ISTITUTO ITALIANO DI TECNOLOGIA
CENTER FOR GENOMIC SCIENCE

Campus IFOM—IEO

Measuring and decoding DNA methylation landscapes

Mattia Pelizzola - Center for Genomic Science of IIT@SEMM

Outline of the presentation

- Background
- Biological relevance
- How to create, maintain, and remove DNA methylation patterns
- Methods to measure DNA methylation
- Data analysis issues

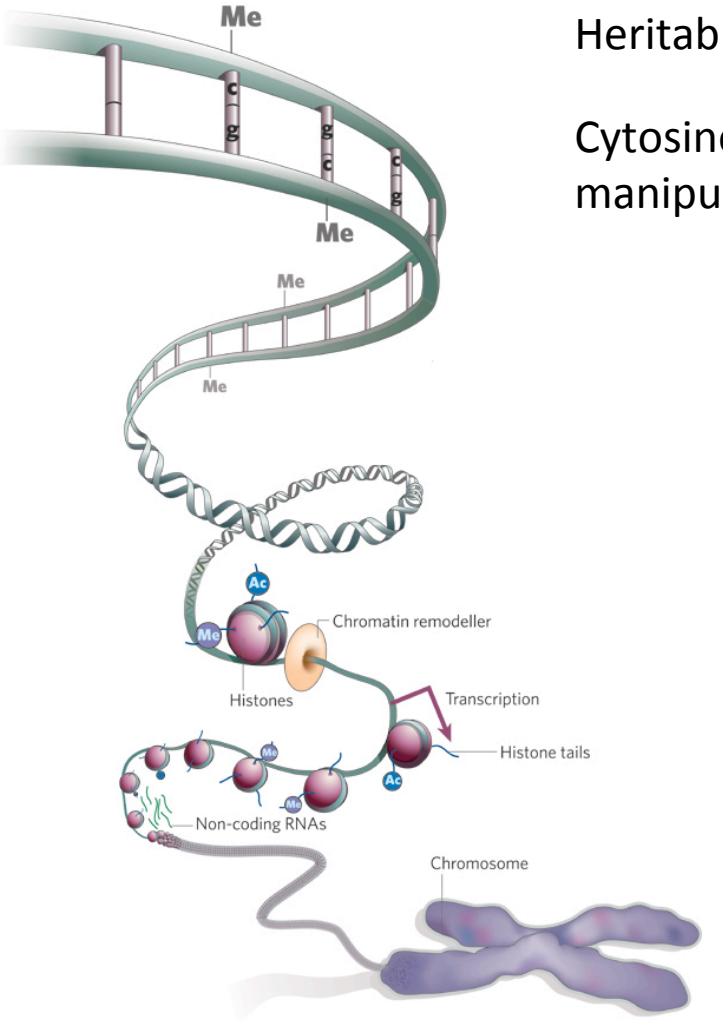
Outline of the presentation

- Background
- Biological relevance
- How to create, maintain, and remove DNA methylation patterns
- Methods to measure DNA methylation
- Data analysis issues

Epigenetics

".. **epigenetics** is the study of heritable changes in gene expression or cellular phenotype caused by mechanisms other than changes in the underlying DNA sequence – hence the name epi- (Greek: επί- over, above, outer) -genetics. It refers to functionally relevant modifications to the genome that do not involve a change in the nucleotide sequence."
Wikipedia (en)

Eukaryotic epigenetics and DNA methylation



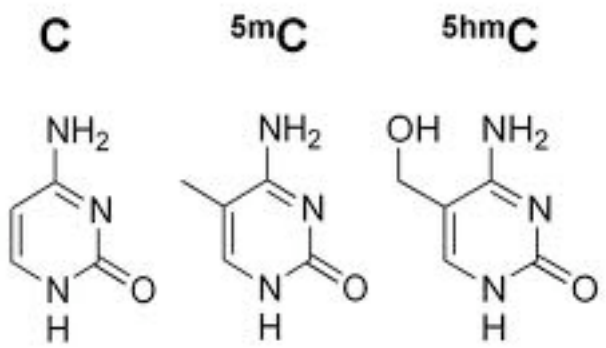
Heritable layer of regulation superimposed on genome

Cytosine DNA methylation (mC) and histone modifications can manipulate the readout of the underlying genetic information.

- Tissue-specific gene regulation
- Cell differentiation
- DNA methylation required for self-renewal and maintenance of pluripotent state
- Transposon silencing
- Modulation of binding of protein to DNA
- responsive to environment / diet
- varying with age
- Tumorigenesis

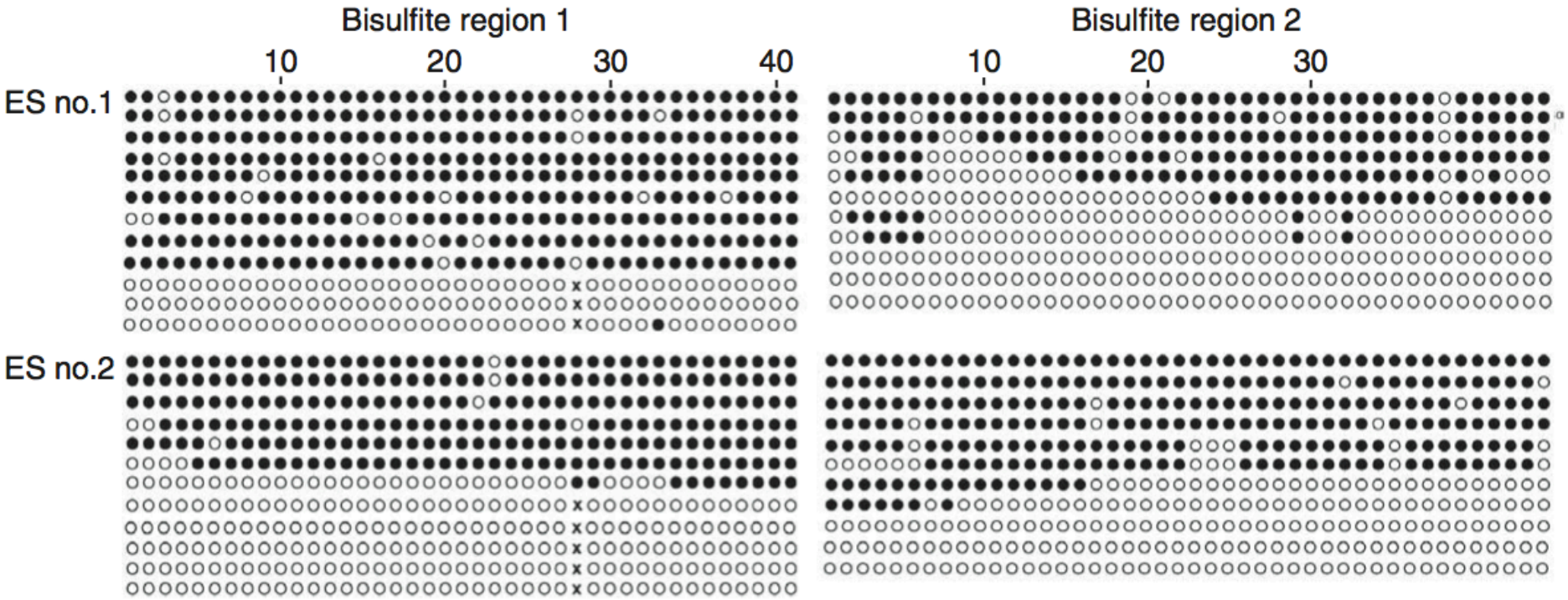
Nature 454, 711-715 (Aug 2008)

DNA methylation

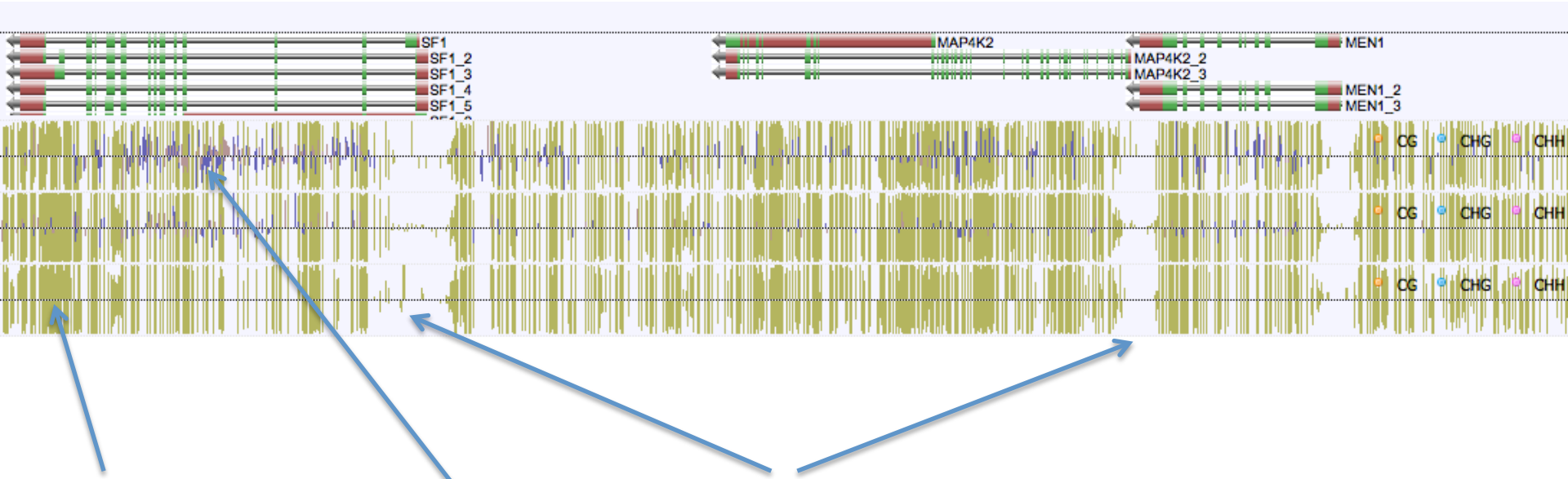


- Only C in specific sequence context(s) can be methylated
- Can be strand specific
- Heterogeneous in cell populations
- Dynamic

DNA methylation heterogeneity in cell populations



DNA methylation: how the data look like



mC in the CpG
Sequence context,
~4e7s mCpG in human

Hypomethylated promoter regions

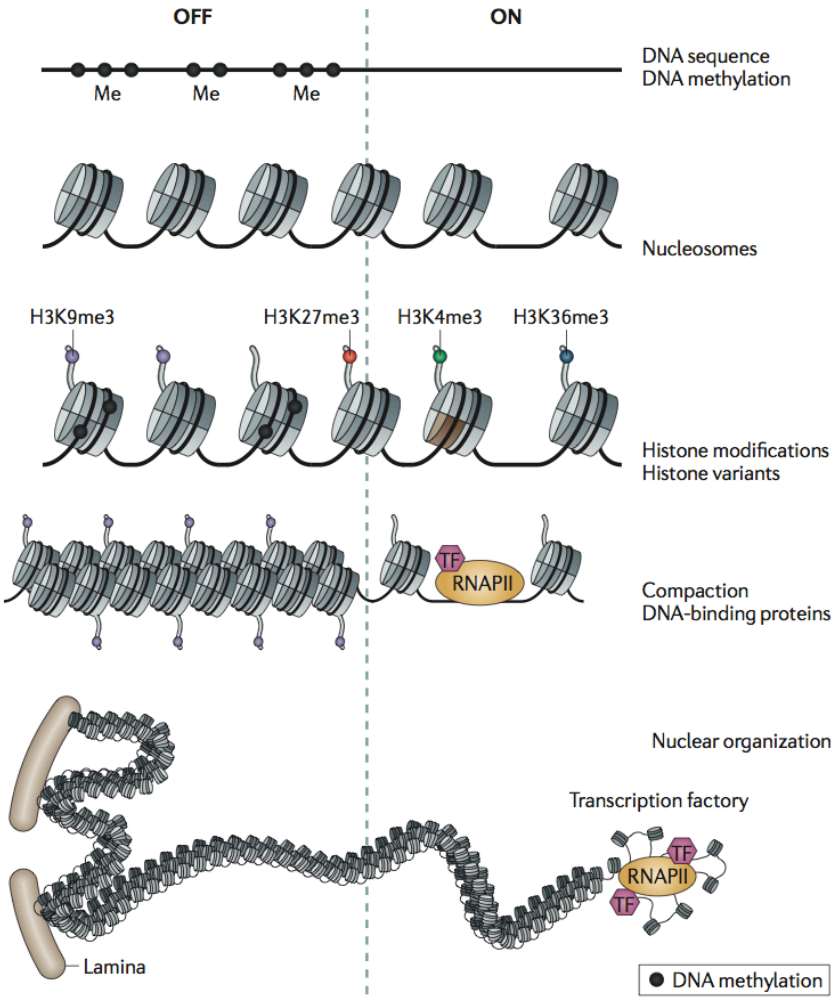
mC in the nonCpG
Sequence context (CHG, CHH)
~1e7 mCpGs in pluripotent human cells

http://neomorph.salk.edu/human_methylome/browser.html

Outline of the presentation

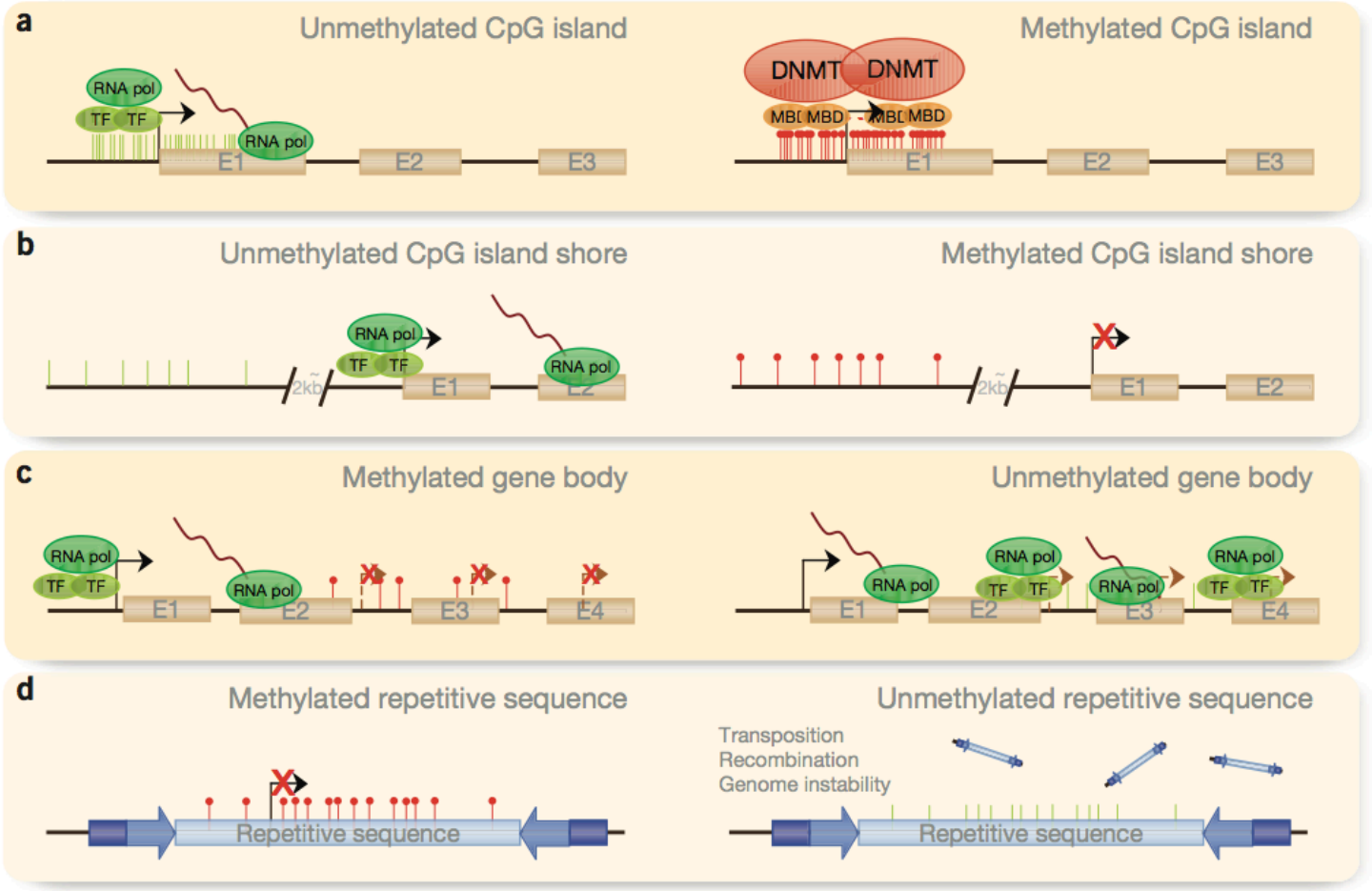
- Background
- Biological relevance
- How to create, maintain, and remove DNA methylation patterns
- Methods to measure DNA methylation
- Data analysis issues

Layers of chromatin organization



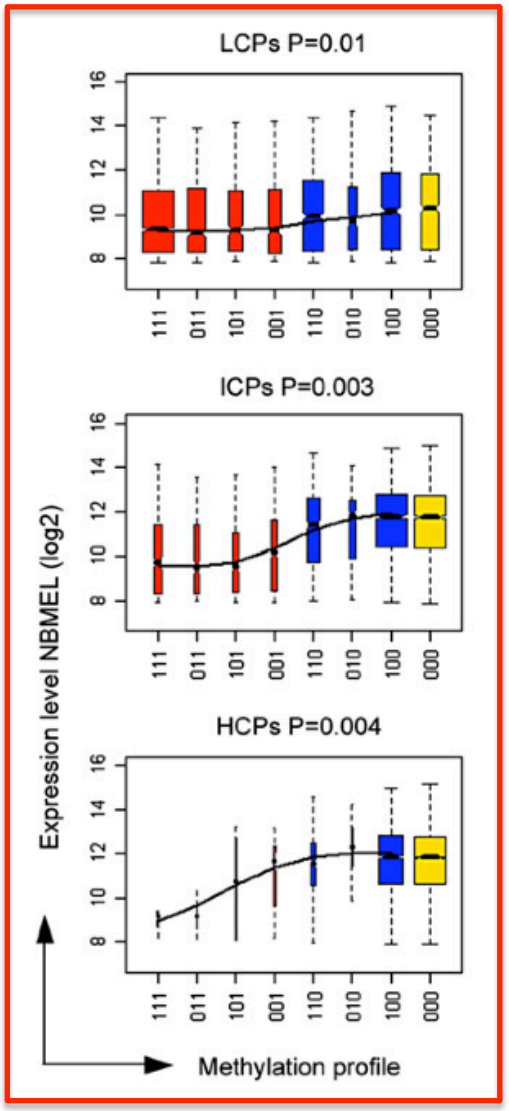
Zhou VW *et al*, Nature Review Genetics 2010

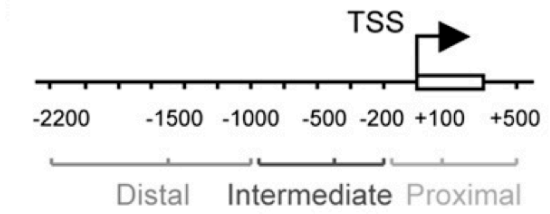
Relevance of DNA methylation



Portela A and Esteller M, Nature Biotechnology 2010

Relevance of DNA methylation



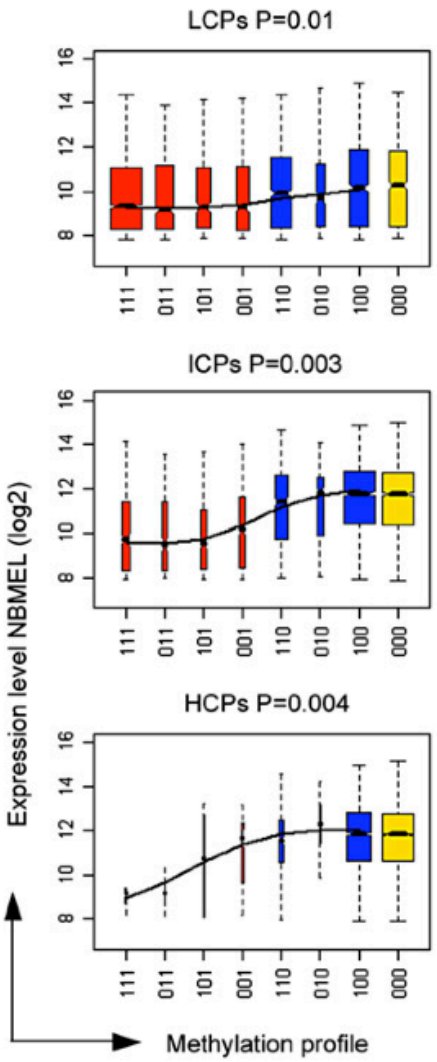


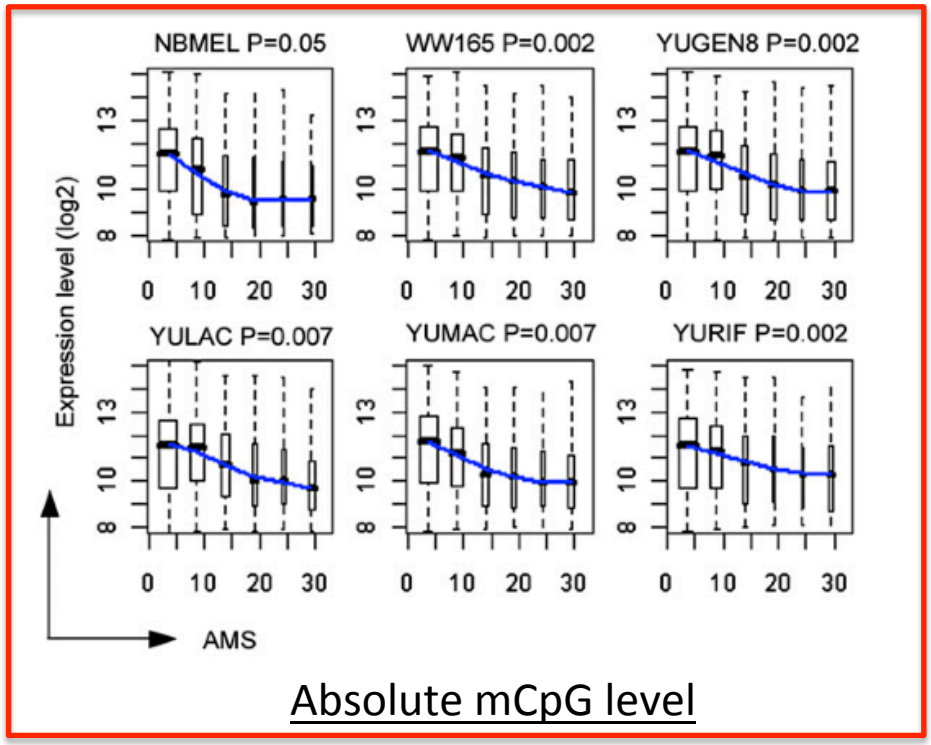
If average *RMS* < 0.5 → 0 (unmethylated)
If average *RMS* >= 0.5 → 1 (methylated)

Distance from TSS

Koga Y *et al*, Genome Research 2009

Relevance of DNA methylation

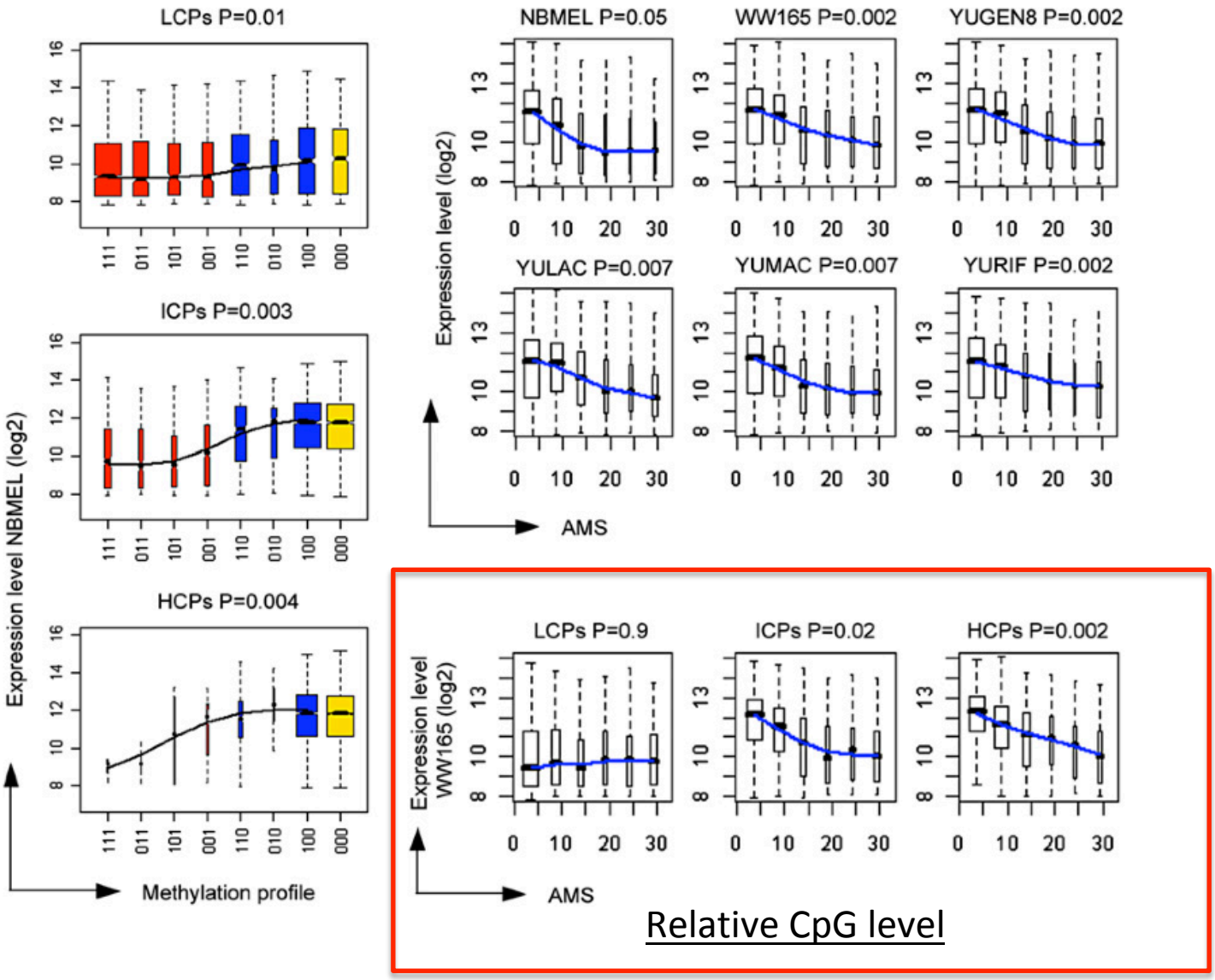




Absolute mCpG level

Koga Y *et al*, Genome Research 2009

Relevance of DNA methylation



Koga Y *et al*, Genome Research 2009

Outline of the presentation

- Background
- Biological relevance
- How to create, maintain, and remove DNA methylation patterns
- Methods to measure DNA methylation
- Data analysis issues

Establishing and maintening DNA methylation patterns

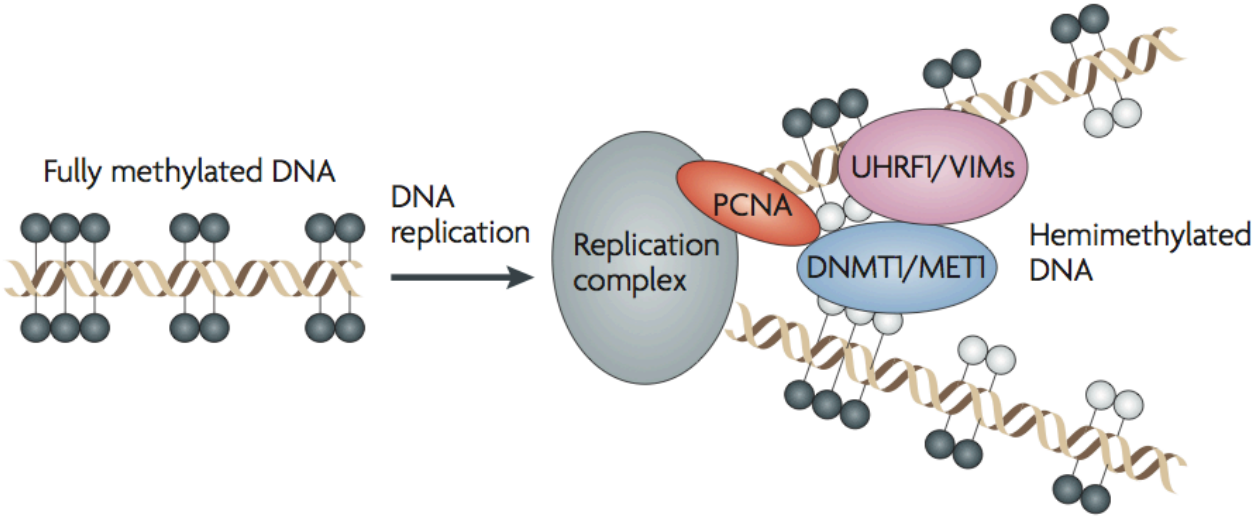
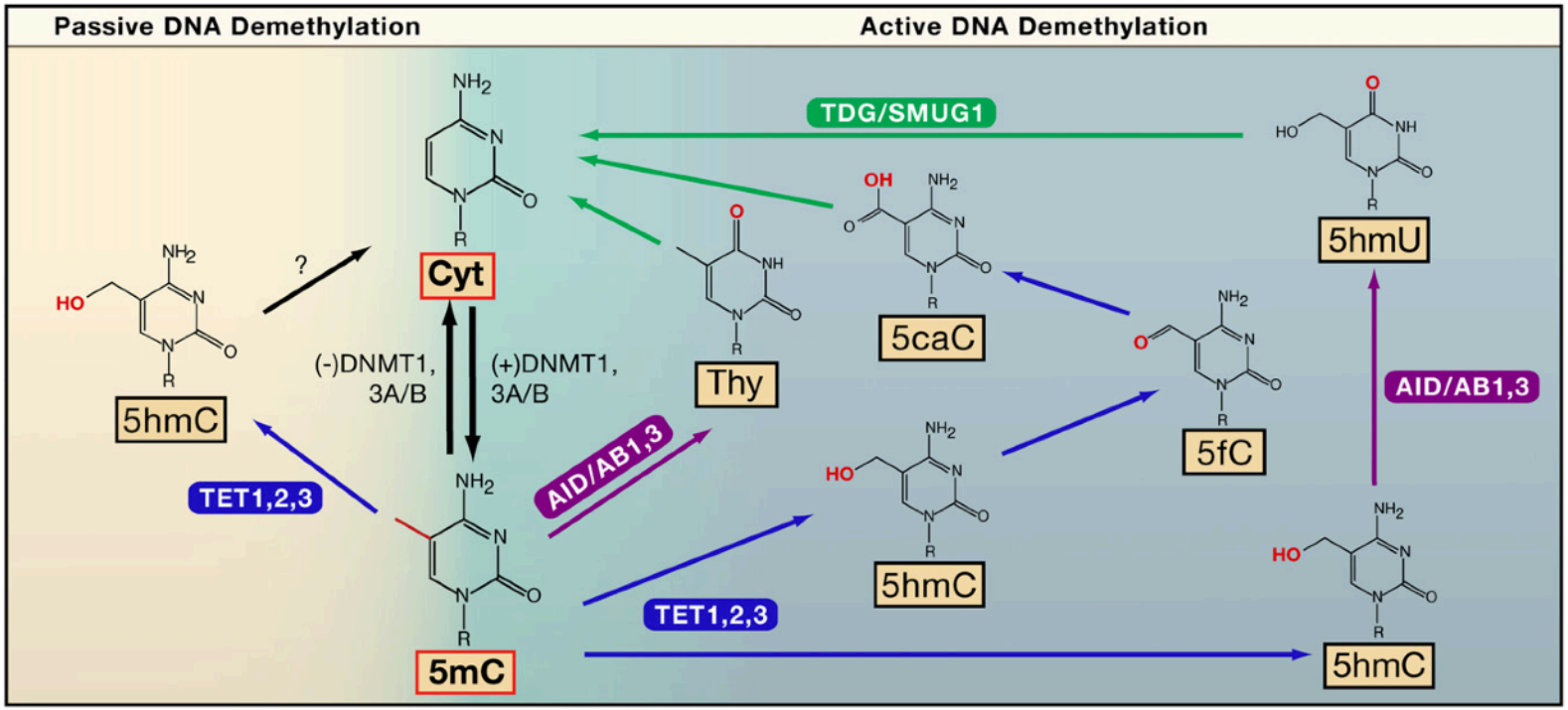


Figure 6 | **Maintenance of DNA methylation in plants and mammals. a** | Model depicting the maintenance of CG methylation during replication. DNA methyltransferase 1 (DNMT1) is proposed to be recruited to replication foci through interactions with ubiquitin-like plant homeodomain and RING finger domain 1 (UHRF1) — a SET- or RING-associated (SRA) domain protein that specifically interacts with hemimethylated DNA — and with proliferating cell nuclear antigen (PCNA). After being recruited, DNMT1 functions to maintain methylation patterns by restoring the hemimethylated DNA to a fully methylated state. In plants, DNA METHYLTRANSFERASE 1

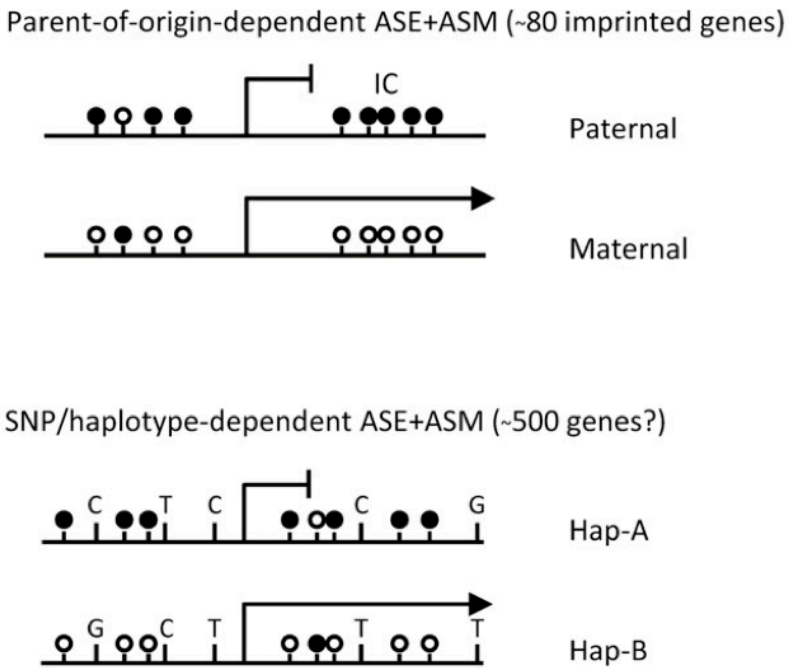
Law JA and Jacobsen SE, Nature Review Genetics 2010

DNA de-methylation



Buthani N *et al*, Cell 2011

Imprinting and allele-specific DNA methylation



In genomic imprinting, the ASM is established in gametogenesis and dictated by the parental origin of the allele, with weak or absent effects of local haplotypes. Some imprinted genes show hypermethylation on the paternal allele as shown here, whereas others show hypermethylation of the maternal allele. In successive generations, the imprint is erased and then reset appropriately in gametogenesis, according to the sex of the transmitting parent. Thus genomic imprinting is non-Mendelian. In contrast, SNP- or haplotype-dependent ASM is dictated in cis by the local DNA sequence, regardless of parent of origin. This type of ASM is transmitted in a Mendelian fashion, and its presence is an indication of nearby regulatory SNPs that function, by mechanisms still largely unknown, to confer the allelic asymmetry. Although the number of imprinted genes is reasonably well established, the number of genes with nonimprinted, sequence-dependent ASM is influenced both by tissue type and by the stringency of the cutoffs utilized for scoring the allelic asymmetry. Black circles indicate methylated CpG dinucleotides; white circles, unmethylated CpGs. IC denotes imprinting center.

Tycko B, The American Journal of Human Genetics 2010

Outline of the presentation

- Background
- Biological relevance
- How to create, maintain, and remove DNA methylation patterns
- Methods to measure DNA methylation
- Data analysis issues

How to measure DNA methylation

Table 1 | **Main principles of DNA methylation analysis**

Pretreatment	Analytical step			
	Locus-specific analysis	**Gel-based analysis**	**Array-based analysis**	**NGS-based analysis**
Enzyme digestion	• *Hpa*II-PCR	• Southern blot • RLGS • MS-AP-PCR • AIMS	• DMH • MCAM • HELP • MethylScope • CHARM • MMASS	• Methyl–seq • MCA–seq • HELP–seq • MSCC
Affinity enrichment	• MeDIP-PCR		• MeDIP • mDIP • mCIP • MIRA	• MeDIP–seq • MIRA–seq
Sodium bisulphite	• MethyLight • EpiTYPER • Pyrosequencing	• Sanger BS • MSP • MS-SNuPE • COBRA	• BiMP • GoldenGate • Infinium	• RRBS • BC–seq • BSPP • WGSBS

AIMS, amplification of inter-methylated sites; BC–seq, bisulphite conversion followed by capture and sequencing; BiMP, bisulphite methylation profiling; BS, bisulphite sequencing; BSPP, bisulphite padlock probes; CHARM, comprehensive high-throughput arrays for relative methylation; COBRA, combined bisulphite restriction analysis; DMH, differential methylation hybridization; HELP, *Hpa*II tiny fragment enrichment by ligation-mediated PCR; MCA, methylated CpG island amplification; MCAM, MCA with microarray hybridization; MeDIP, mDIP and mCIP, methylated DNA immunoprecipitation; MIRA, methylated CpG island recovery assay; MMASS, microarray-based methylation assessment of single samples; MS-AP-PCR, methylation-sensitive arbitrarily primed PCR; MSCC, methylation-sensitive cut counting; MSP, methylation-specific PCR; MS-SNuPE, methylation-sensitive single nucleotide primer extension; NGS, next-generation sequencing; RLGS, restriction landmark genome scanning; RRBS, reduced representation bisulphite sequencing; –seq, followed by sequencing; WGSBS, whole-genome shotgun bisulphite sequencing.

Laird PW, Nature Review Genetics 2010

Methods to profile genome-wide DNA methylation

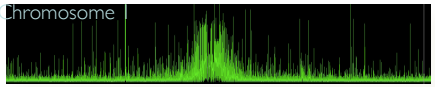
2006 Microarrays

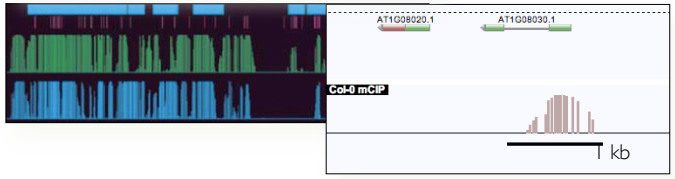
Methylcytosine IP (meDIP)

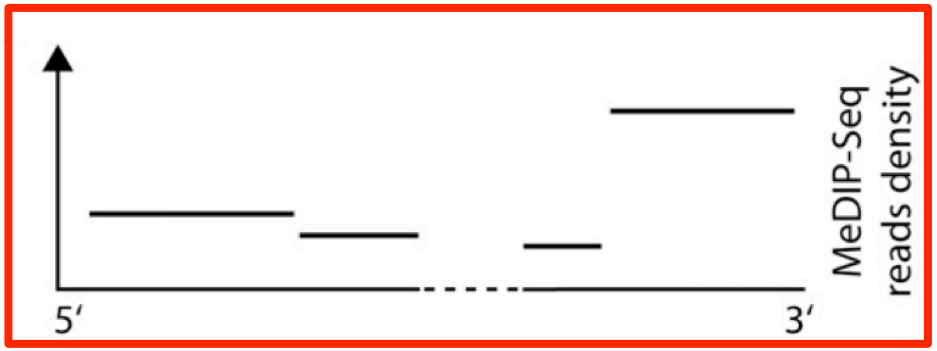
Methyl-sensitive restriction digest

Whole genome amplification

Hybridization to tiling array

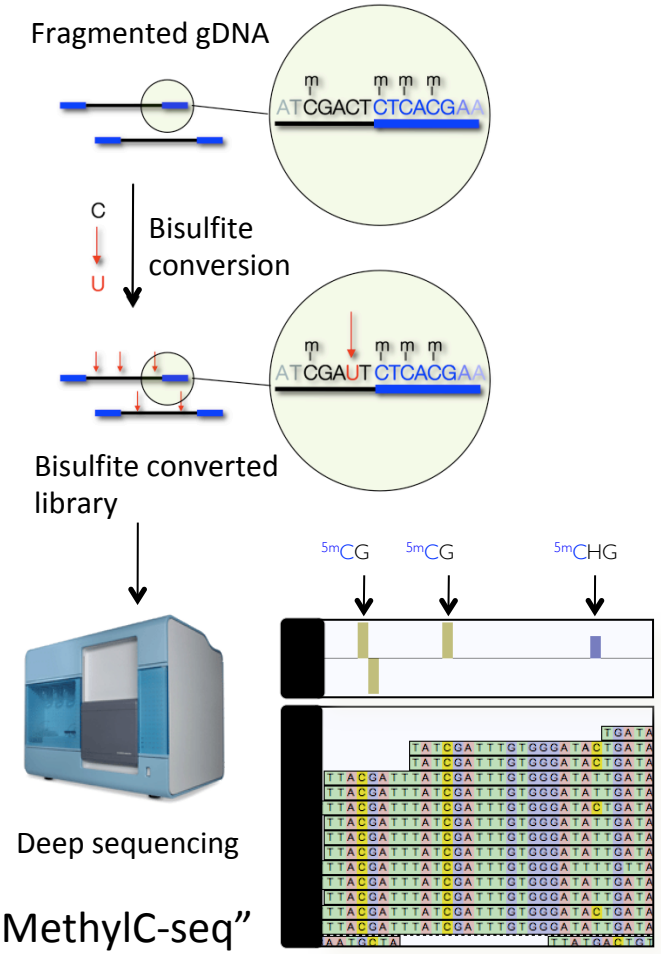




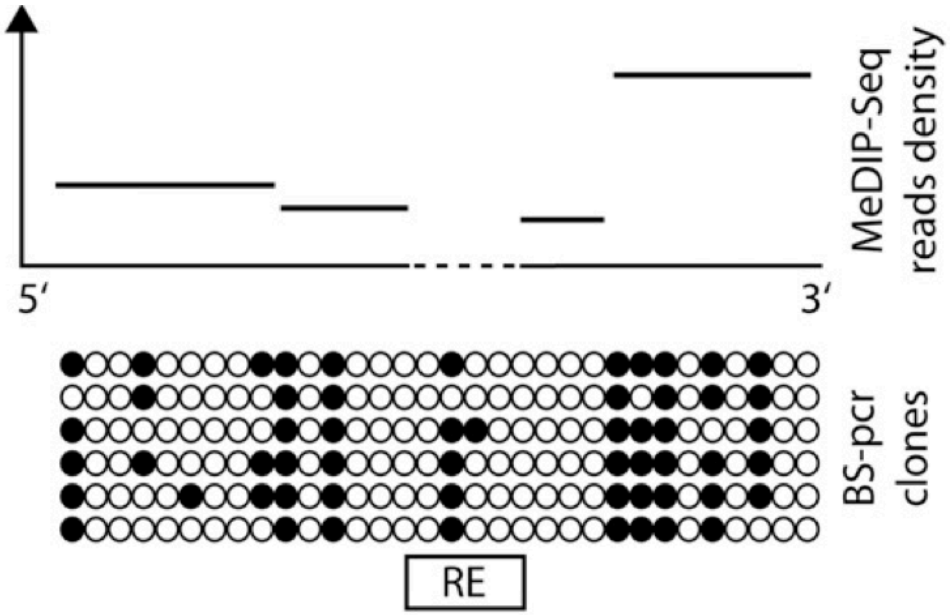


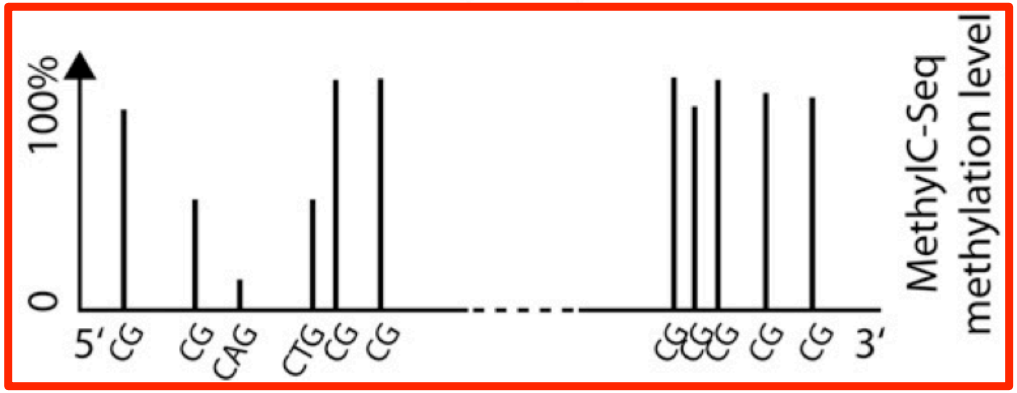
Methods to profile genome-wide DNA methylation

2007-9 Bisulfite sequencing



"MethylC-seq"





How to measure DNA methylation

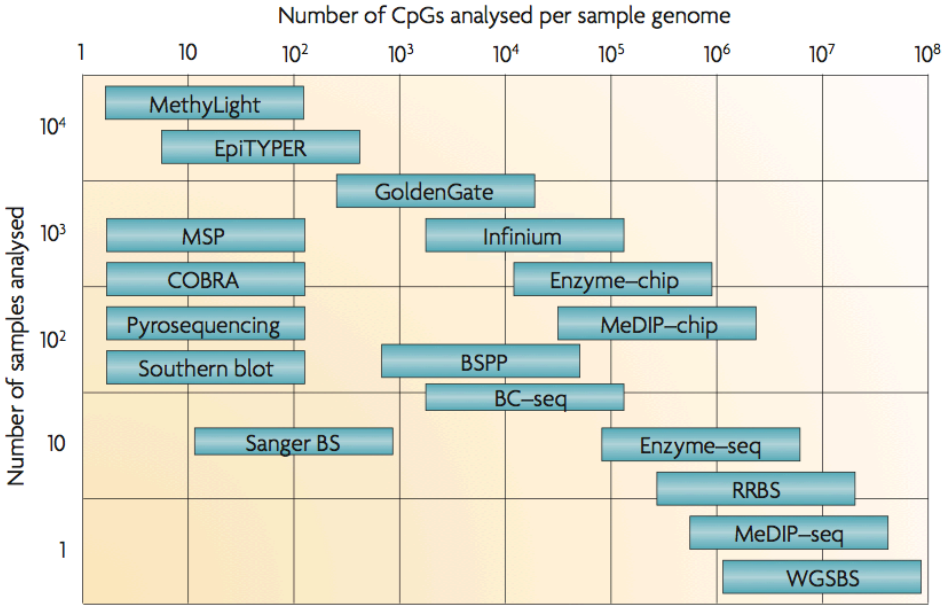
Table 2 | **Features and sources of bias for various techniques**

Technology	Features							Potential sources of bias							
	Unambiguous identification of CpG measured	In *cis* co-methylation information	Non-CpG methylation information	Allele-specific measurement capability	Good coverage of regions with low CpG density	Compatible with low amounts of input DNA	Full repeat-masked genome coverage	Copy-number variation bias	Fragment size bias	Incomplete bisulphite conversion bias	Bisulphite PCR bias	Cross-hybridization bias	DNA methylation status bias	GC content bias	CpG density bias
Infinium	(•)					•				•	•	•			
Enzyme–chip	(•)	(•)			(•)				•			•		•	
MeDIP–chip							•	•				•		•	•
BSPP	•	•	•	•						•	•		•		
BC–seq	•	•	•	•						•	•		•		
RRBS	•	•	•	•		•				•	•				
Enzyme–seq	•	•		•	(•)	•			•						
MeDIP–seq				•			•	•						•	•
WGSBS	•	•	•	•	•	•	•			•	•				

'•' indicates that the method has this feature or potentially has this bias; '(•)' indicates that the method has this feature to a limited extent or in some circumstances. BC–seq, bisulphite conversion followed by capture and sequencing; BSPP, bisulphite padlock probes; –chip, followed by microarray; MeDIP, methylated DNA immunoprecipitation; RRBS, reduced representation bisulphite sequencing; –seq, followed by sequencing; WGSBS, whole-genome shotgun bisulphite sequencing.

Laird PW, Nature Review Genetics 2010

Sample throughput versus genome coverage



Laird PW, Nature Review Genetics 2010

Table 1 Critical parameters in sequencing-based DNA methylation profiling

Method	H1 DNA sample no.	Total bases generated (Gbp)	Total high quality bases (Gbp)	Total bases in map (Gbp)	Maximum resolution (bp)	1-read coverage of CpGs in repeats (no.,%)	Percentage of assayed CpGs in repeats (%)
MethylC-seq	no. 3	172.49	115	87.5	1	13,303,415 (91.8)	49.7
RRBS	no. 3	1.58	1.43	1.28	1	1,646,649 (11.4)	47.5
MeDIP-seq	no. 1	3.42	2.07	1.95	150	10,004,670 (68.3)	52.9
MeDIP-seq	no. 2	3.02	1.84	1.73	150	10,101,868 (68.9)	53.2
MeDIP-seq	nos.1 + 2	6.44	3.91	3.68	150	11,693,059 (79.8)	53.5
MBD-seq	no. 2	5.67	3.71	2.21	150	10,080,007 (68.8)	59.1
MRE-seq	no. 1	3.61	1.31	0.96	1	306,635 (2.07)	21.7
MRE-seq	no. 2	4.03	1.69	1.3	1	232,885 (1.59)	18.6

Harris RA, Nature Biotechnology 2010

Outline of the presentation

- Background
- Biological relevance
- How to create, maintain, and remove DNA methylation patterns
- Methods to measure DNA methylation
- Data analysis issues

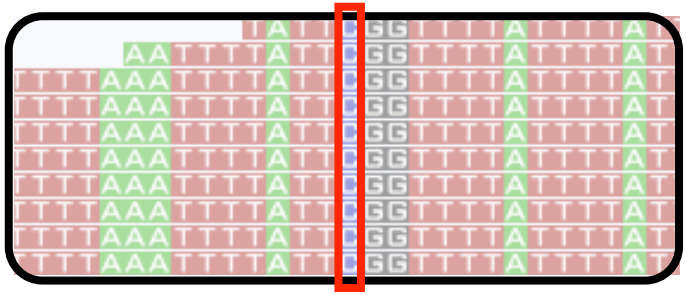
Data analysis pipeline

Illumina NGS - SE 87 nt

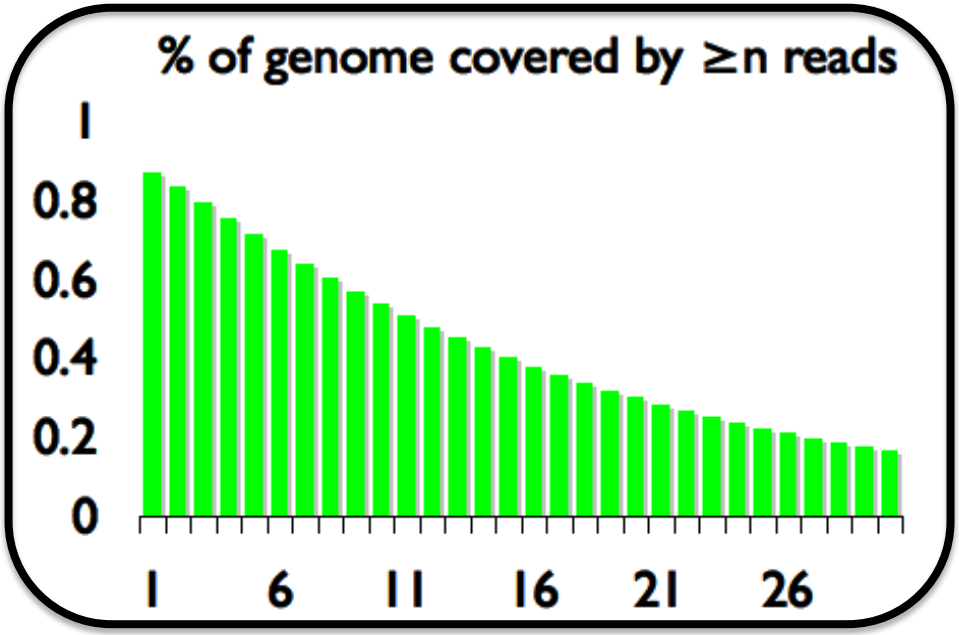
Map reads against computationally BS converted + / - strands

Post-processing & stack

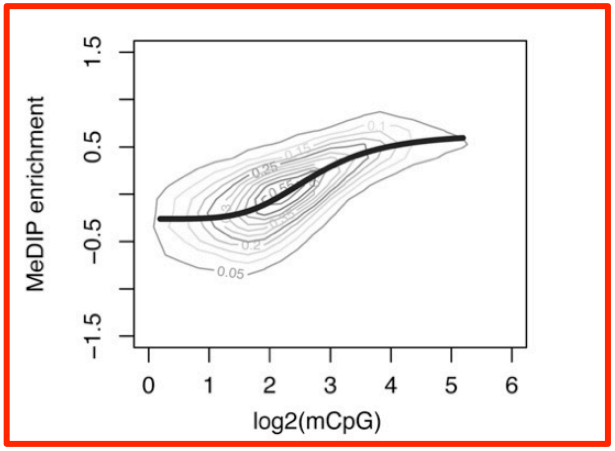
Call mC at 1% FDR with binomial test



mC MySQL DB



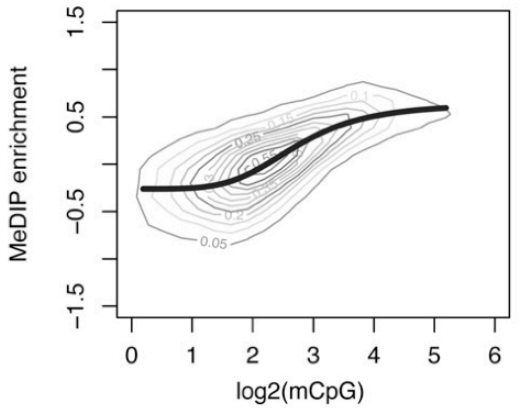
Summary of pitfalls in the analysis of DNA methylation

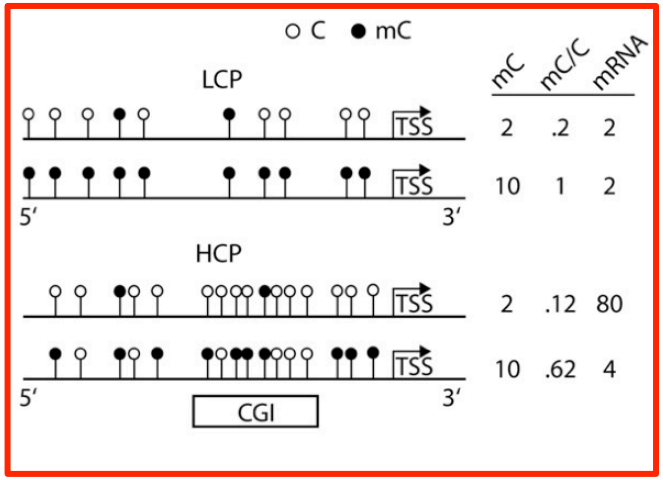


Pay attention to what you are measuring

Pelizzola M and Ecker JR, FEBS Letters 2010

Summary of pitfalls in the analysis of DNA methylation

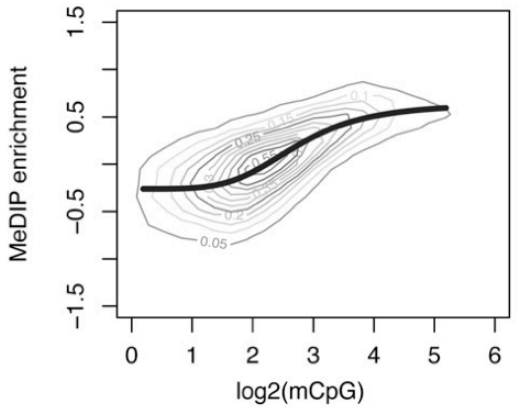


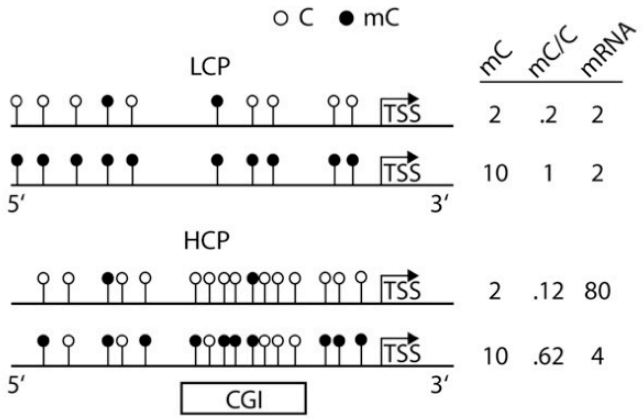


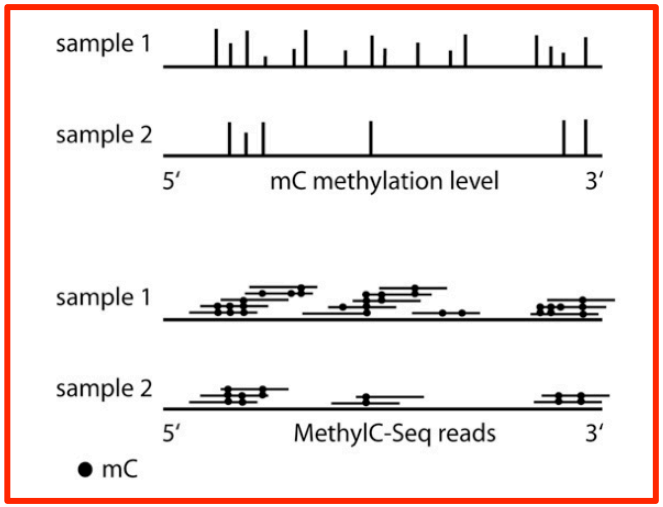
Put it in the right genomic context

Pelizzola M and Ecker JR, FEBS Letters 2010

Summary of pitfalls in the analysis of DNA methylation



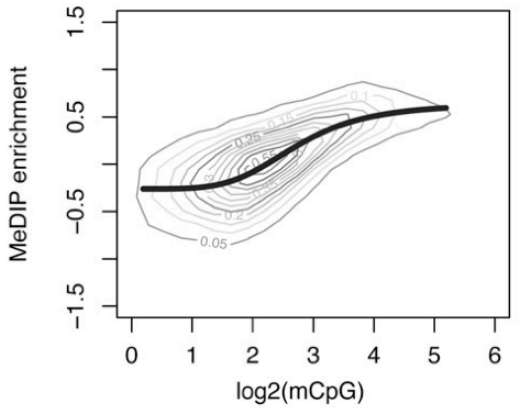


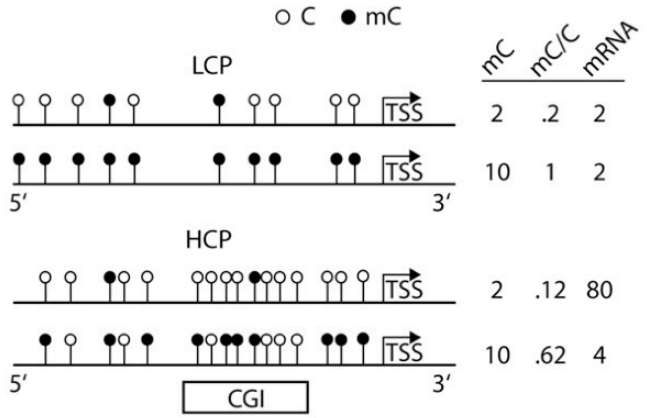


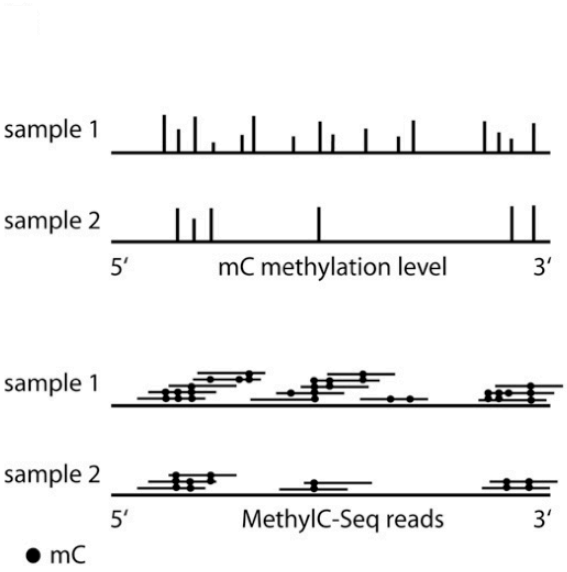
Watch out of coverage issues

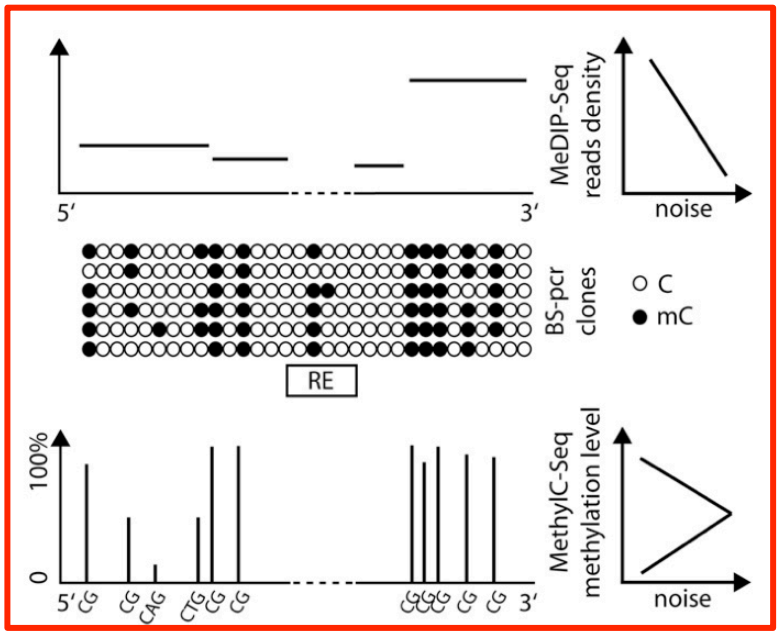
Pelizzola M and Ecker JR, FEBS Letters 2010

Summary of pitfalls in the analysis of DNA methylation





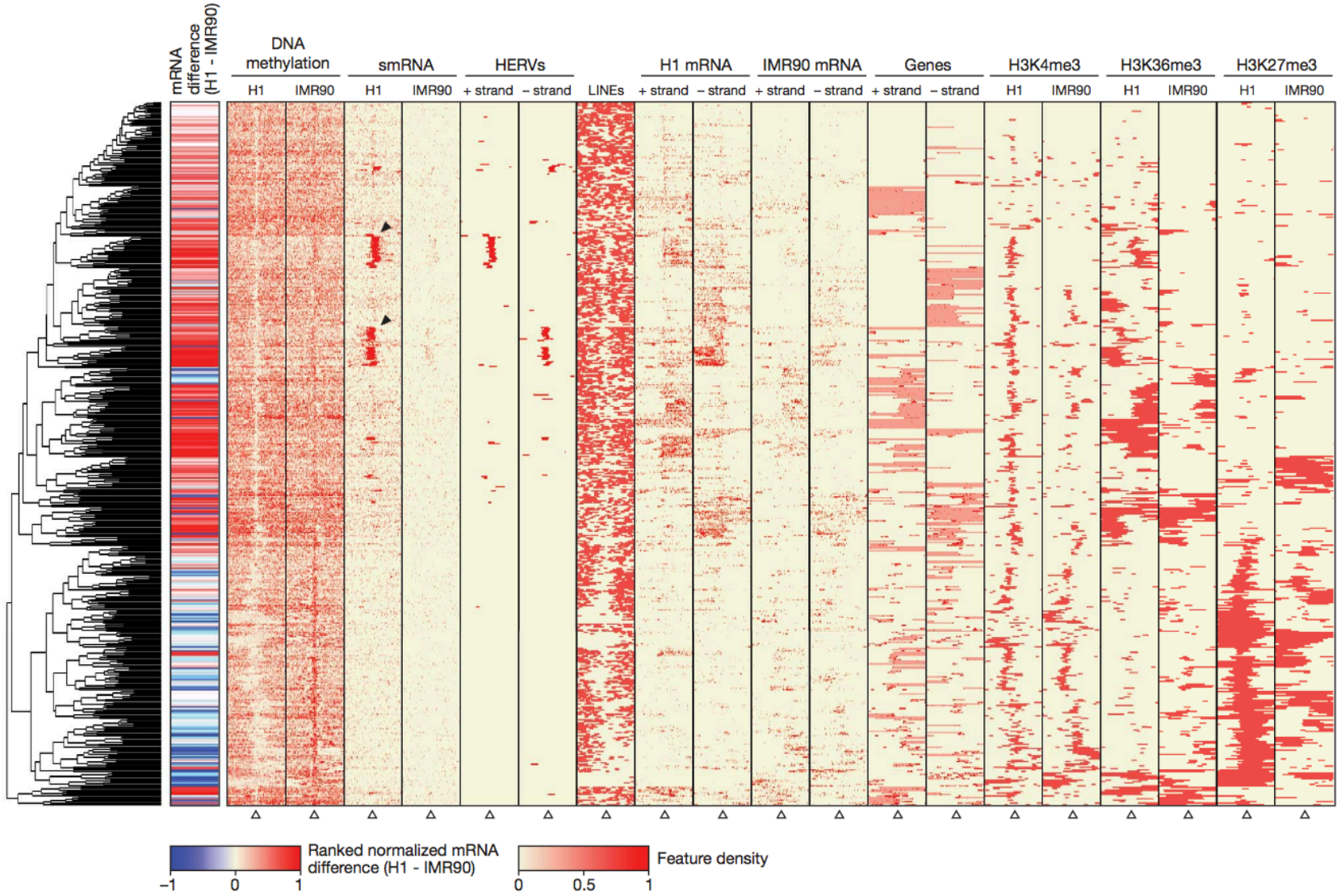




Model the noise

Pelizzola M and Ecker JR, FEBS Letters 2010

Integration with other data types



Lister R *et al*, Nature 2009