Washington University in St. Louis
School of Medicine

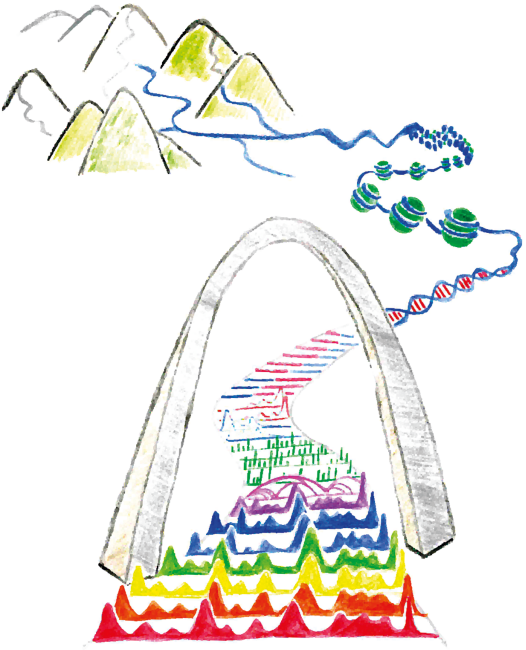
Department of Genetics

系统生物学与生物信息学

海外学者短期讲学系列课程

Current Topics in Epigenomics

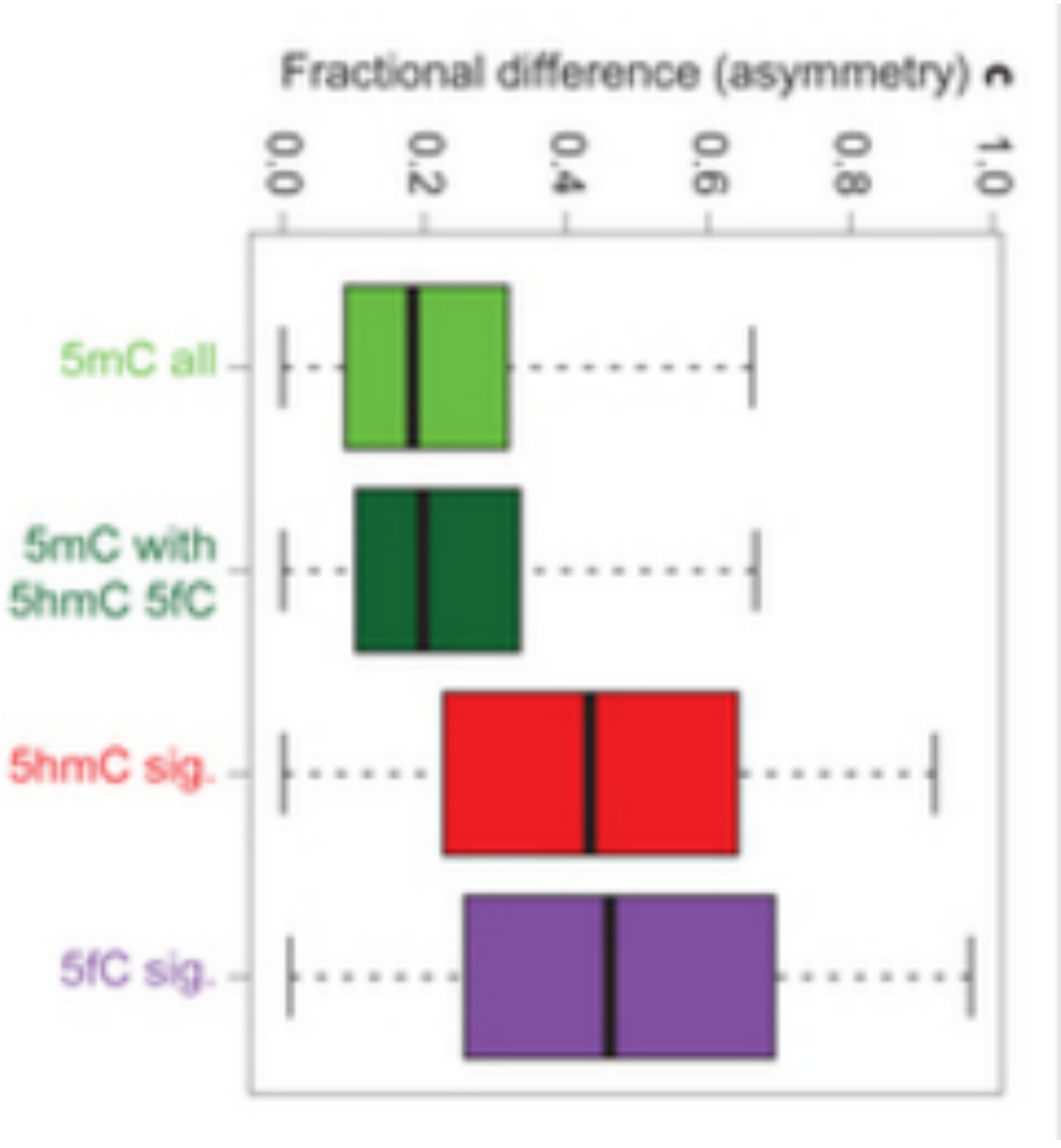
表观基因组学前沿



Ting Wang
Department of Genetics
Center for Genome Sciences and Systems Biology
Washington University School of Medicine

Tsinghua University
April 15-27

Epigenomic technology


c
Fractional difference (asymmetry)
0.0
0.2
0.4
0.6
0.8
1.0
5mC all
5mC with 5hmC 5fC
5hmC sig.
5fC sig.

Trends in Genetics

CellPress

Review

Evolution of Epigenetic Regulation in Vertebrate Genomes

Rebecca F. Lowdon,[1,*] Hyo Sik Jang,[1] and Ting Wang[1,*]

Empirical models of sequence evolution have spurred progress in the field of evolutionary genetics for decades. We are now realizing the importance and complexity of the eukaryotic epigenome. While epigenome analysis has been applied to genomes from single-cell eukaryotes to human, comparative analyses are still relatively few and computational algorithms to quantify epigenome evolution remain scarce. Accordingly, a quantitative model of epigenome evolution remains to be established. We review here the comparative epigenomics literature and synthesize its overarching themes. We also suggest one mechanism, transcription factor binding site (TFBS) turnover, which relates sequence evolution to epigenetic conservation or divergence. Lastly, we propose a framework for how the field can move forward to build a coherent quantitative model of epigenome evolution.

Trends

Epigenome evolution is characterized by variable conservation and divergence across the genome; within a clade (here vertebrates), rates of conservation or divergence are highly genome feature-specific.

TFBS turnover can mediate epigenome conservation or divergence.

Developmental genes are enriched in loci with divergent chromatin features, suggesting that rapid epigenome evolution may contribute novel regulatory mechanisms for lineage-

Questions on Epigenome Evolution:

1）细胞治疗是否能启动甲基化。我们实验室做细胞治疗，有时注射细胞后，发现比对照有增重，注射的细胞而非质粒，肯定不能改变基因，那这种增重究竟是不是epigenetic的。如何检测？直接抽血？

2） 现已知的表观遗传进化的现象，以及它们在表观遗传进化中所起的作用

3） epigenetic revolution的研究，取得了哪些重要或有趣的成果？举个例子？

4）epigenetic revolution的研究的难点在哪？比较同一物种内的epigenetic和/或genetic差异，而不是物种之间的差异，是否对我们研究基因调控等方面更直接和更有帮助？

Questions on Big Bio Data:

a) 基于已知数据库，对于新的数据，我们用怎样的方法将这些数据归类并识别的。比如基因组里有很多的信息是我们无法归类的，那么我们应该对这些不确定的数据怎样处理，归类到编码区，非编码区，启动子或增强子等等？

b) 对于big data processing的入门学习，我应该从哪些方面入手？

c) 怎样将多重数据如表达，mutation,甲基化，Hi-C,chip-seq以及其他数据进行有效整合来阐述表观中重要的assumptions（比如表达与表观遗传修饰之间的定量关系）？

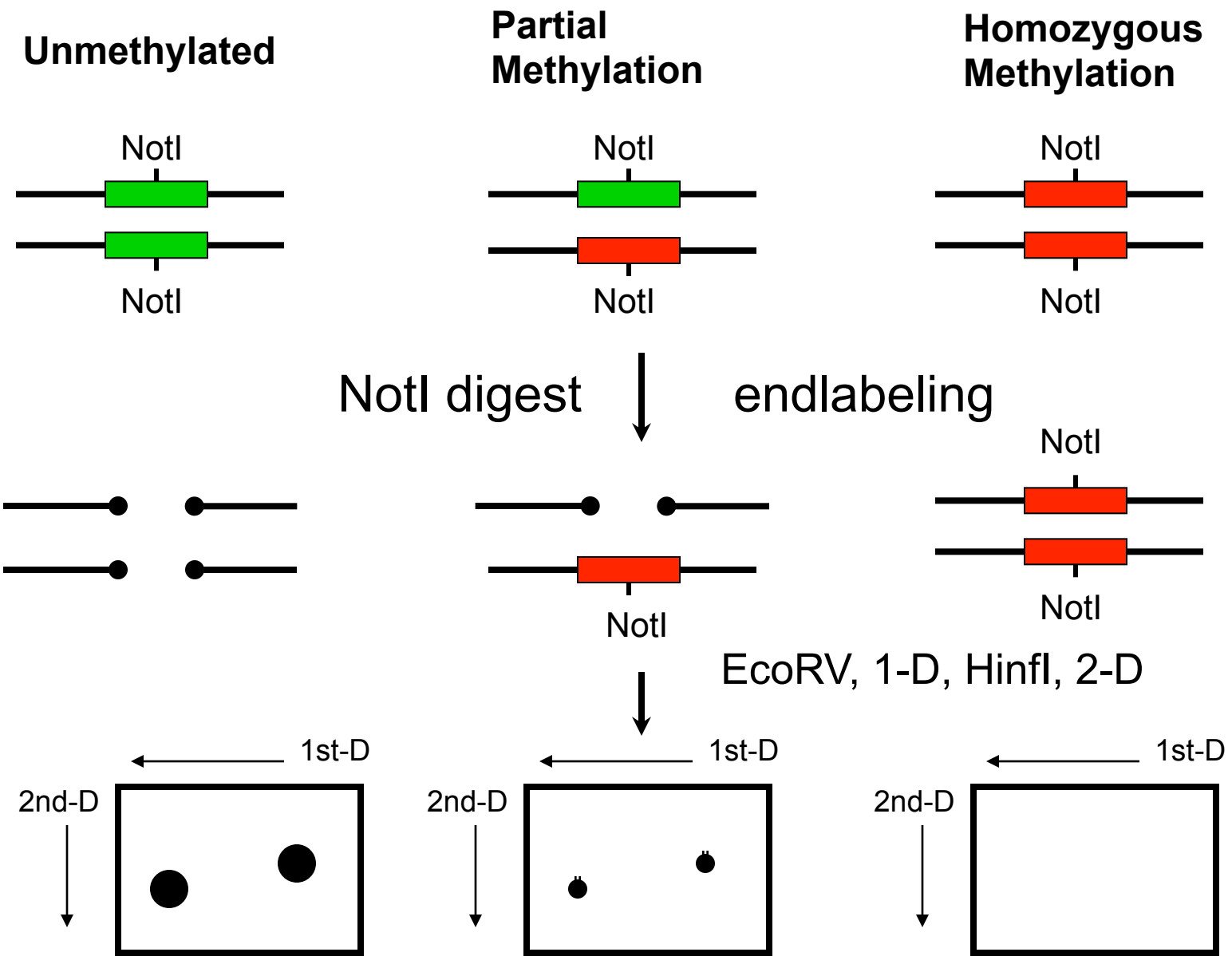
d) 表观的修饰通常都是genome-wide的形式进行，怎么样通过genome-wide的大数据分析来阐述如reader,writer,eraser间作用具有特异性(specificity)和选择性(selectivity)?

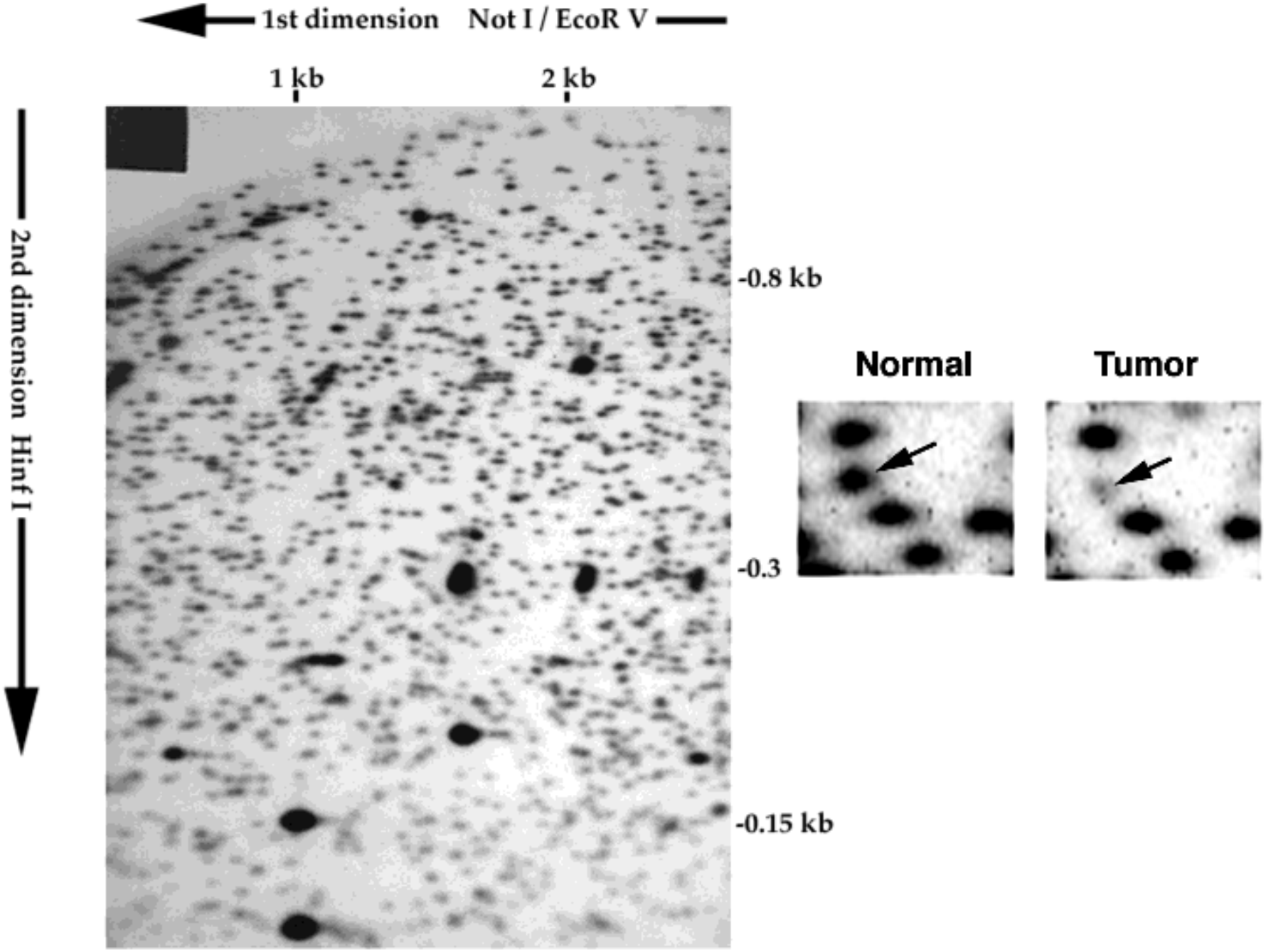
e) 在大数据时代，数据一致性将会是个问题。不同实验室发布的数据之间一致性到底有多大？single cell得到的数据和mix cell 得到的数据一致性有多大？如果出现不一致，怎么从数据分析的角度评判哪个数据好？

f) 在同一种细胞群体中和一个细胞的增殖生长周期中，表观遗传学的各检测方法和检测的指标的波动（方差）有多大？能否充分利用现有的大数据信息对这个方差做出估计？

How to detect epigenetic marks?

Restriction Landmark Genome Scanning (RLGS)



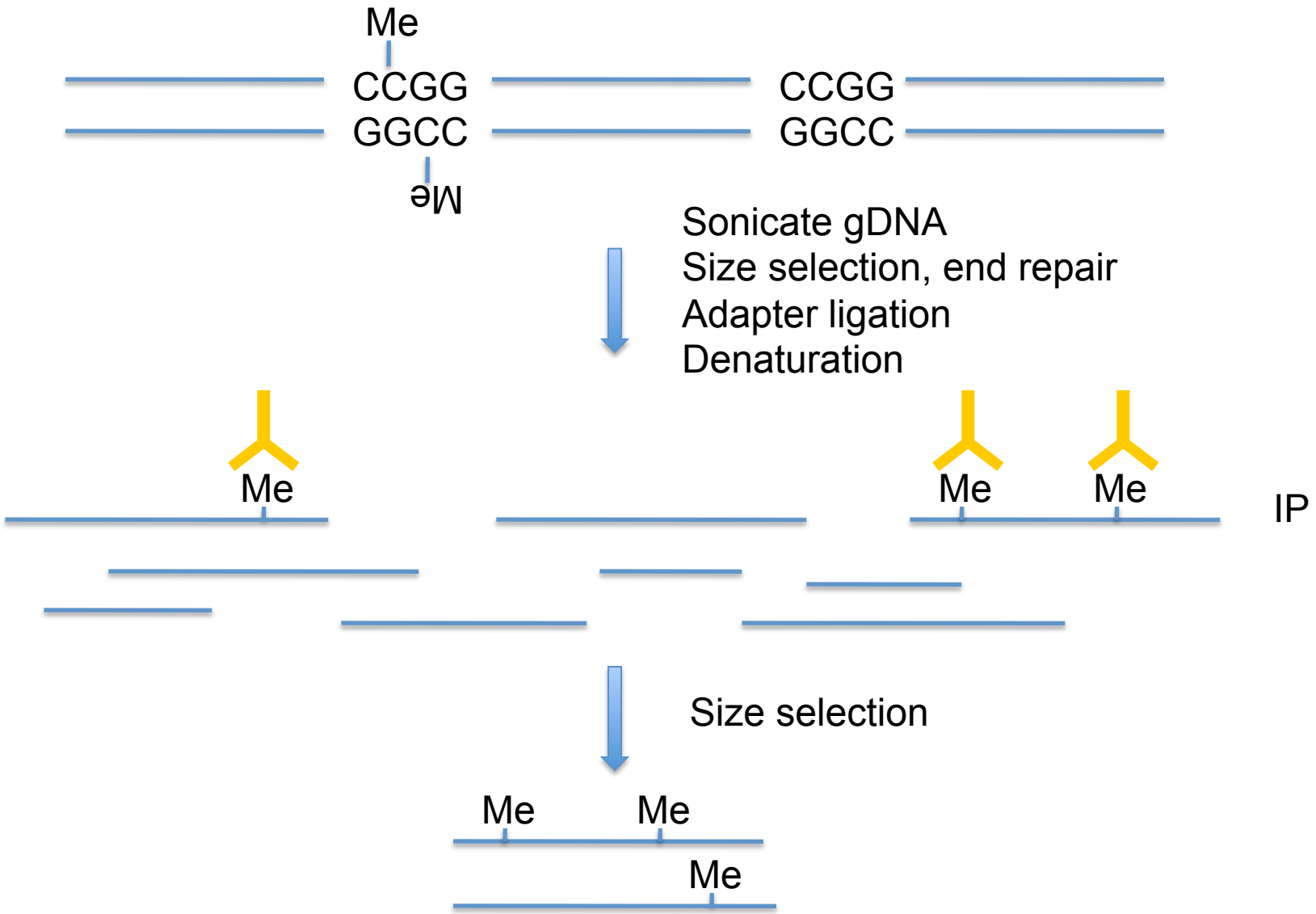

1st dimension Not I / EcoR V
1 kb
2 kb
2nd dimension Hinf I
-0.8 kb
-0.3
-0.15 kb
Normal
Tumor

Scaling to high coverage, high resolution

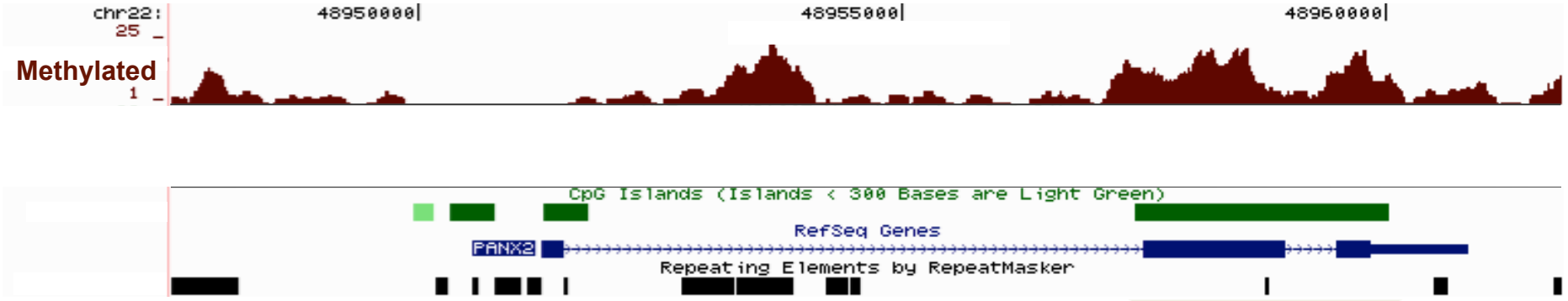
- Enrichment based methods
  - MeDIP-seq
  - MBD-seq/MethylMiner
- Restriction enzyme based methods
  - MRE-seq
  - HELP, Methyl-MAPS, Methyl-seq
- Bisulfite based methods
  - MethylC-seq
  - RRBS, bisulfite padlock
- Direct reading of modified nucleotides
  - SMRT
  - Nanopore sequencing

Enriching for methylated DNA targets

MeDIP-seq and MBD-seq



Typical MeDIP data on a genome browser

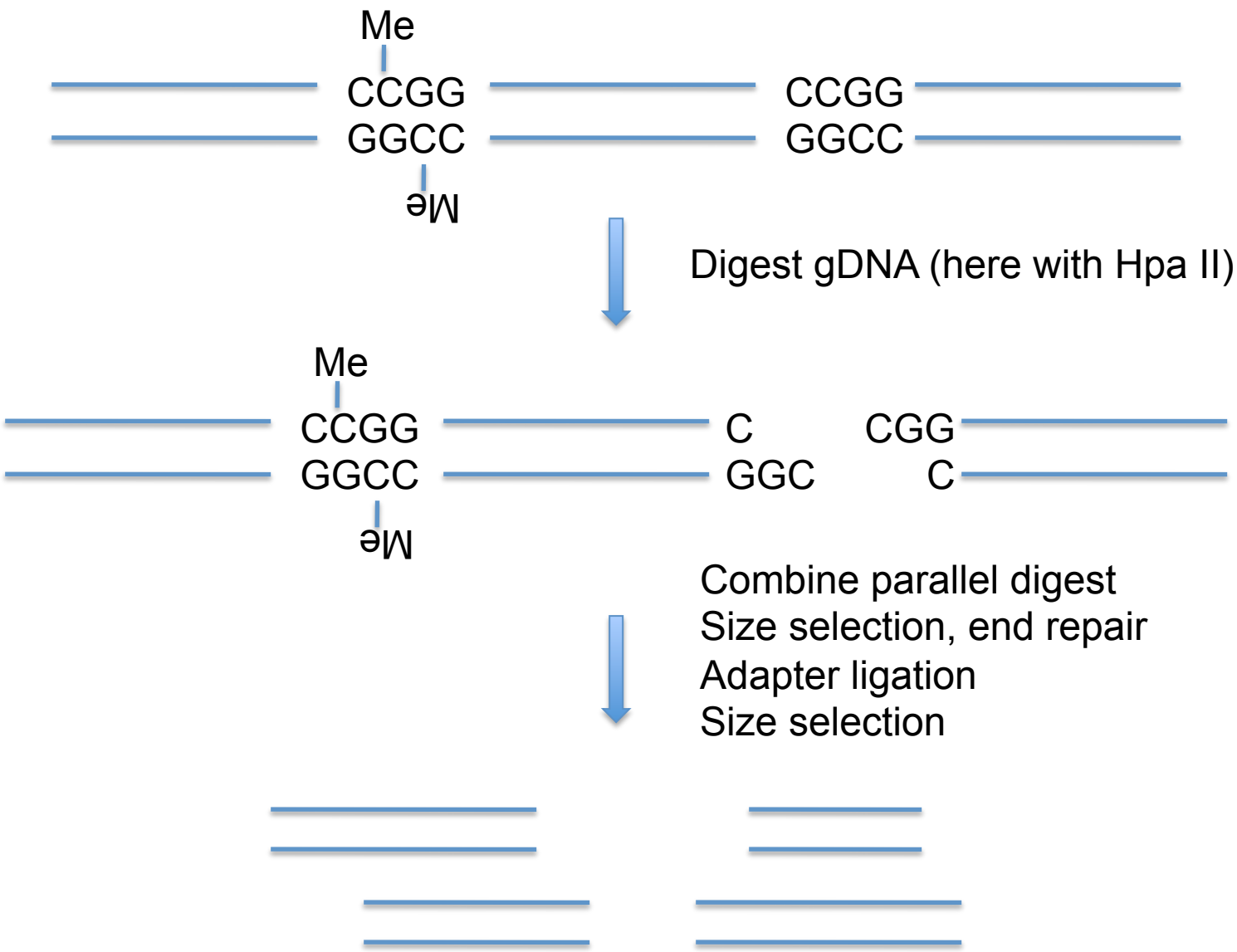


5' CpG islands are unmethylated

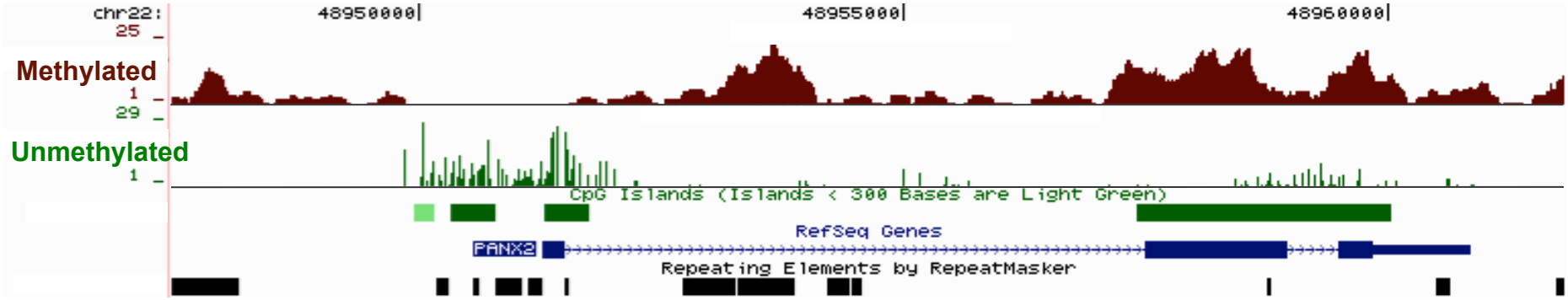
3' CpG island is partially methylated

Taking advantage of methylation dependent restriction enzymes

MRE-seq


Me
CCGG
GGCC
Me
CCGG
GGCC
Digest gDNA (here with Hpa II)
Me
CCGG
GGCC
Me
C
GGC
CGG
C
Combine parallel digest
Size selection, end repair
Adapter ligation
Size selection

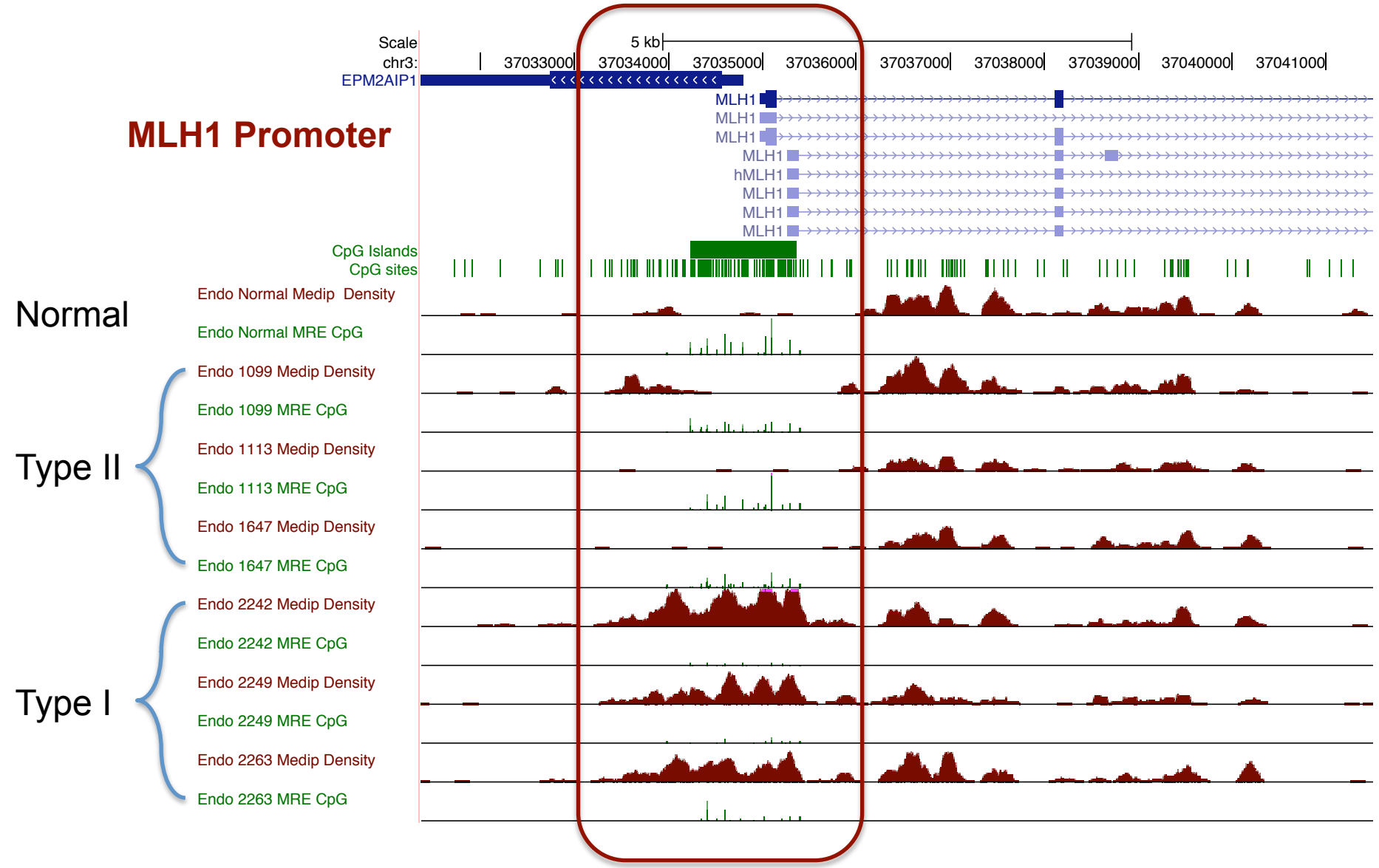
Typical MeDIP/MRE data on a genome browser



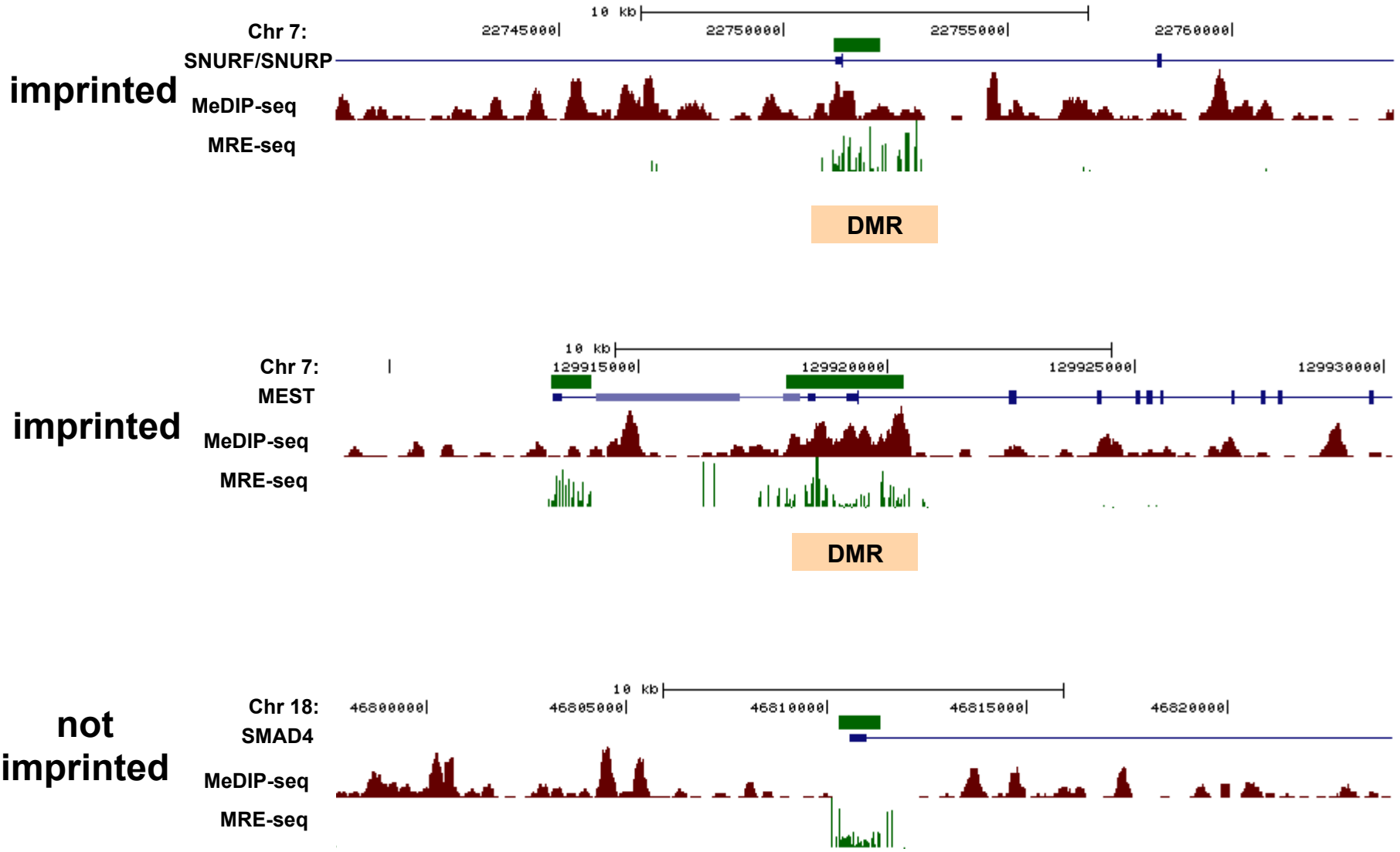
5’ CpG islands are unmethylated

3’ CpG island is partially methylated

MLH1 Promoter

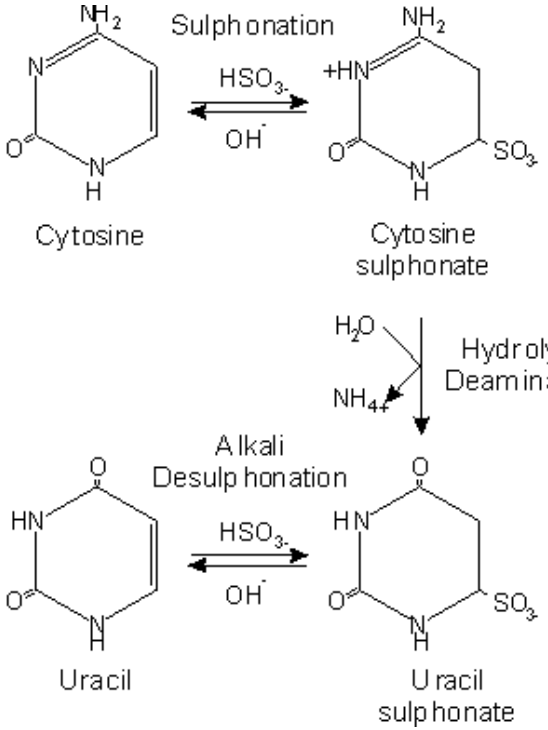


Allele-specific methylation at imprinted genes



Gold standard: bisulfite sequencing

Bisulfite sequencing



Sonication

Watson >>AC^{m}GTTCGCTTGAG>>

Crick <<TGC^{m}AAGCGAACTC<<

C^{m} methylated
C Un-methylated

Denaturation

Watson >>AC^{m}GTTCGCTTGAG>> Crick <<TGC^{m}AAGCGAACTC<<

Bisulfite Treatment

BSW >>AC^{m}GTTUGUTTGAG>> BSC <<TGC^{m}AAGUGAAUTU<<

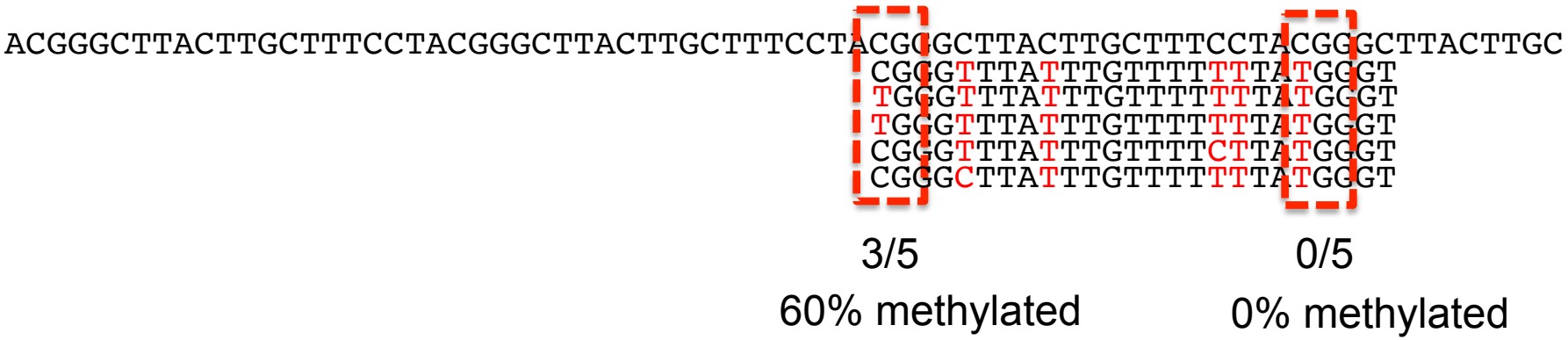
PCR Amplification

BSW >>AC^{m}GTTTGTTTGAG>> BSC <<TGC^{m}AAGTGAATTT<< C-poor stands

BSWR <<TG CAAACAAACTC<< BSCR >>ACG TTCACTTAAA>> G-poor stands

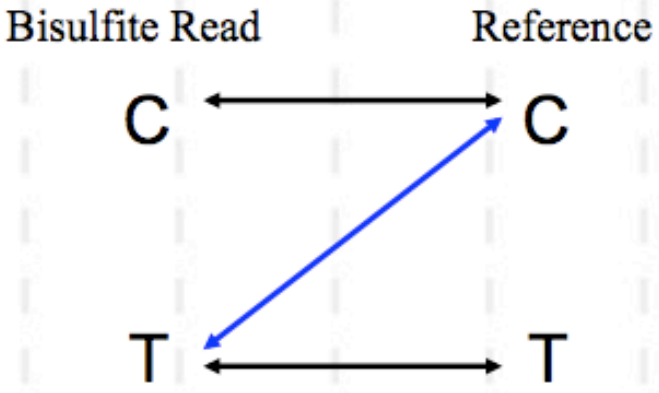
Sequencing and Mapping

Methylation call

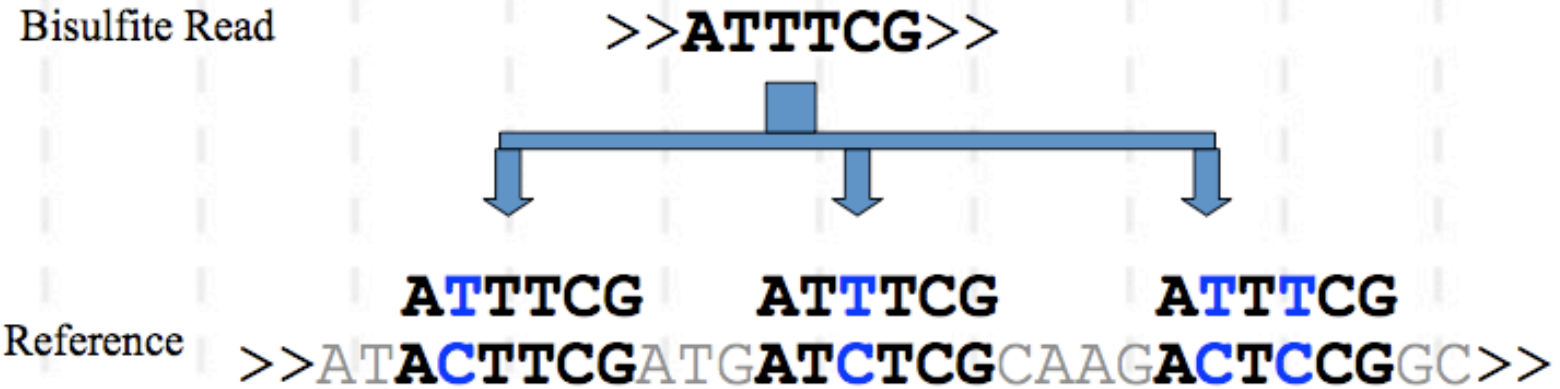


Bioinformatics challenges in mapping high throughput bisulfite reads

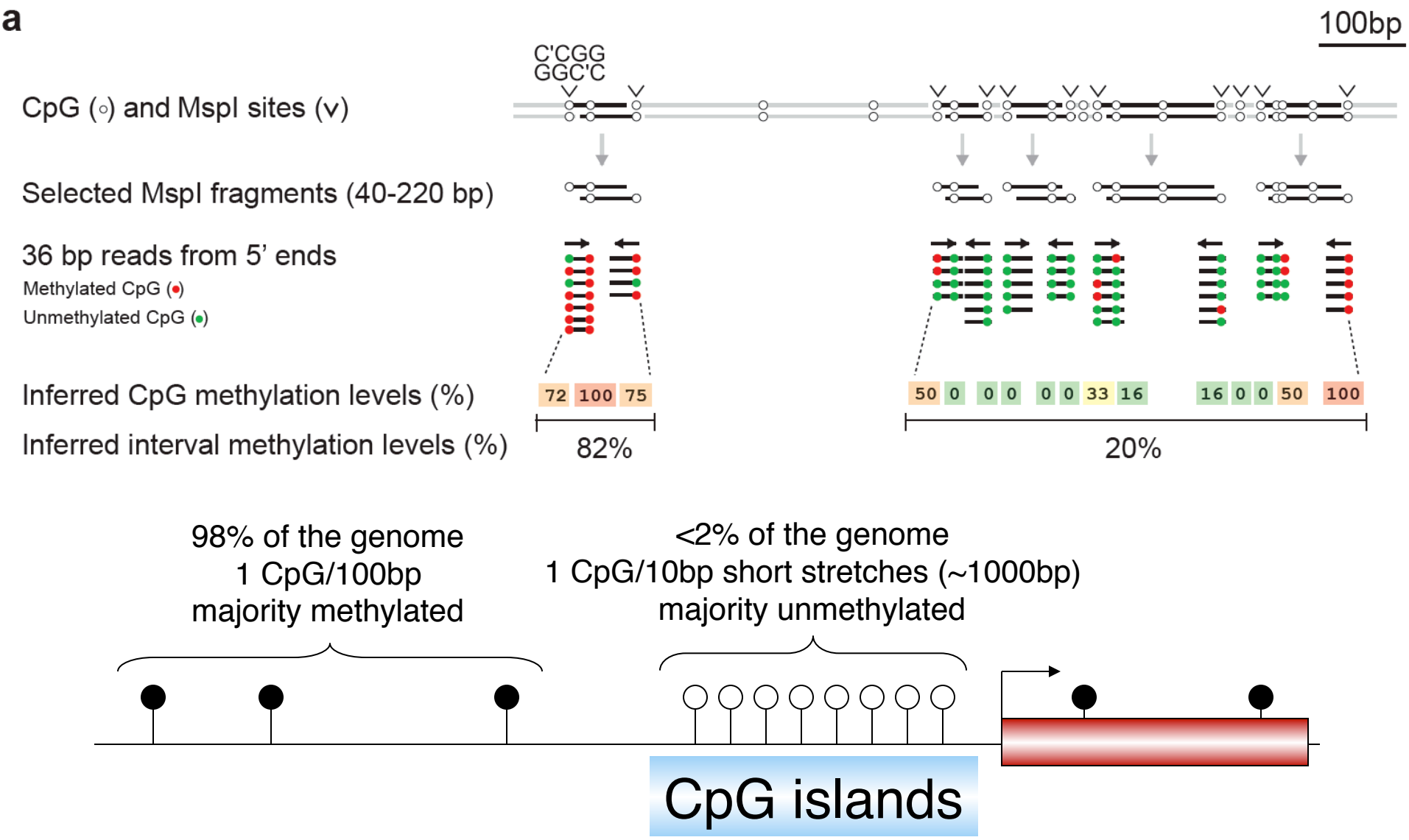
- Mapping Asymmetry



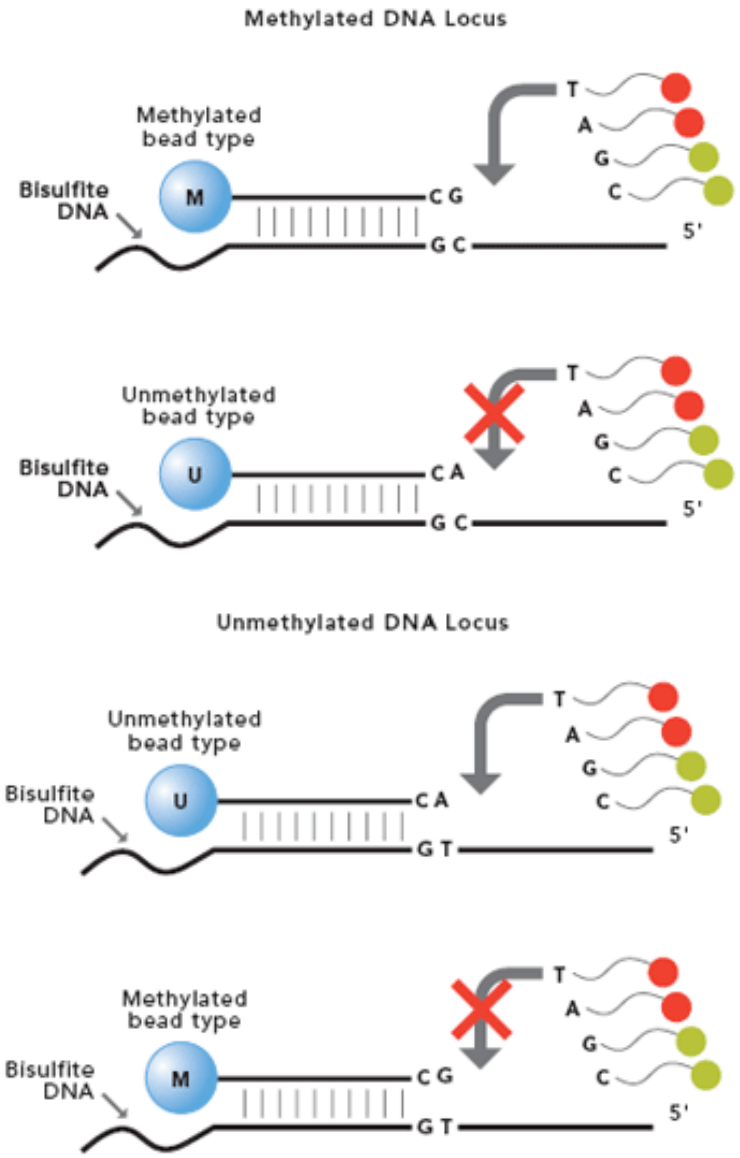
- Increased Mapping Space



RRBS: Reduced Representations Allow Enrichment of CpG Dinucleotides



5-mC detection with the Infinium BeadChip



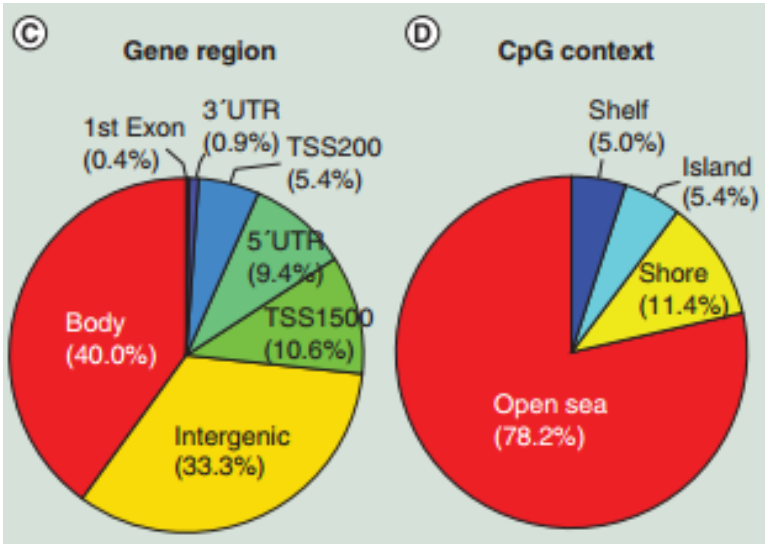
- Pros
  - Relatively inexpensive
  - Single-base resolution
  - Internal quality controls
  - Highly reproducible (r >0.98)
  - PCR-free protocol
  - Works with FFPE samples
- Cons
  - Only covers a subset of the methylome
  - Bisulfite treatment can damage DNA
  - Dependent on bisulfite conversion

www.illumina.com

5-mC detection with the Infinium BeadChip

Array	Year released	# of Sites	Targeted sites
HumanMethylation27k	2008	> 27k	• >14K RefSeq genes
HumanMethylation450k	2011	> 450k	• 99% of RefSeq genes
Infinium MethylationEPIC	2015	> 850k	• > 90% of sites in the 450k • FANTOM5 and ENCODE enhancers (350k sites) • CpG sites outside of CpG islands • ENCODE open chromatin

Context of the new EPIC probes

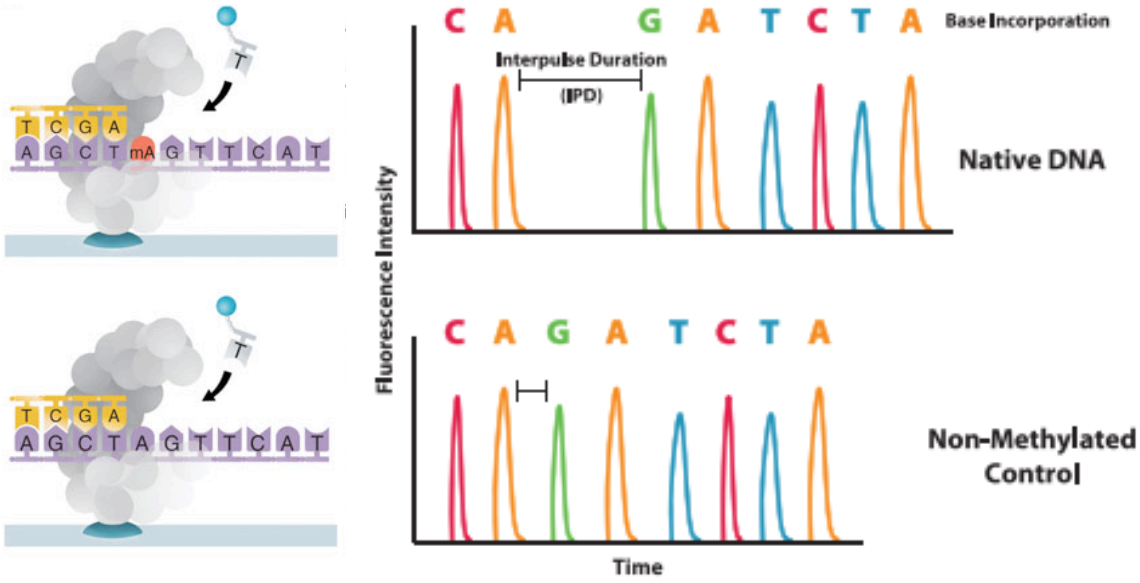


www.illumina.com | Moran et al., *Epigenomics* (2016)

Direct detection of modified nucleotides

Single-molecule real-time (SMRT) sequencing

SMRT sequencing discriminates between different bases by analyzing variations in polymerase kinetics



- Pros
  - Single-base resolution
  - Measures absolute levels of many modified nucleotides
  - “Raw” DNA is used
  - Long reads

Cons

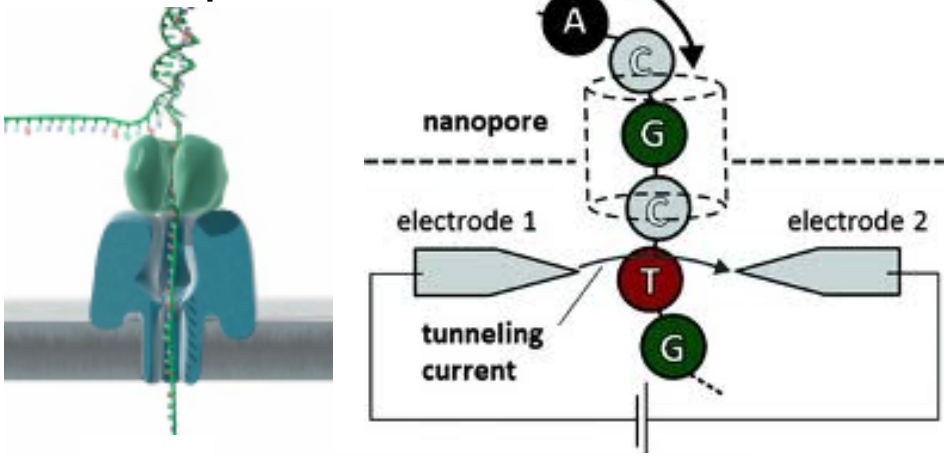
- Suboptimal accuracy
- Low throughput

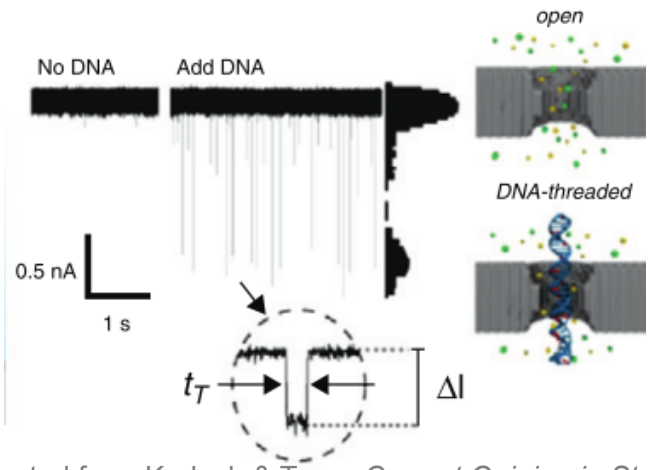
Figure adapted from Clark et al., *Nucleic Acids Research* (2011) | Kinney et al., *Epigenetic Alterations in Oncogenesis* (2013)

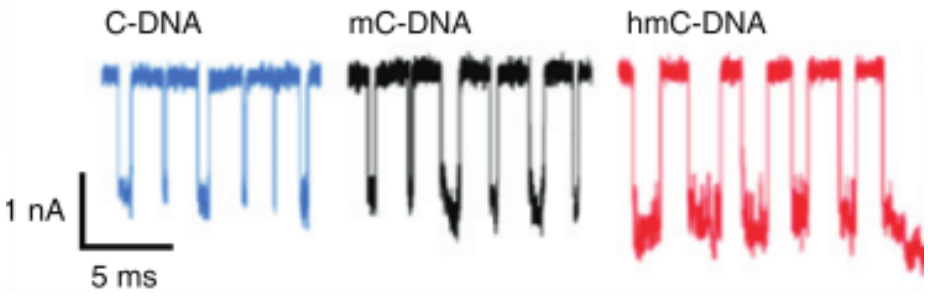
Nanopore sequencing

Nanopore amperometry methods can discriminate between C, 5-mC, and 5-hmC due to differences in current profiles.

- Pros
  - Single-base resolution
  - Measures absolute levels of many modified nucleotides
  - “Raw” DNA is used
  - Long reads
- Cons
  - Suboptimal accuracy
  - Low throughput

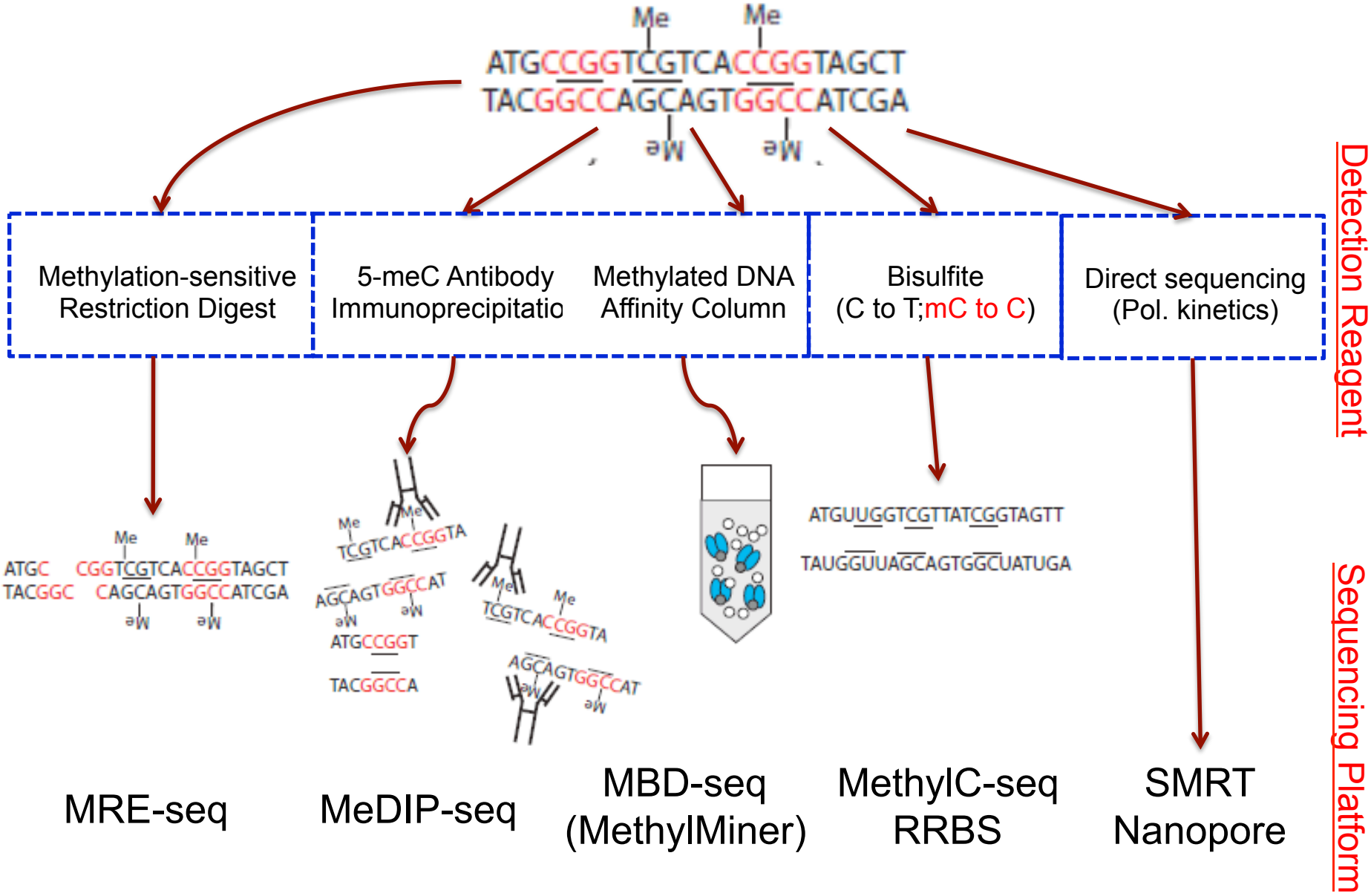






Figures adapted from Korlach & Tuner *Current Opinion in Structural Biology* (2012) | Kinney et al., *Epigenetic Alterations in Oncogenesis* (2013)

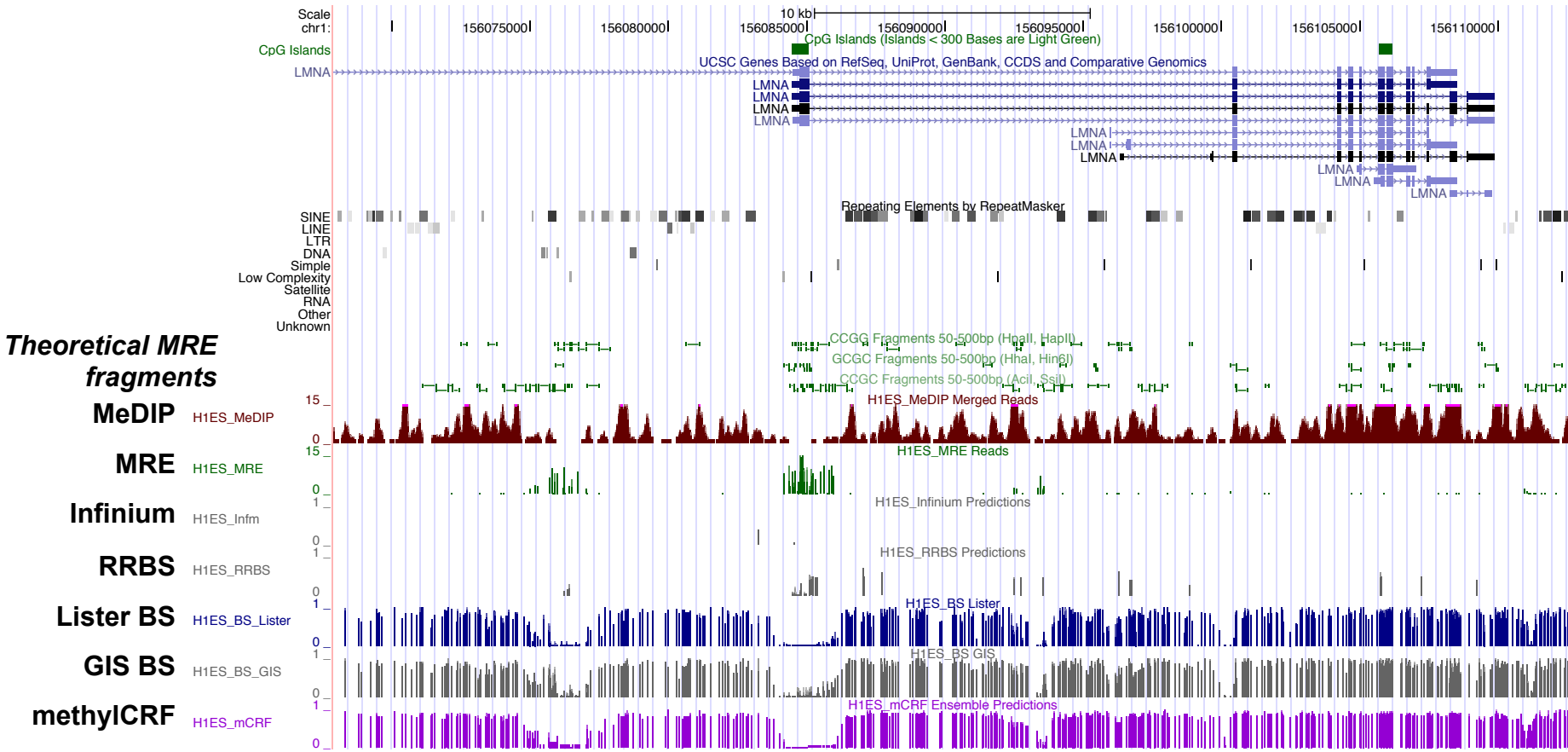
Modern DNA Methylomics



Comparison of Methylome coverage

	CpG coverage	CpG island coverage	Resolution (bp)	Illumina Lanes
Genome total	28 M	28 K	NA	NA
BS shot gun	26 M	27 K	1	207 (2009) 3 (now)
RRBS	0.2-1M	15 K	1	1
MRE-seq and MeDIP-seq	25 M	27 K	1 and 200	8 (2009) half (now)
Golden-Gate	1,500	800	1	NA
Infinium	27,500 480,000 (2012) 850,000 (2016)	12 K 27 K 27 K	1	NA

Predicting single CpG methylation level with Conditional Random Field (methylCRF)

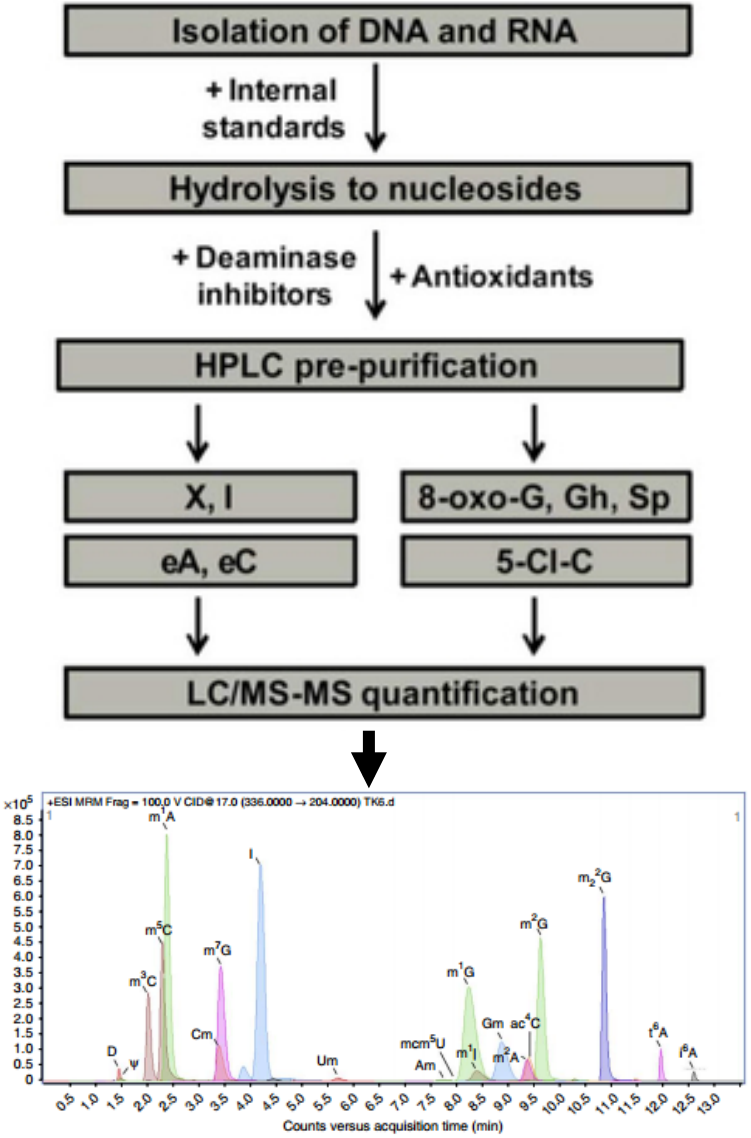


Technologies for quantifying the methylome & hydroxymethylome

DNA modification	Measurement	Non-targeted enrichment	Targeted enrichment	Whole genome	Arrays
5mC	Absolute (single base)	RRBS, mRRBS, LCM-RRBS or scRRBS	Microdroplet PCR Bisulphite, Patch PCR, mTACL, BSPP, LHC-BS (pre- and post-conversion) or RSMA	WGBS, T-WGBS or PBAT	Infinium BeadChip
	Relative (peak)	MRE-seq, MeDIP-seq, MBD-seq or MethylCap-seq			CHARM or MeKL-ChIP
5mC oxidation derivatives	Absolute (single base)	RRHP	Locus-specific sequencing with TAB-seq, oxBS-seq, CAB-seq, fCAB-seq or redBS-seq	TAB-seq, oxBS-seq, CAB-seq, fCAB-seq, redBS-seq	
		Reduce representation sequencing with TAB-seq, oxBS-seq, CAB-seq, fCAB-seq or redBS-seq			
	Relative (peak)	DIP-seq, anti-CMS, hMe-Seal,fC-Seal, GLIB, JBP1, EpiMark or Aba-seq			

Plongthonkum et al., *Nature Reviews Genetics* (2014)

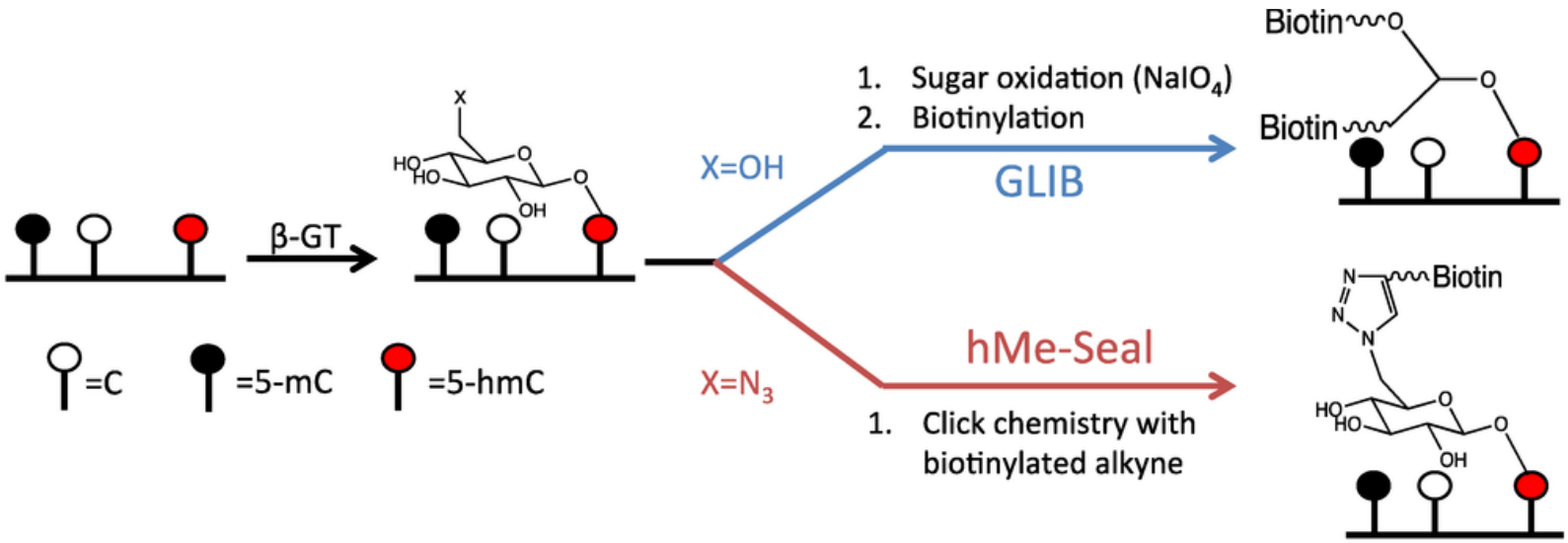
Liquid chromatography and mass spectrometry (LC-MS)



- Pros
  - Well established method
  - A gold standard
  - High sensitivity
- Cons
  - Requires extensive experience
  - Only measures global levels

Figures adapted from Mangerich et al. *PNAS* (2012) and Su et al., *Nature Protocols* (2014) | Branco et al., *Nature Reviews Genetics* (2012)

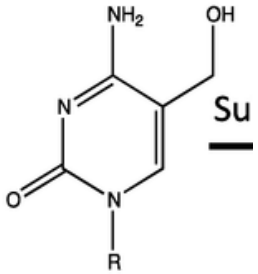
5-hmC detection with GLIB- & hME-Seal-seq



- Pros
  - Covalent chemical labeling and biotin-based purification is sensitive and specific
- Cons
  - Biotinylated adduct can affect PCR so amplification-free library and sequencing preferred

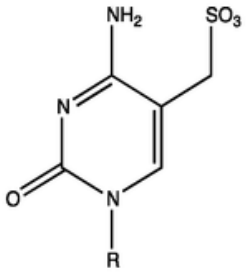
Pastor et al., *Nature* (2011) | Mariani et al., *Cancers* (2013)

5-hmC detection with CMS-seq



5-hmC

Sulfonation



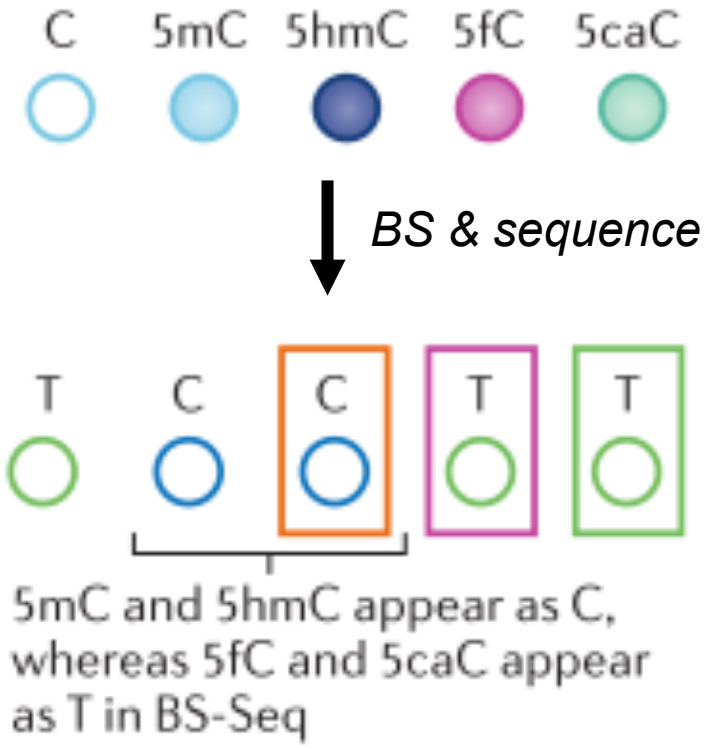
5-methylsulphonate (CMS)

Pull-down with anti-CMS antibody & sequence

- Pros
  - Anti-CMS antibodies are more sensitive and less density dependent than anti-5-hmC antibodies
- Cons
  - Low resolution
  - Limited to specific sequences

Pastor et al., *Nature* (2011) | Figure adapted from Mariani et al., *Cancers* (2013)

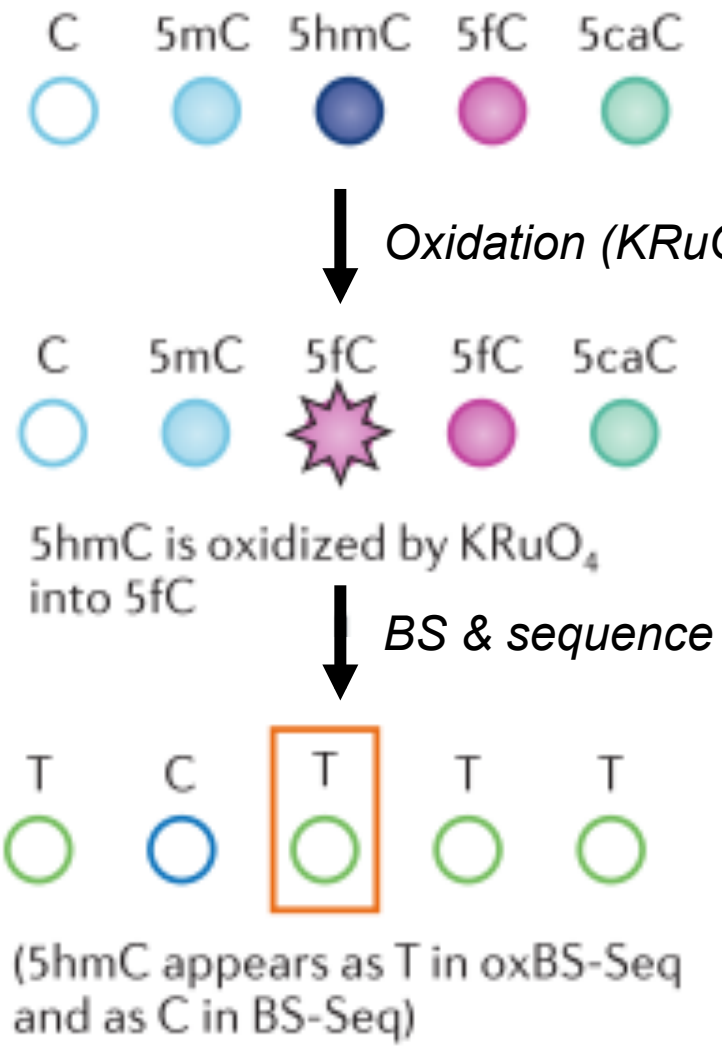
Whole genome bisulfite sequencing (WGBS)



- Pros
  - "Gold standard"
  - Single-base resolution
  - Profiles a large % of the "CpG-ome"
  - Measures absolute levels
- Cons
  - ¥¥¥
  - Does not distinguish between 5-mC and 5-hmC
  - Does not distinguish between C, 5-caC, & 5-fC
  - Bisulfite treatment can damage DNA
  - Dependent on bisulfite conversion
  - Reduced complexity of sequencing reads creates alignment challenges
  - Requires a large amount of DNA

Figures adapted from Plongthonkum et al., *Nature Reviews Genetics* (2014)

5-mC & 5-hmC detection with oxBS-Seq

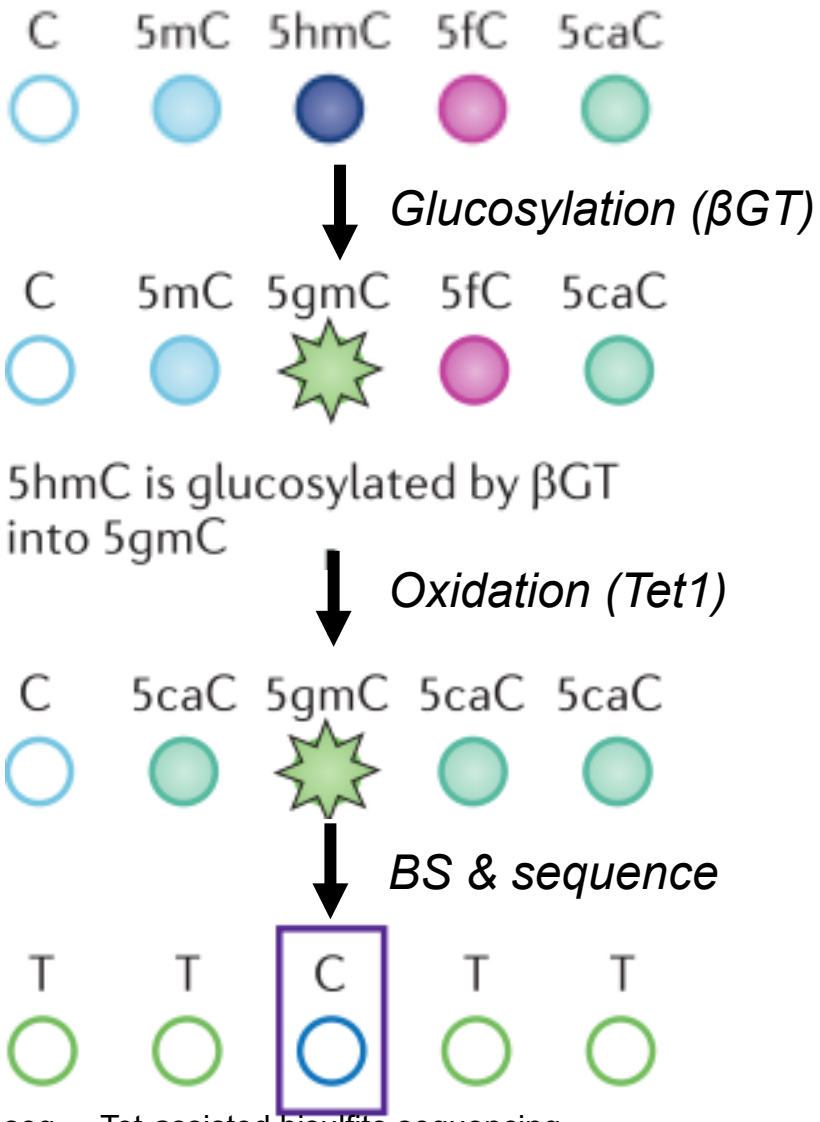


- Pros
  - A gold standard
  - Single-base resolution
  - Measures absolute levels of 5-hmC and 5-mC
- Cons
  - ¥¥¥
  - Requires a large amount of DNA
  - Bisulfite treatment can damage DNA
  - Dependent on bisulfite & oxidative efficiencies
  - Requires a large amount of DNA
  - Reduced complexity of sequencing reads creates alignment challenges

oxBS-seq =Oxidative bisulfite sequencing
KRuO4 = potassium perruthenate
BS = Bisulfite treatment

Figures adapted from Plongthonkum et al., *Nature Reviews Genetics* (2014) | Booth et al., *Science* (2012)

5-mC & 5-hmC detection with TAB-Seq

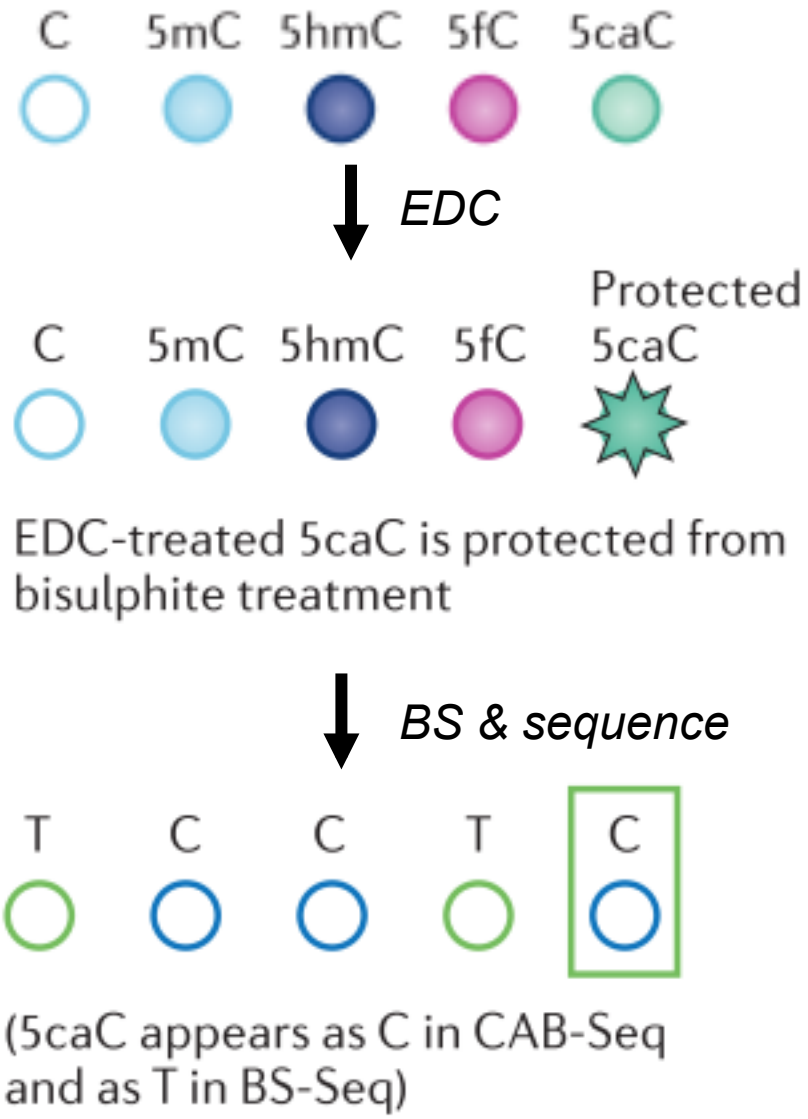


- Pros
  - Single-base resolution
  - Enzymatic treatments are more mild than oxBS-seq
  - Measures absolute levels of 5-hmC and 5-mC
- Cons
  - ¥¥¥
  - Bisulfite treatment can damage DNA
  - Dependent on bisulfite and glycosylation efficiencies
  - Requires a large amount of DNA
  - Reduced complexity of sequencing reads creates alignment challenges

TAB-seq = Tet-assisted bisulfite sequencing
BS = Bisulfite treatment

Figures adapted from Plongthonkum et al., *Nature Reviews Genetics* (2014) | Yu et al., Cell (2012)

5-caC detection with CAB-Seq

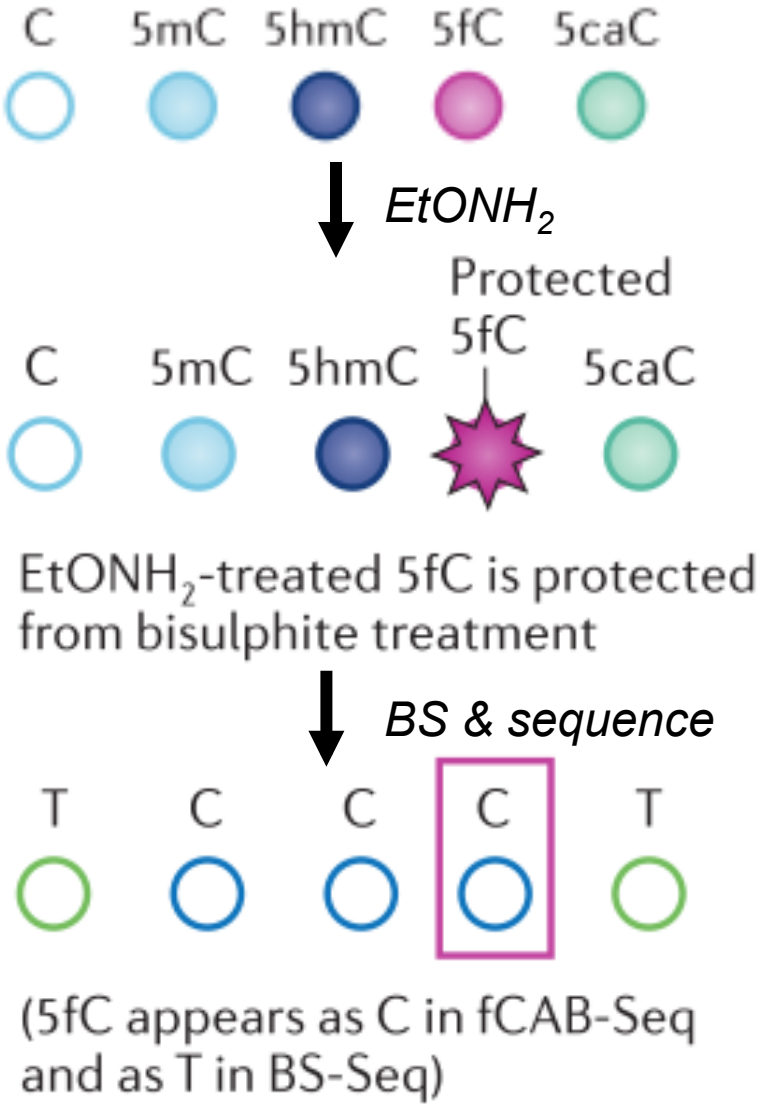


- Pros
  - Single-base resolution
- Cons
  - ¥¥¥
  - Bisulfite treatment can damage DNA
  - Dependent on bisulfite conversion
  - Requires a large amount of DNA
  - Reduced complexity of sequencing reads creates alignment challenges

CAB-seq = Chemical modification-assisted bisulfite sequencing
BS = Bisulfite treatment

Figures adapted from Plongthonkum et al., *Nature Reviews Genetics* (2014) | Lu et al., *Journal American Chemical Society* (2013)

5-fC detection with fCAB-Seq

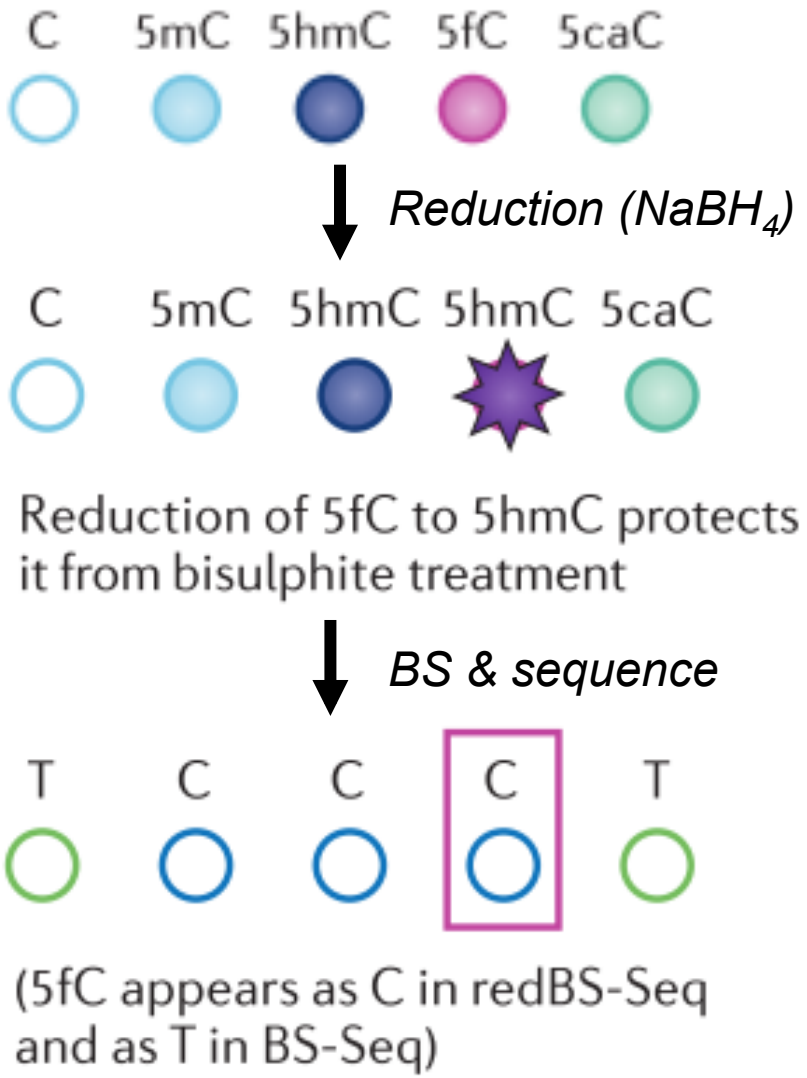


- Pros
  - Single-base resolution
- Cons
  - ¥¥¥
  - Bisulfite treatment can damage DNA
  - Dependent on bisulfite conversion
  - Requires a large amount of DNA
  - Reduced complexity of sequencing reads creates alignment challenges

fCAB-seq = 5fC Chemical modification-assisted bisulfite sequencing
$EtONH_2$ = O-ethylhydroxylamine
BS = Bisulfite treatment

Figures adapted from Plongthonkum et al., *Nature Reviews Genetics* (2014) | Song et al., *Cell* (2013)

5-fC detection with redBS-Seq



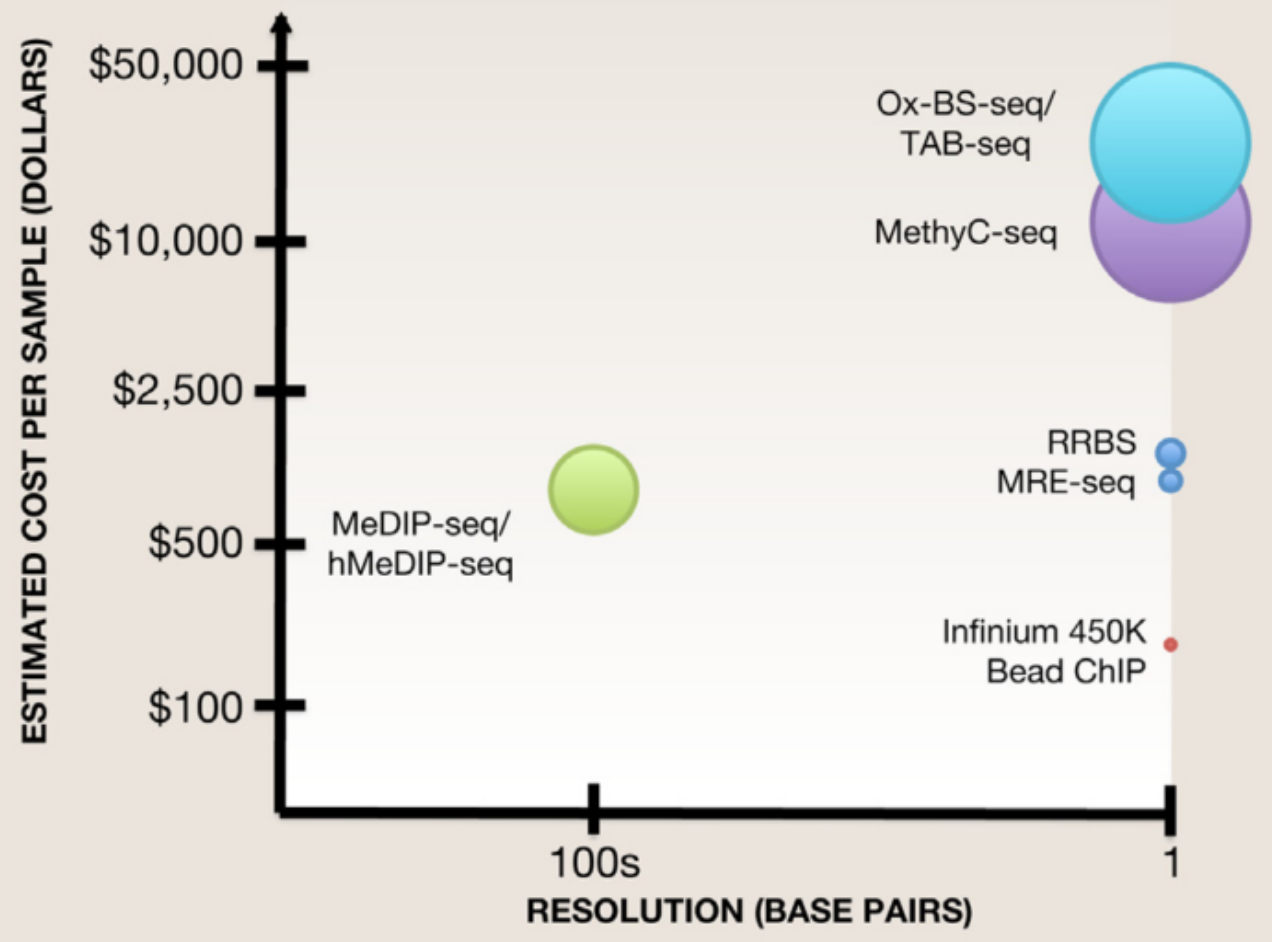
- Pros
  - Single-base resolution
- Cons
  - ¥¥¥
  - Bisulfite treatment can damage DNA
  - Dependent on bisulfite conversion
  - Requires a large amount of DNA
  - Reduced complexity of sequencing reads creates alignment challenges

redBS-seq = Reduced bisulfite sequencing
$NaBH_4$ = sodium borohydride
BS = Bisulfite treatment

Figures adapted from Plongthonkum et al., *Nature Reviews Genetics* (2014) | Booth et al., *Nature Chemistry* (2014)

Comparison of DNA 5-mC and 5-hmC mapping technologies

Major considerations when selecting a profiling method are: 1) cost, 2) resolution, & 3) genome coverage



Rivera & Ren *Cell* (2013)

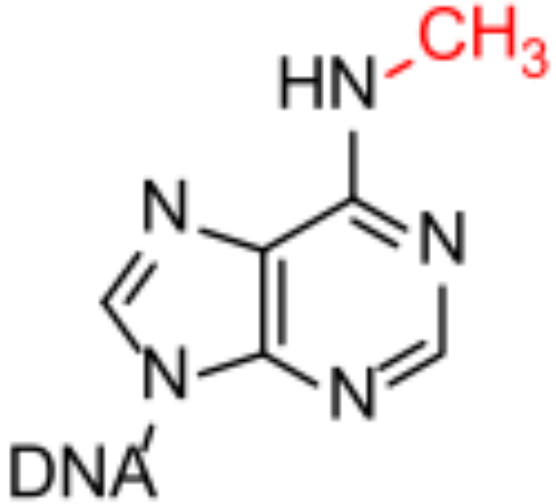
Sources of bias in 5-mC and 5-hmC mapping technologies

	Method	Amplification	Effect of copy number variation on quantification	Incomplete treatment (chemical or enzyme digestion)	Background signals	Batch to batch variation	Cross hybridization	Sensitivity to sequence context
5mC assays	BS-seq	Low*	None	Low	None	None	None	Medium‡ (RRBS only)
	Low input BS-seq	High§	None	Low	None	None	None	Medium (RRBS only)
	BS-based arrays	Low	None	Low	None	Medium	Medium	None
	Non-BS-seq	Low	Medium	Low	High	None	None	Medium (MeDIP/MBD)
	Low input non-BS-seq	High	Medium	None	High	None	None	Medium (MeDIP/MBD)
	Non-BS- arrays	Low	Medium	None	High	Medium	Medium	None
5mC oxidation derivatives assays	BS-based seq	Low	None	Low	None	None	None	None
	Non-BS-based seq	Low	Medium	Low	High	None	None	Medium
	RE-based 5hmC seq	Low	None	Low	None	None	None	Medium
All	Third generation sequencing	Low	None	None	None	None	None	None

*Low, can be ignored in most cases. ‡Medium, should be considered when performing analyses. § High, must be considered when performing analyses.

Plongthonkum et al., *Nature Reviews Genetics* (2014)

N^6-methyladenine



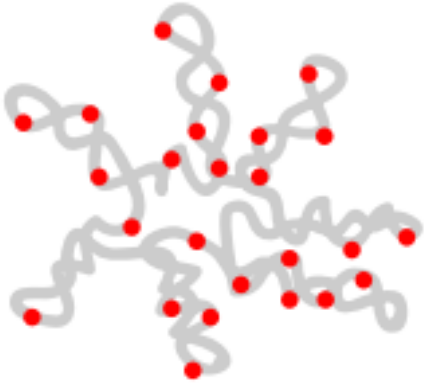
N^6-methyladenine

Bacteria

Foreign DNA

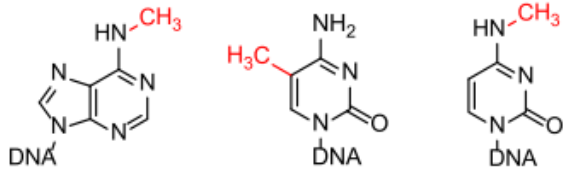
Restriction and modification system; Package of phage DNA

Domestic DNA



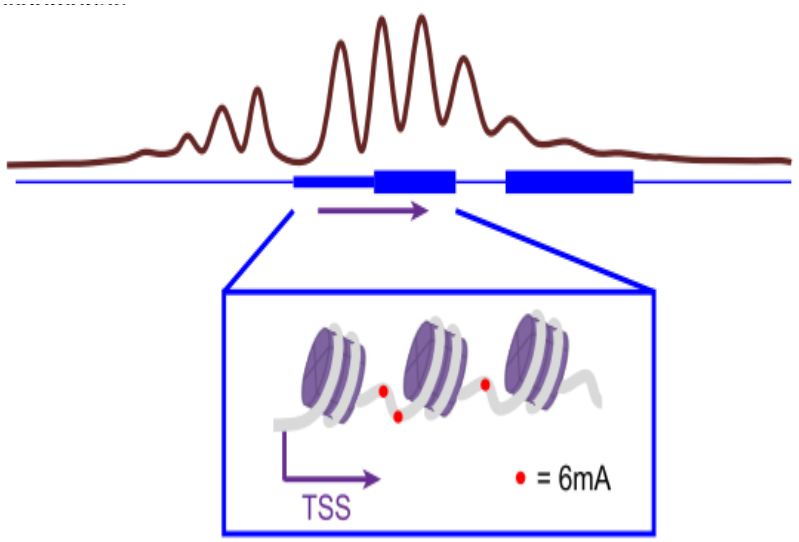
Nucleoid structure; Replication; Transposition; Mismatch repair

• = 6mA or 5mC or 4mC



6mA 5mC 4mC

C. reinhardtii

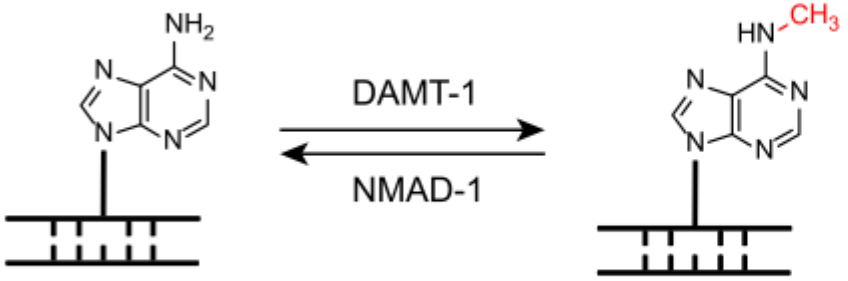


- 6-mA depleted at TSS
- 6-mA enriched at linker regions between nucleosomes
- 6-mA marks actively transcribed genes

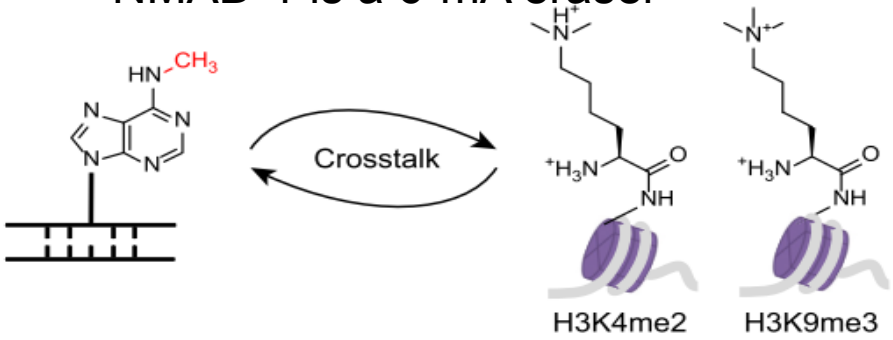
Figures adapted from Chen et al., *Cell Chemical Biology* (2016)

N^6-methyladenine (cont.)

C. elegans

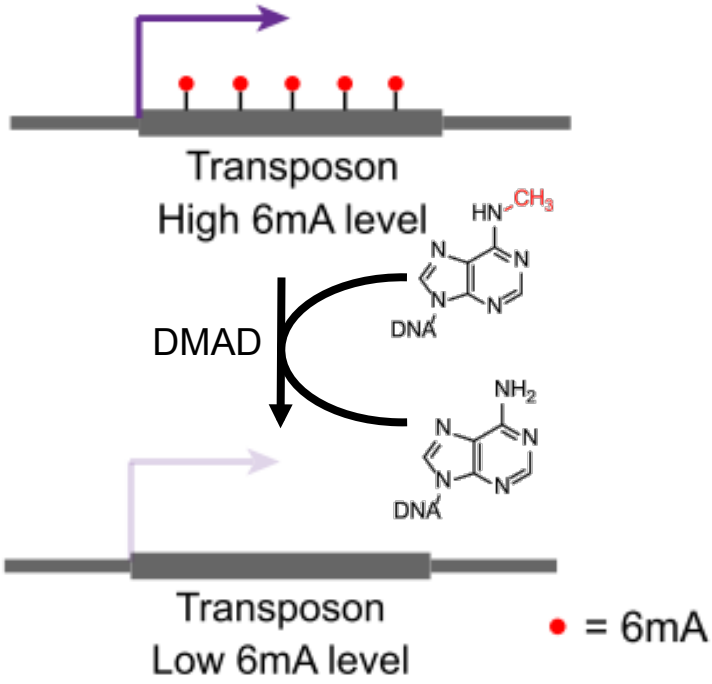


- DAMT-1 is a 6-mA writer
- NMAD-1 is a 6-mA eraser



- 6-mA may play a role regulating gene expression
- *C. elegans* lacks 5-mC

D. melanogaster



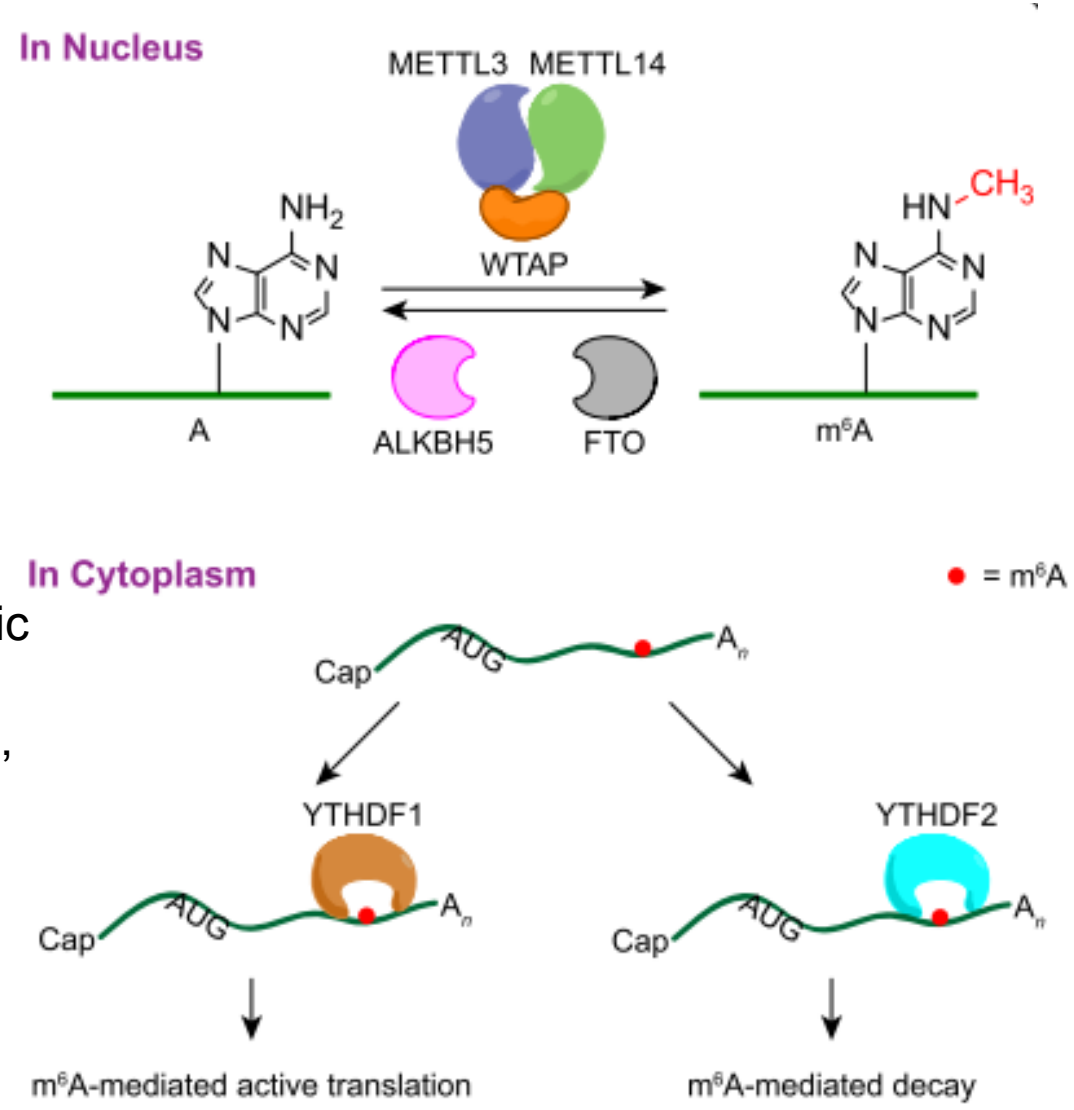
- DMAD, a Tet homolog, may be an eraser of 6mA
- Dynamic 6-mA demethylation associated with transposon expression
- 6-mA plays a critical roles in development

Figures adapted from Chen et al., *Cell Chemical Biology* (2016)

RNA modifications

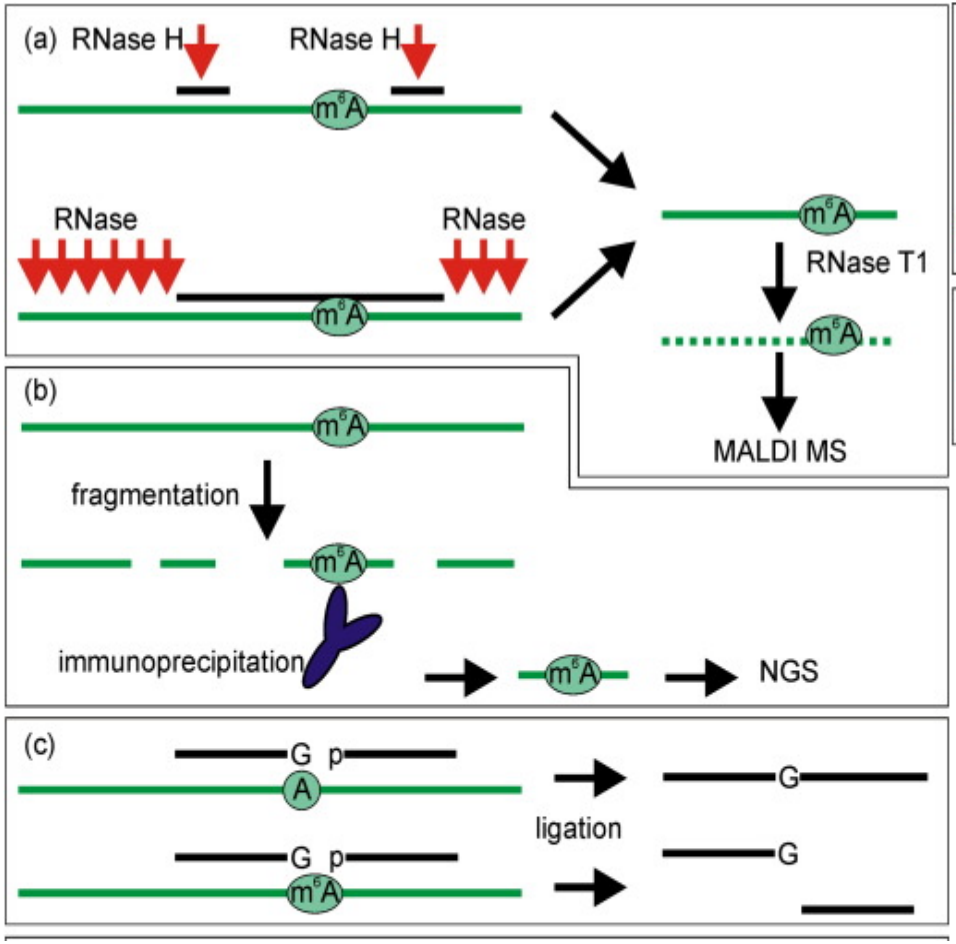
N^6-methyladenosine (m^6A) in mRNA

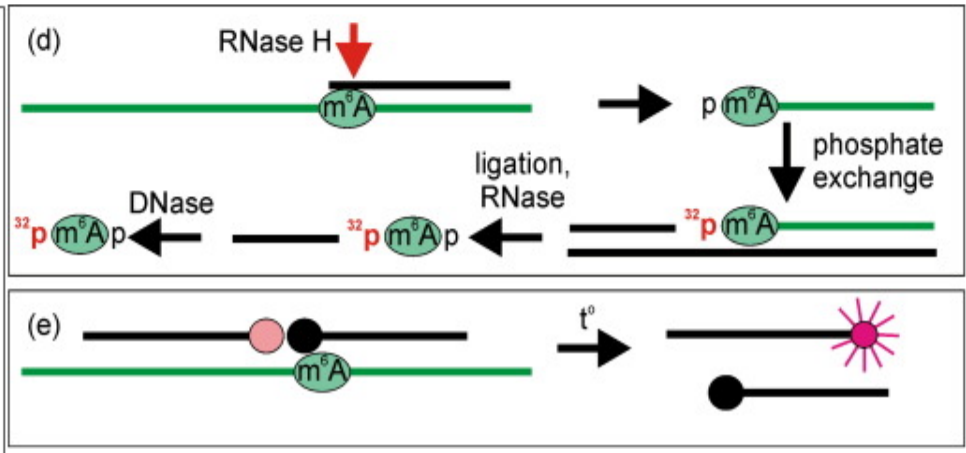
- Discovered in eukaryotic and viral mRNAs in 1974
- Most prevalent modification in mRNAs and lncRNAs in higher eukaryotes
- m^6A plays a critical role in development & disease
  - Deficiencies in methyltransferases is embryonic lethal in mouse and leads to developmental defects in yeast, flies, zebrafish & plants
  - m^6A dysregulation associated with obesity, cancer, and other diseases in humans



Figures adapted from Chen et al., *Cell Chemical Biology* (2016) | Sergiev et al., *Journal of Molecular Biology* (2015)

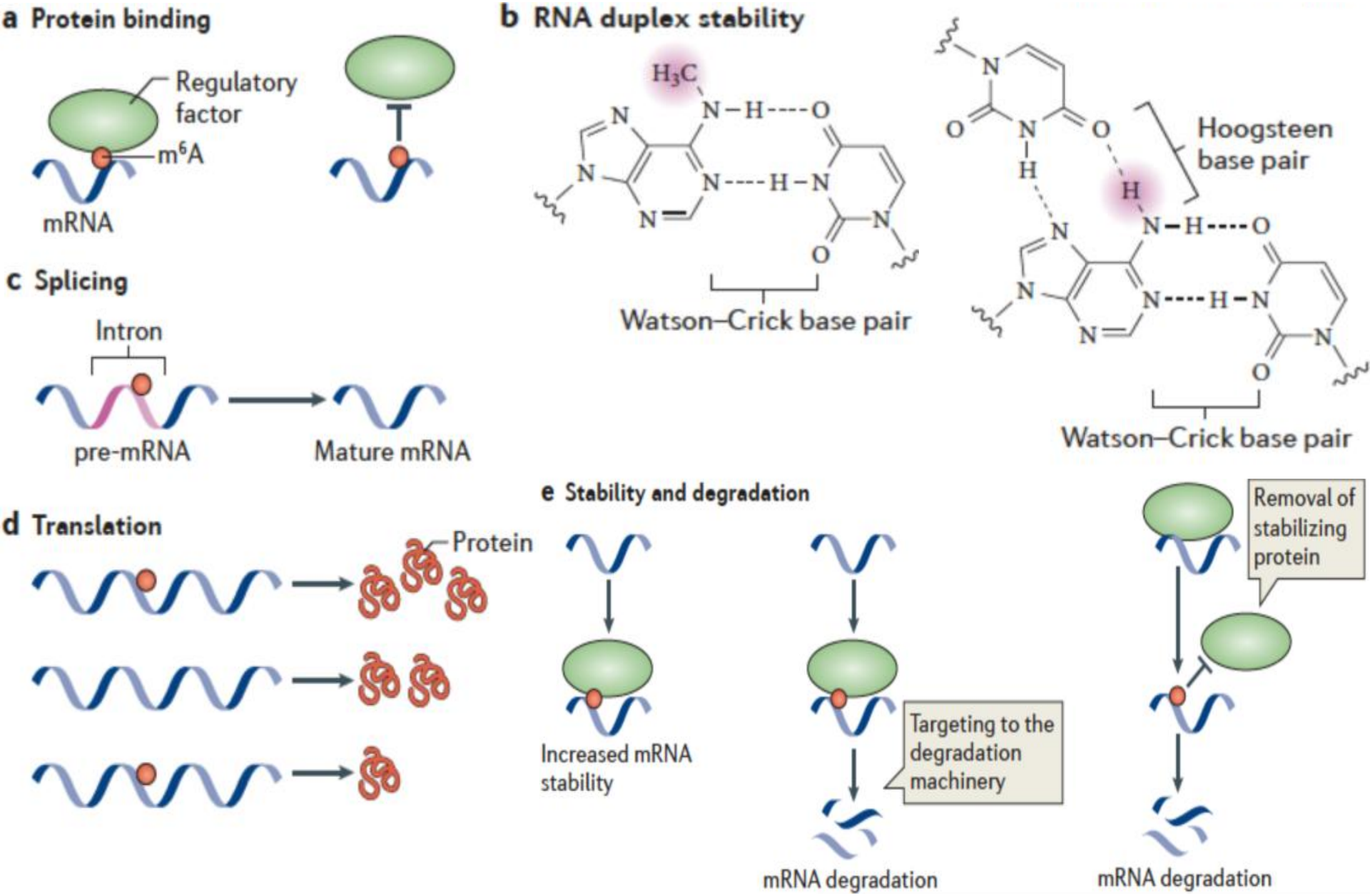
Methods for detecting m^{6}A in RNA





Figures adapted from Sergiev et al., *Journal of Molecular Biology* (2015)

Mechanisms and functions of m^{6}A



Figures adapted from Meyer & Jaffrey *Nature Reviews Molecular Cell Biology* (2015)

Other RNA modifications: Pseudouridine (Ψ)

- Most abundant of over 100 post-transcription modifications
  - Several hundred to thousands of Ψ sites genome-wide
- Found on rRNAs, snRNAs, mRNAs, & tRNAs
- Functions
  - rRNA: required for ribosome biogenesis and translation fidelity
  - snRNA: alters the structural stability snRNAs and plays critical roles in snRNP assembly, spliceosome formation, and splicing
  - mRNA: may play a role in response to cellular stress and differentiation

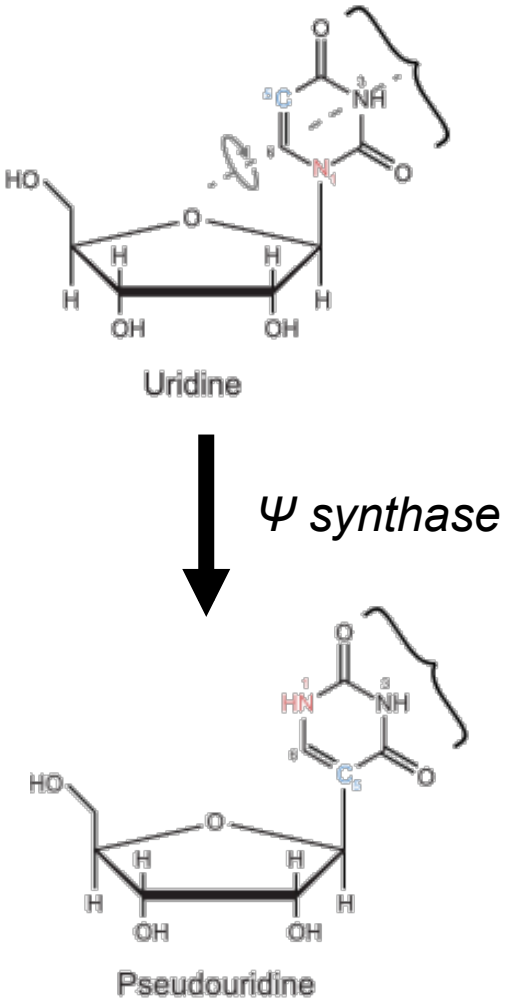
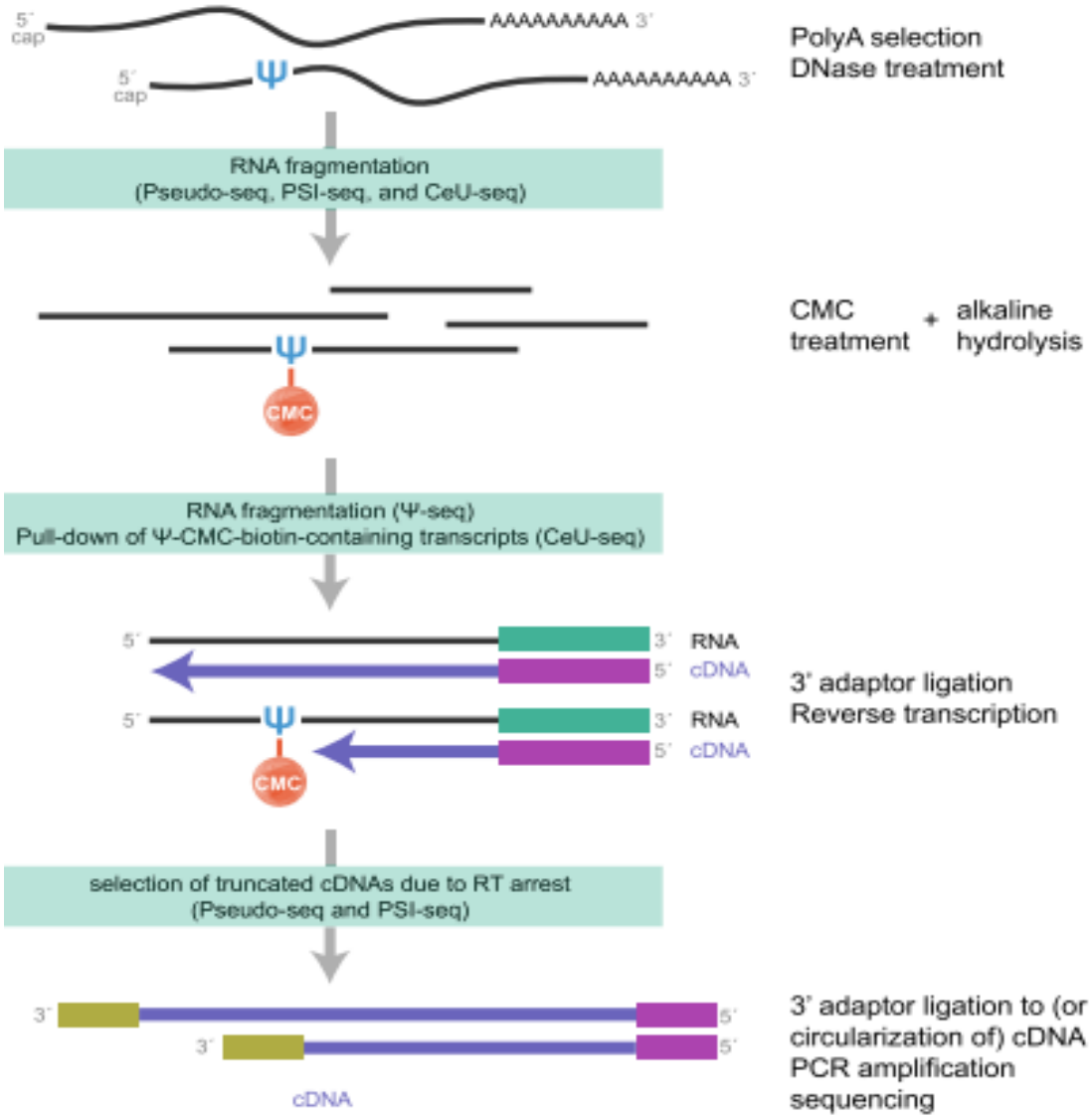


Figure adapted from Zaringhalam & Papavasiliou *Methods* (2016) | Chen et al., *Cell Chemical Biology* (2016)

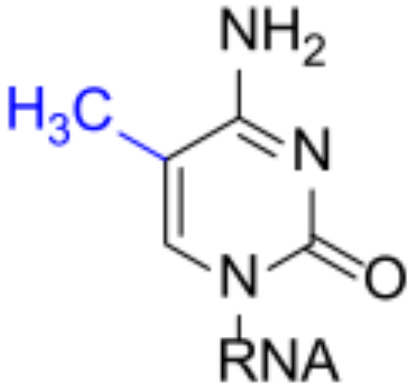
Methods for detecting Ψ

- Methods to map Ψ modifications recently invented
  - Pseudo-seq (Carlie et al., 2014)
  - Ψ-seq (Schwartz et al., 2014)
  - PSI-seq (Lovejoy et al., 2013)
  - CeU-seq (Li et al., 2015)



Chen et al., *Cell Chemical Biology* (2016) | Zaringhalam & Papavasiliou *Methods* (2016)

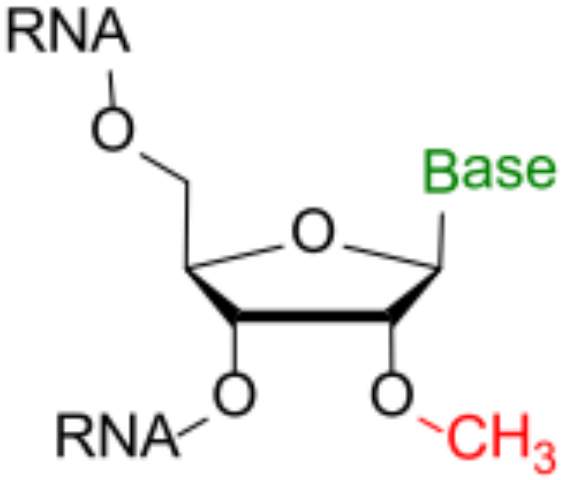
Other RNA modifications: 5-methylcytidine



- Found on mRNAs, tRNAs, & rRNAs
- Functions
  - tRNAs: increases the stability of tRNAs
- Plays a critical role in development & disease
  - Deficiencies RNA cytosine methyltransferases (RCMT) associated with embryonic lethality in mouse
  - Dysregulation of RCMTs associated with cancer infertility, & mental retardation in humans

Chen et al., *Cell Chemical Biology* (2016)

Other RNA modifications: 2'-O-methylation



- Found on miRNAs, snRNAs, tRNAs, & rRNAs
- Functions
  - miRNAs: increases the stability of miRNAs
  - snRNAs: alters the structural stability snRNAs & plays critical roles in snRNP assembly, spliceosome formation, and splicing

Chen et al., *Cell Chemical Biology* (2016)

tRNA modifications

- Modifications in tRNAs affect tRNA stability, protein translation, and play key roles in stress response and immune recognition
- In *S. cerevisiae*, there are ~25 different tRNA modifications
  - Dysregulation leads to lethality, poor growth, & temperature sensitivity
- Methods to map tRNA modifications recently invented
  - ARM-seq (Cozen et al., 2015)
  - DM-tRNA-seq (Zheng et al., 2015)

Chen et al., *Cell Chemical Biology* (2016)

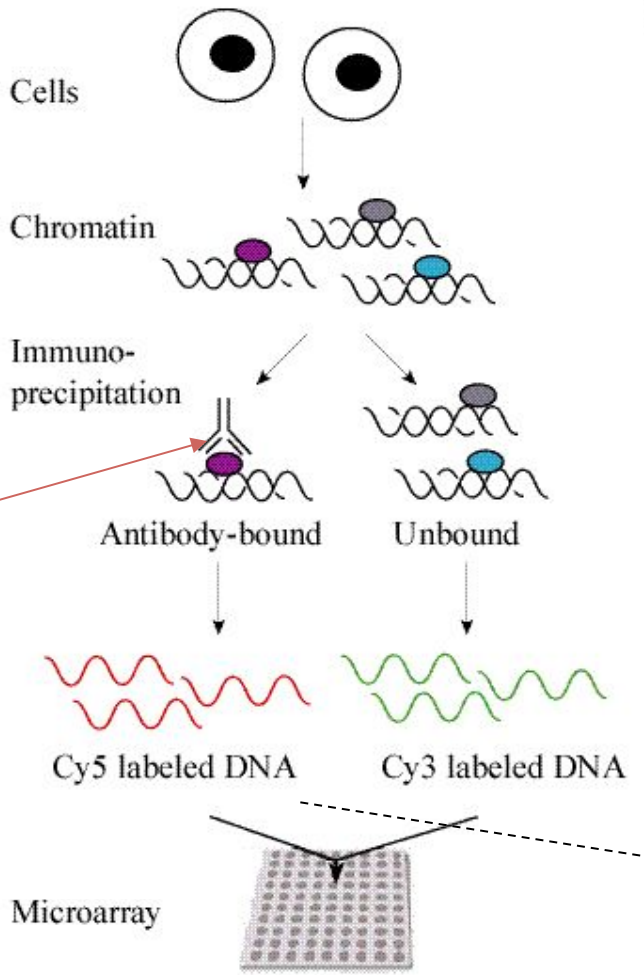
Summary

- Nucleic acid modifications is an exciting and evolving area of research
- Nucleic acid modifications expand the complexity of gene regulation
- DNA 5-mC is found in almost all kingdoms of life
- Dysregulation of DNA 5-mC is associated with cancer and other human diseases
- DNA 6-mA is a recently discovered epigenetic mark and may play a complementary role to 5-mC
- RNA modifications are diverse and have many functions
- Many technologies have been developed for mapping DNA and RNA modifications

Technologies for Interrogating Chromatin States

Histone Modifications

ChIP-chip

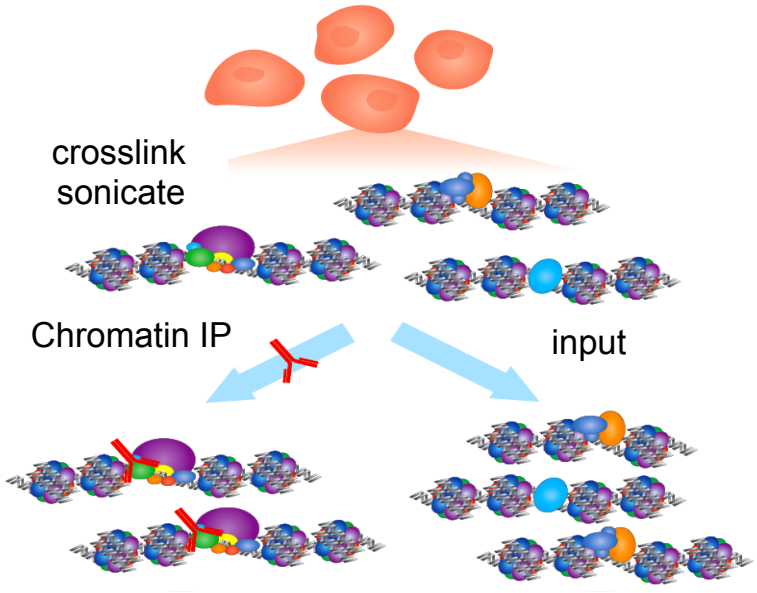


Antibody specific to one type of histone modification

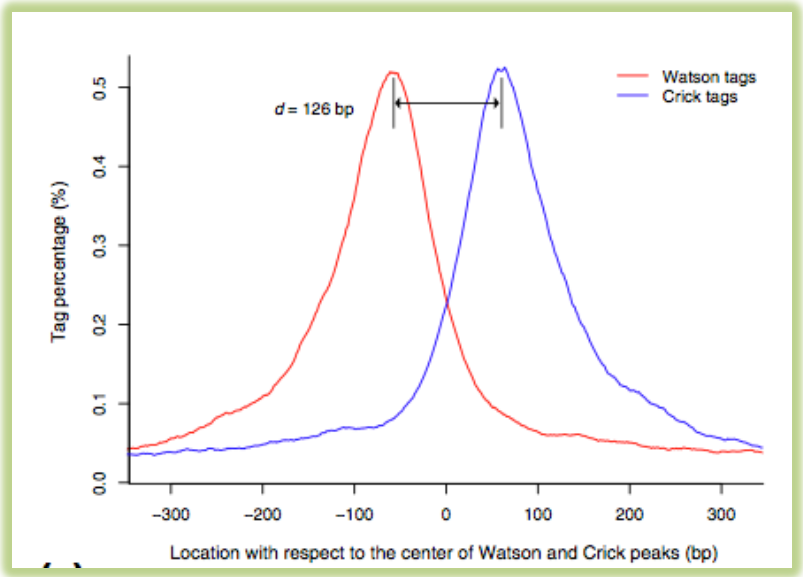
ChIP-seq

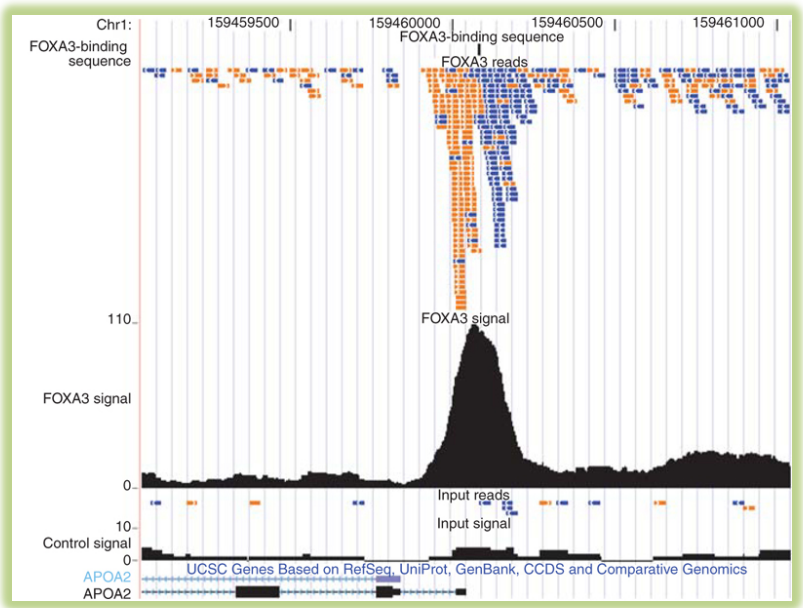
Deep sequencing

ChIP-seq

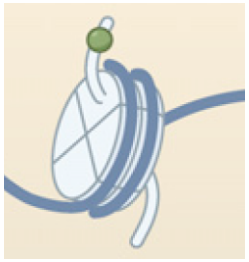


Zhang et al. Genome Biol 2008



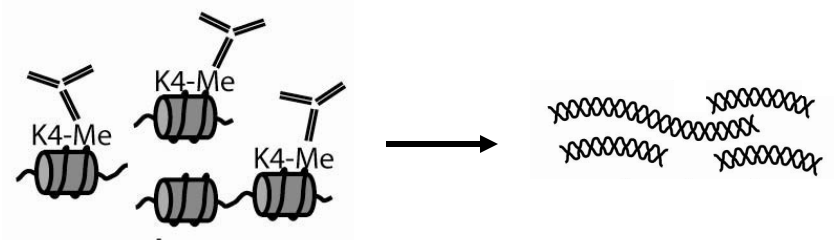


Chromatin-IP Sequencing

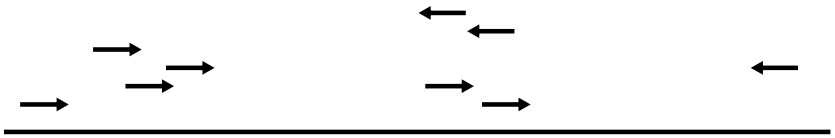


K4me1
K4me2
K4me3 } "active"

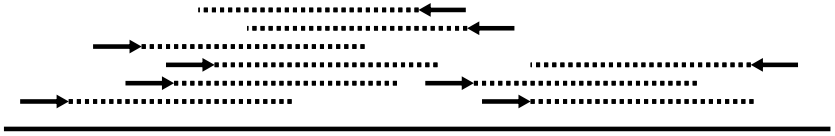
K27me3 "repressive"



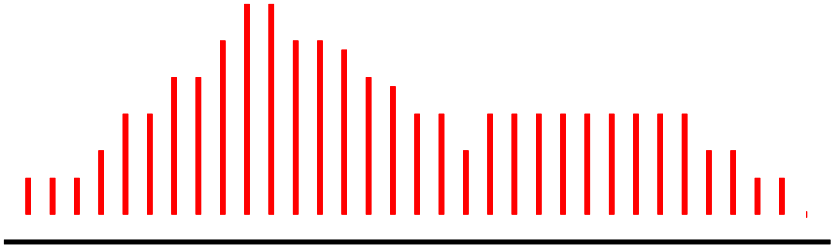
A. Align reads

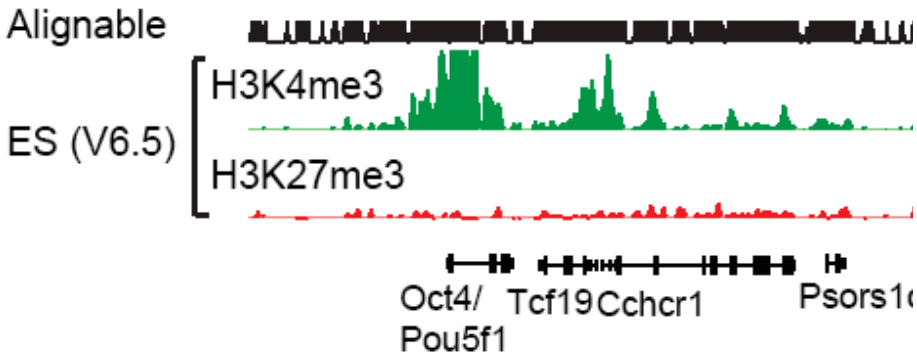


B. Infer positions of ChIP fragments



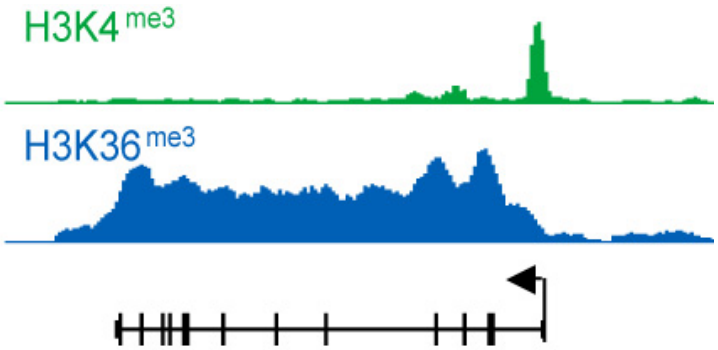
C. Count fragments at each genomic position





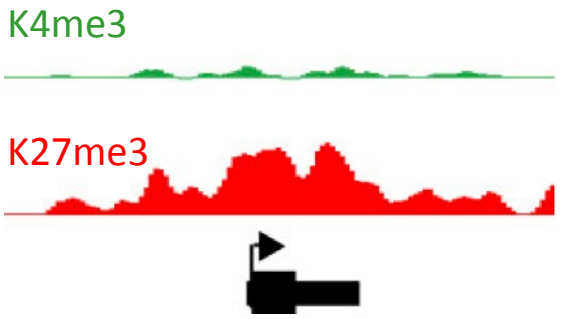
Histone methylation and transcriptional state

Transcribed gene



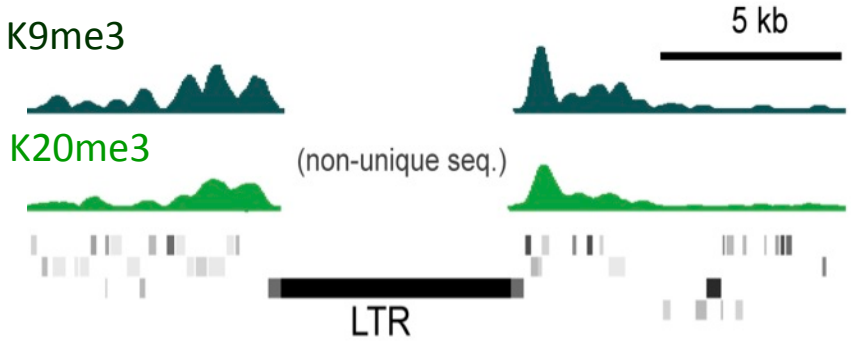
FoxP1

Silent developmental gene

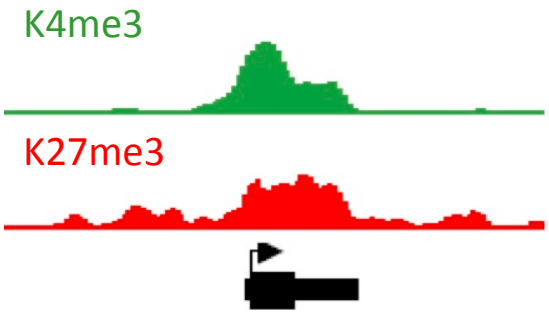


Olig1

Constitutive heterochromatin

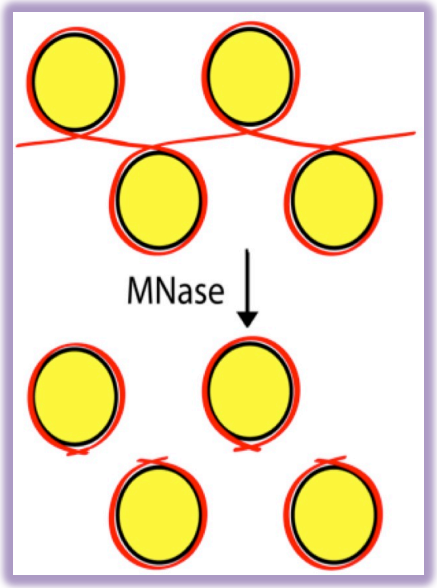


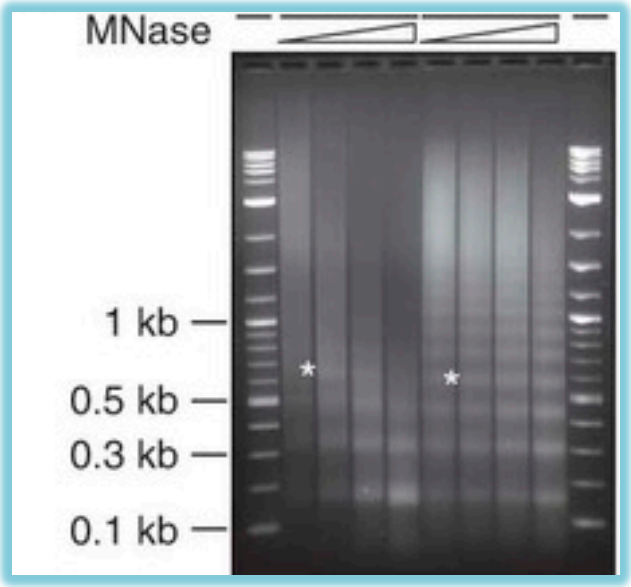
'Poised' developmental gene

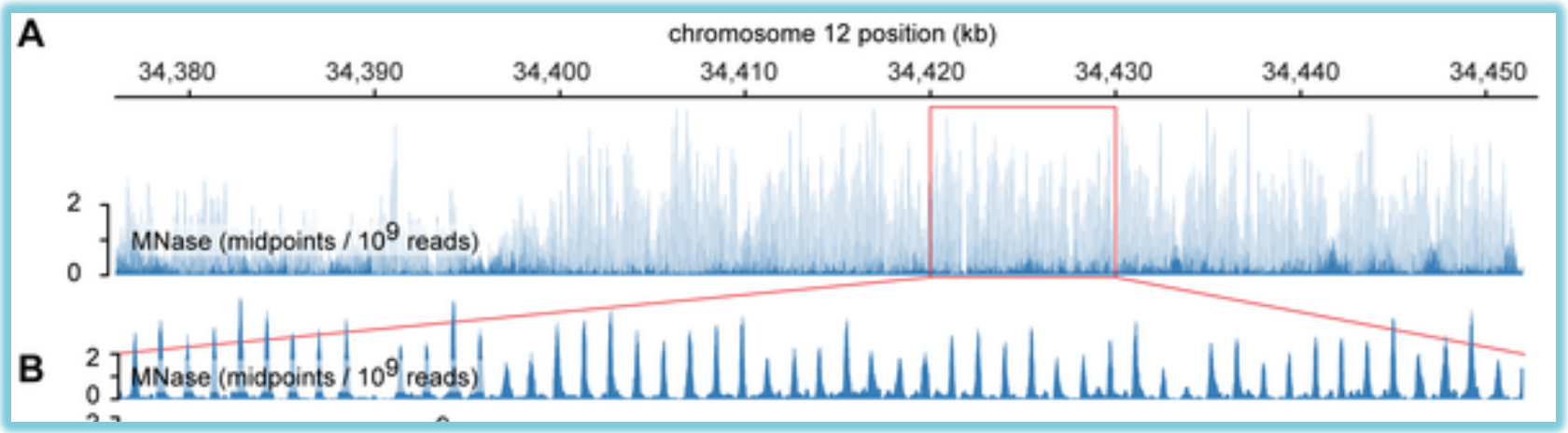


Olig1

Mapping nucleosome positions

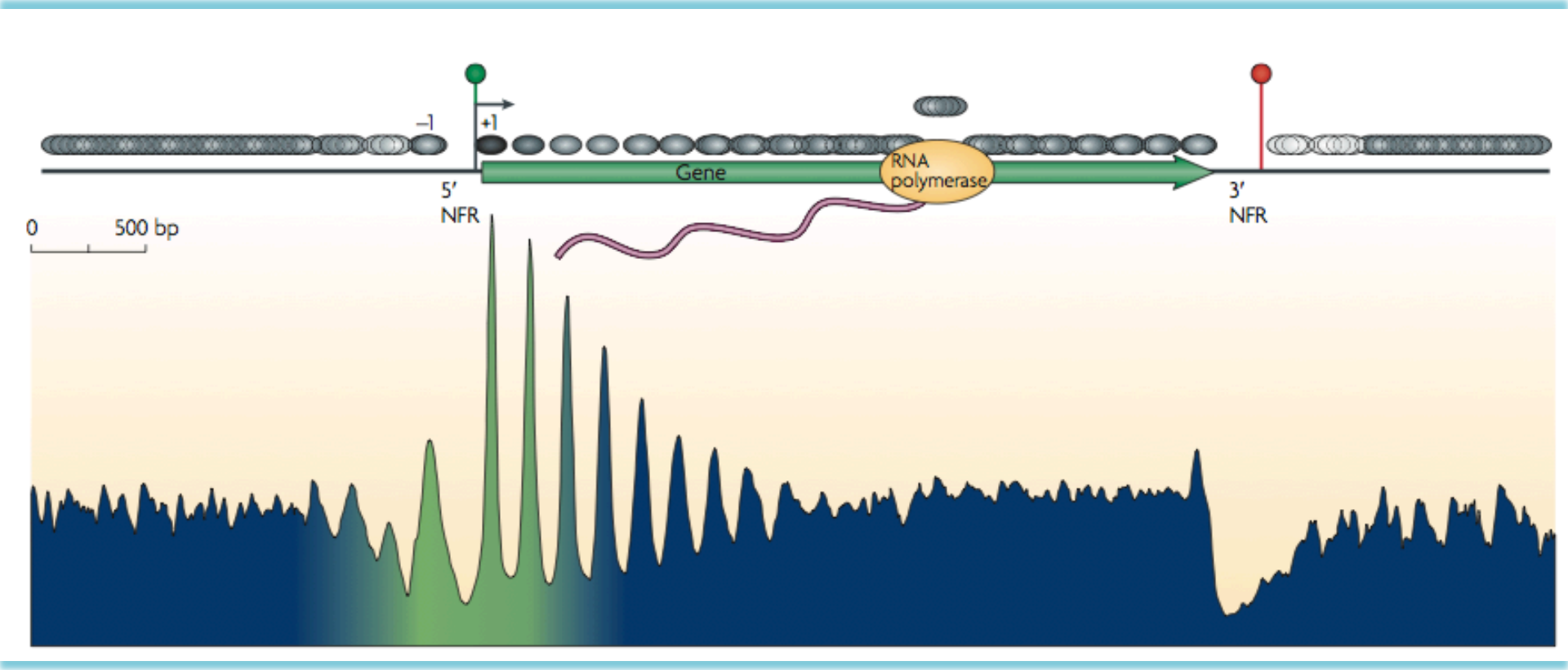






Gaffney DJ, McVicker G, Pai AA, Fondufe-Mittendorf YN, et al. (2012) Controls of Nucleosome Positioning in the Human Genome. PLoS Genet 8(11): e1003036. doi:10.1371/journal.pgen.1003036
http://www.plosgenetics.org/article/info:doi/10.1371/journal.pgen.1003036

Genomic distribution of nucleosomes



- The presence of NFRs demonstrated that open promoter states are stable and common, even at genes that are transcribed so infrequently

Cizhong Jiang, B Franklin Pugh, Nature Review Genetics, 2009 vol. 10 (3) pp. 161-72

Nucleosome Positioning from Histone ChIP-seq

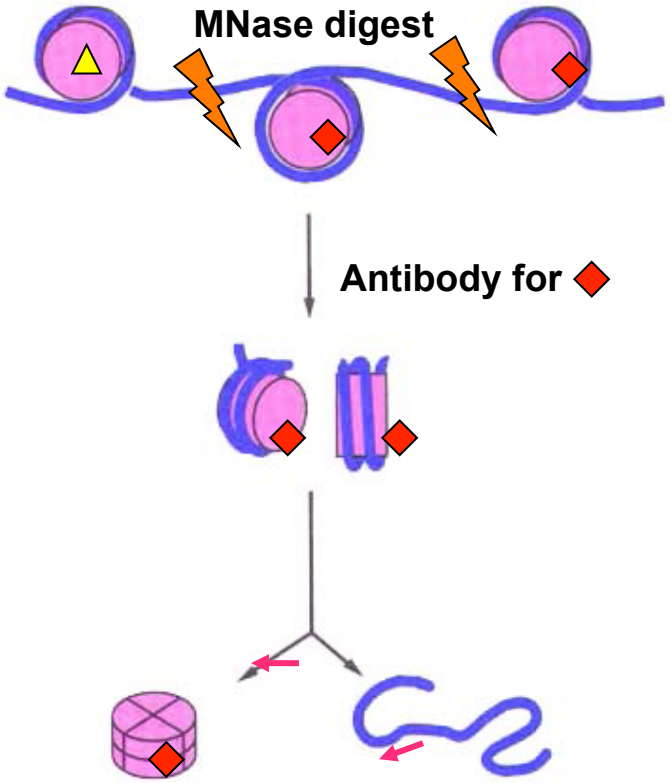
- Barski et al, Cell 2007
  - Nucleosome resolution ChIP-seq of 21 histone marks in $CD4^+$ T-cells
  - Total 185.7 M

    25 nt tags sequenced
  - Analysis not at

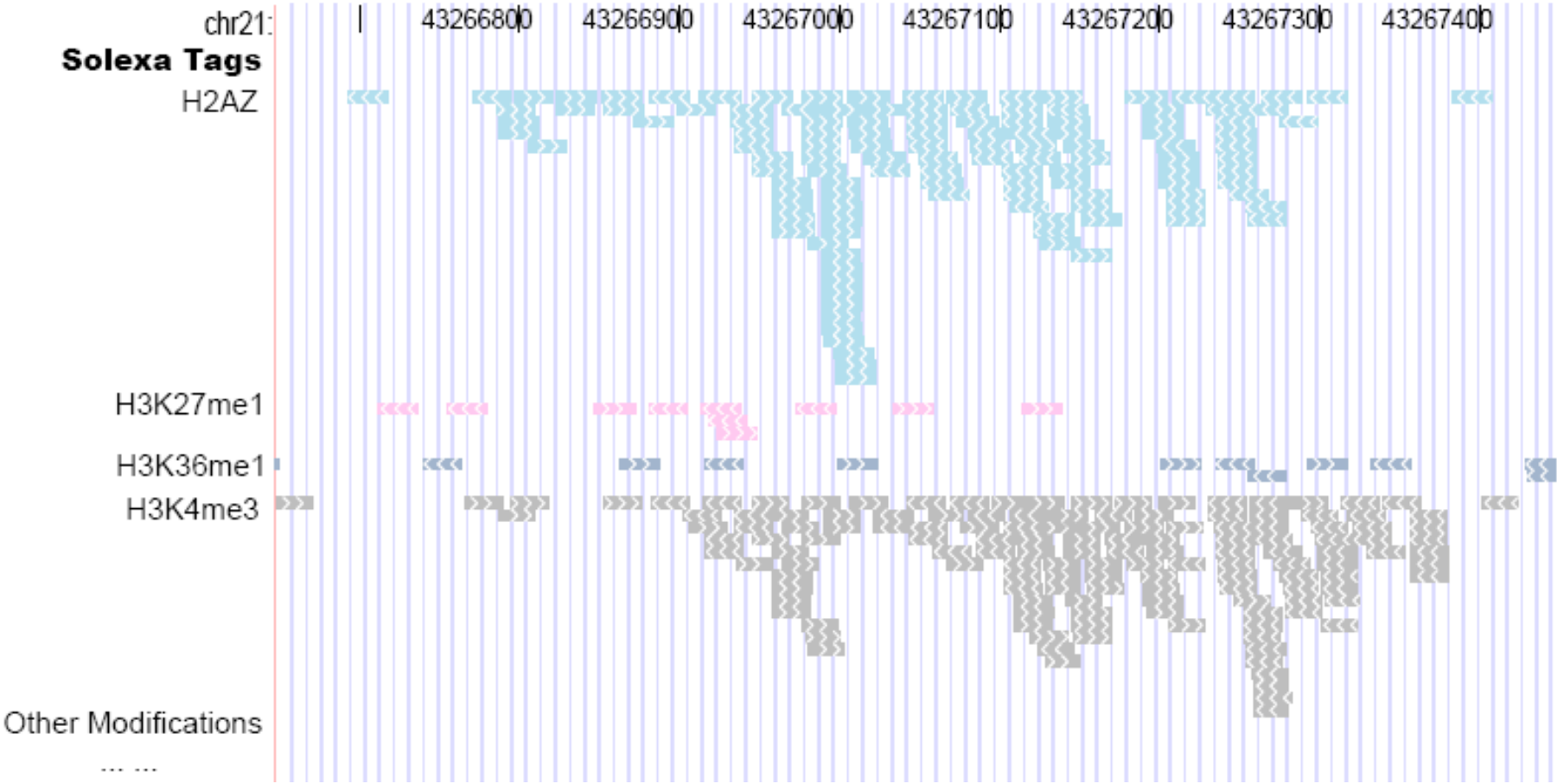
    nucleosome resolution

    to map nucleosomes

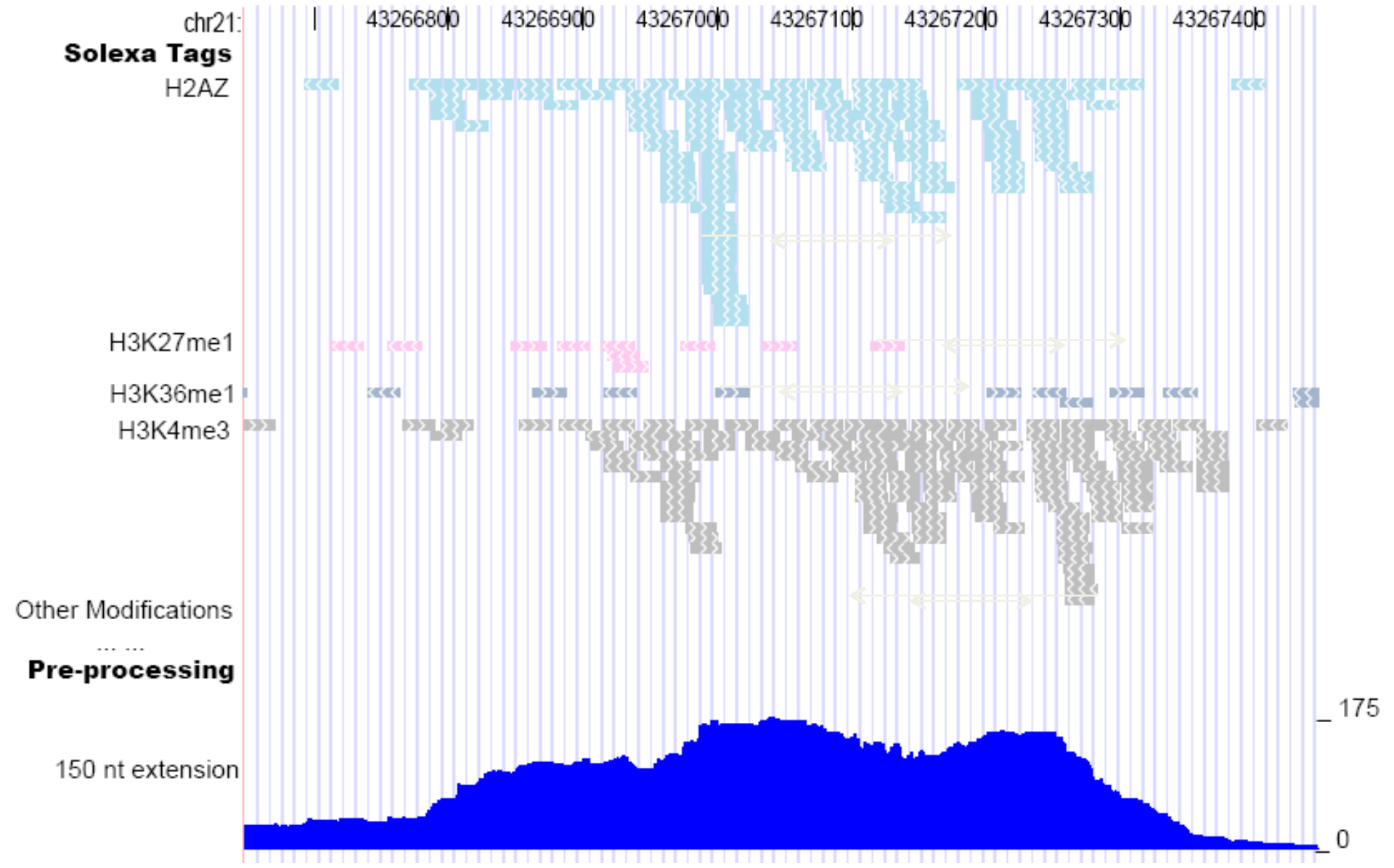
    at specific regions



Combine Tags From All ChIP-Seq



Extend Tags 3' to 150 nt
Check Tag Count Across Genome



Take the middle 75 nt