



DNA Methylation Methods and Technologies

Jessica Nordlund, PhD
SNP&SEQ Technology Platform
National Genomics Infrastructure
Uppsala University



Outline

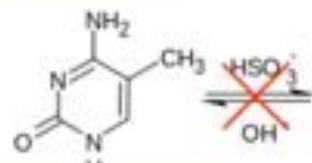
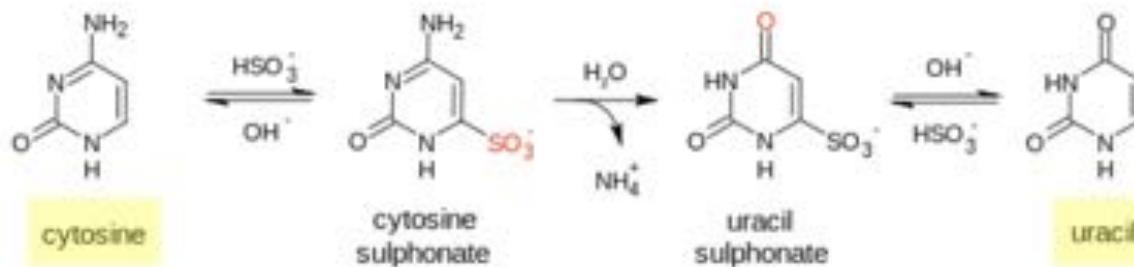
- i. Overview of methods for interrogation of DNA methylation
 - Overview of important concepts
 - Enrichment & targeted-based methods
 - Genome-wide methods
- ii. Questions/break
- iii. How to access epigenomics services for your research project at Sweden's National Genomics Infrastructure (NGI)



Short intro: bisulfite conversion

"Gold standard" for DNA methylation analysis. Single nucleotide resolution.

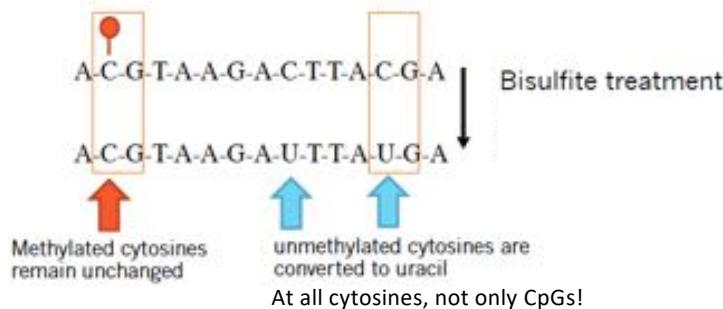
The Chemistry of Bisulfite Conversion of Cytosine to Uracil:



!!!!5-methyl-cytosine is resistance
to conversion!!!!

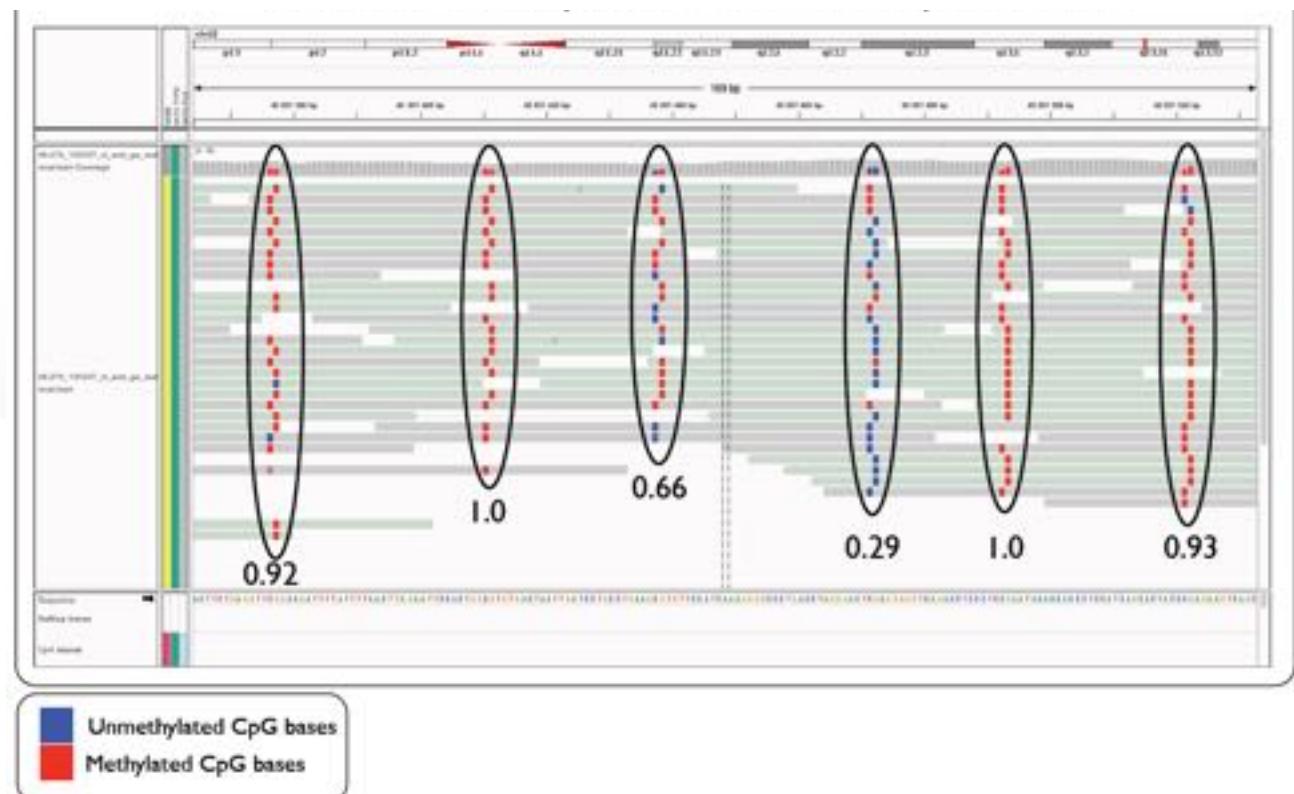


Base-pair resolution and quantitative measurement of methylation levels



C = methylated

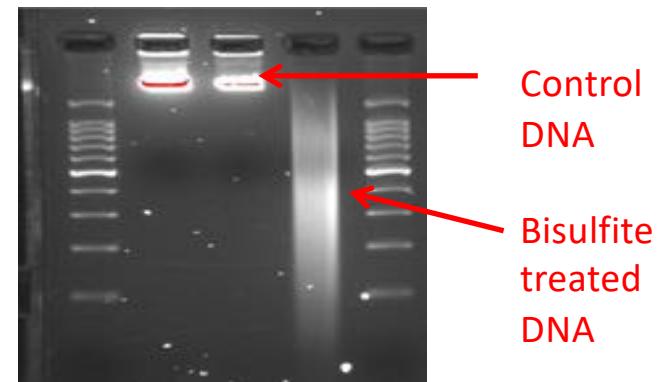
U->T = unmethylated





What you need to know about bisulfite conversion

- Gold standard for DNA methylation analysis
- Very harsh chemical that degrades and fragments DNA
- 5mC and 5hmC are indistinguishable in standard bisulfite protocols

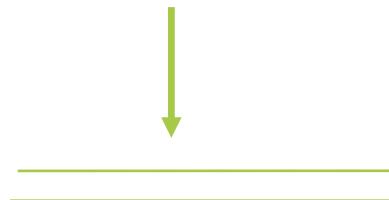




Short intro: "NGS" libraries



Double stranded genomic DNA



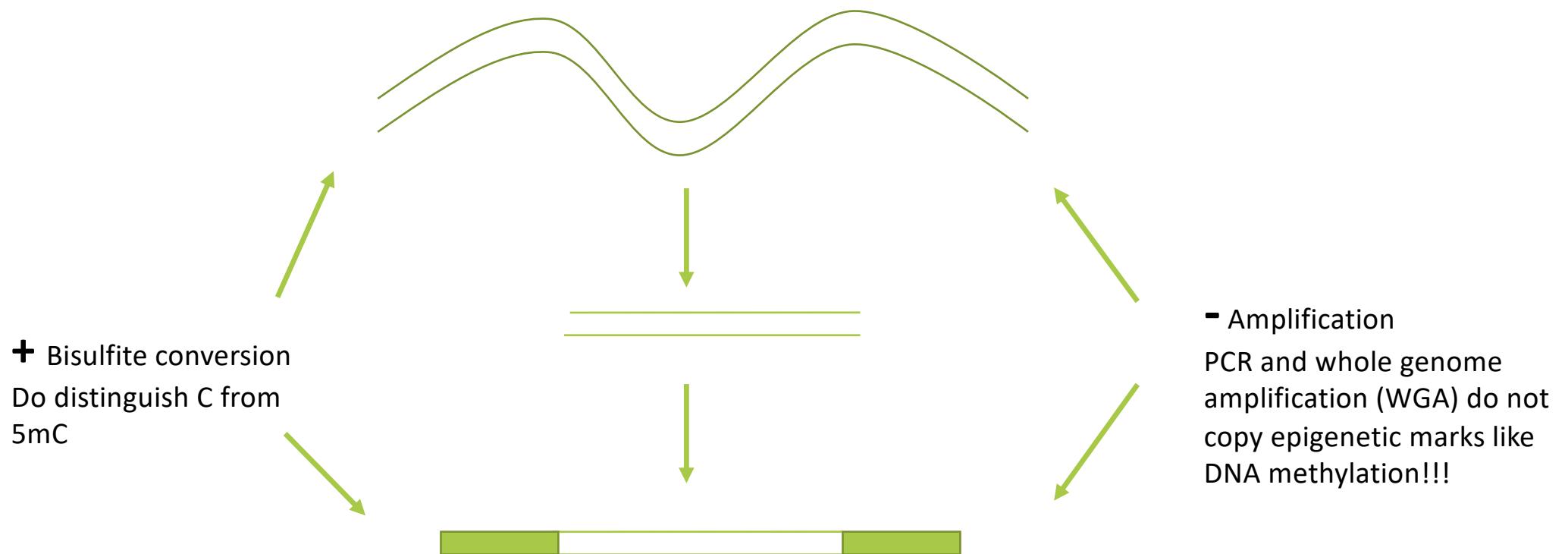
Shearing to make DNA
fragments shorter (optional)



Ligate platform-specific
sequencing adapters



Short intro: "NGS" libraries



Enrichment & targeted-based methods



Different approaches to reduce the genome to regions of interest
(typically those with many CpG sites)

- Cost saving (less sequencing required)
- Less computationally intensive (less data generated)
- High throughput (some approaches)



Enrichment-based methods

Capture of methylated DNA fragments using methyl-binding protein or a anti-methyl-cytosine antibody

- MeDIP-seq (Methylated DNA immunoprecipitation):

- ✓ Genome-wide coverage
- ✓ ~150bp resolution.
- ✓ Anti-body against 5-Hydroxy-methyl-cytosine
- ✓ Relatively cost-efficient

- MBD-seq (Methylated DNA binding domain):

- ✓ Genome-wide coverage
- ✓ ~150bp resolution.
- ✓ Only capture CpG methylation not CHH
- ✓ Relatively cost-efficient

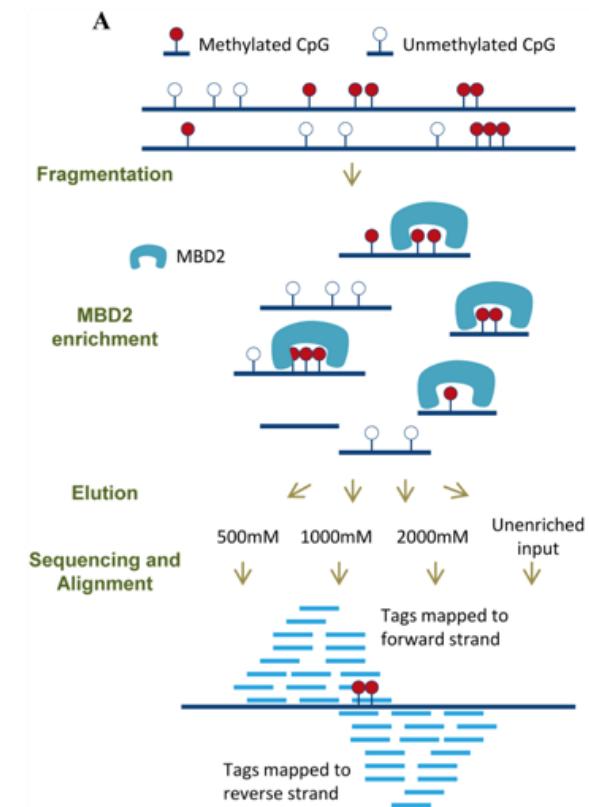
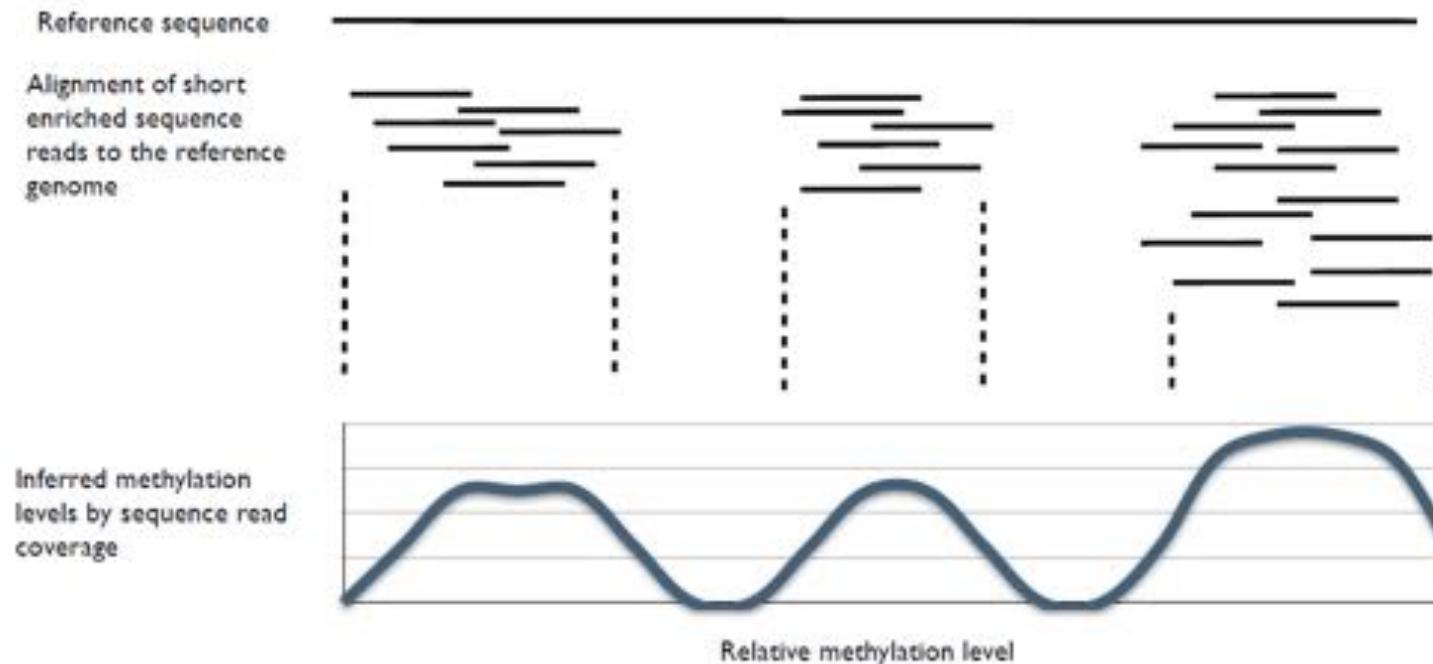


Figure from:
Lan, et al. (2011) High Resolution Detection and Analysis of CpG Dinucleotides Methylation Using MBD-Seq Technology.
<https://doi.org/10.1371/journal.pone.0022226>



Enrichment-based methods



The depth of sequence reads is taken as an indirect measurement of Methylation levels

Cons:

- Not base-pair resolution
- Indirect measurement of DNA methylation can be difficult to interpret
- Lab-intensive and not easily automated



Target-Capture

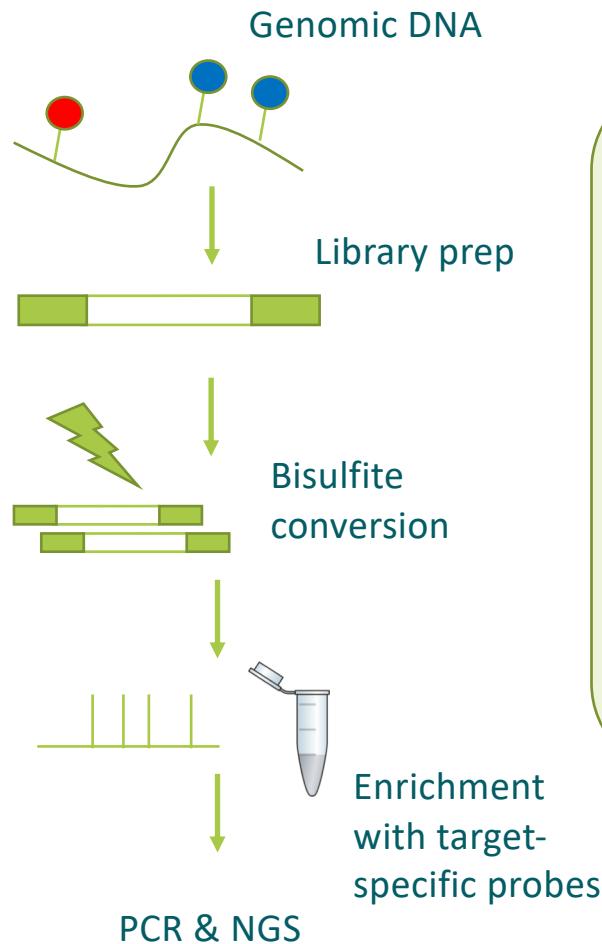
Target-capture of pre-defined genomic regions, NGS library preparation, uses bisulfite conversion.

Pros:

- Focused set of targets regions: can achieve high coverage on target
- “Cost-effective”
- Captures millions of CpG sites (3-5M)

Cons:

- Typically only for Human, other species possible on some platforms
- Bisulfite conversion cannot distinguish between 5mC and 5hmC
- Labor intensive wetlab protocols



Seq-Cap enrichment
(Roche) –
5M CpGs

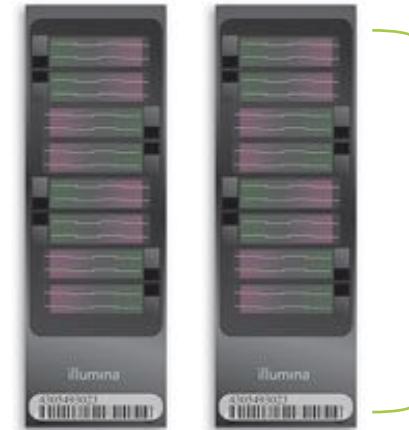
SureSelect^{XT} Methyl-Seq
Target Enrichment Kit
(Agilent Technologies) –
3.7M CpGs

TruSeq[®] Methyl Capture
EPIC kit (Illumina) –
3.3M CpGs



DNA methylation arrays

- Bisulfite converted DNA
- >800,000 CpG sites
- 96% CpG islands
- 99% Refseq genes
- CpG sites outside of CpG islands
- Non-CpG methylated sites identified in human stem cells
- Differentially methylated sites found in cancer and several tissue types
- FANTOM 4 promoters
- DNase hypersensitive sites
- miRNA promoters



$$\text{Beta value } (\beta) = \frac{M}{M + U + 100}$$



DNA methylation arrays

Pros:

- The most popular method on the market
 - Base-pair resolution
 - Compatible with FFPE DNA
 - Compatible with 5hmC detection
 - Many **R packages** available for data analysis and publically available datasets
-
- Cons
 - Human only* (Mouse Methylation BeadChip available with 285k CpG sites)
 - 850k out of 29M CpG sites





Reduced Representation Bisulfite Sequencing (RRBS)

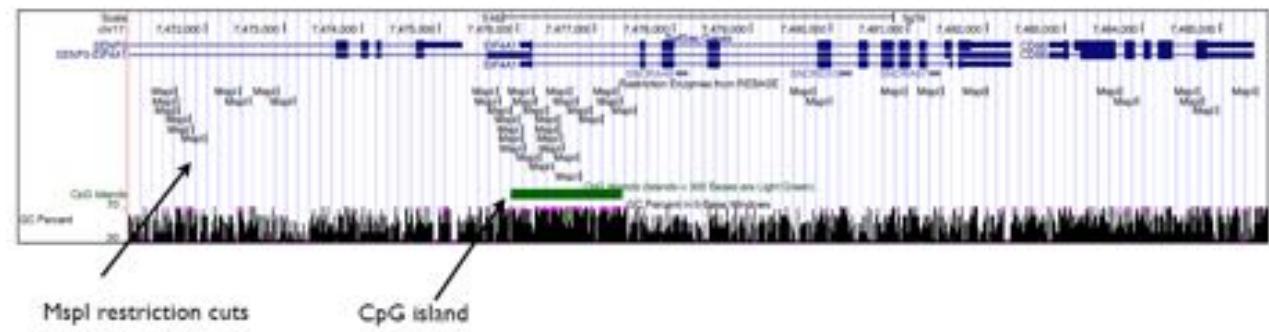
- "reduces" the genome to informative regions with high CG content
- Based on restriction digestion with an enzyme that cuts at CCGG (MspI)

Pros:

- Compatible with most species
- Low cost
- Base-pair resolution (bisulfite)
- reads are heavily concentrated to CpG islands
- High throughput

- MspI (C^CGG)
Methylation insensitive restriction enzymes.
- TaqI (T^CGA)

• Base-pair resolution



Cons

- Does not capture all promoters or CpG islands
- Results can vary depending on input DNA quality / contaminants in the sample



RRBS

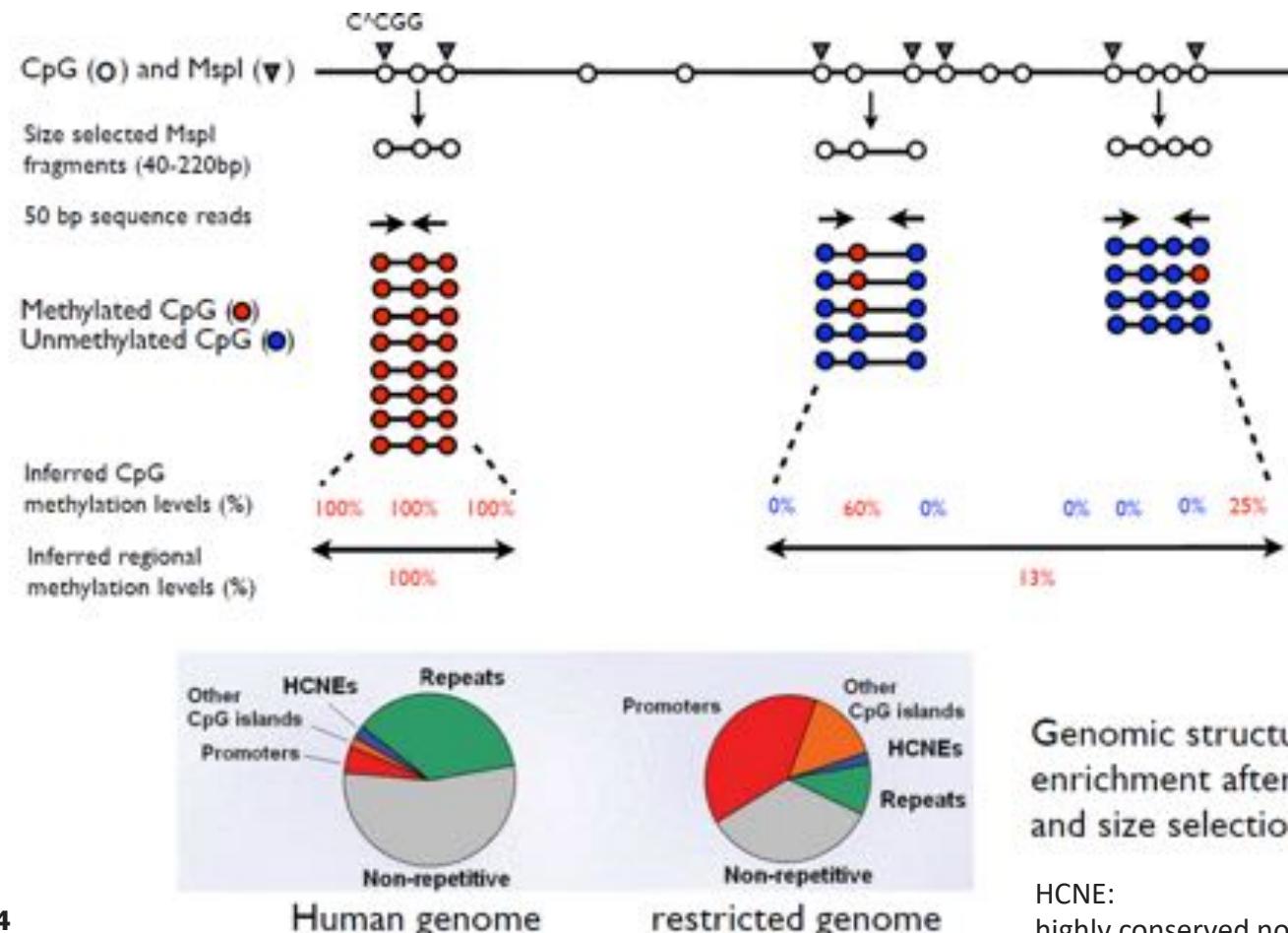
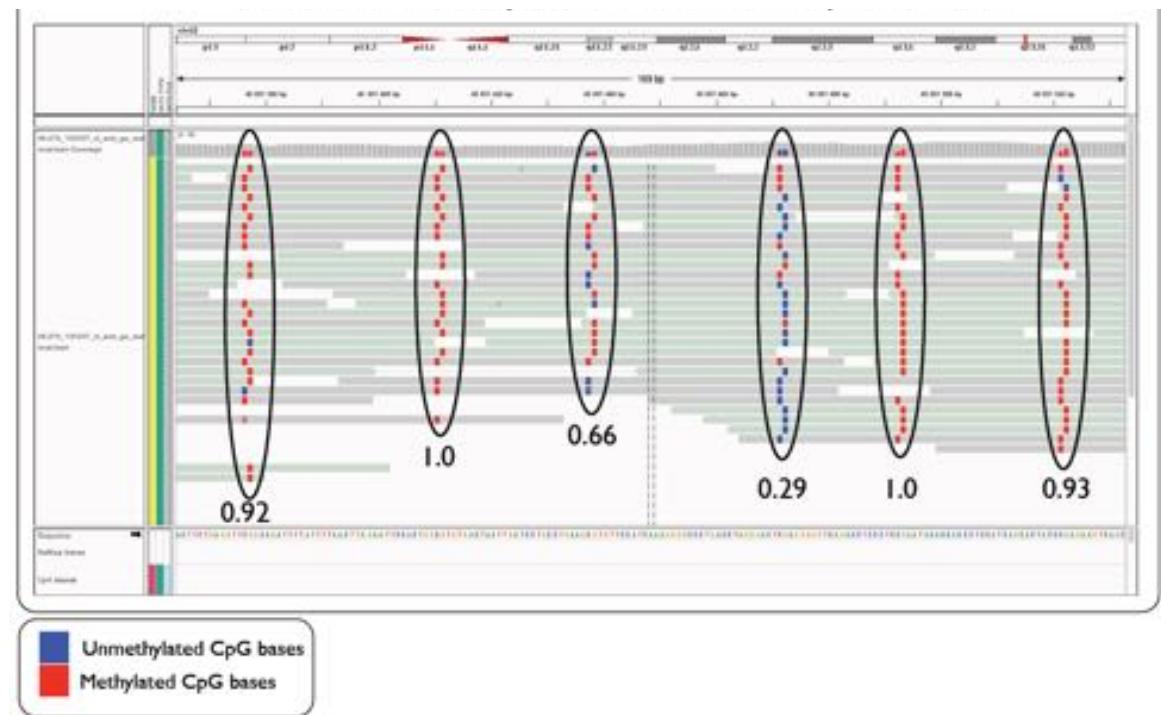


Figure adapted from Meissner et al. *Nature* **454** (2008) <https://doi.org/10.1038/nature07107>

Whole Genome Bisulfite Sequencing (WGBS)

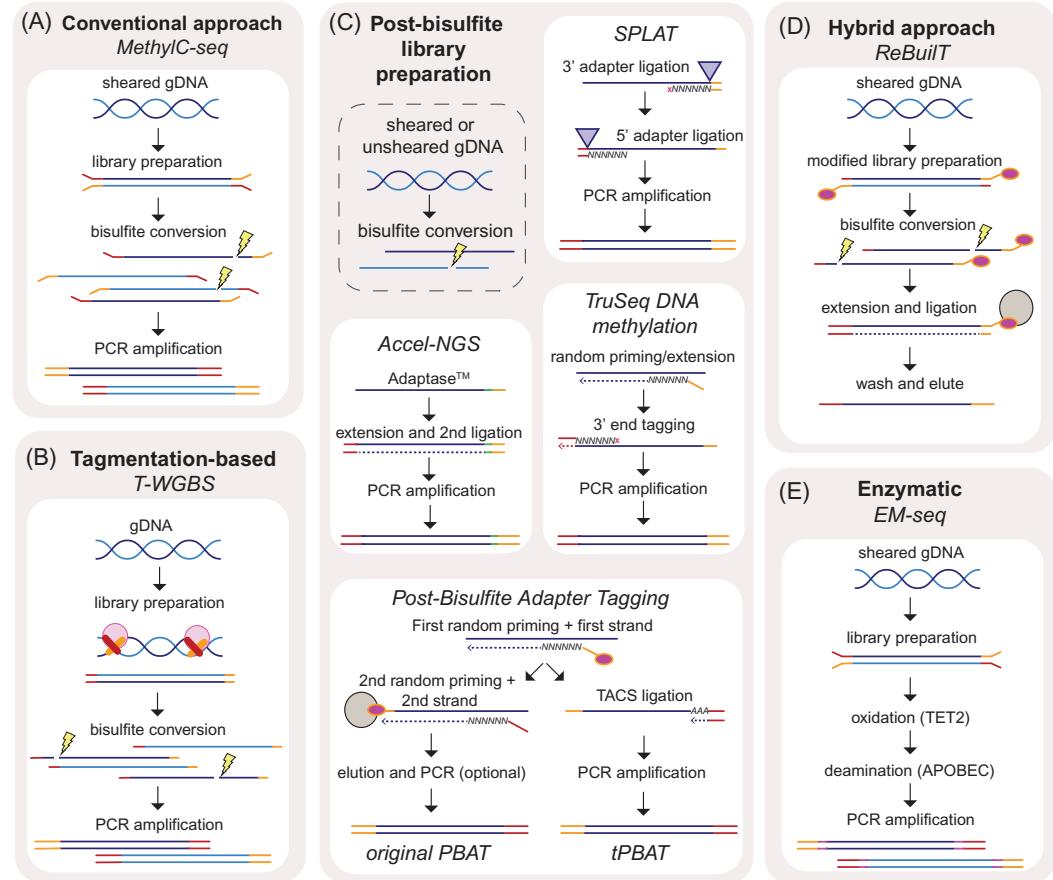
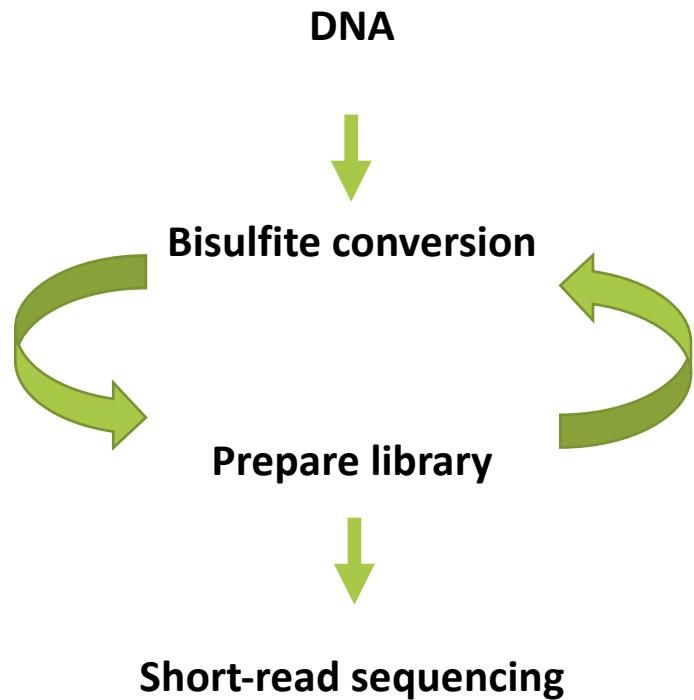


- Many acronyms ; WGBS, MethylC-seq, BS-seq
- “Unbiased” – no selection or enrichment
- Genome-wide coverage of all cytosines
- Base-pair resolution
- Uses bisulfite conversion to distinguish methylated from unmethylated cytosines





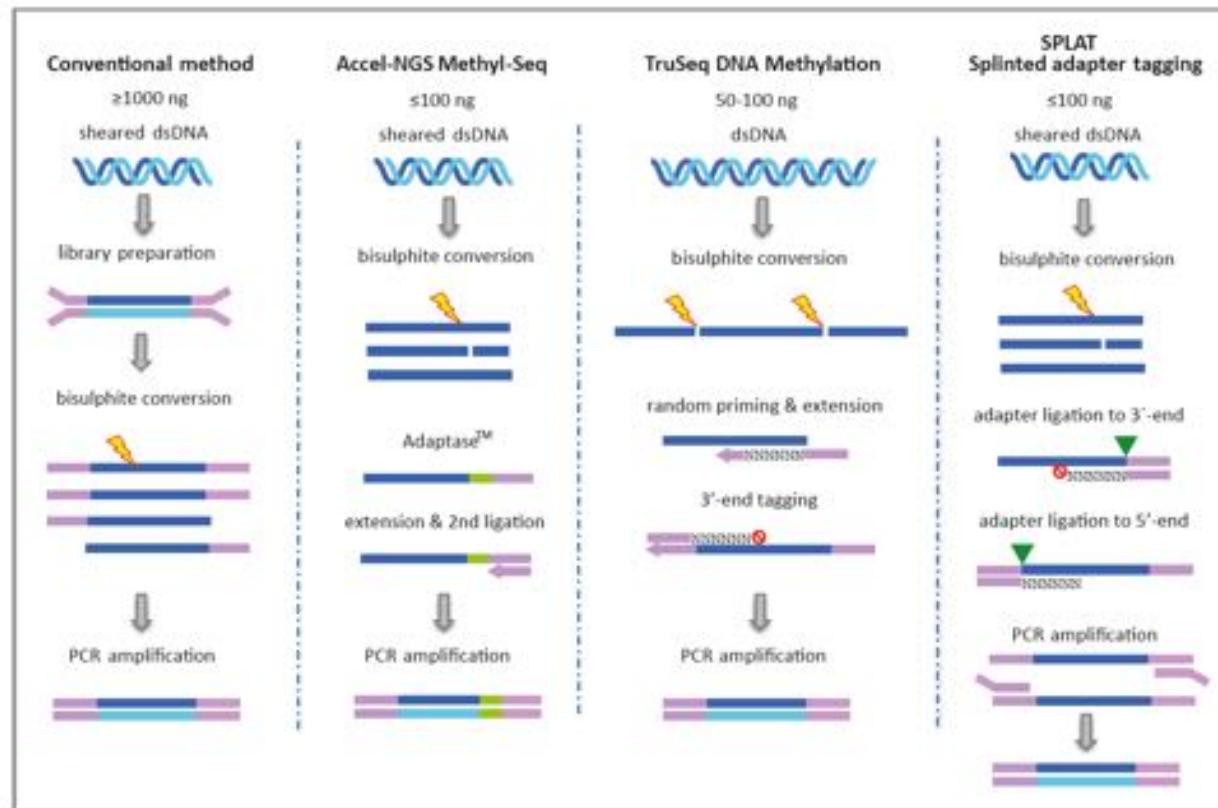
Many different approaches ...



J Nordlund, Chapter Eleven - Advances in whole genome methylomic sequencing, Epigenetics Methods, Academic Press (2020), <https://doi.org/10.1016/B978-0-12-819414-0.00011-2>.



"Post bisulfite" library preparation

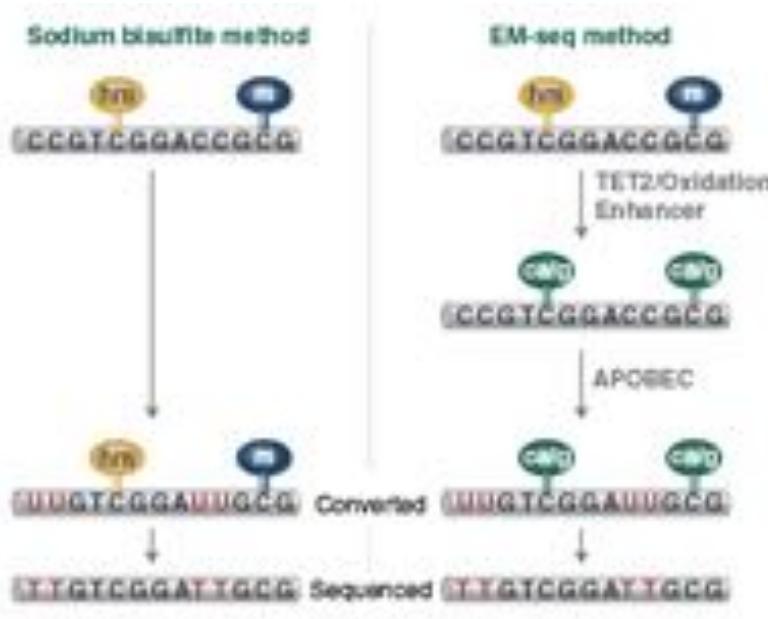


New methods for WGBS library prep enables sequencing of samples using much lower DNA quantities (even single cells)

Raine et al, NAR 2017
<https://doi.org/10.1093/nar/gkw1110>



New innovation! Bisulfite-free!



Gentle with little/no strand breakage!

WGBS is the gold standard for methylome analysis, but the chemical bisulfite reaction:

- I. Damages / degrades DNA
- II. Results in fragmentation / loss
- III. Can result in CG bias and uneven genome coverage

Enzymatic methylation sequencing (EM-seq)

TET2 enzymatically oxidizes 5mC and 5hmC through a cascade reaction into 5-carboxycytosine (5caC)

*5-methylcytosine (5mC) → 5-hydroxymethylcytosine (5hmC)
→ 5-formylcytosine (5fC) → 5-carboxycytosine (5caC)*

A second enzymatic step uses APOBEC to deaminate cytosine to uracil, but does not affect 5caC.



Genome-wide methylome sequencing

Direct read out of DNA modifications by single molecule, long read technologies (PacBio, Oxford Nanopore)

PacBio
SMRT seq

DNA passes thru
polymerase in an
illuminated volume



Raw output is fluorescent signal
of the nucleotide incorporation,
specific to each nucleotide

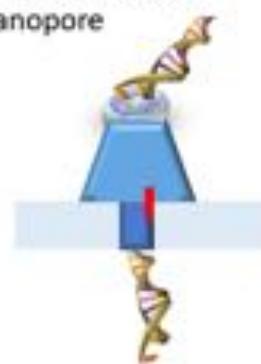


A,C,T,G have known pulse
durations, which are used to
infer methylated nucleotides



Oxford
Nanopore

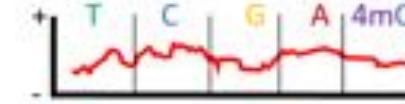
DNA passes thru
nanopore



Raw output is electrical signal
caused by nucleotide blocking
ion flow in nanopore



Each nucleotide has a specific
electric "signature"



In theory can detect all sorts of DNA modification-Challenge is to train models to correctly detect specific modifications

Cons; need a lot of native DNA for sequencing + compute

Pros: Phased information! Allele specific methylation. Imprinting

Short vs long-read sequencing, what's the difference?



Short-read

Illumina

Pros:

- Low cost
- High throughput
- Detect 5mC & 5hmC *depending on library prep applied
- Species agnostic

Cons:

- Requires conversion of (un)modified bases DNA with chemicals or enzymes
- 5mC cannot be distinguished from 5hmC (and other types of marks) without specific conversion approaches

Long-read

PacBio/ONT

Pros:

- Base modification can be read directly from sequencing
- Maintain phasing information
- Detect 4mC, 5mC, 5hmC, 5fC, 5caC, 6mA, etc
- Species agnostic

Cons:

- Cost (high coverage needed) – limiting for large genomes
- Difficult to detect signals
- Low throughput



Reproducibility & quality

Foox J, Nordlund J, et al. The SEQC2 Epigenomics Quality Control (EpiQC) Study: Comprehensive Characterization of Epigenetic Methods, Reproducibility, and Quantification. BioRxiv: <https://doi.org/10.1101/2020.12.14.421529>

EPIC arrays

- duplicate/triplicate at 3 labs

WGBS

- TruSeq DNA methylation (Illumina)
- Accel-NGS methyseq (Swift)
- SPLAT (Raine et al, NAR 2017)

OXBS

- TrueMethyl oxBs-seq (NuGEN)

Enzymatic deamination

- EM-seq (NEB)

ONT: direct methylation calling

7 cell lines

Alignment and methylation calling:

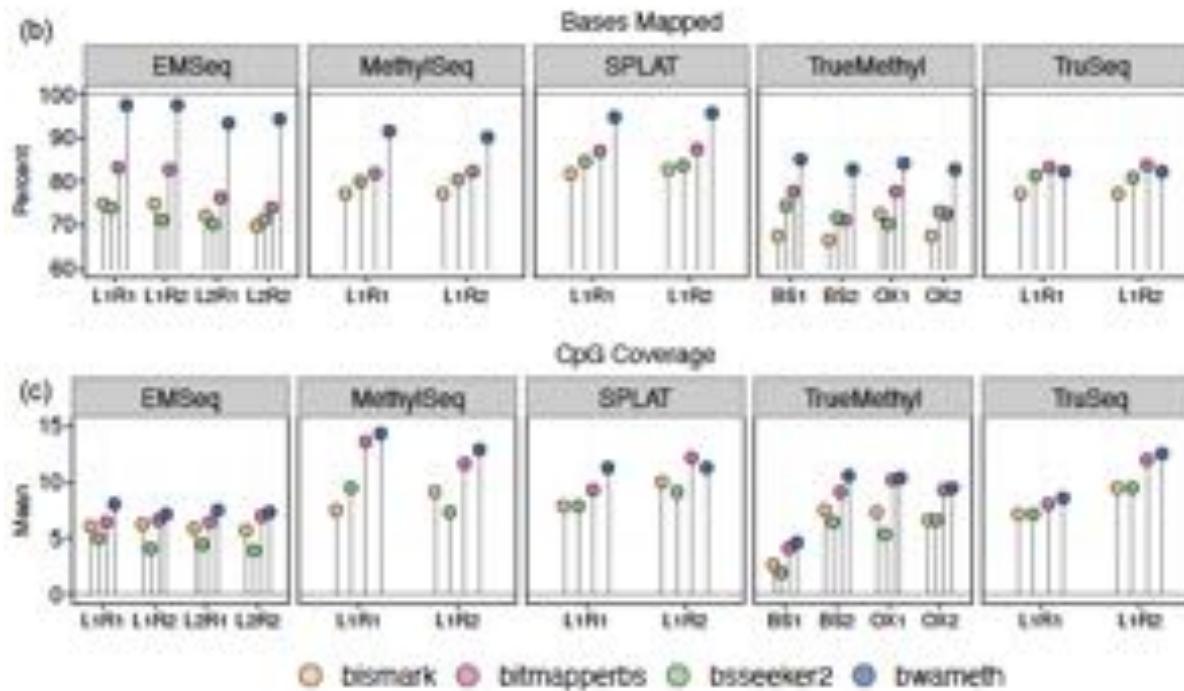
- BISMARK
- BitMapperBS
- BSSeeker2
- Bwa-meth
- Gem-bs

Microarray normalization

- 26 between-array and within-array normalization methods



Reproducibility & quality



Overall, no major quantitative difference between pipelines but bwa-meth was easiest to implement and retained most data.

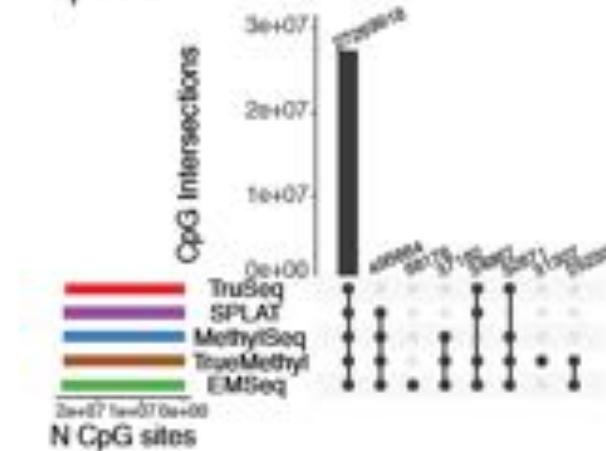
Noticeable inter- and intra-library differences



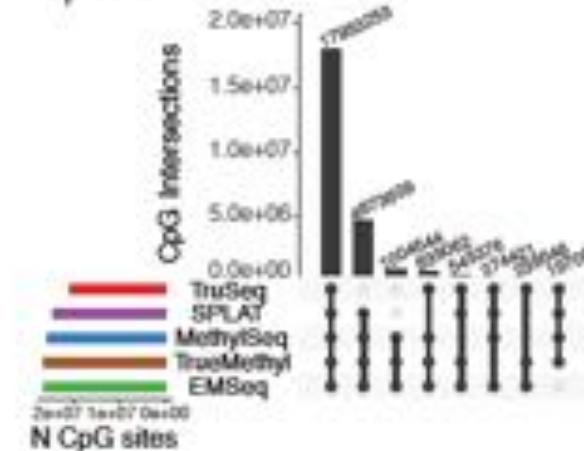
Reproducibility & quality

Average 20x GC coverage:

CpGs $\geq 1x$



CpGs $\geq 10x$

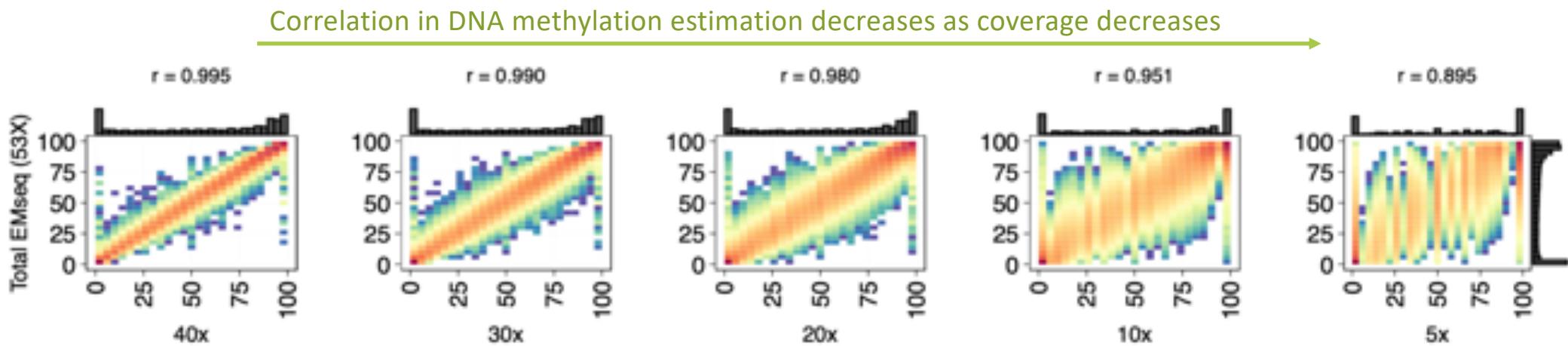


Overall, no major quantitative difference between methylation (beta-values) called after libraries were normalized for nr reads mapped (see next slide).

But they did differ in number of CpG sites detected!



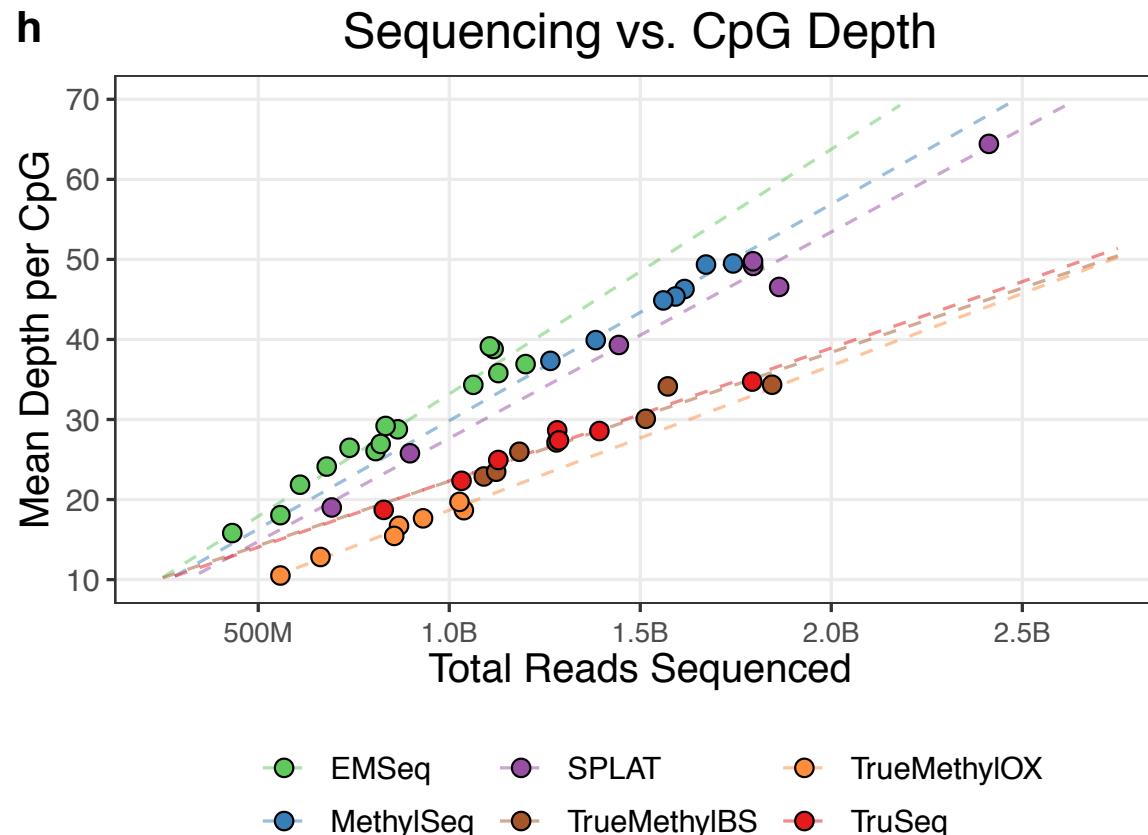
Reproducibility & quality



Foox J, Nordlund J, et al. The SEQC2 Epigenomics Quality Control (EpiQC) Study: Comprehensive Characterization of Epigenetic Methods, Reproducibility, and Quantification. BioRxiv: <https://doi.org/10.1101/2020.12.14.421529>.



Reproducibility & quality

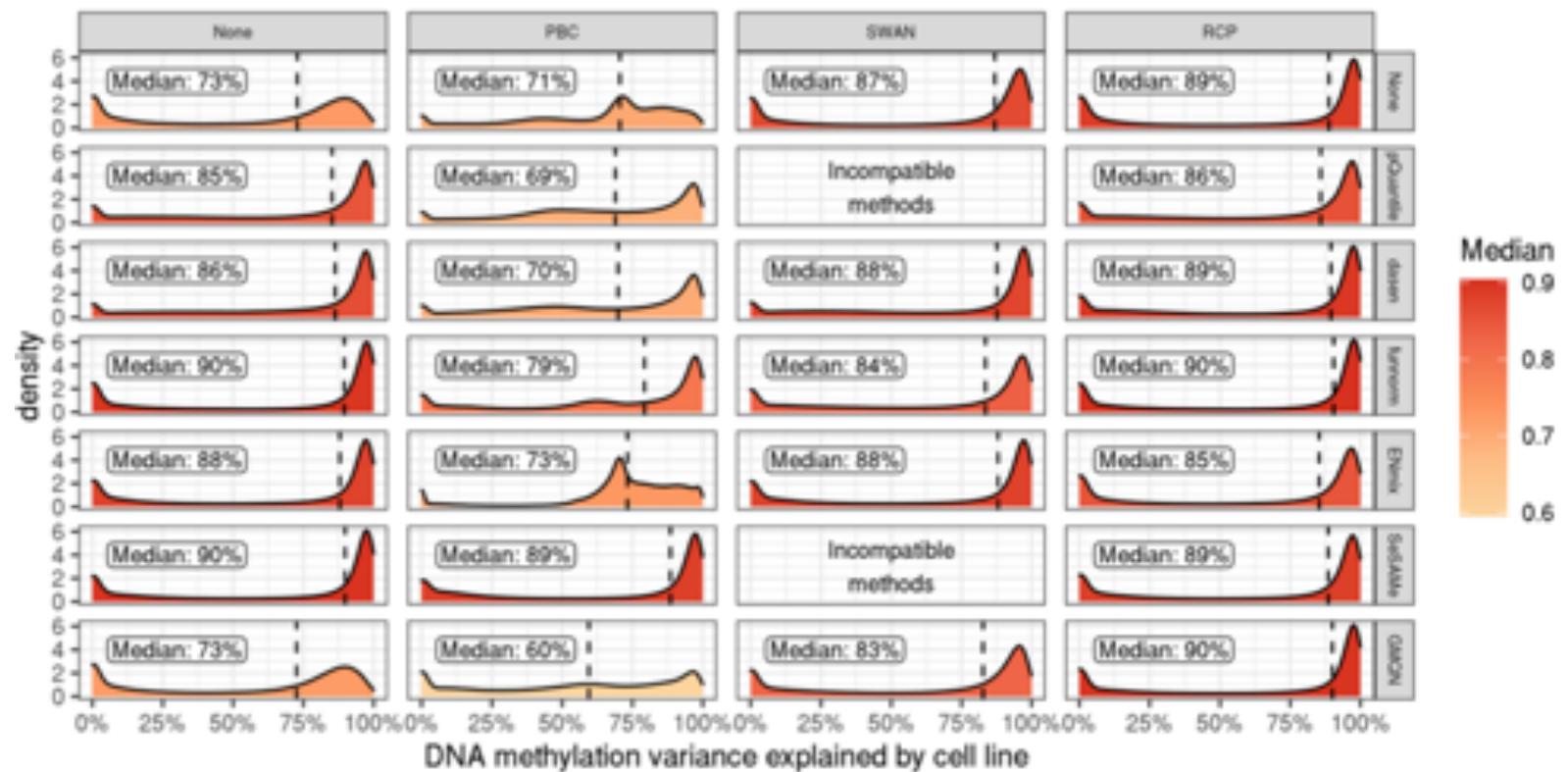


Foxx J, Nordlund J, et al. The SEQC2 Epigenomics Quality Control (EpiQC) Study: Comprehensive Characterization of Epigenetic Methods, Reproducibility, and Quantification. BioRxiv: <https://doi.org/10.1101/2020.12.14.421529>.



Reproducibility & quality

(a) Concordance between microarray replicates across the epigenome, by normalization pipeline

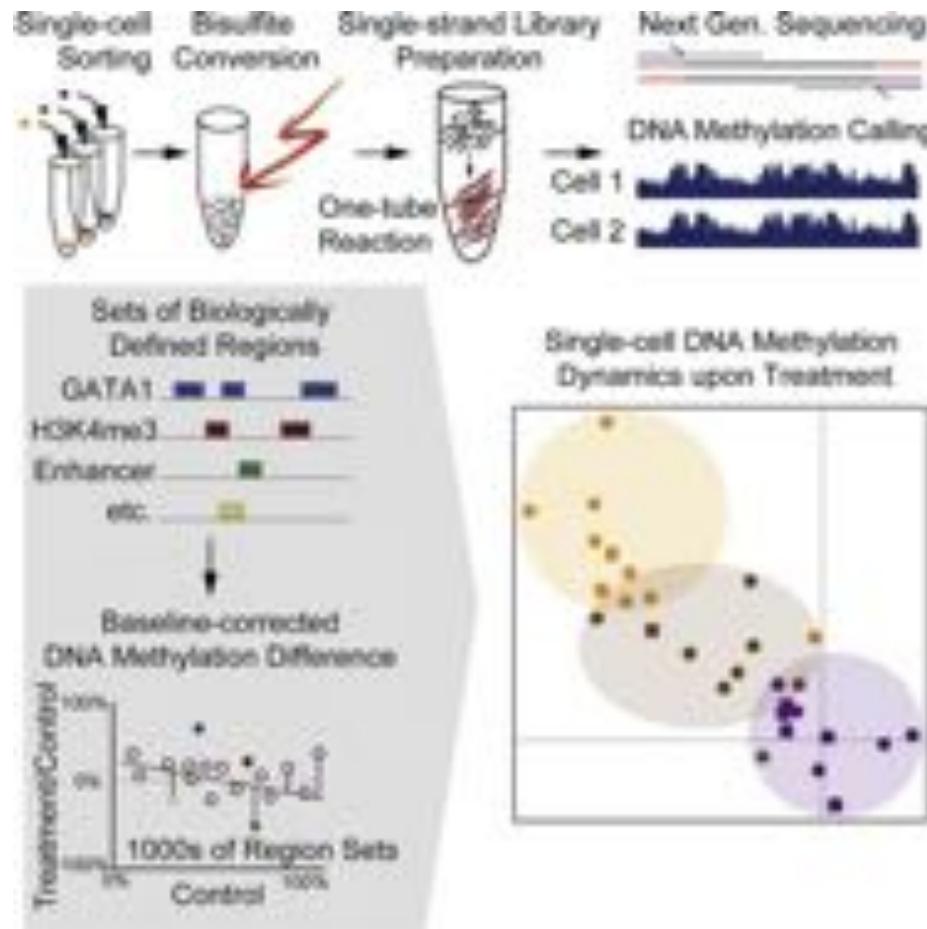


funnorm + RCP
worked best on
these samples

Foox J, Nordlund J, et al. The SEQC2 Epigenomics Quality Control (EpiQC) Study: Comprehensive Characterization of Epigenetic Methods, Reproducibility, and Quantification. *BioRxiv*: <https://doi.org/10.1101/2020.12.14.421529>.



Single-cell WGBS



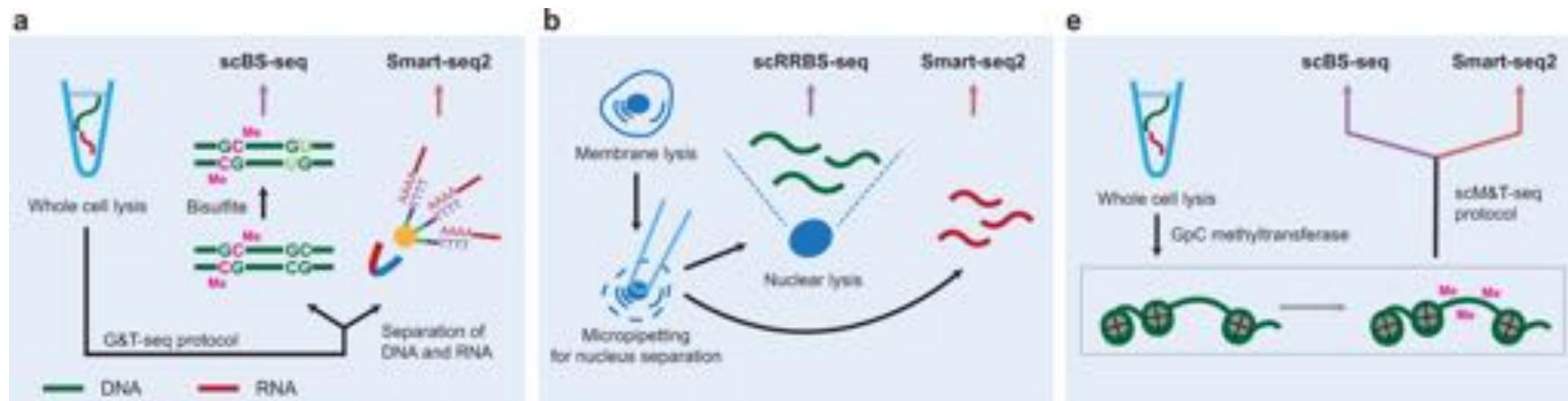
Single cell WGBS

- ✓ Single stranded library prep
- ✓ FACS sorting required (384 plates)
- ✓ Plate- based low throughput (although automation enable throughput of >1000 cells/exp)
- ✓ Expensive
- ✓ Sparse information-At most 50% CpG sites coverage, usually a lot less



From "bulk" to single cells

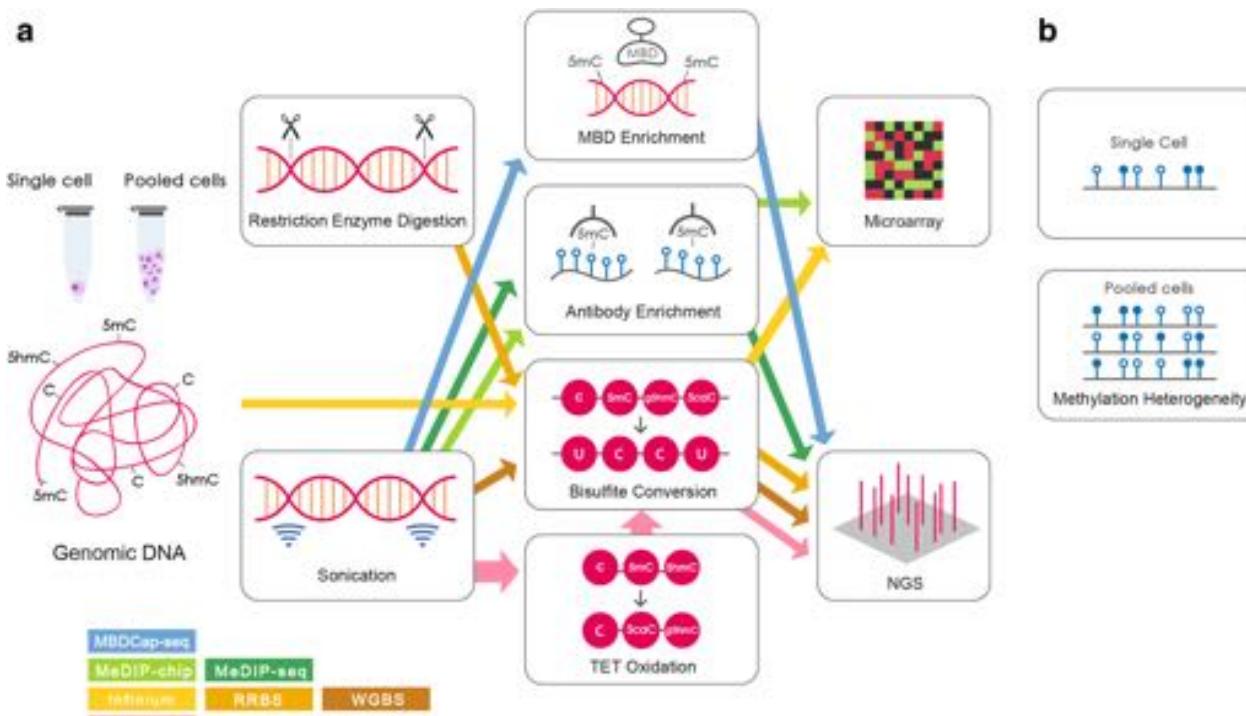
Numerous protocols exist for scWGSB, RRBS, etc – and even integrate transcriptomics in and DNA methylation measurements from the same cell!





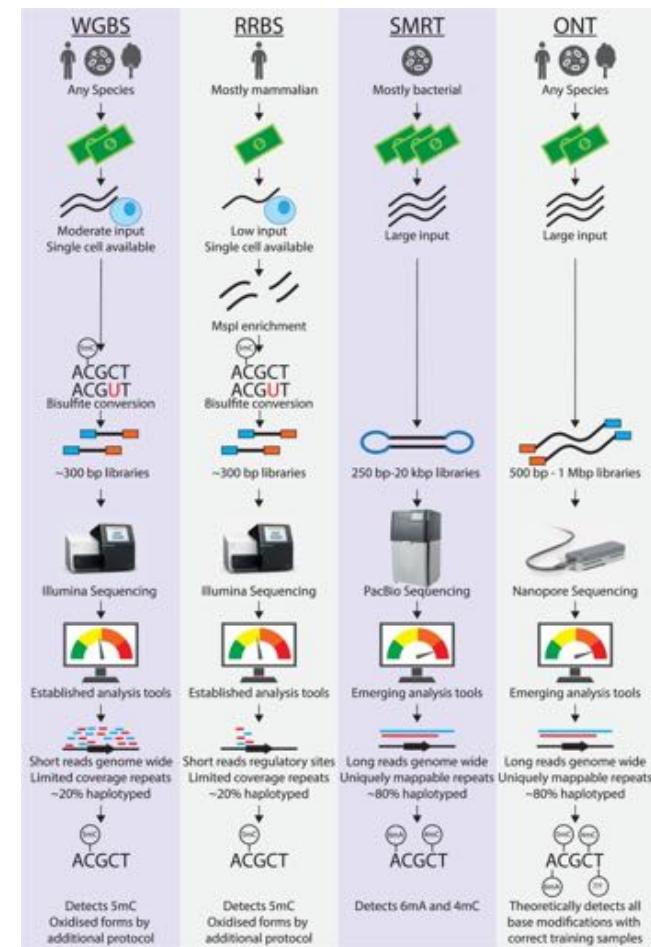
In summary, there are many approaches for studying DNA methylation

a



Yong et al. *Epigenetics & Chromatin* (2016) 9:26
DOI 10.1186/s13072-016-0075-3

b



Gouil and Keniry, *Essays in Biochemistry*, 2019. <https://doi.org/10.1042/EBC20190027>



So which method should I choose?

**nature
biotechnology**

ANALYSIS

Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications

R Alan Harris^{1,*}, Ting Wang², Cristian Coarfa¹, Raman P Nagarajan³, Chibo Hong³, Sara L Downey³, Brett E Johnson³, Shaun D Fouse³, Allen Delaney⁴, Yongjun Zhao⁴, Adam Olsen³, Tracy Ballinger⁵, Xin Zhou⁶, Kevin J Forberg², Junchen Gu², Lorigail Echipare⁶, Henriette O'Geen⁶, Ryan Lister⁷, Mattia Pelizzetti⁷, Yuanxin Xi⁸, Charles B Epstein⁹, Bradley E Bernstein^{9–11}, R David Hawkins¹², Bing Ren^{12,13}, Wen-Yu Chung^{14,15}, Hongcang Gu⁹, Christoph Bock^{9,16–18}, Andreas Gnirke⁹, Michael Q Zhang^{14,15}, David Haussler⁹, Joseph R Ecker⁷, Wei Li³, Peggy J Farnham⁴, Robert A Waterland^{1,19}, Alexander Meissner^{9,16,17}, Marco A Marra⁴, Martin Hirst⁴, Aleksandar Milosavljevic¹ & Joseph F Costello³

- Species
- Sample availability
- DNA quality
- Scientific question(s)
- Budget

Essays in Biochemistry (2019) **63** 639–648
<https://doi.org/10.1042/EBC20190027>

PORTLAND
PRESS

Review Article

Latest techniques to study DNA methylation

Quentin Gouil^{1,2} and Andrew Keniry^{1,2}

¹Epigenetics and Development Division, Walter and Eliza Hall Institute of Medical Research, Parkville, Australia; ²Department of Medical Biology, University of Melbourne, Parkville, Australia

Lee et al. *Experimental & Molecular Medicine* (2020) 52:1428–1442
<https://doi.org/10.1038/s12276-020-0420-2>

Experimental & Molecular Medicine

REVIEW ARTICLE

Open Access

Single-cell multiomics: technologies and data analysis methods

Jeongwoo Lee¹, Do Young Hyeon¹ and Daehee Hwang¹

**nature
genetics**

PERSPECTIVE
<https://doi.org/10.1038/s41588-018-0290-x>

Single-cell and single-molecule epigenomics to uncover genome regulation at unprecedented resolution

Efrat Shema^{1,2*}, Bradley E. Bernstein^{1,2} and Jason D. Buenrostro^{1,2,3*}

Yong et al. *Epigenetics & Chromatin* (2016) 9:26
DOI 10.1186/s13072-016-0075-3

Epigenetics & Chromatin

REVIEW

Open Access

Profiling genome-wide DNA methylation

Wai-Shin Yong^{1†}, Fei-Man Hsu^{2‡} and Pao-Yang Chen^{1*}

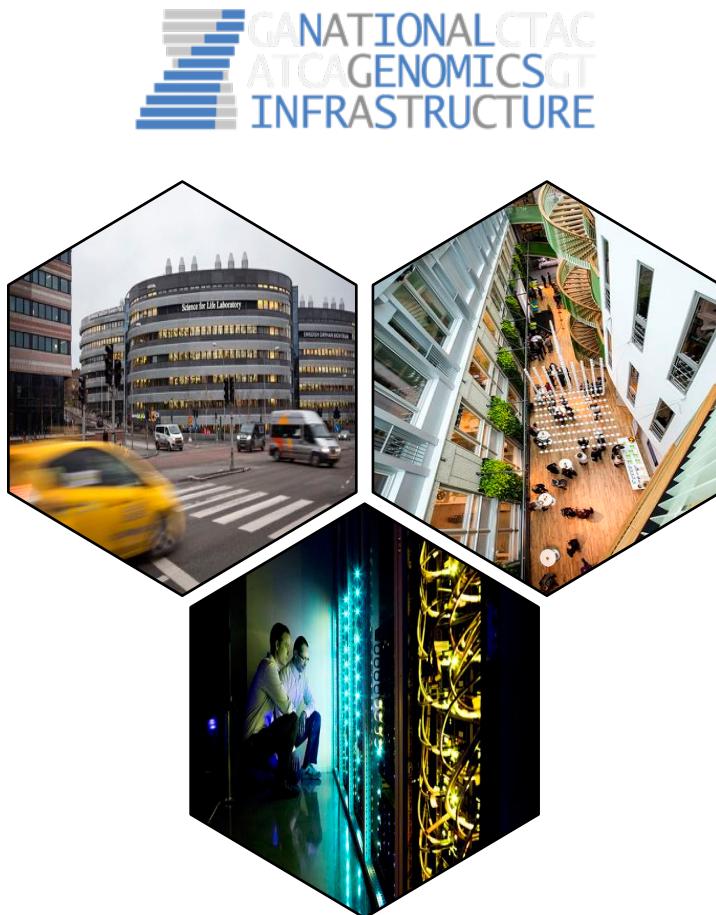


BREAK





Epi^{genomics} services offered by the National Genomics Infrastructure (NGI)



NGI is a facility within the
SciLifeLab Genomics Platform located at two nodes:

NGI-Uppsala

- SNP&SEQ Technology Platform (UU)
- Uppsala Genome Centre (UU)

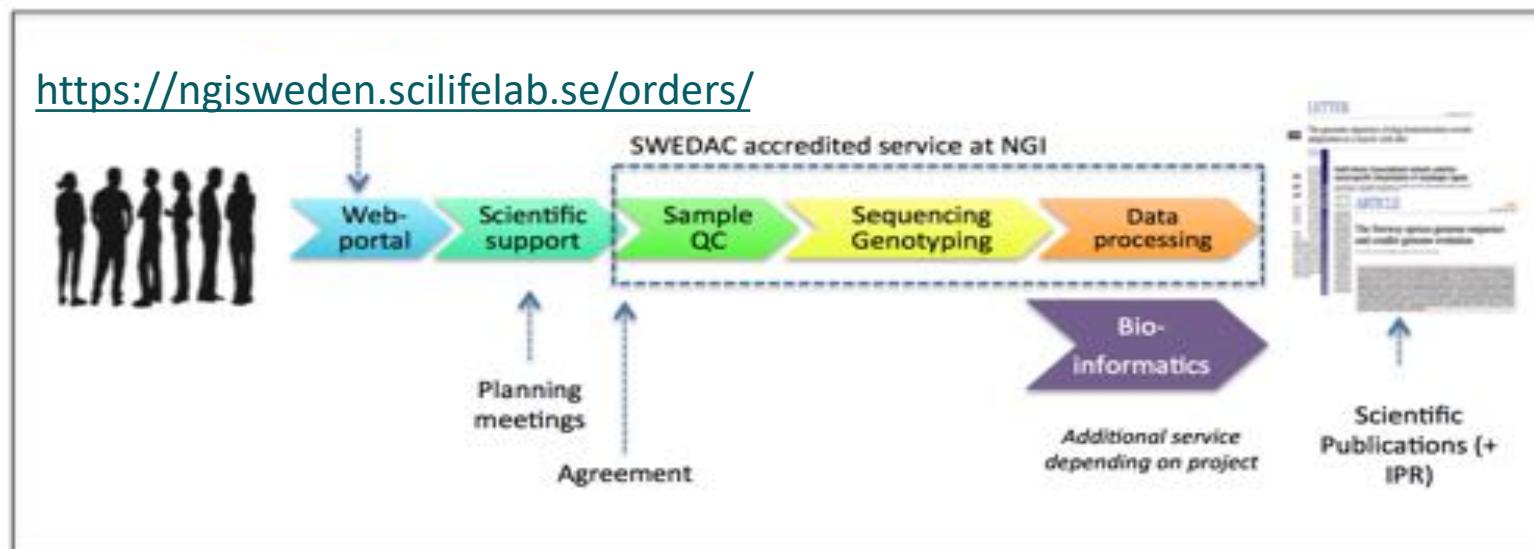
NGI-Stockholm

- SciLifeLab Solna (KTH, KI, SU)



NGI's project portal

- All projects submitted through a **common order system**
- Projects are dynamically allocated between Stockholm/Uppsala depending on type of application, queue situation, or request by researcher

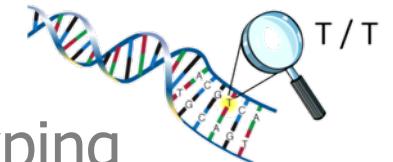




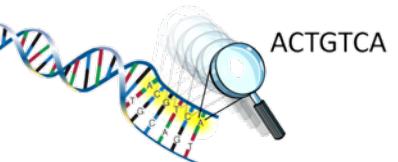
Genotyping and sequencing on all scales



Genotyping



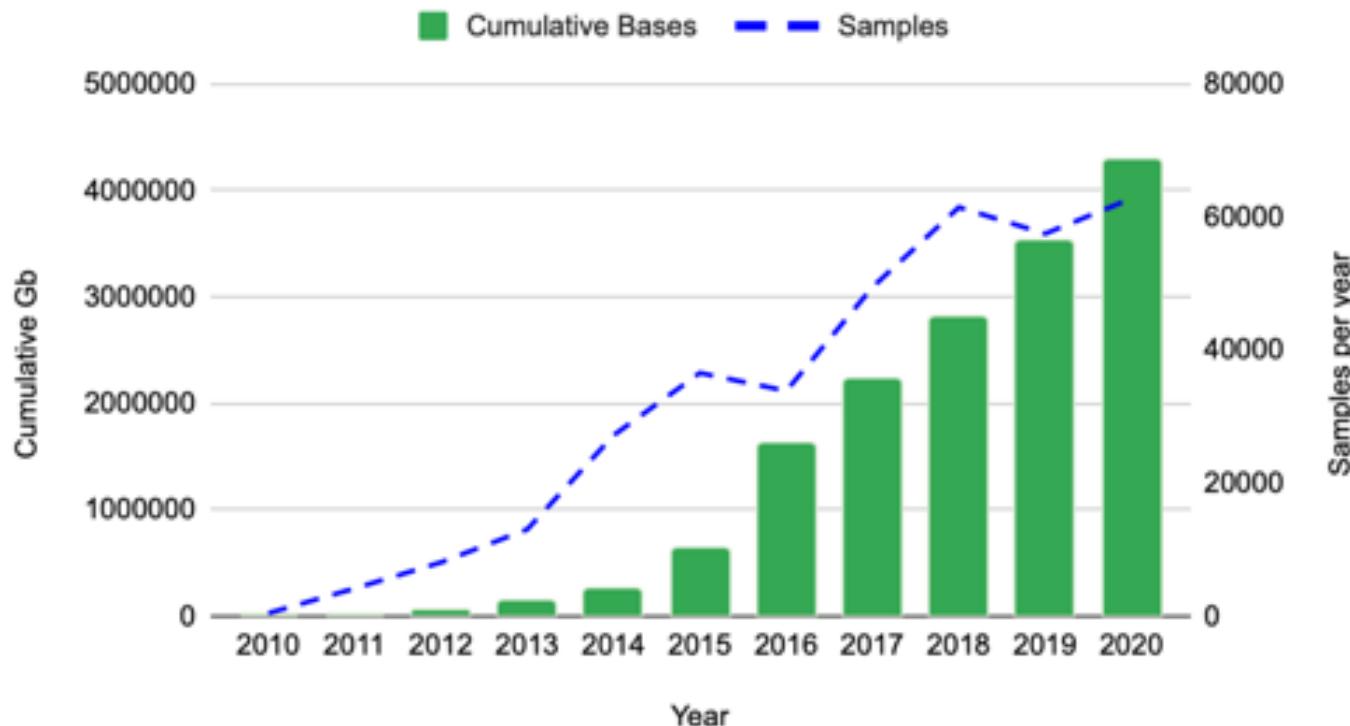
Short-reads



Long-reads



A decade of NGS at NGI



Statistics for 2020:

- 1112 projects with 63,000 samples
- 765 Terabases (10^{12}) of sequence data

As of Jan 1, 2021 NGI has delivered a total of 4.3 Petabases (10^{15}) of sequencing data



Support

Pre support

- **Project design** via discussions with expert project coordinators
- **Advise** in sample collection and/or preparation
- **DNA extraction services available** for specific applications
- **Sample quality (QC)** for all incoming samples and user-made libraries

Post support

- Control over produced data: making sure data meet our **high standards** in terms of quality and yield.
- Open source Bioinformatic pipelines for a wide range of applications: *NF-core lecture*
- Data delivered via **UPPMAX**



Epigenetic methods available at NGI



EPIC Arrays:

500 ng DNA
~2000 SEK/sample

Minimum sample size 15 samples: lower cost per sample for large projects

RRBS:

500 ng DNA
~2000 SEK/sample

**limited availability*

Short-read

Whole genome methylome sequencing with SPLAT (WGBS) or EM-Seq

500 ng DNA
~10,000 SEK/sample at 30x CpG coverage

Long-read

whole genome sequencing (+base modifications)

PacBio Sequell II / Oxford Nanopore PromethION

Cost depends on genome size and epigenetic marks analyzed

Single-cell:

scATAC-seq (10x Genomics): **~25,000** SEK/sample

scWGBS with SPLAT:
contact us for pricing

Epigenetic methods available at NGI



ATAC-seq

>50.000 cells
~**2000** SEK/sample

ChIP-seq, MeDIP-Seq, MBD-seq, etc

You ChIP it,
we'll sequence it!

Contact us for pricing

Non-coding RNAs

miRNA & other short RNAs
ncRNAs
Full-length transcripts

Contact us for pricing

HiC

method for mapping
genome-wide DNA contacts

contact us for pricing

Single-cell

scATAC-seq (10x
Genomics): ~**25,000**
SEK/sample

scWGBS: ***contact us for
pricing***

Epigenetic methods available at NGI



ChIP-seq, MeDIP-Seq, MBD-seq, Target Capture, etc:

You ChIP it,
(or prepare any compatible library)
And we'll sequence it!

Contact us for pricing



Contact information:

Additional information on the types of sequencing applications
that NGI supports:

<https://ngisweden.scilifelab.se>

Don't hesitate to reach out to NGI's project coordinators:
support@ngisweden.se

-or me-

jessica.nordlund@medsci.uu.se / seq@medsci.uu.se