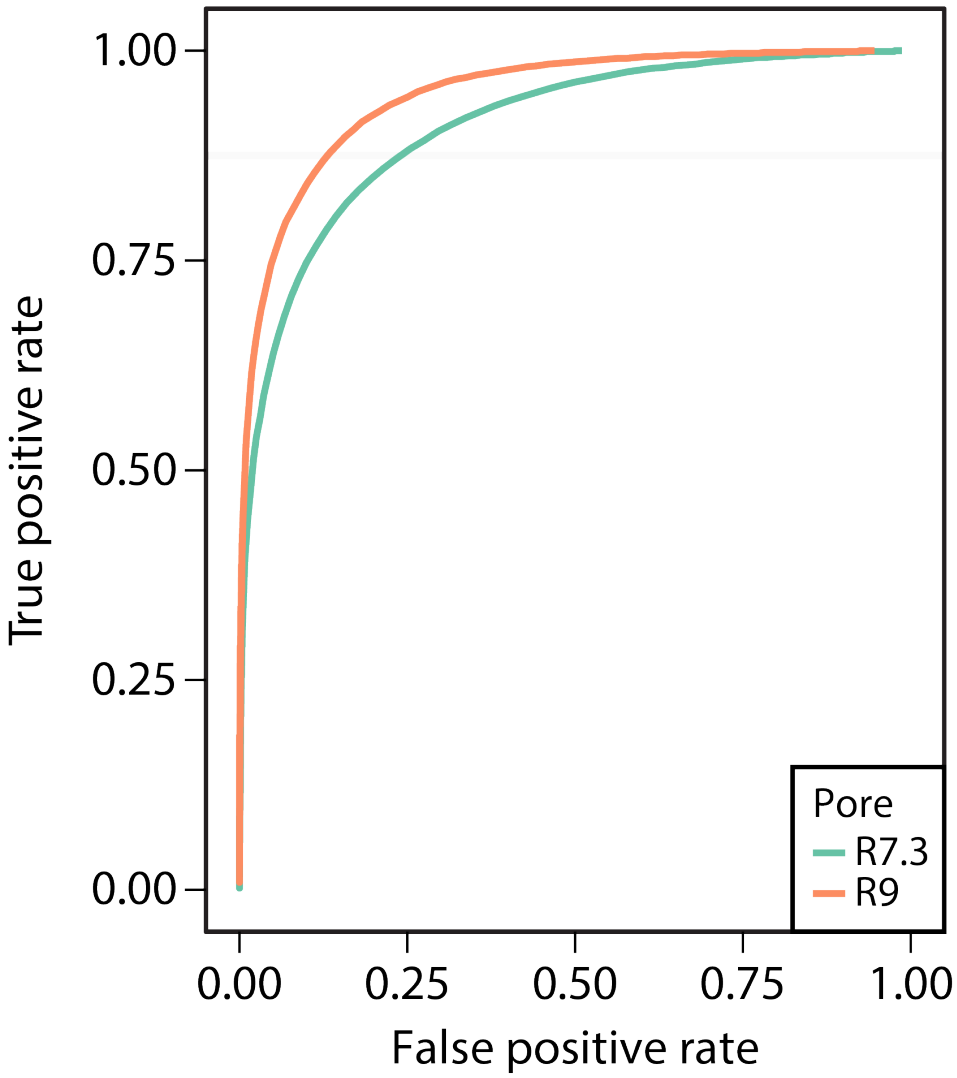
Simpson et. al *Nature* Methods 2017

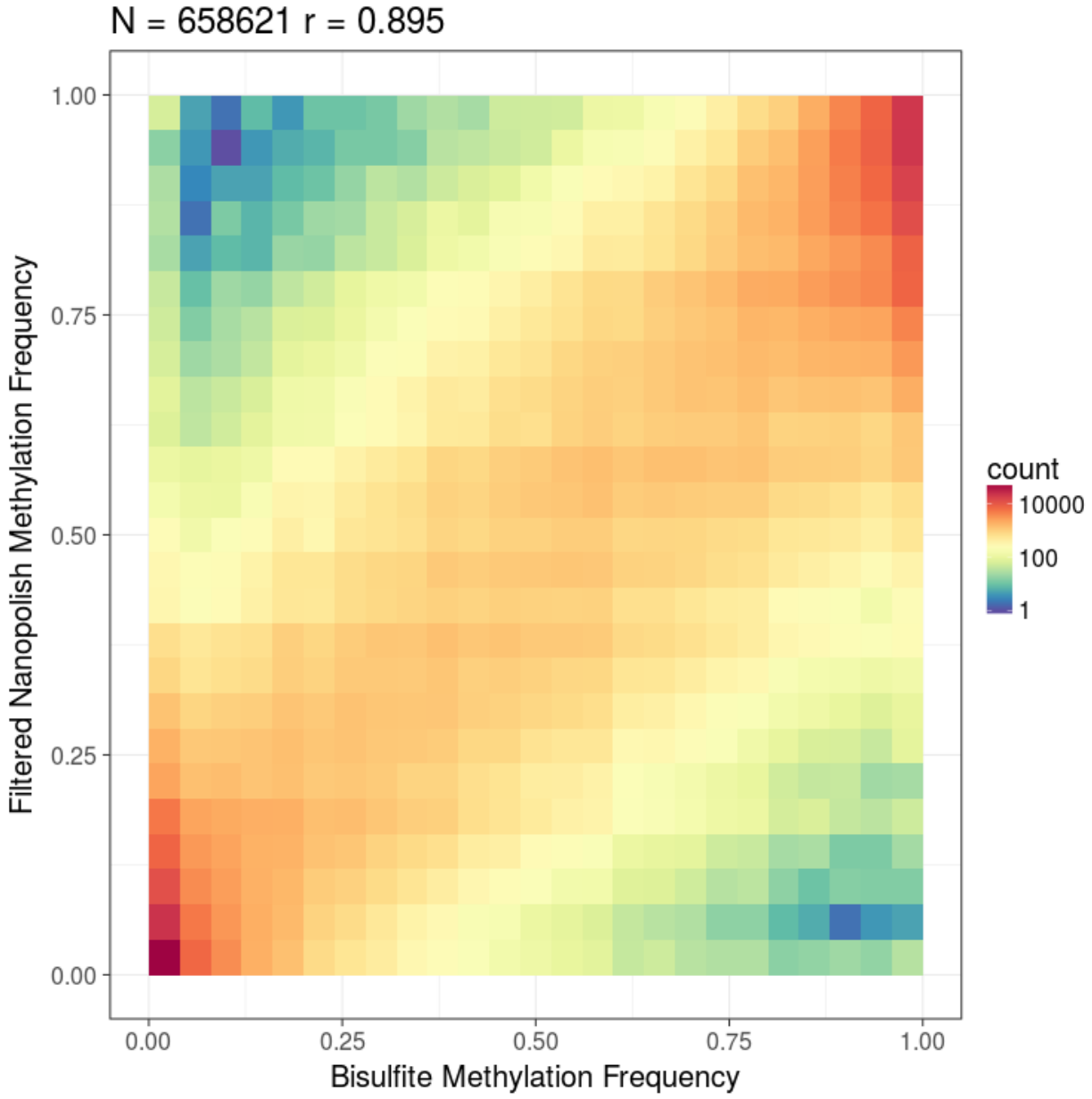
Nanopore: nanopolish methyltrain



- 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, and threshold for >2.5 as methylated; <-2.5 as unmethylated

Nanopolish Methylation

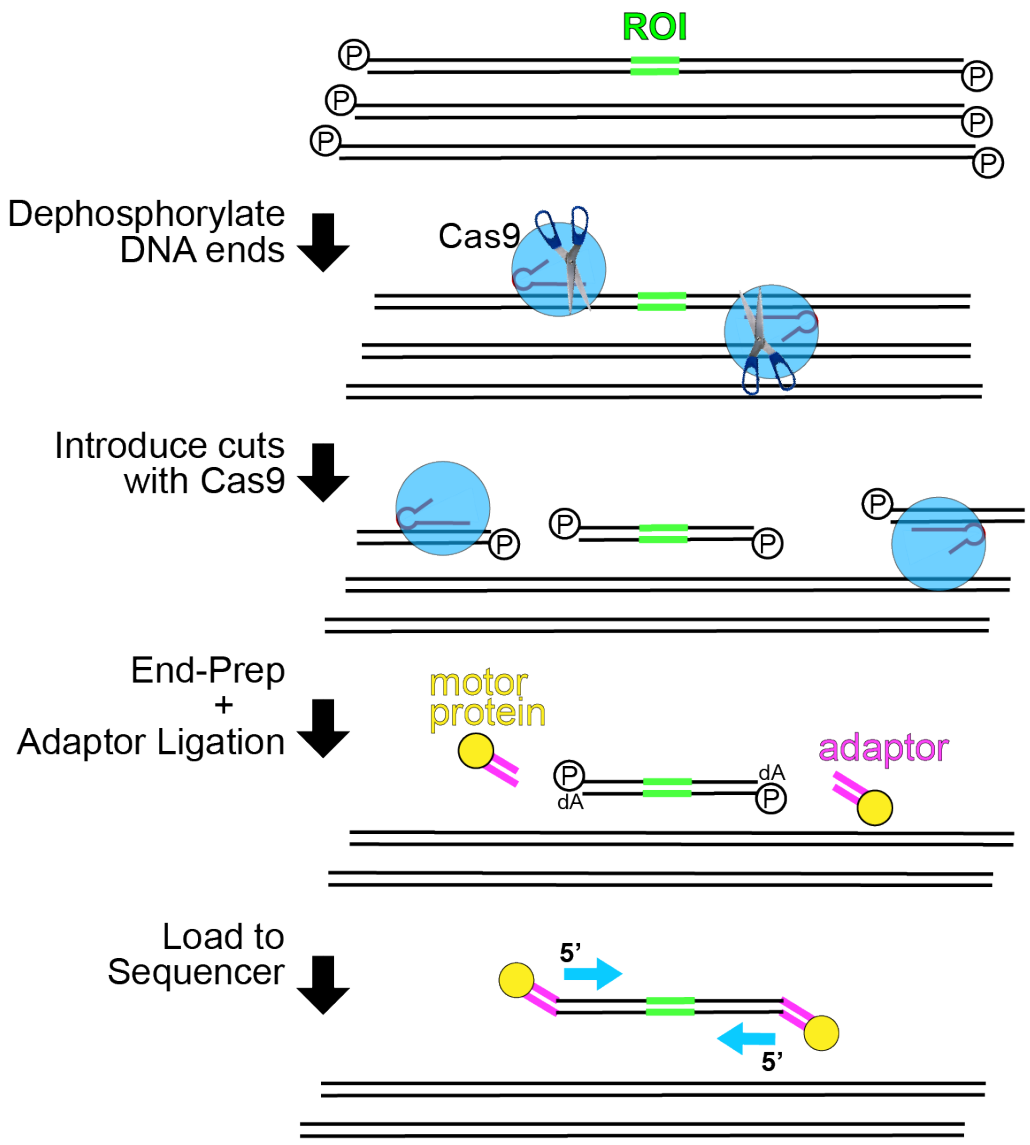


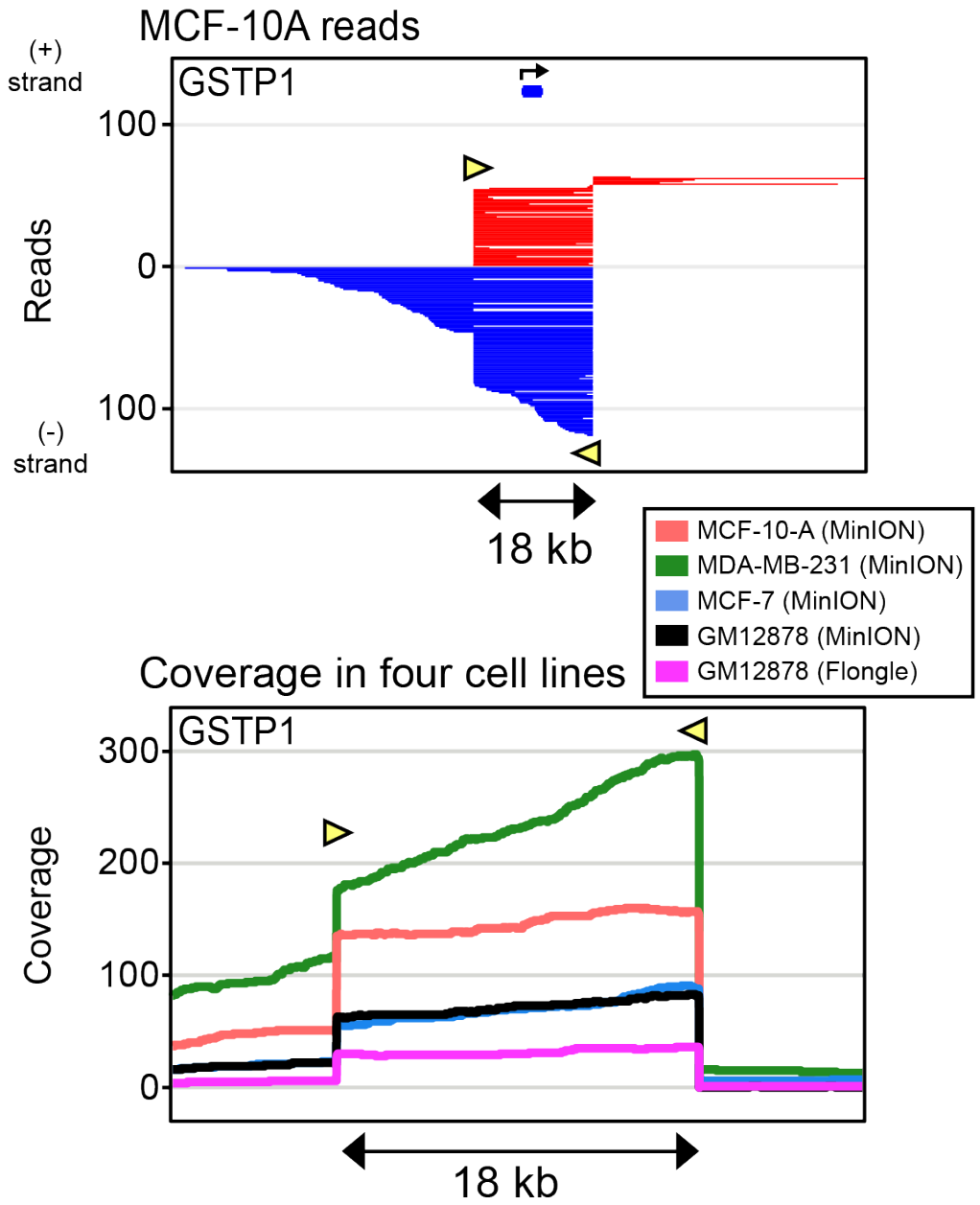


R9 calculates methylation 94% accurate at 77% of sites
NA12878 data shows .895 correlation with bisulfite

Jain et al *Nat Biotech* (2018)
Simpson et al *Nat Methods* (2017)

Cas9 enrichment Method

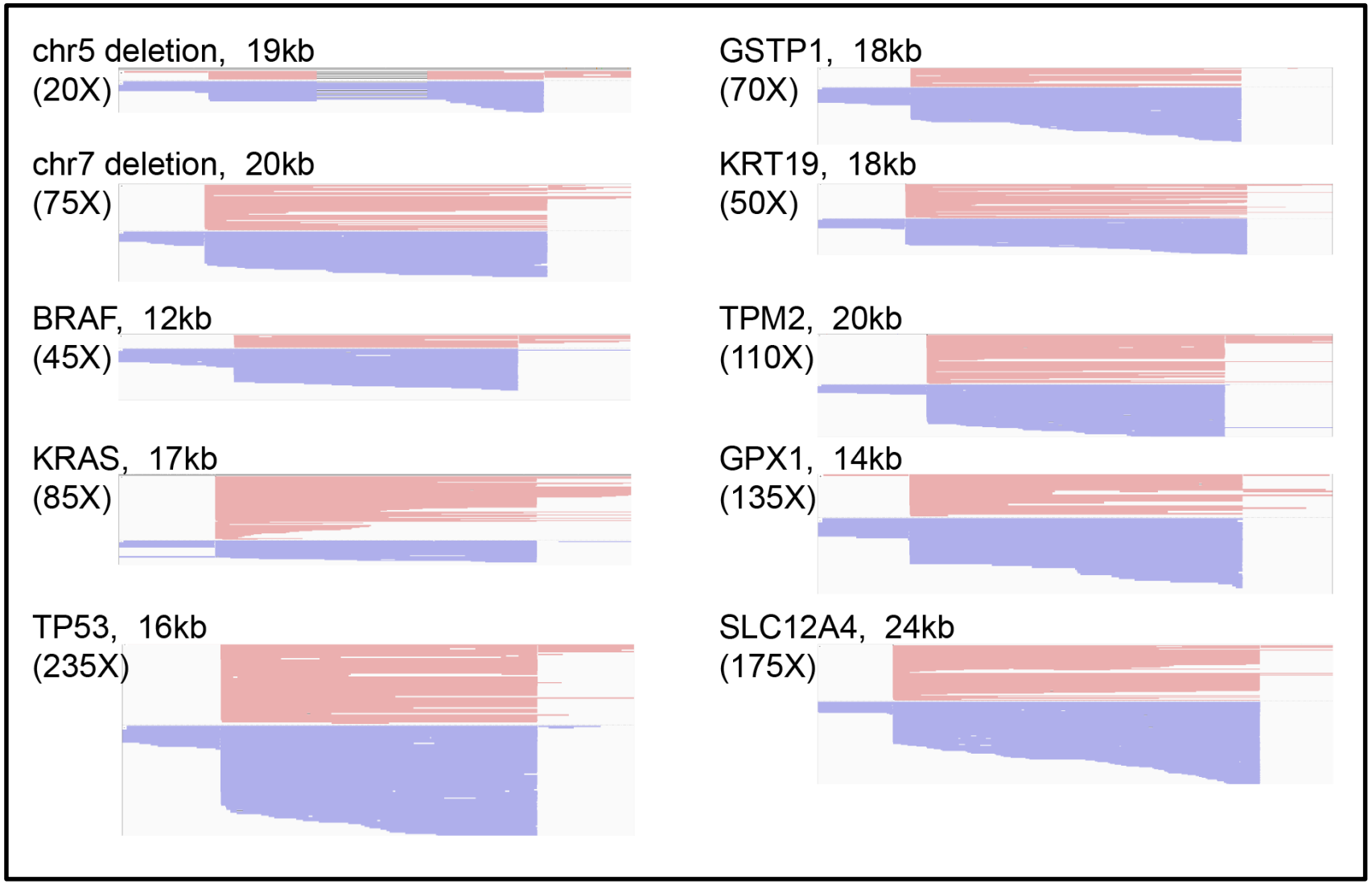




Gilpatrick et al bioRxiv (2019)

Using a panel of guideRNAs

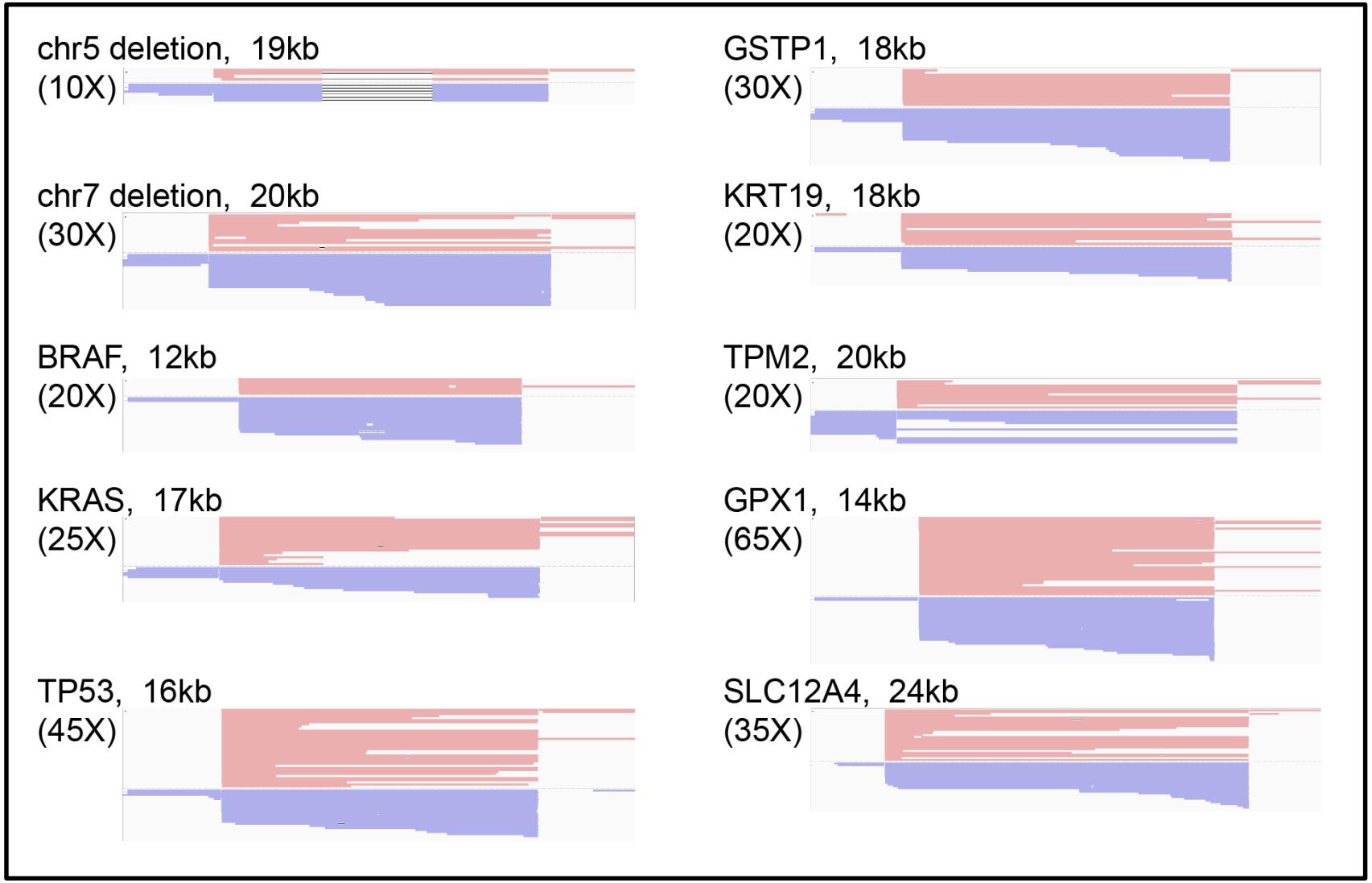
- Yield from
- 3ug GM12878 gDNA
- MinION Flow cell



Gilpatrick et al bioRxiv (2019)

Using a panel of guideRNAs

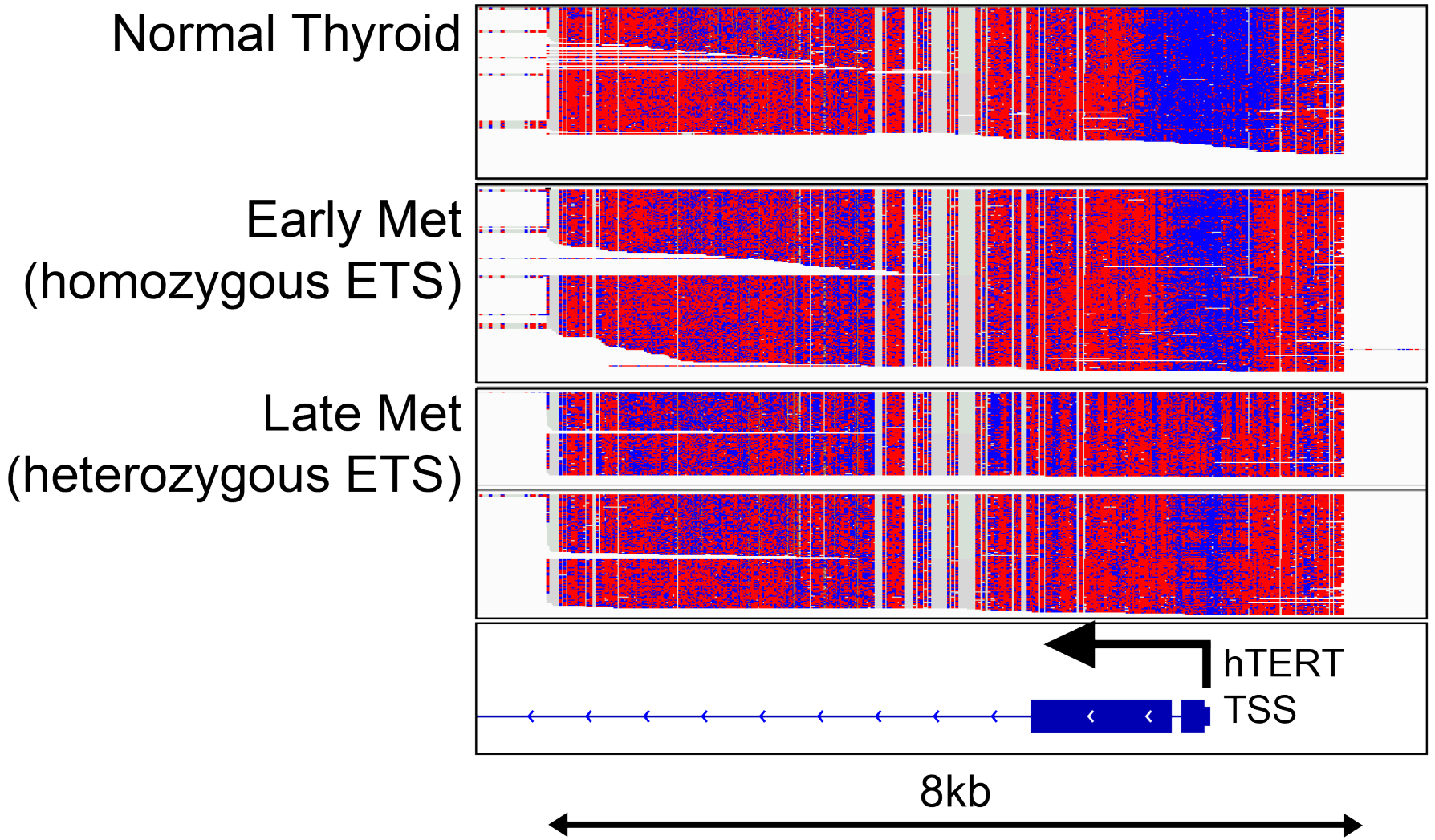
- Yield from
- 3ug GM12878 gDNA
- Flongle Flow cell



Gilpatrick et al bioRxiv (2019)

Enrichment of hTERT region

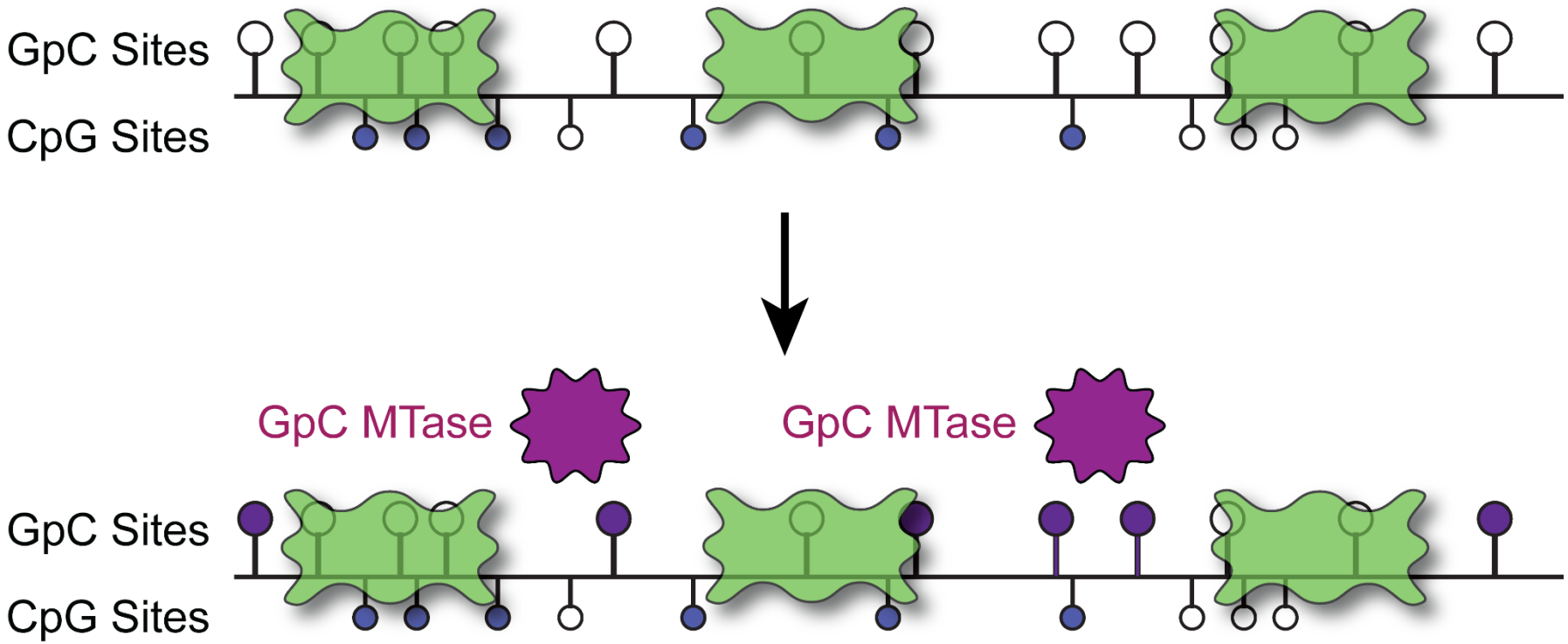
- We observe an “erosion” of the unmethylated (blue) CpG island in the promoter of hTERT in progressive cancer samples
- In the late metastasis a mutation in the ETS binding site of the promoter occurs in one of the alleles
- The mutant allele appears to have a more unmethylated island than the WT allele



NanoNOMe: Chromatin Accessibility with Nanopore

- NOMe-seq : Nucleosome Ocupancy and Methylome sequencing (Kelly et. al. *Genome Res.* 2012)

Simultaneously measures DNA methylation (CpG) and nucleosome occupancy (GpC)

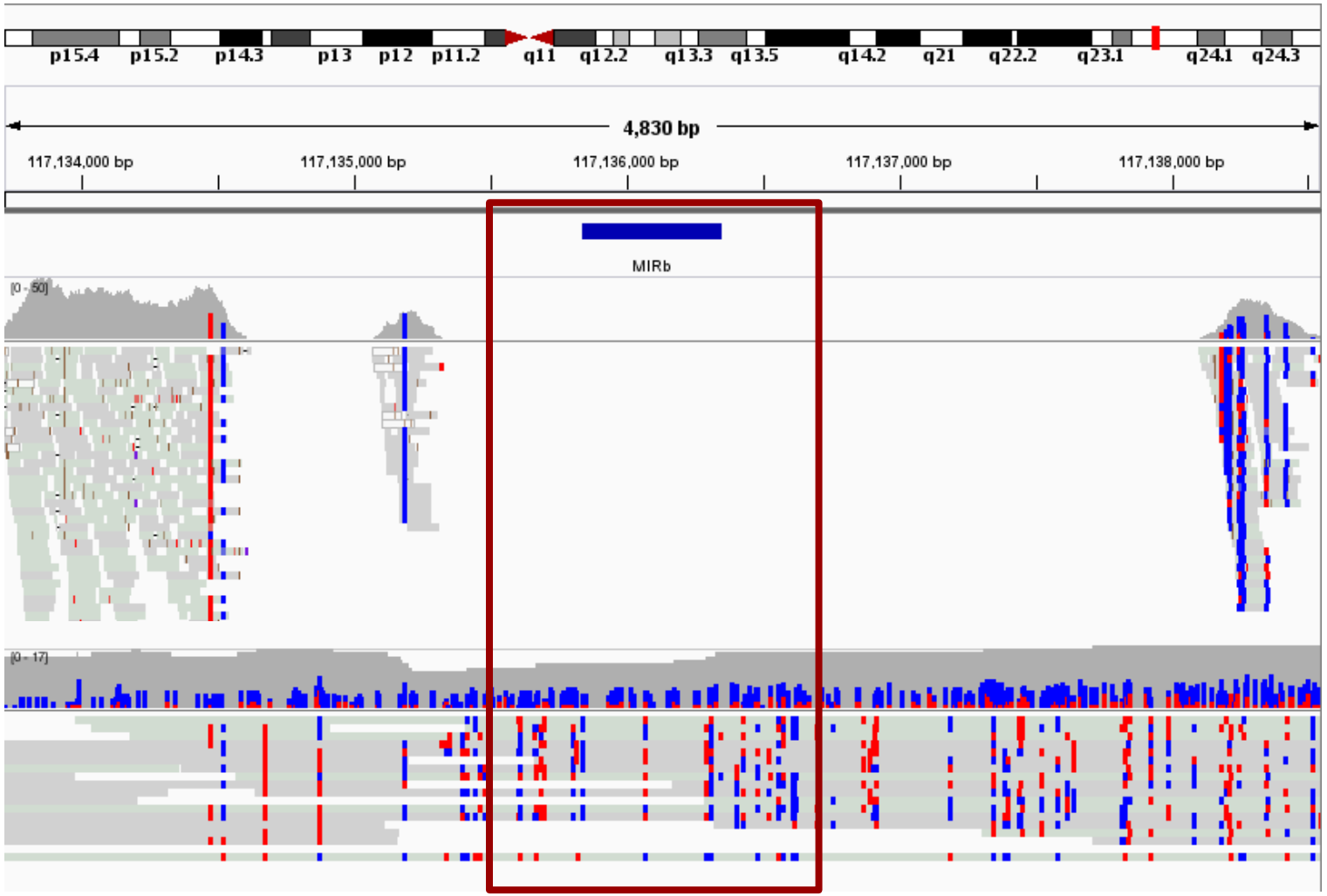


Lee et al. bioRxiv 2018

Methylation in Repetitive Regions

Bisulfite Sequencing (Illumina)

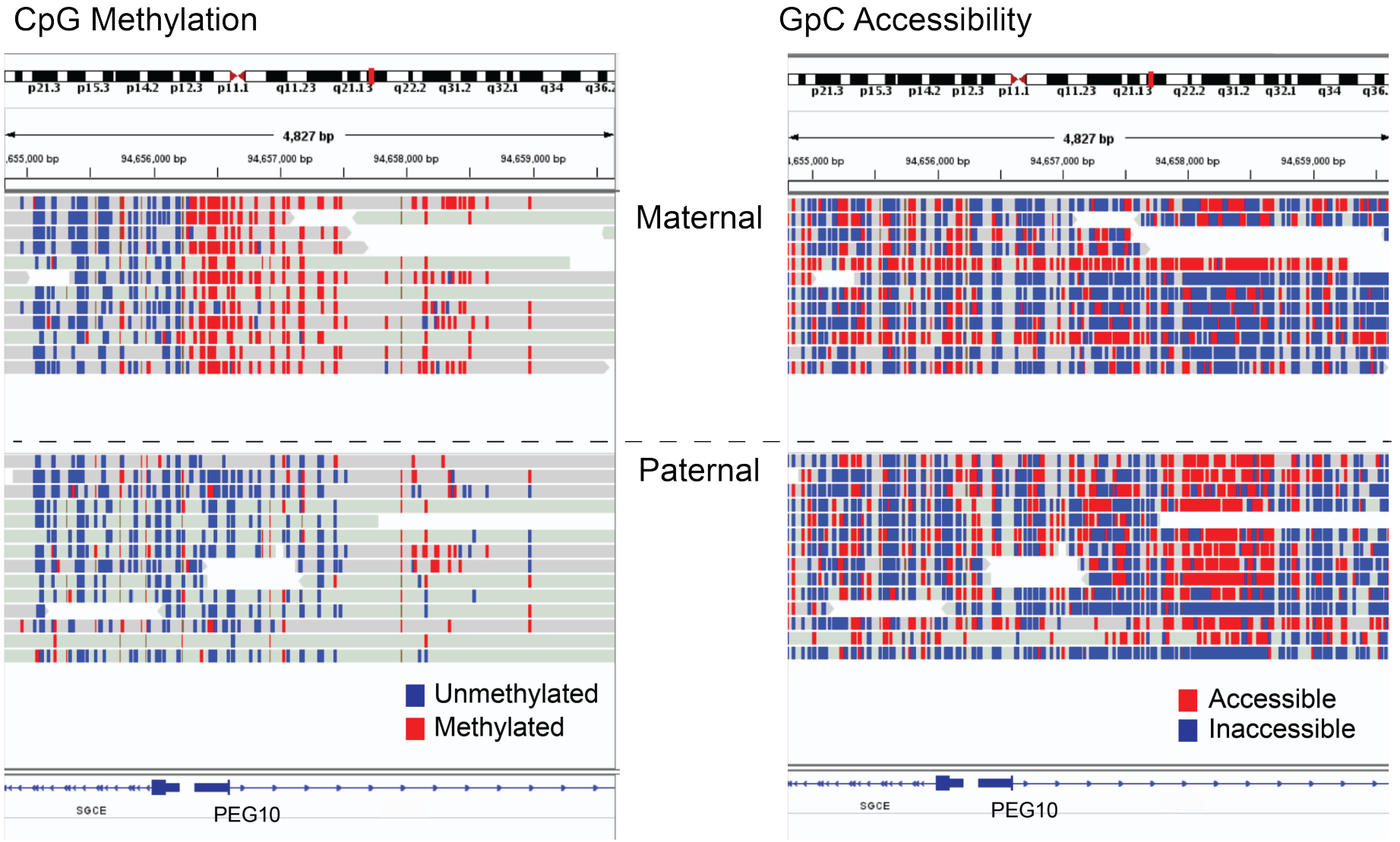
nanopolish methcall (Nanopore)



Regions unmappable by NGS are mappable with long reads

Allele Specific Chromatin and Methylation

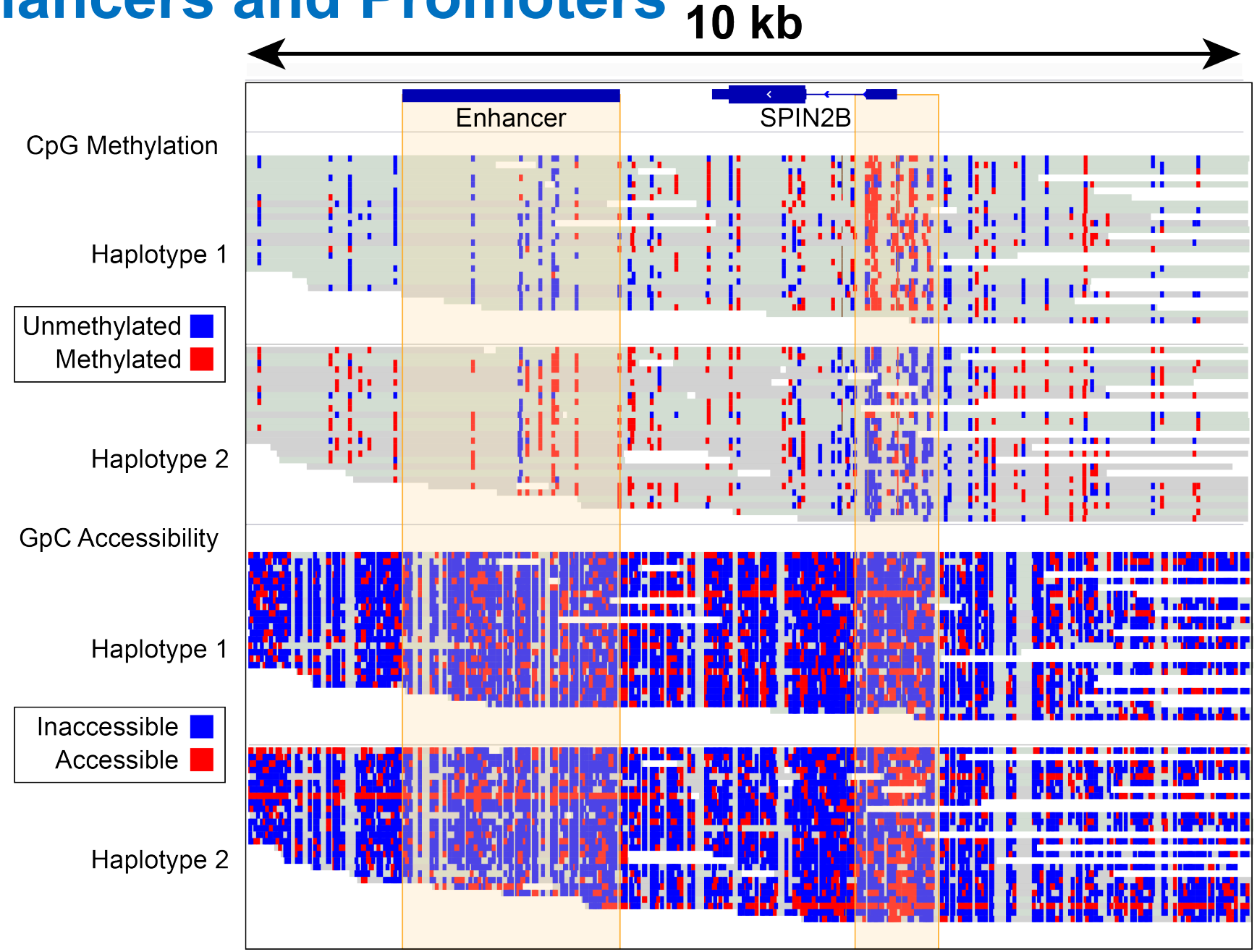
- Using long reads, we are likely to encounter a SNP
- This allows for phased methylation and chromatin data
- Near *PEG10* (imprinted gene):
  - Maternal copy is methylated and inaccessible
  - Paternal copy is unmethylated and accessible



Lee et al. bioRxiv 2018

Coordinated Enhancers and Promoters

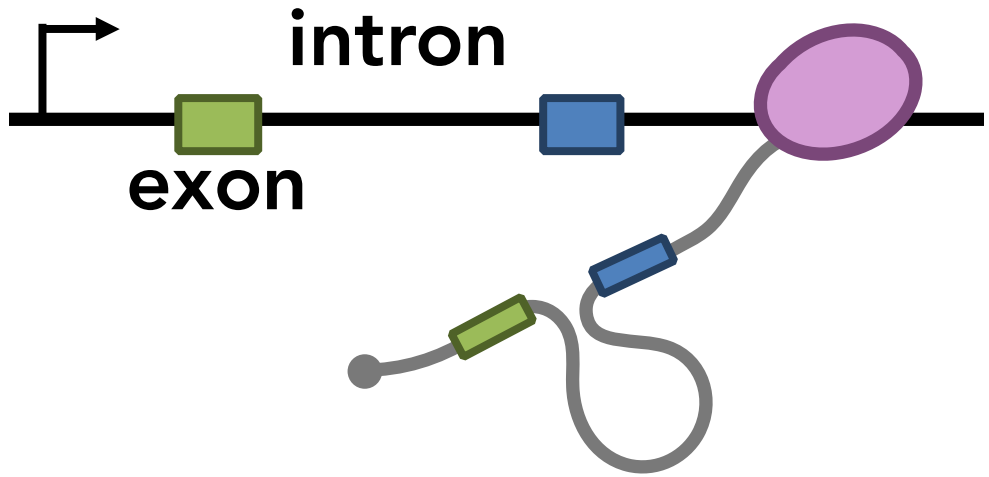
Using long reads, we can examine methylation and chromatin at some promoters and enhancers at the same time

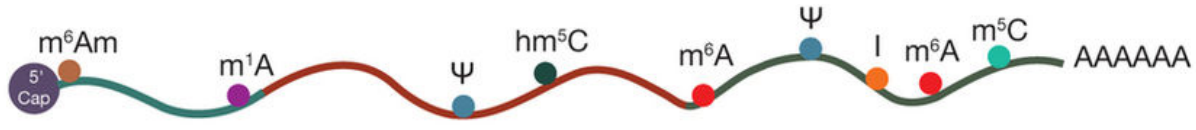


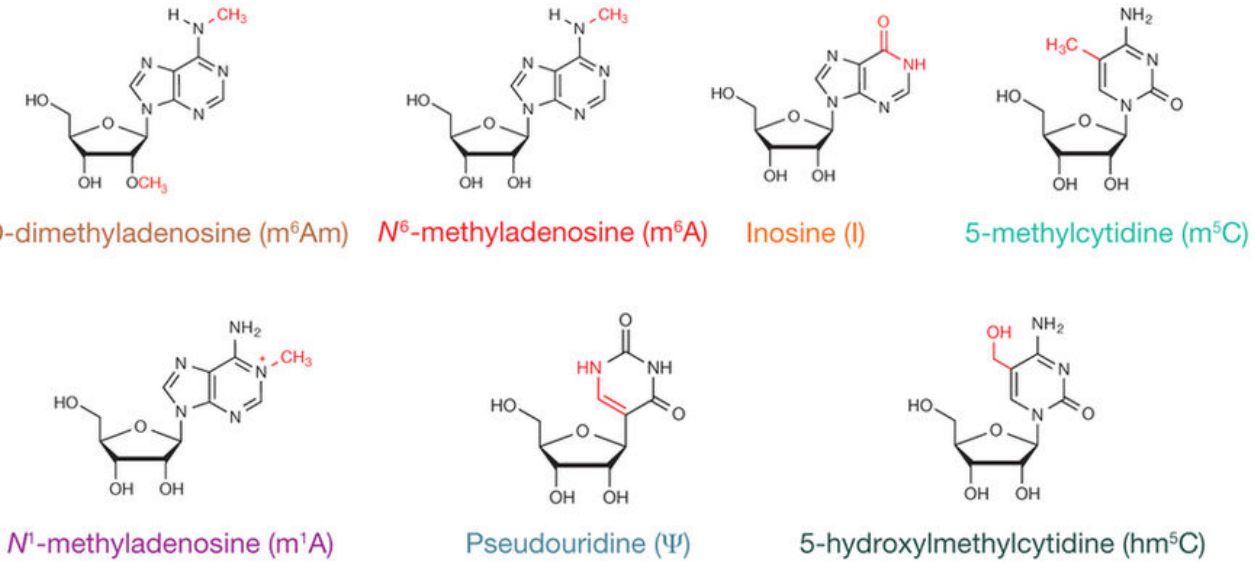
Lee et al. bioRxiv 2018

Questions in RNA:

Pol II

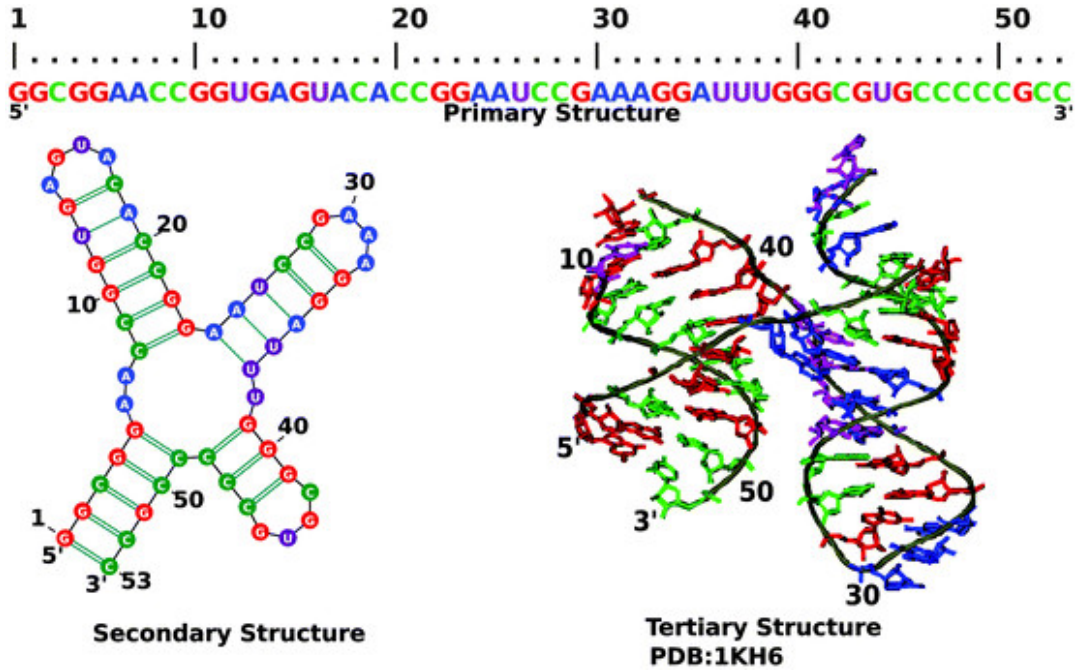






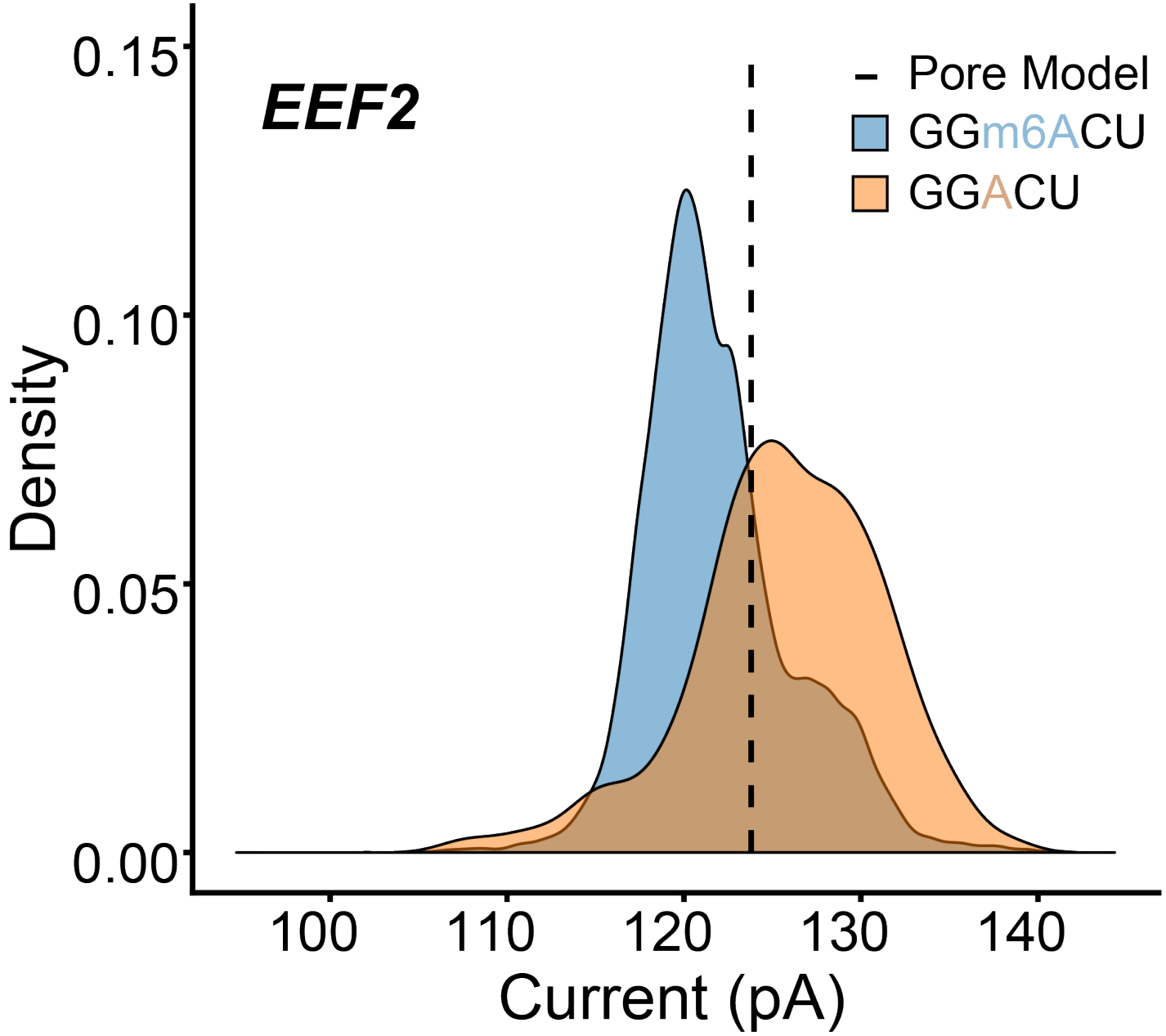
Li, Xiong, Yi, Nature Methods (2017)

- RNA modifications
- RNA dynamics
- RNA structure



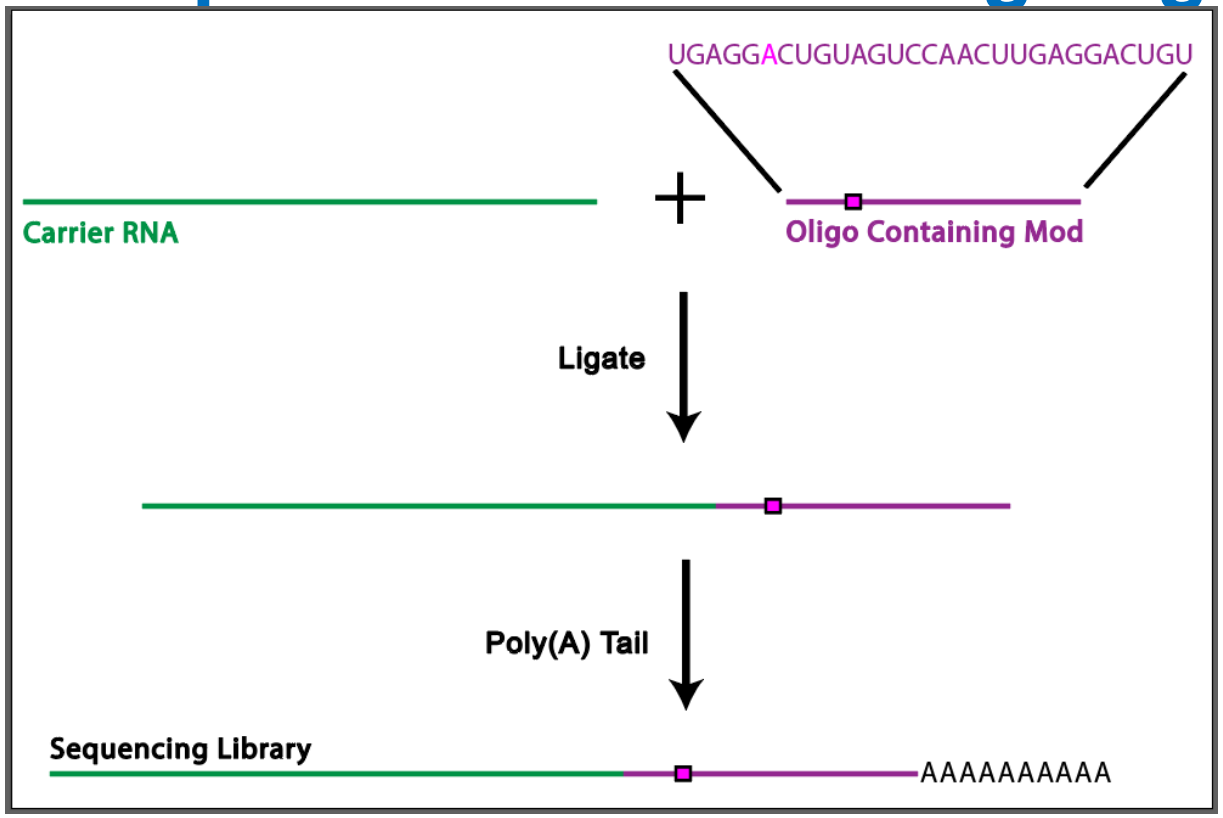
Exploring the dRNA for m6A

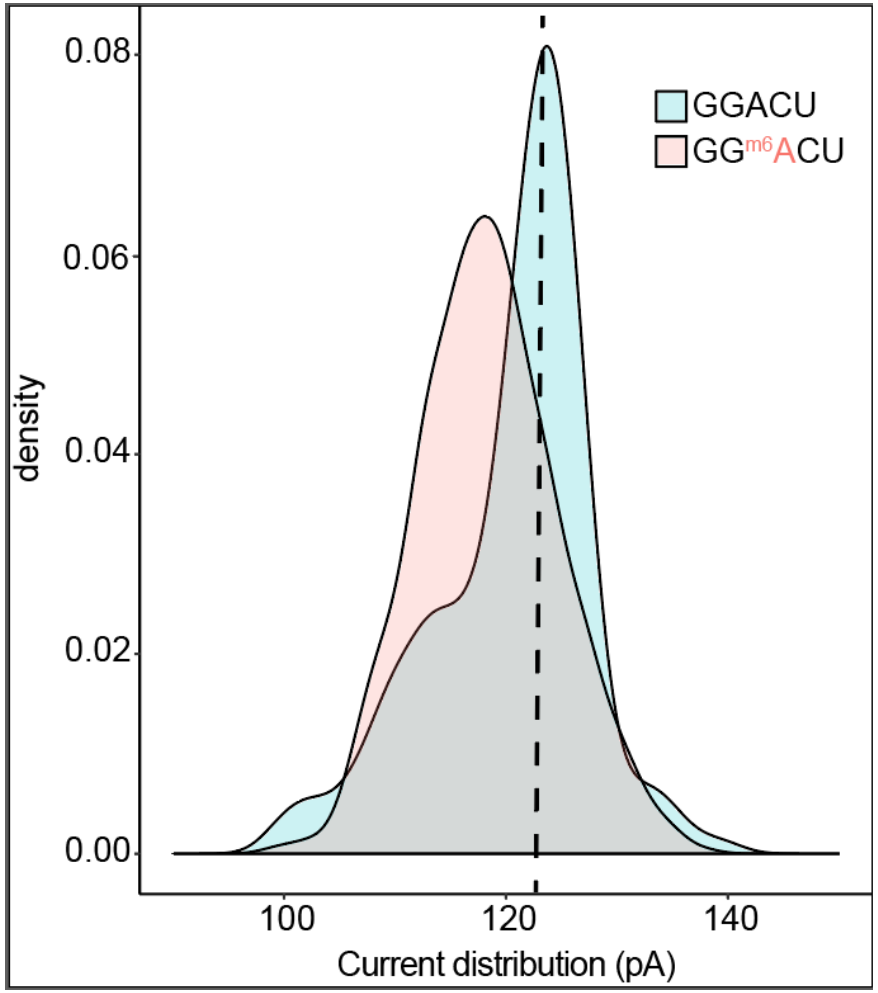
- Eukaryotic elongation factor 2 has a METTL3 motif GGACU (m6A writer) in the mRNA sequence
- Has been shown to have m6A via IP-seq methods (Meyer et al Cell 2012)
- Compared dRNA data with IVT'd dRNA signal



Workman et al bioRxiv (2018)

Training RNA basecaller to recognize modified sites requires truth sets: oligo ligation



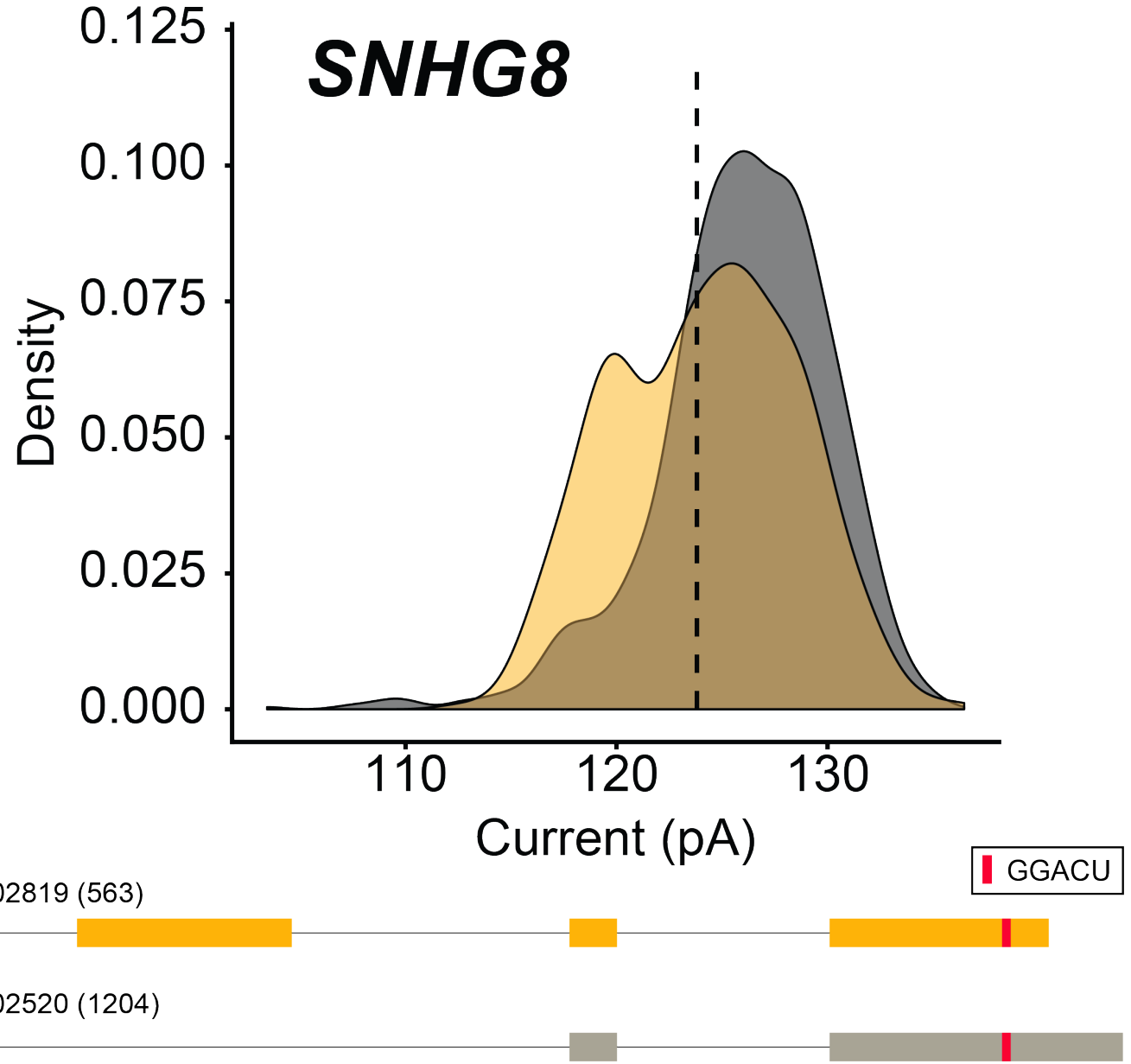


Short oligo (cheaper) containing METTL3 motif GGACU (m6A writer) ligated to handle (to achieve ~100b desired for seq)

Workman et al bioRxiv (2018)

Isoform Specific m6A modifications: *SNHG8*

- Examining isoform dependence of modification signal: METTL3 motif in *SNHG8* isoforms
- Different % of transcripts are modified dependent on isoform



Workman et al bioRxiv (2018)

Summary

- Nanopore technology is full of potential for sequencing, but always choose the right tool for the right job. Often multiple approaches with complementary data yield the best results.
- Multiple bases affect the electrical signal from nanopores; rather than a problem, this can be an advantage, as each base is interrogated multiple times.
- Modifications to the primary DNA sequence (e.g. cytosine methylation) can be detected directly using nanopores
- Exogenous labeling allows simultaneous detection of chromatin and methylation state using nanopore sequencing
- Targeted sequencing with Cas9 allows for long reads in targeted regions, sidestepping issues of cost.
- Direct RNA sequencing suggests we can measure isoforms, poly (A) tail lengths and even RNA modifications

Acknowledgments

National Human Genome Research Institute
1R01HG010538
1R01HG009190

OICR
Ontario Institute for Cancer Research

- Jared Simpson
- Paul Tang
- P.C. Zuzarte
- Michael Molnar

Nanopore RNA Consortia

- UCSC (Akeson, Brooks)
- UBC (Snutch, Tyson)
- OICR (Simpson)
- JHU (Timp)
- Nottingham (Loose)
- Birmingham (Loman)

JOHNS HOPKINS
BIOMEDICAL ENGINEERING

- Jawara Allen
- Brittany Avin
- Sheridan Cavalier
- Yunfan Fan
- Ariel Gershman
- Timothy Gilpatrick
- Isac Lee
- Brittany Pielstick
- Roham Razaghi
- Norah Sadowski
- Rachael Workman

BCM
Baylor College of Medicine

- Fritz Sedlazeck