Breakout: DNA Methylation

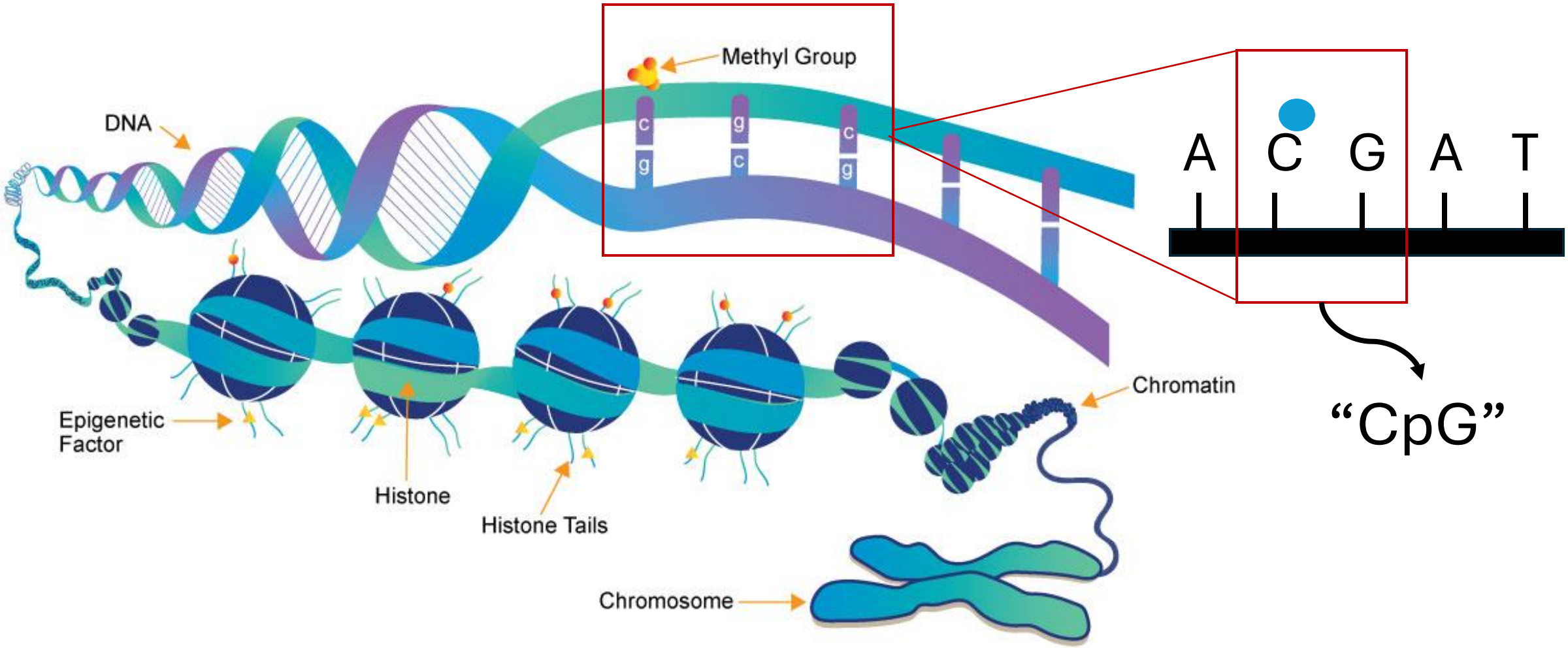
2024-10-17

Demo code and slides

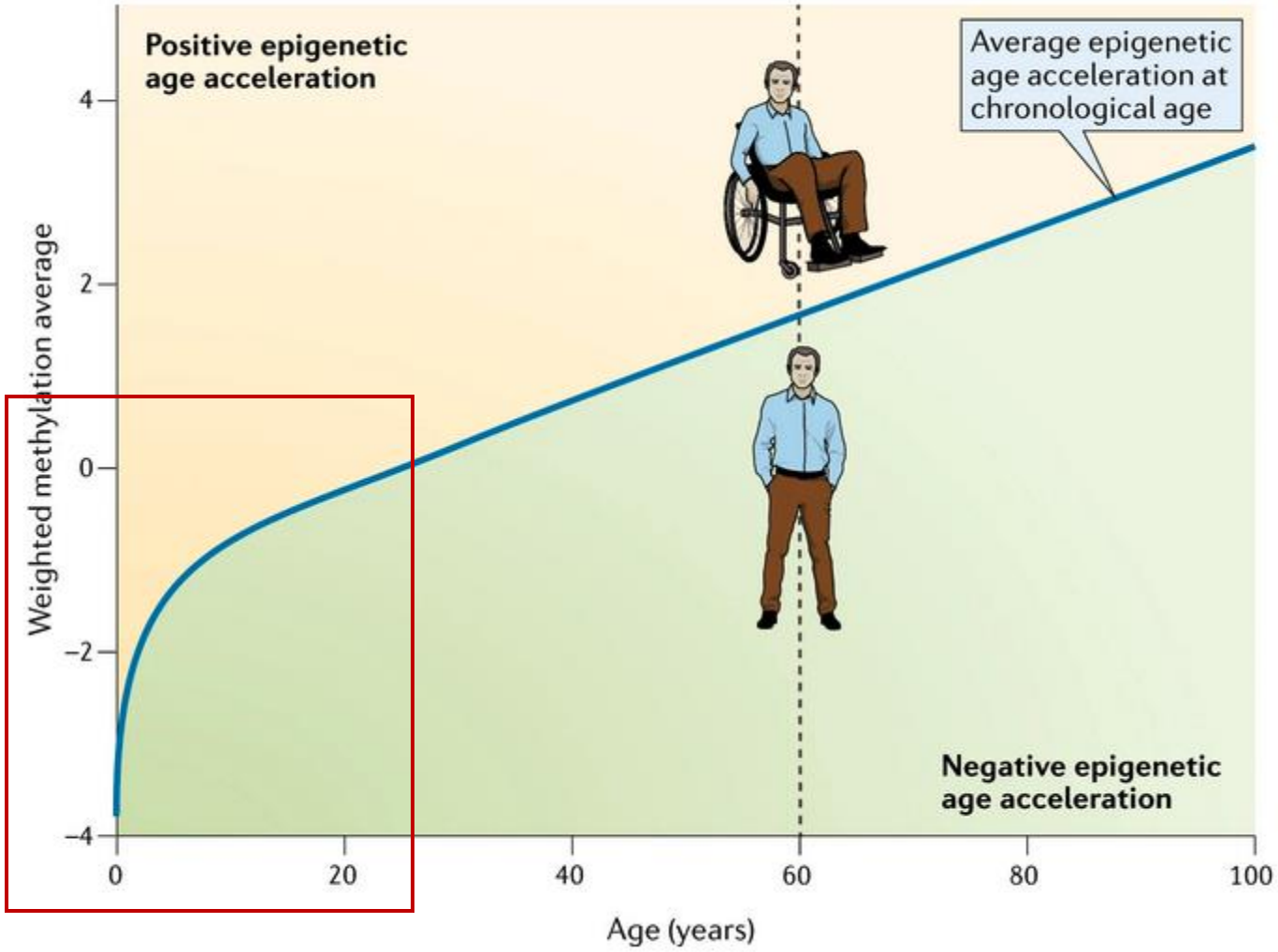
https://github.com/kelseykeith/dbhi_days_meth_breakout

Background

DNA Methylation

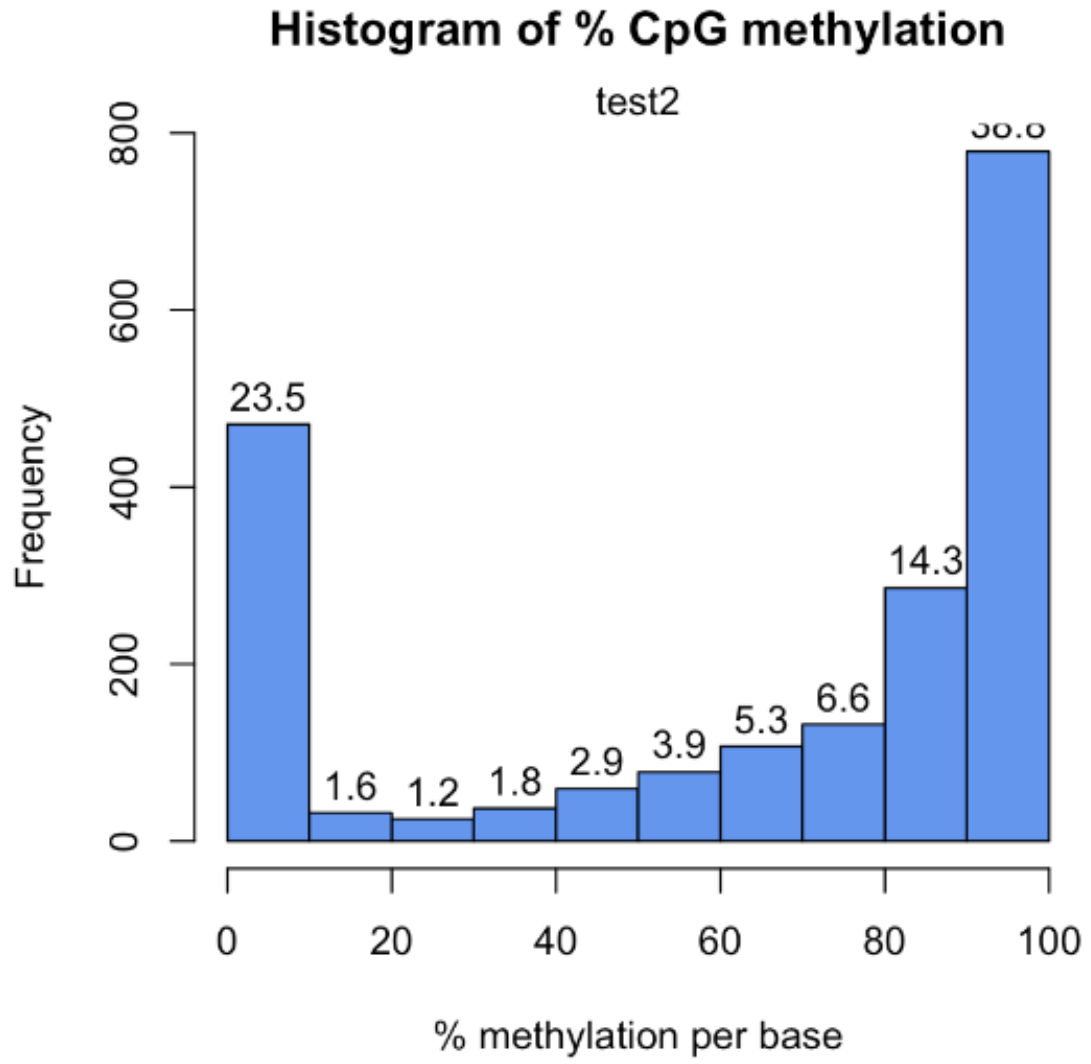


Aging and DNA Methylation

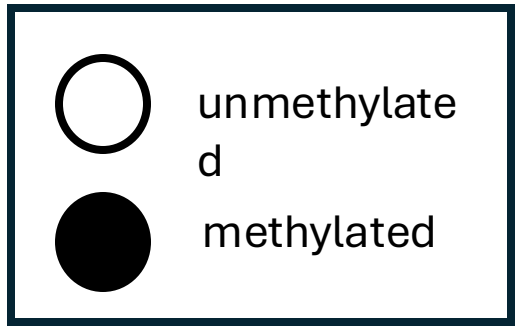


Horvath 2018

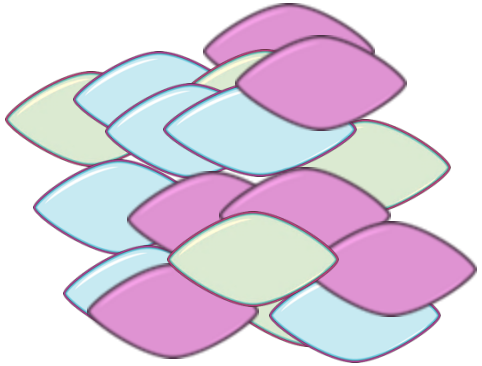
Methylation is Reported as a Percentage



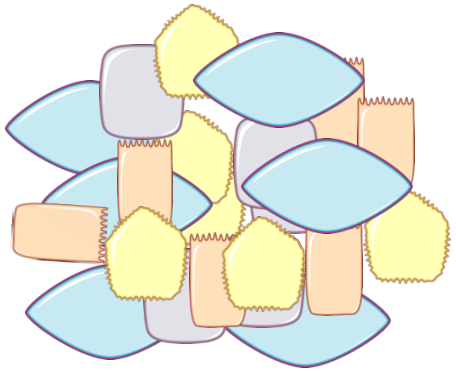
Methylation is binary



Some variation b/t cells of the same type



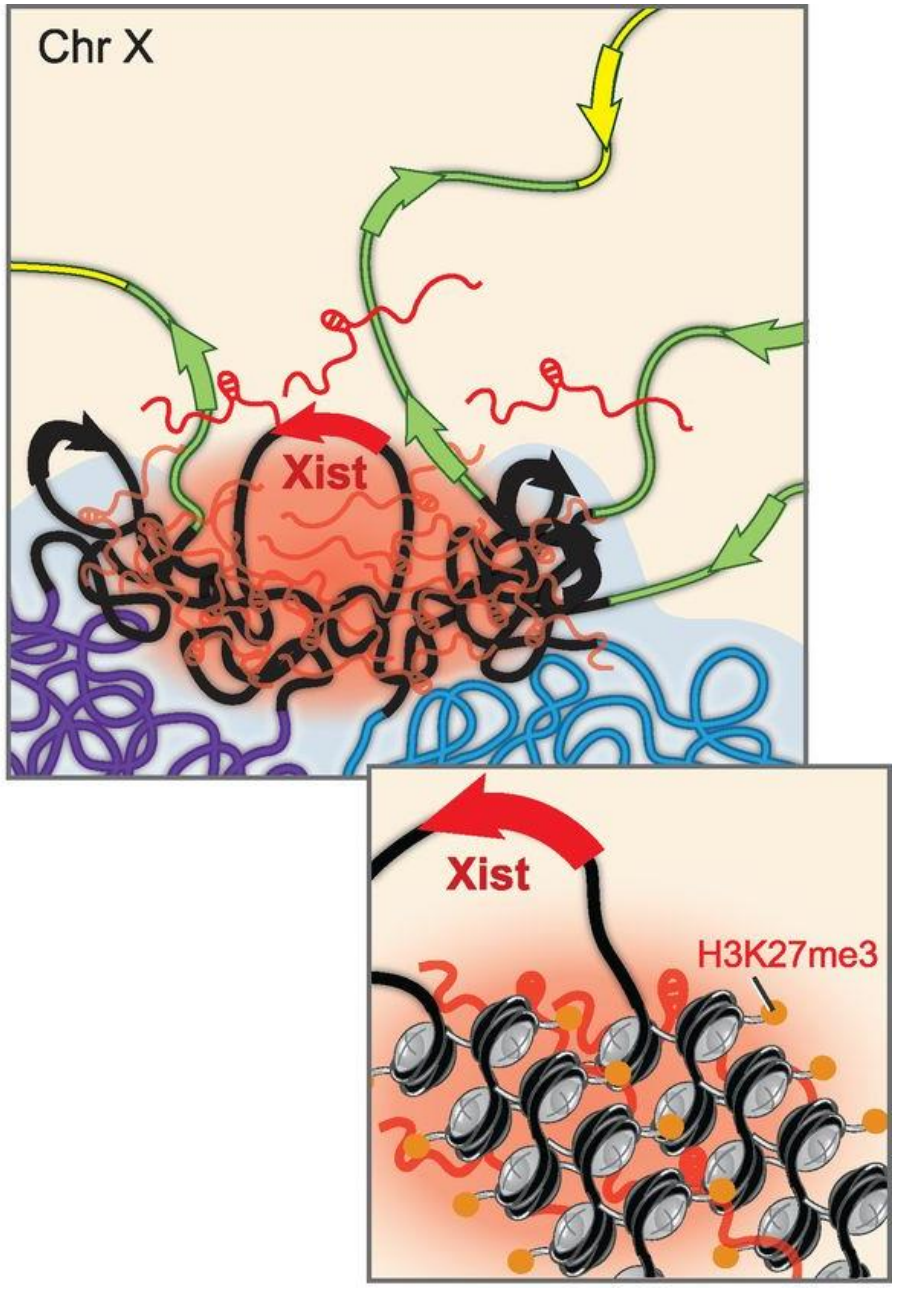
Population of different cells from the same tissue



Histogram: methylKit, Cells: Servier Medical Art

Sex Differences in Methylation: The X chromosome is the problem child

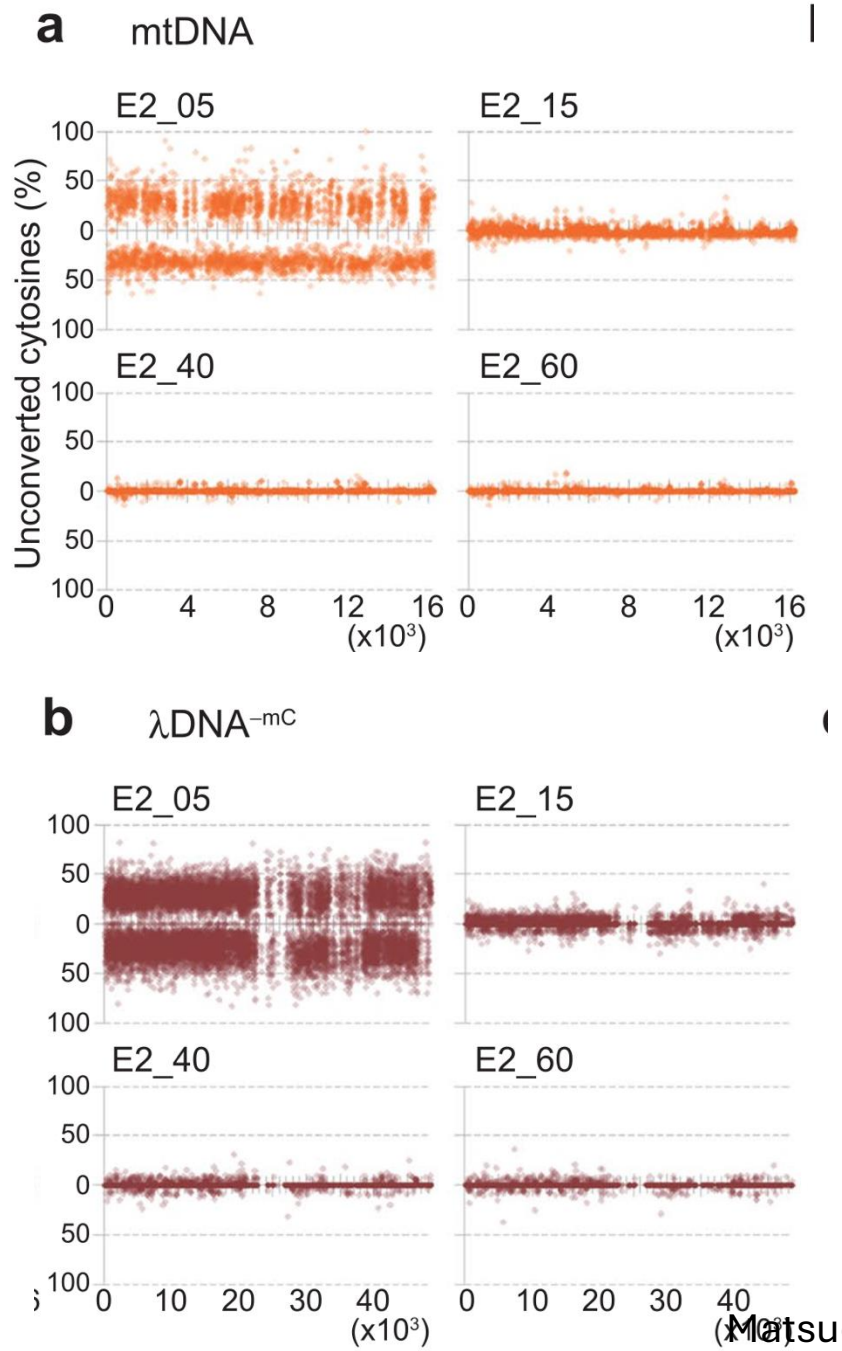
- XX vs XY
- X chromosome inactivation
    1. Xist inactivates X chromosome
    2. DNA methylation
    3. Histone acetylation
- XX samples have methylation ranging from 50-100%
- Separate autosomes and sex chromosomes



Jacobsen 2022, Image: Engreitz 2013

Mitochondrial DNA Methylation

- Matsuda 2018 exhaustively looked for mitochondrial methylation using biological mtDNA, synthetic mtDNA, and standard unmethylated control λ phage DNA
  - Bisulfite sequencing
  - Methylation specific endonuclease
  - Mass spectrometry
- Many sources of error
  - Cytosine bias towards the light strand of mitochondrial DNA
  - Difficulty linearizing and denaturing mitochondrial DNA for bisulfite treatment which isn't included in standard protocols
  - Insufficient controls in antibody-based methods
- DNMTs and TETs have been observed in the mitochondria



Matsuda 2018

Methods of Detecting Methylation

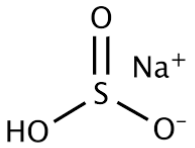
Array

Illumina 250K, 450K, 850K/EPIC arrays

mass spec

Sequencing

Bisulfite (BS-seq)



WGBS

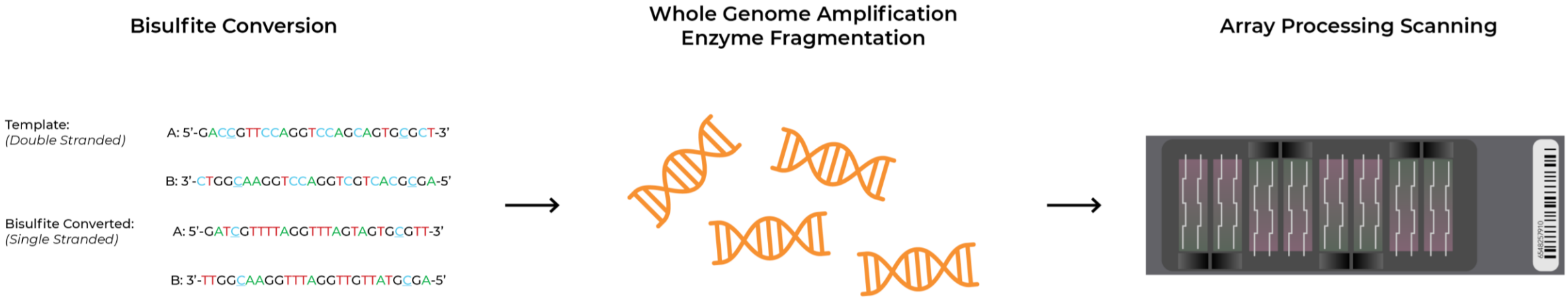
RRBS

MeDIP

Chemically modified 5hmC, 5fc, 5caC can only be detected by this method

Illumina Methylation Arrays

Illumina Methylation Arrays: Function



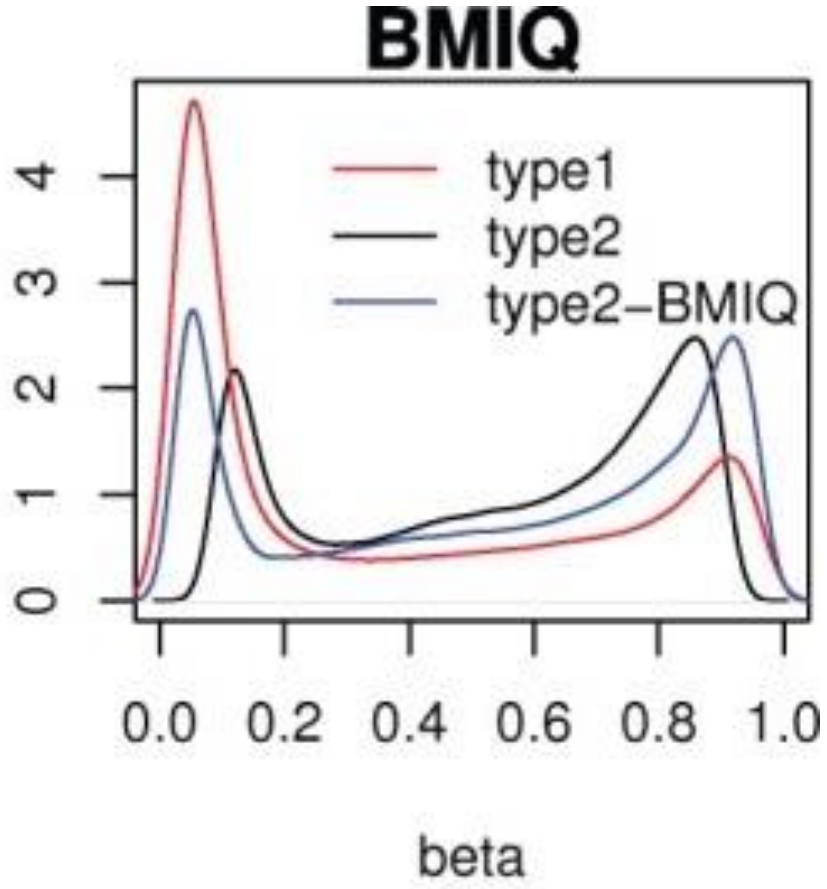
Illumina® Methylation Arrays Workflow (Wetlab)

- Measures methylation through the ratio of red/green fluorescence
- Measure specific loci, primarily around genes

Image: Zymo

Illumina Methylation Arrays: Processing

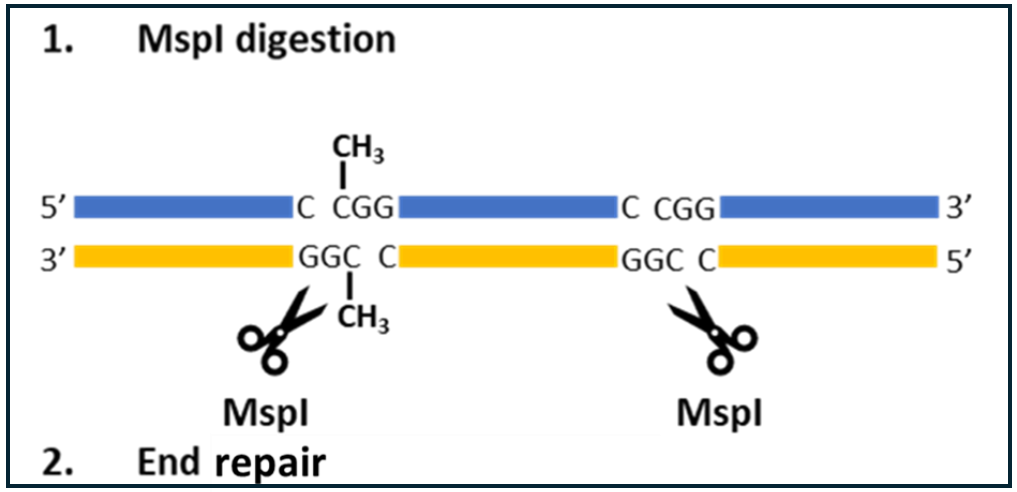
1. Filter
   - **Bad quality probes (array provides a detection p-value for the probability of a false result)**
   - **There are multiple beads per site, filter out the site if its detected by too few beads**
   - **Filter probes at known SNPs that cause false results**
   - Filter out chrX , chrY, chrMT
2. Check for batch effects or other associations with unwanted covariates
3. **Probe normalization (type I vs type II)**
4. Calculate differential methylation

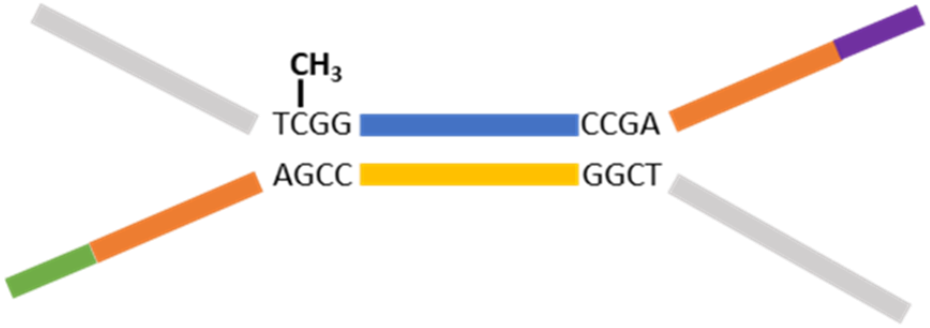


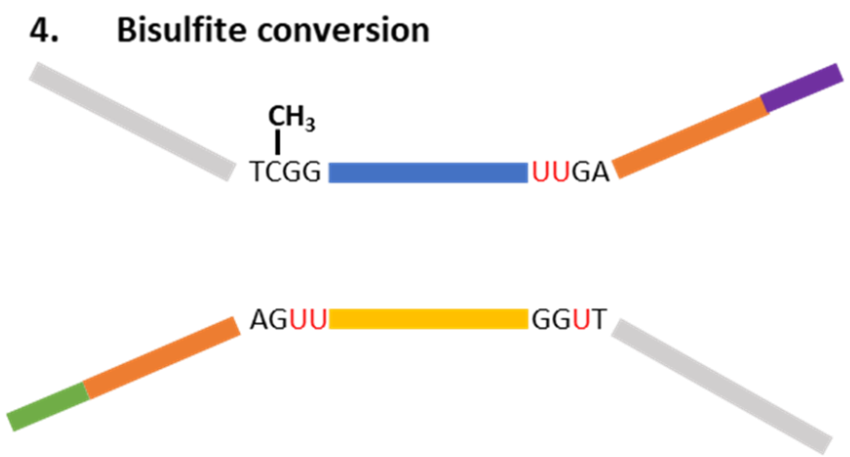
Bisulfite Sequencing

Constructing Bisulfite Sequencing Libraries

Reduced Representation







Bisulfite Sequencing

1. Quality Check

   ```
   fastqc *.fq.gz
   ```

2. Trim

   ```
   trim_galore --rrbs --paired --fastqc -q 30 --illumina
   --output ../01_trim file_R1.fq.gz file_R2.fq.gz
   ```

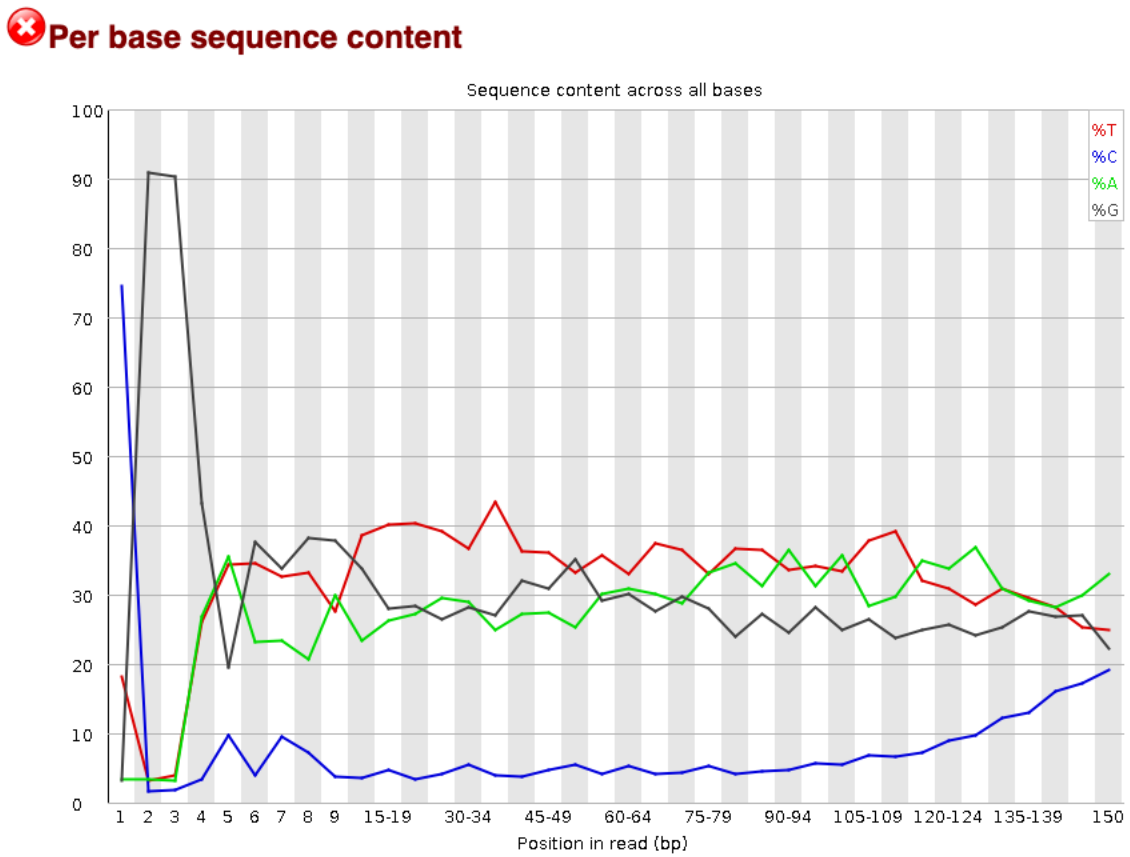
3. Align

   ```
   bismark --bowtie2 /path/to/bismark/index --output
   ../02_align -1 file_R1.fq.gz -2 file_R2.fq.gz
   ```

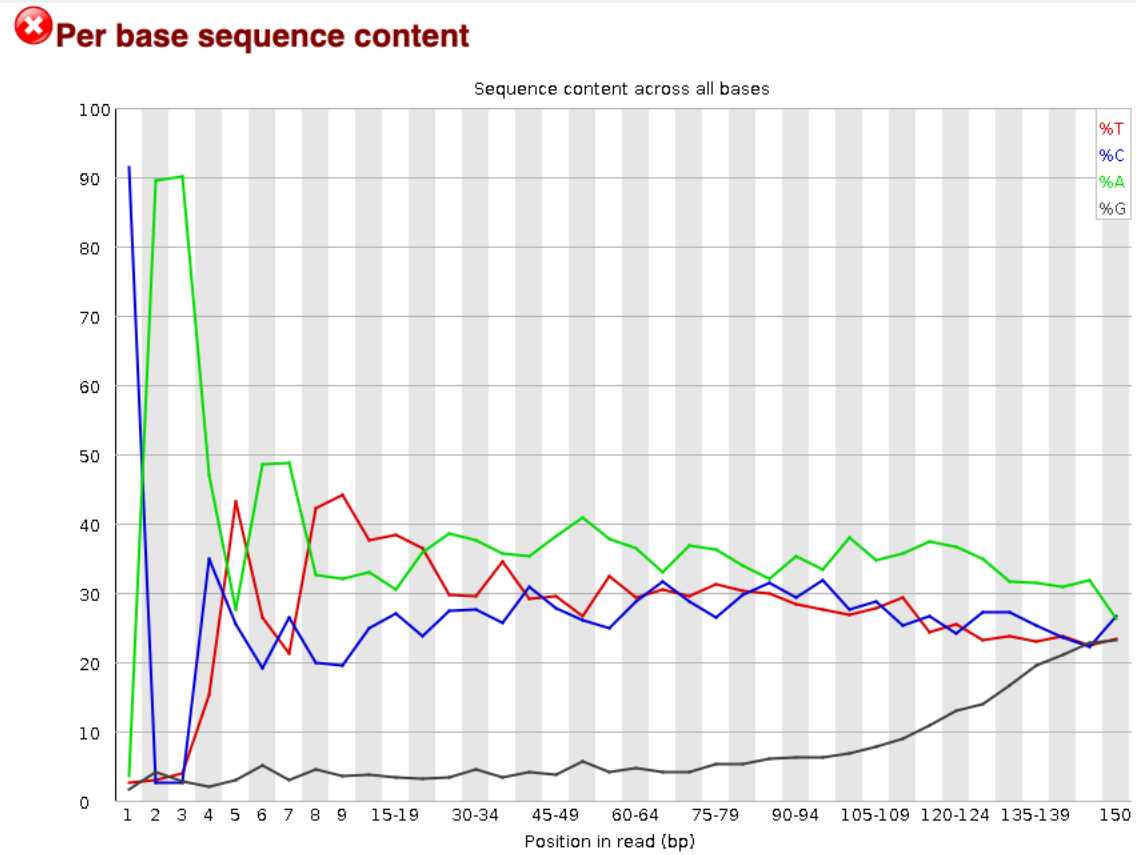
4. Count

FastQC: Per base sequence content

Read 1

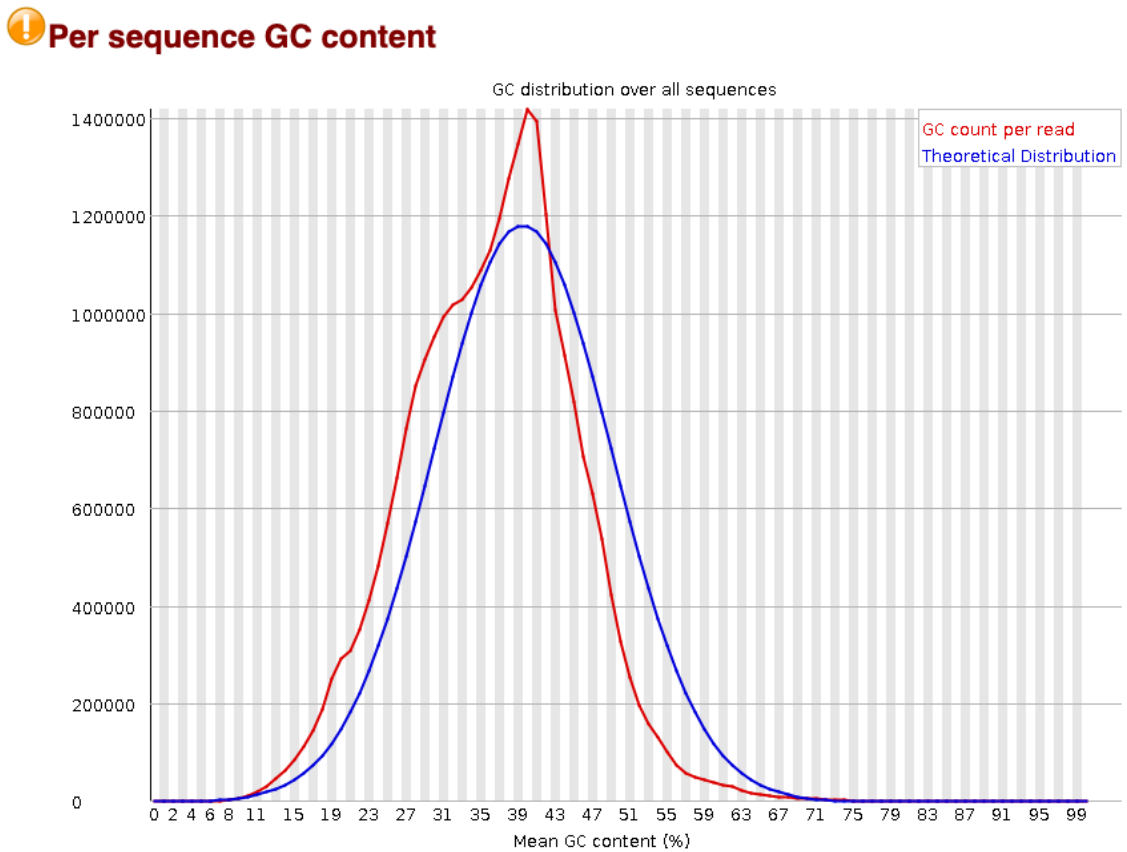


Read 2

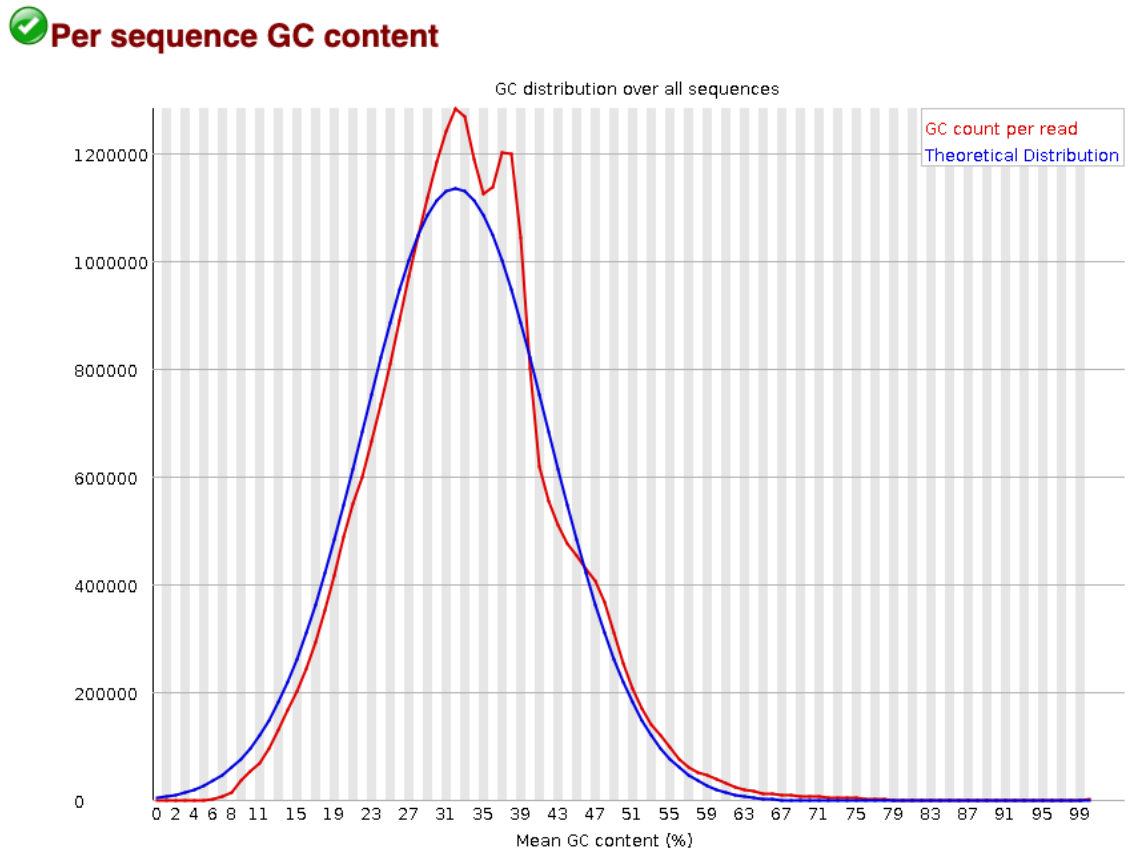


FastQC: Per sequence GC content

Read 1

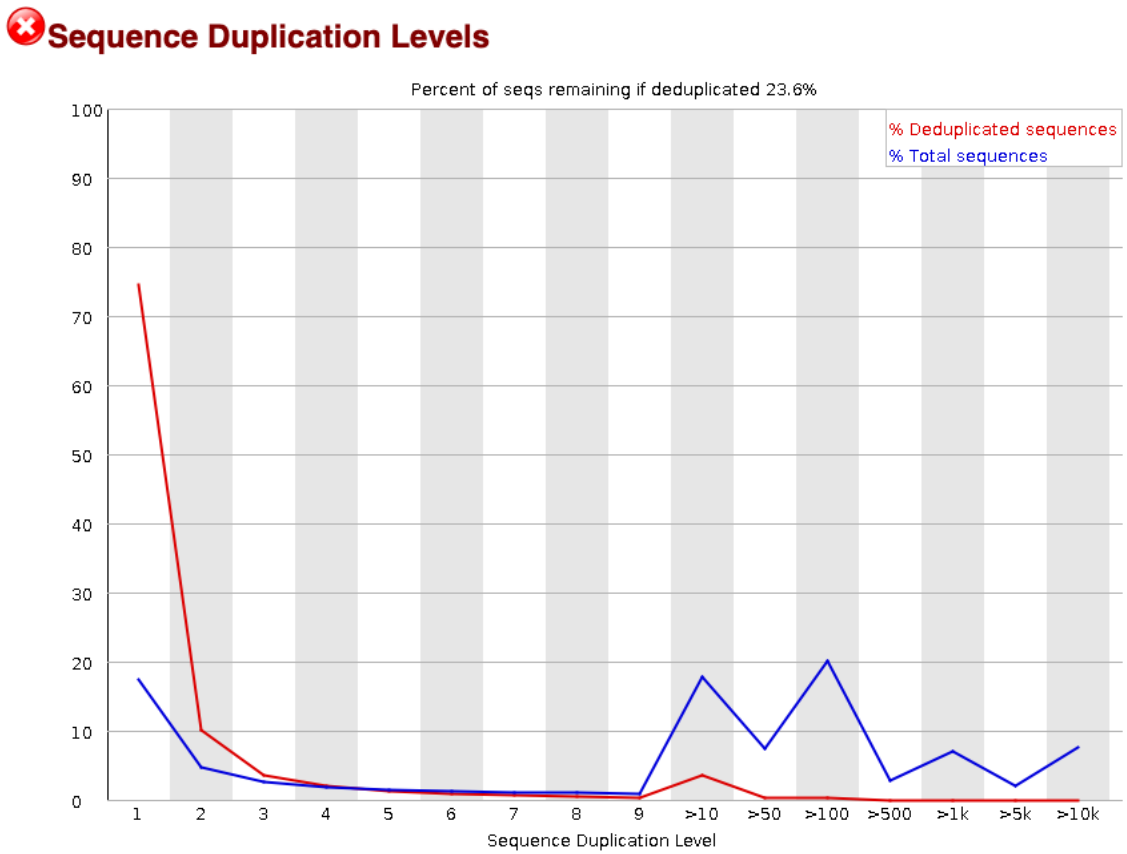


Read 2

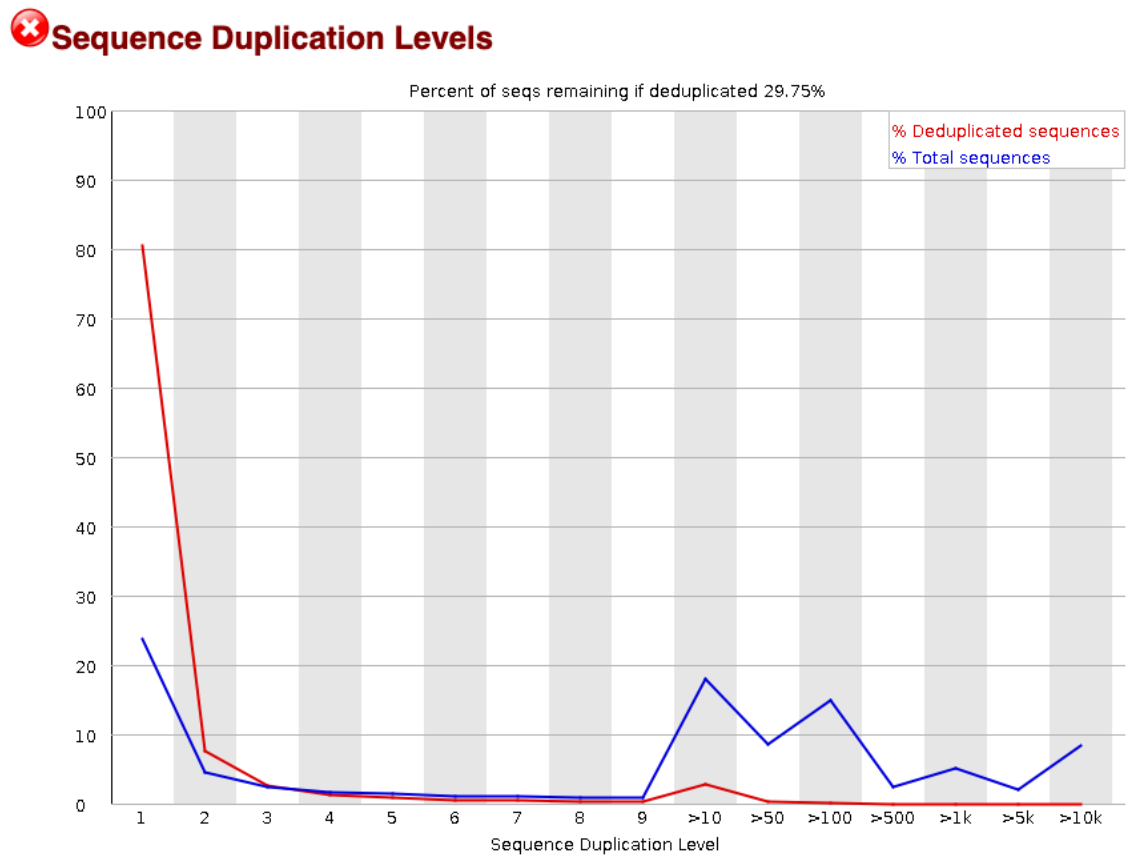


FastQC: Per sequence GC content

Read 1



Read 2



FastQC: Overrepresented sequences

Read 1

Overrepresented sequences

Sequence	Count	Percentage	Possible Source
CGGAATAGAATGGAATGGAATGGAATGGAACGGAATGGAATGGAATGGAA	285907	1.008042814151108	No Hit
CGGGCGCGGTGGTTTACGTTTGTAATTTTAGTATTTTGGGAGGTCGAGGC	207030	0.7299405184682567	No Hit
CGGGTGGAGTGGAATGGAATGTAATGGAGTGGAATGTAATGGAATTTAGT	96288	0.3394895070389388	No Hit
CGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAA	93849	0.3308901498223804	No Hit
CGGAATAGAATGGAATGTAATGGAATGGAACGGAATGGAATGGAATGGAA	81960	0.2889722498848395	No Hit
CGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATTAAT	70703	0.2492826376721304	No Hit
CGGAACGGAATGGAATGGAATGGAATGGAATTAATTCGATTGTAATGGAA	65183	0.22982037779701675	No Hit
CGGTGGATTTTTCGGTTTAAGTTTTGGTAATACGGTGAAATTTCGTTTTA	61728	0.2176388365164882	No Hit
CGGAATAGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAA	53744	0.18948907513190352	No Hit
CGGAATGGAATGGAATGGAATGGAATGGAATGAAATGTAATGGATTTAAT	49586	0.17482891633467118	No Hit
CGGGCGCGGTGGTTTACGTTTGTAATTTTAGTATTTTGGGAGGTCGAGGT	47280	0.16669848675640814	No Hit
CGGGAGGCGGAGTTTGTAGTGAGTCGAGATCGCGTTATTGTATTTTAGTT	41059	0.1447646609080237	No Hit
CGGAATGGAATGGAATGGAATGGAATTAATTTTATTGTAATGGAATGGAA	31979	0.11275065372214838	No Hit
CGGGAGGCGGAGTTTGTAGTGAGTCGAGATTTCGTTATTGTATTTTAGTT	31331	0.11046595364985244	No Hit
CGGAATAGAATGGAATGGAATGGAATGGAACGGAACGGAATGGAATGGAA	29976	0.10568853297398667	No Hit
CGGAATGGATTGGATTGGAATAAAACGGATTCGAATGTAAAGTATTGTAA	28528	0.10058321552848587	No Hit

Read 2

Overrepresented sequences

Sequence	Count	Percentage	Possible Source
CAAATAATTCCATTCCATTACATTCCATTCCATTCCATTCCCCTACACTC	627432	2.2121819996308516	No Hit
CAAATAATTCCATTCCATTACATTCCGTTCCATTACATTCCCCTACACTC	120581	0.42514107934802137	No Hit
CAACTAATTTTTTATATTTTTAATAAAAACGAAATTTCACCGTATTAACC	112378	0.3962191739575218	No Hit
CAAAAAATTCCATTCCATTACATTCCATTCCATTCCATTCCCCTACACTC	96550	0.3404132592286634	No Hit
CAACCTATTTATCTATTTATTAACTTTAAATCCAAATTATAAAACCAATT	85170	0.300289977094824	No Hit
CAACTAATTTTTTATATTTTTAATAAAAACGAAATTTCACCTTATTAACC	68317	0.24087014635654683	No Hit
CAAATAATTCCATTCCATTAAATTCCATTCCATTCCATTCCCCTACACTT	56942	0.20076449308129002	No Hit
CAAACGCGATAACTCACGCCTATAATCCCAACACTTTAAAAAACCGAAAC	50011	0.17632736931418627	No Hit
CAAATTCACGCCATTCTCCTACCTCAACCTCCCGAATAACTAAAACTACA	45832	0.1615931693109073	No Hit
CAAAAAATTCCATTCCATTACATTCCATTCCATTCCATTCCATTCCATTC	39358	0.13876732321824686	No Hit
CAACTAATTTTTTATATTTTTAATAAAAACGAAATTTCACCATATTAACC	36925	0.13018912063198754	No Hit
CAAATAATTCCATTCCATTACATTCCATTCCATTCCATTCCCCAACACTC	34825	0.12278500002732473	No Hit
CAAATAATCAATACTATATCCGACAATATTCGTATTTTATTTTCTTCCTA	32826	0.11573698236602904	No Hit
CAATTAATATTTACTATCCTTCATCATAACATCTTATTTTCTAAAAATCC	32734	0.11541261136811048	No Hit
CAAATACTATAAACTTTTTCTTTAACTTTTTACTATTTCTTTCCTTTTCT	28468	0.1003716692254955	No Hit

Bisulfite Sequencing: Other Considerations

- Alignment efficiency will be bad
  - 50-60%
  - Any lower probably need to repeat sample
- lambda bisulfite conversion efficiency
- bisulfite conversion kits

LIVE DEMO

Other things you can do

- Analyze data as differentially methylation regions (DMRs) https://bioconductor.org/packages/release/bioc/html/DMRcate.html
- Pathway analysis – use
- Estimate age/“biological” age https://www.bioconductor.org/packages/release/bioc/html/methylclock.html
- Cell type deconvolution https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9947552/
- DNA methylation entropy
  - Shannon entropy most commonly used https://cran.r-project.org/web/packages/shannon/shannon.pdf
  - Example paper https://academic.oup.com/nar/article/51/5/2046/7033790

References

- Biology/Sequencing Methodology
  - mtDNA is not methylated https://www.nature.com/articles/s41598-018-24251-z
  - IgG controls in MeDiP (and other antibody-based assays) https://www.nature.com/articles/s41592-018-0038-7
- arrays https://www.sciencedirect.com/science/article/pii/S2001037024001624
  - minfi https://bioconductor.org/packages/release/bioc/html/minfi.html
  - limma https://www.bioconductor.org/packages/release/bioc/html/limma.html
  - **ChAMP** https://bioconductor.org/packages/release/bioc/html/ChAMP.html
- Bisulfite sequencing
  - bisulfite sequencing
    - Trim Galore https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
    - Bismark aligner https://felixkrueger.github.io/Bismark
  - Analysis tools
    - **methylSig** https://www.bioconductor.org/packages/release/bioc/html/methylSig.html
    - methylKit https://www.bioconductor.org/packages/release/bioc/html/methylKit.html
    - edgeR https://www.bioconductor.org/packages/devel/bioc/vignettes/edgeR/inst/doc/edgeRUsersGuide.pdf