nature biotechnology

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

R Alan Harris[1,*], Ting Wang[2], Cristian Coarfa[1], Raman P Nagarajan[3], Chibo Hong[3], Sara L Downey[3], Brett E Johnson[3], Shaun D Fouse[3], Allen Delaney[4], Yongjun Zhao[4], Adam Olshen[3], Tracy Ballinger[5], Xin Zhou[2], Kevin J Forsberg[2], Junchen Gu[2], Lorigail Echipare[6], Henriette O'Geen[6], Ryan Lister[7], Mattia Pelizzola[7], Yuanxin Xi[8], Charles B Epstein[9], Bradley E Bernstein[9–11], R David Hawkins[12], Bing Ren[12,13], Wen-Yu Chung[14,15], Hongcang Gu[9], Christoph Bock[9,16–18], Andreas Gnirke[9], Michael Q Zhang[14,15], David Haussler[5], Joseph R Ecker[7], Wei Li[8], Peggy J Farnham[6], Robert A Waterland[1,19], Alexander Meissner[9,16,17], Marco A Marra[4], Martin Hirst[4], Aleksandar Milosavljevic[1] & Joseph F Costello[3]

Alan Harris

3rd Epigenome Informatics Workshop
March 5, 2012

DNA Methylation

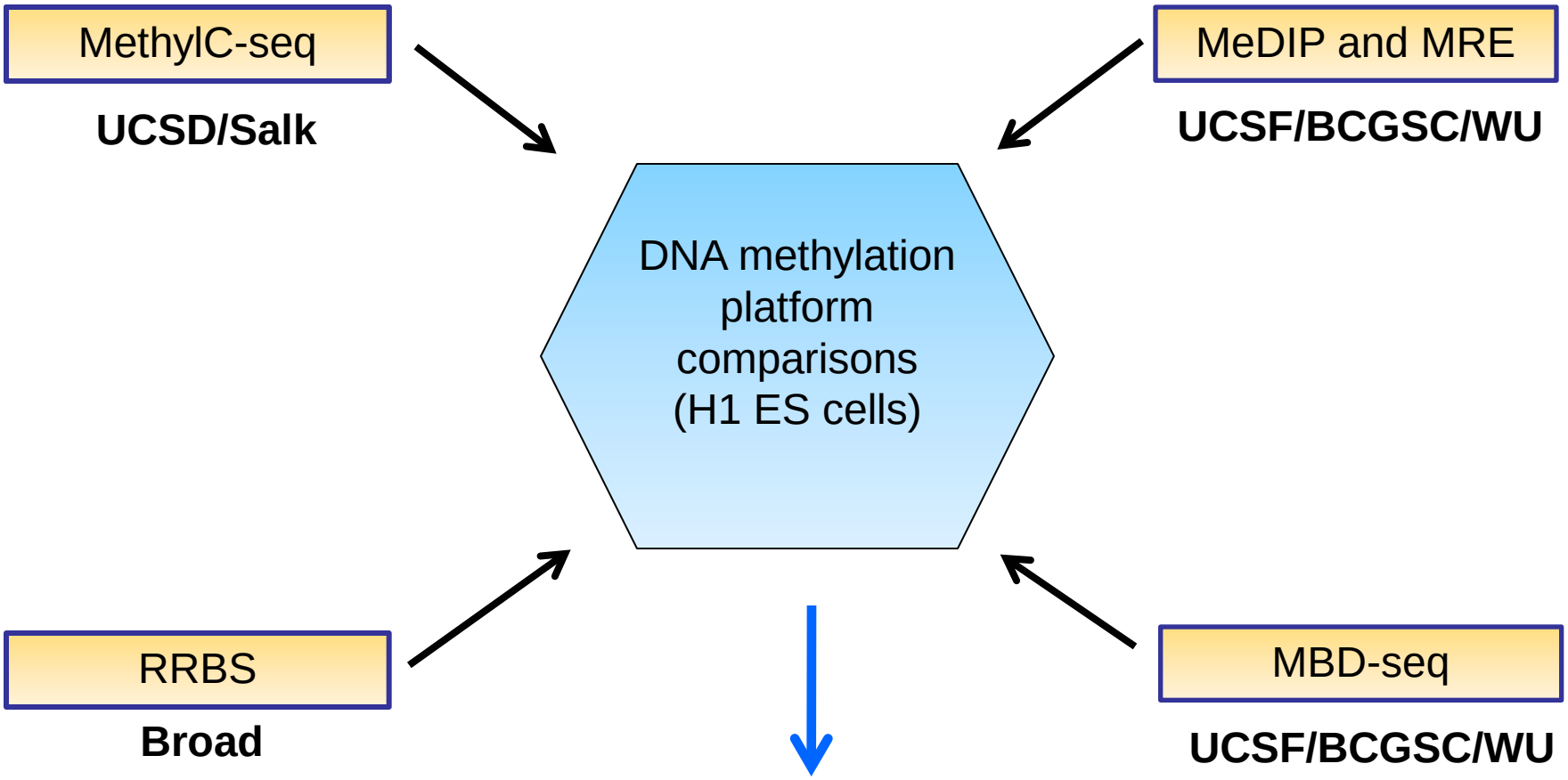
- C5 position of cytosines primarily in CpGs, but also in non-CpGs primarily in embryonic stem cells
- CpG Islands
- Regulation of cellular processes
  - Transcription
  - Defense against endogenous retroviruses
  - Embryonic development
  - X chromosome inactivation
  - Imprinting

DNA Methylation Methods

- **Bisulfite-based Methods**
  - MethylC-seq: whole genome shotgun bisulfite sequencing
  - RRBS: Reduced Representation Bisulfite Sequencing
    - MspI digestion

- **Enrichment-based Methods**
  - MeDIP-seq: Methylated DNA Immunoprecipitation
    - 5-methylcytosine antibody
  - MBD-seq: Methyl-Binding Domain
    - MBD2 protein methyl-CpG binding domain
  - MRE-seq: Methylation-sensitive Restriction Enzyme
    - Parallel HpaII, Hin6I, and AciI digestion

DNA Methylation Methods



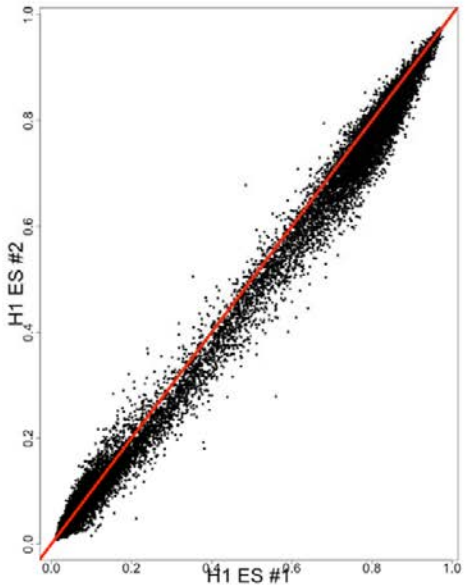
Joint REMCs-EDACC et al, publication

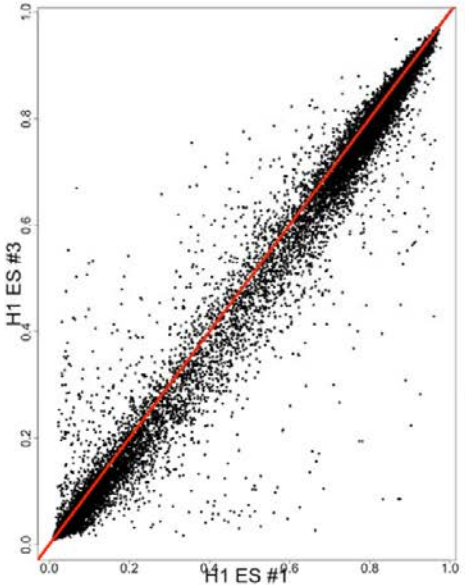
H1 ESC Samples

- 3 Biological Replicates Assayed
  - #1 – Passage 30
  - #2 – Passage 32
  - #3 – Passage 27
- Establish reproducibility across replicates

Infinium 27K Array Reproducibility

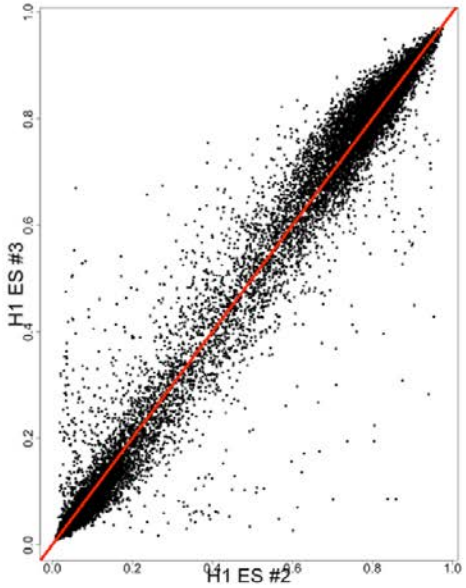
CCC **0.996**





CCC **0.992**

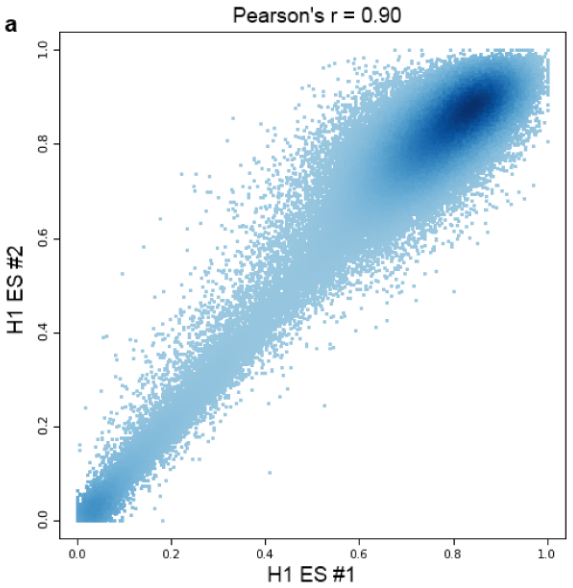
CCC **0.992**

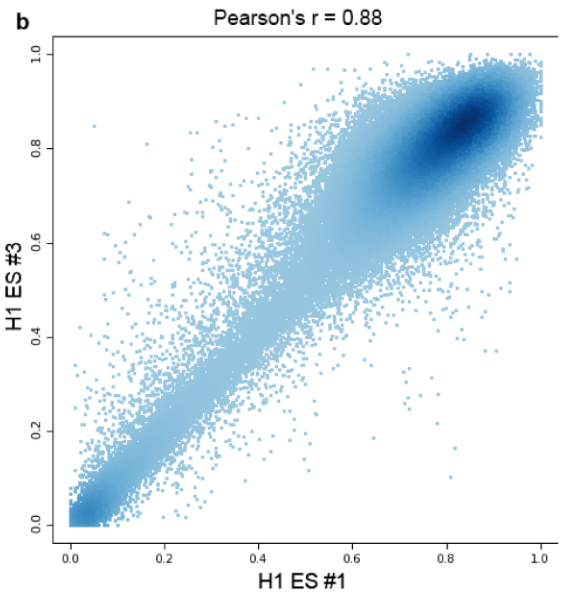


CCC – concordance correlation coefficient
Linear relation relative to 45° line

RRBS Reproducibility

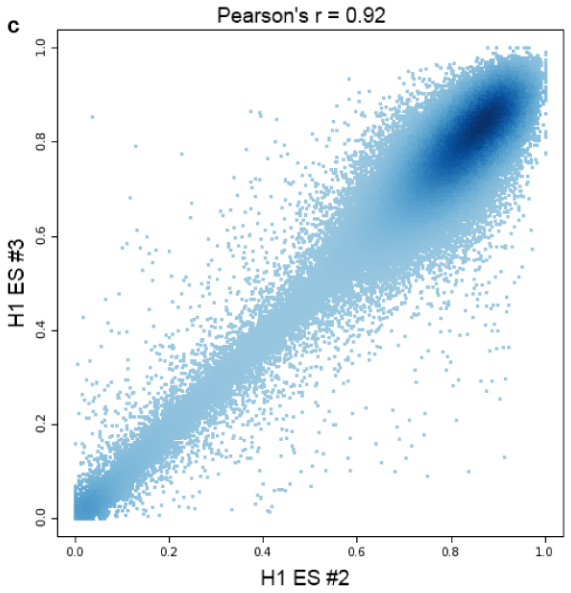
Pearson's r
0.90



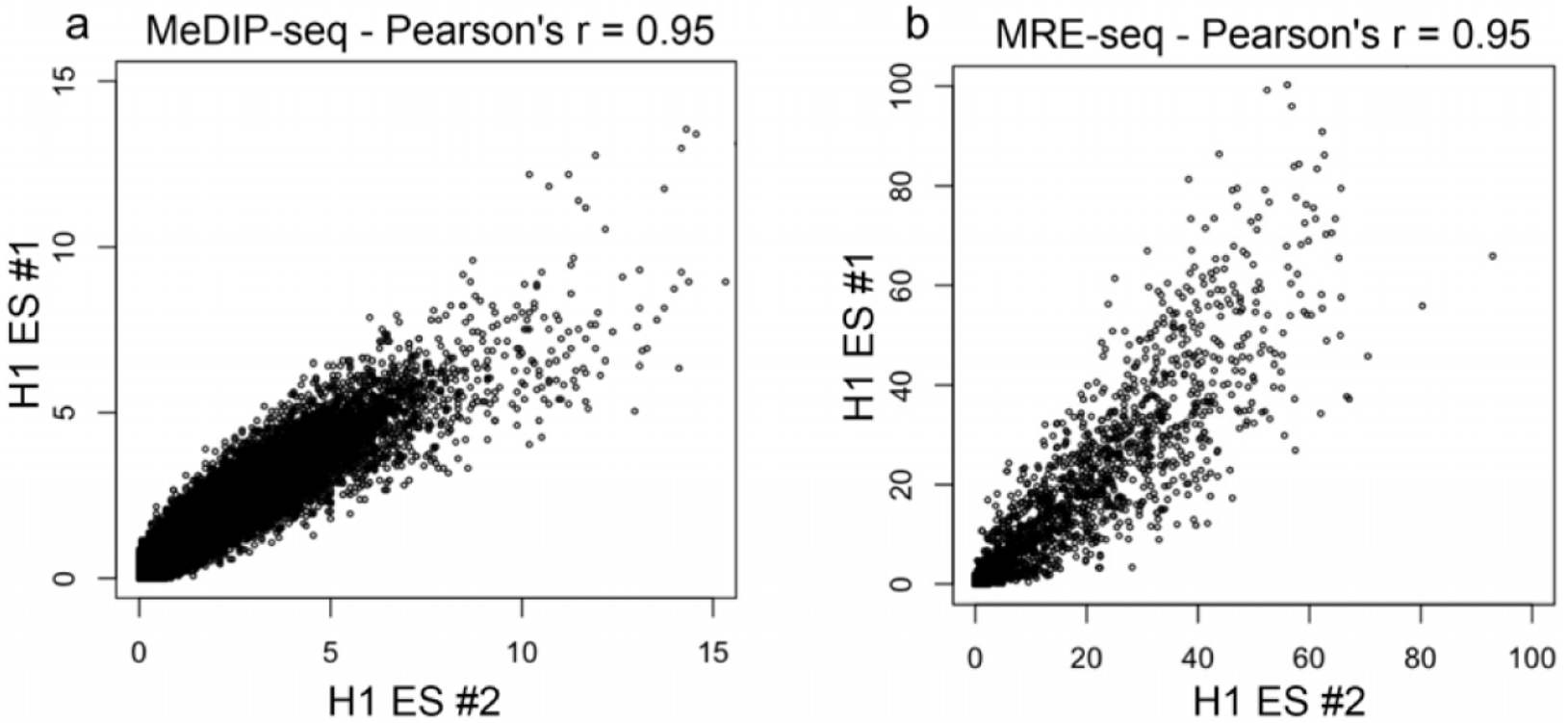


Pearson's r
0.92

Pearson's r
0.92



MeDIP-seq and MRE-seq Reproducibility

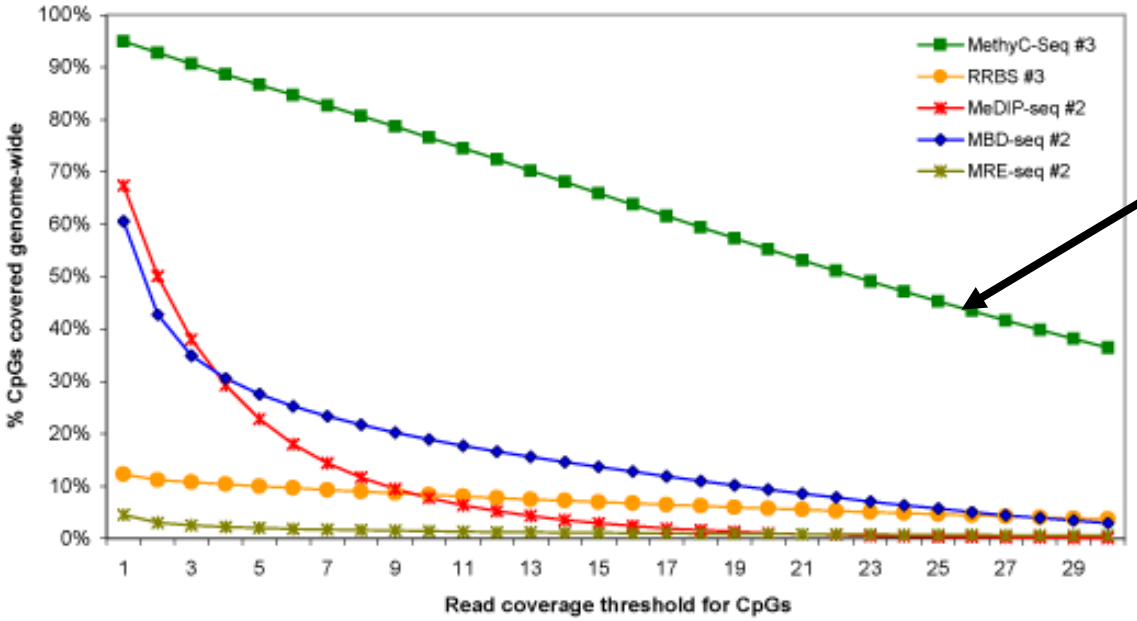


Sequencing and Repeat Coverage

Method	H1 DNA sample #	Total bases generated (Gbp)	Total high quality bases (Gbp)	Total bases in map (Gbp)	Maximum resolution (bp)	1-read coverage of CpGs in repeats (# ,%)	Percentage of assayed CpGs in repeats (%)
MethylC-seq	#3	172.49	115	87.5	1	13,303,415 (91.8)	49.7
RRBS	#3	1.58	1.43	1.28	1	1,646,649 (11.4)	47.5
MeDIP-seq	#1	3.42	2.07	1.95	150	10,004,670 (68.3)	52.9
MeDIP-seq	#2	3.02	1.84	1.73	150	10,101,868 (68.9)	53.2
MeDIP-seq	#1+#2	6.44	3.91	3.68	150	11,693,059 (79.8)	53.5
MBD-seq	#2	5.67	3.71	2.21	150	10,080,007 (68.8)	59.1
MRE-seq	#1	3.61	1.31	0.96	1	306,635 (2.07)	21.7
MRE-seq	#2	4.03	1.69	1.3	1	232,885 (1.59)	18.6

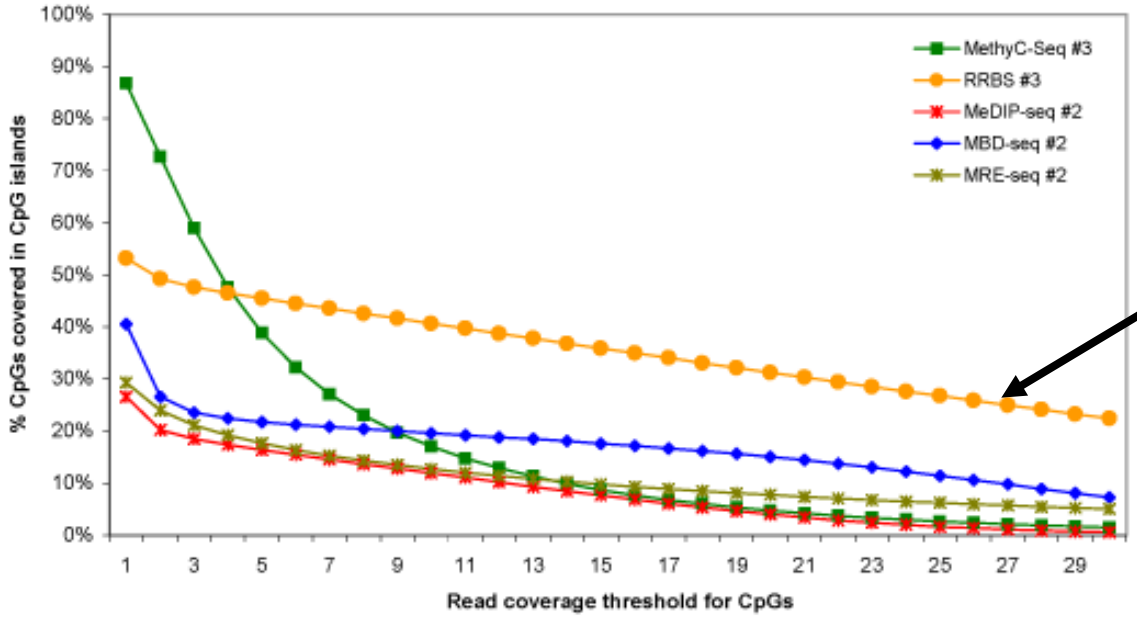
Genome-wide and CGI Coverage

Genome-wide



MethylC-seq

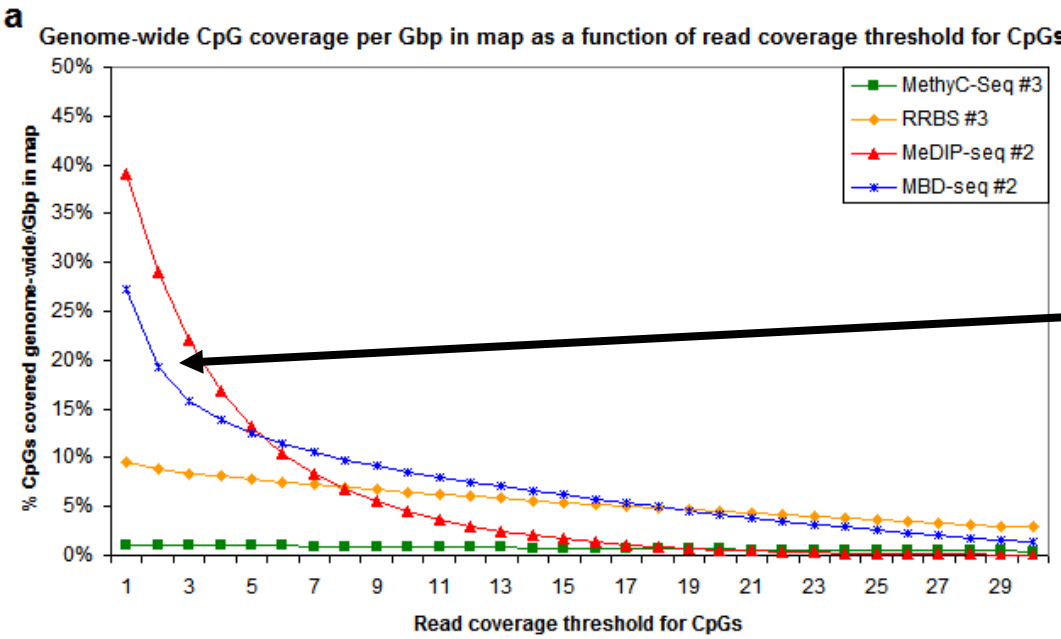
CpG Islands



RRBS

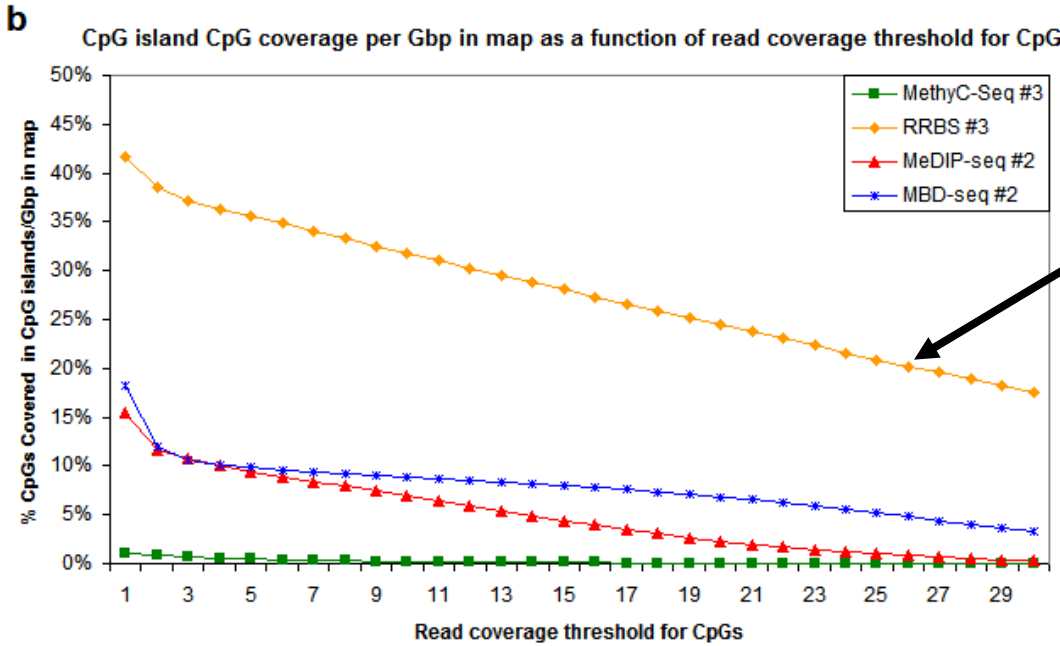
Coverage per Gbp of sequence

Genome-wide



MeDIP-seq
MBD-seq

CpG Islands



RRBS

Overlap of Methylome Coverage

Method(s)	Genome-wide CpGs covered by method(s)
Coverage by 4 methods	
MethylC, RRBS, MeDIP, MBD	6.32%
Coverage by 3 methods	
MethylC, RRBS, MeDIP	0.81%
MethylC, RRBS, MBD	1.46%
MethylC, MeDIP, MBD	39.09%
RRBS, MeDIP, MBD	0.31%
Coverage by 2 methods	
MethylC, RRBS	2.30%
MethylC, MeDIP	19.95%
MethylC, MBD	10.27%
RRBS, MeDIP	0.03%
RRBS, MBD	0.68%
MeDIP, MBD	0.61%
Coverage by 1 method	
MethylC	14.73%
RRBS	0.37%
MeDIP	0.09%
MBD	1.77%
No coverage	
None	1.21%

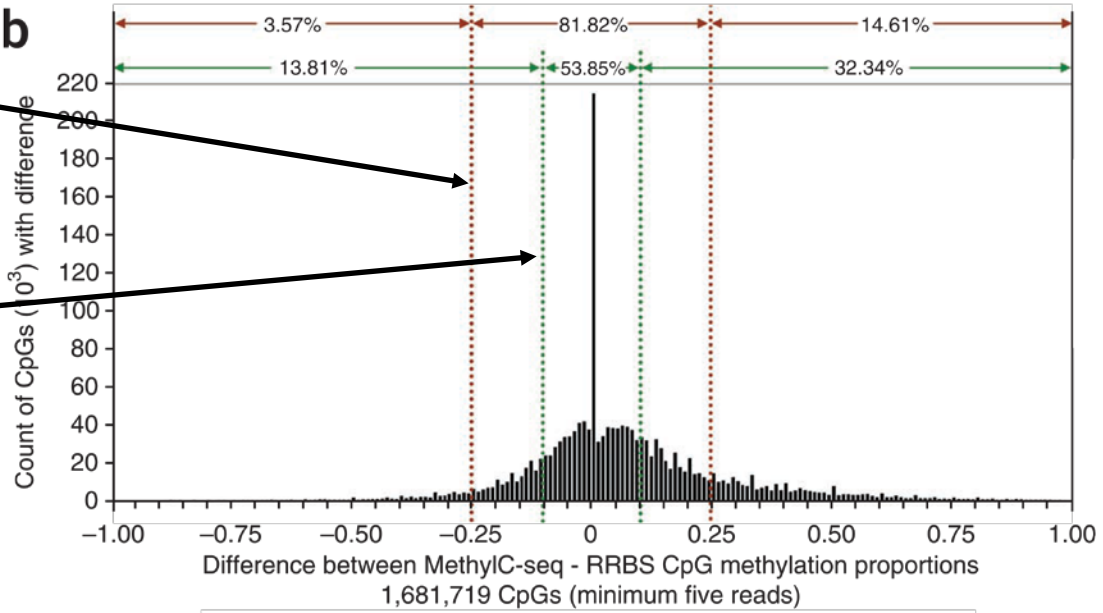
MethylC-seq RRBS CpG Comparison

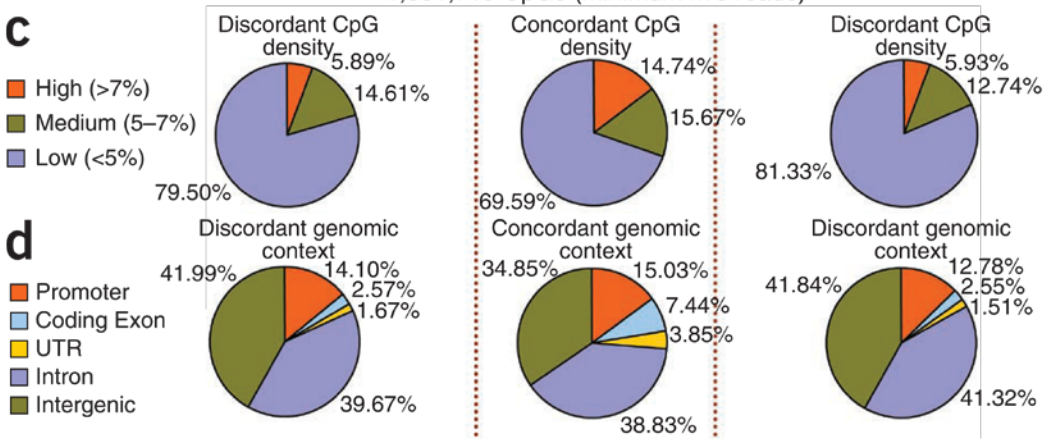
a

Minimum read depth	CpGs covered	Percent genome-wide CpGs	0.80–0.20 Methylation cutoff % concordant	0.75–0.25 Methylation cutoff % concordant	0.20 Methylation cutoff % concordant
2	2,542,763	9.03	68.35	72.86	94.14
5	1,681,719	5.97	67.40	72.28	96.15
10	913,230	3.24	67.79	73.20	97.13

0.25 Difference
81.82%

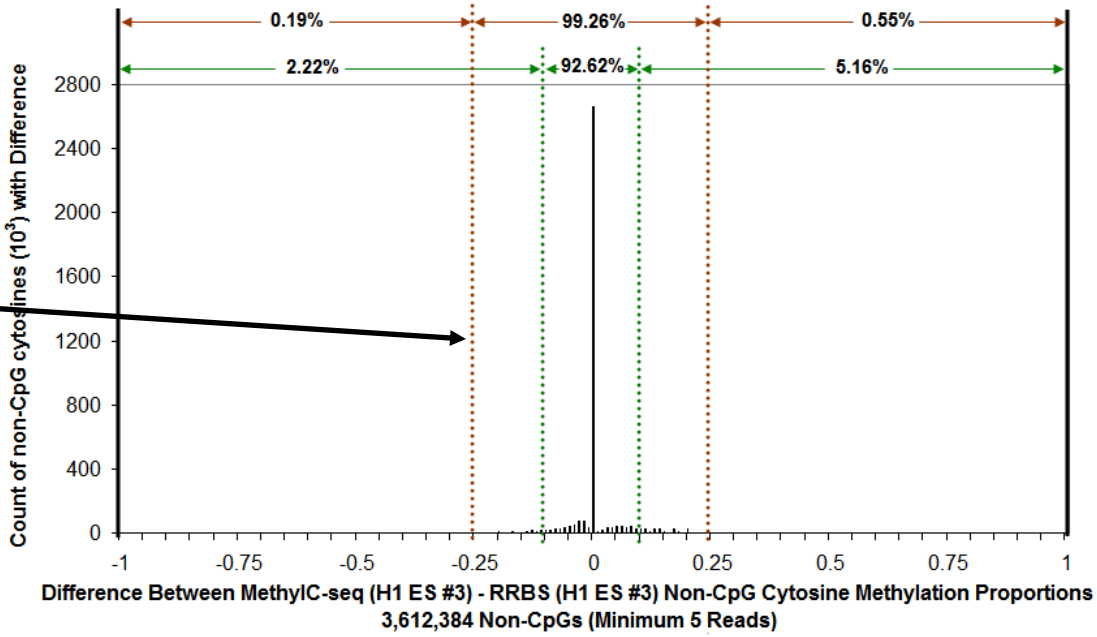
0.10 Difference
53.85%



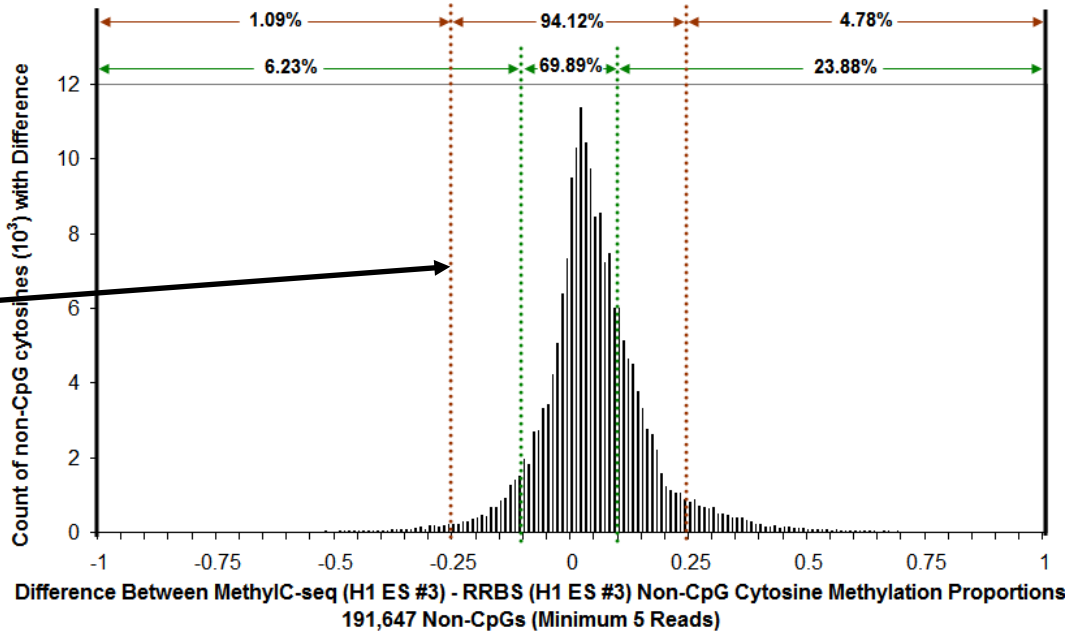


MethylC-seq RRBS Non-CpG Comparison

All Non-CpG Sites
0.25 Difference
99.26%



Methylated Non-CpG Sites
0.25 Difference
94.12%

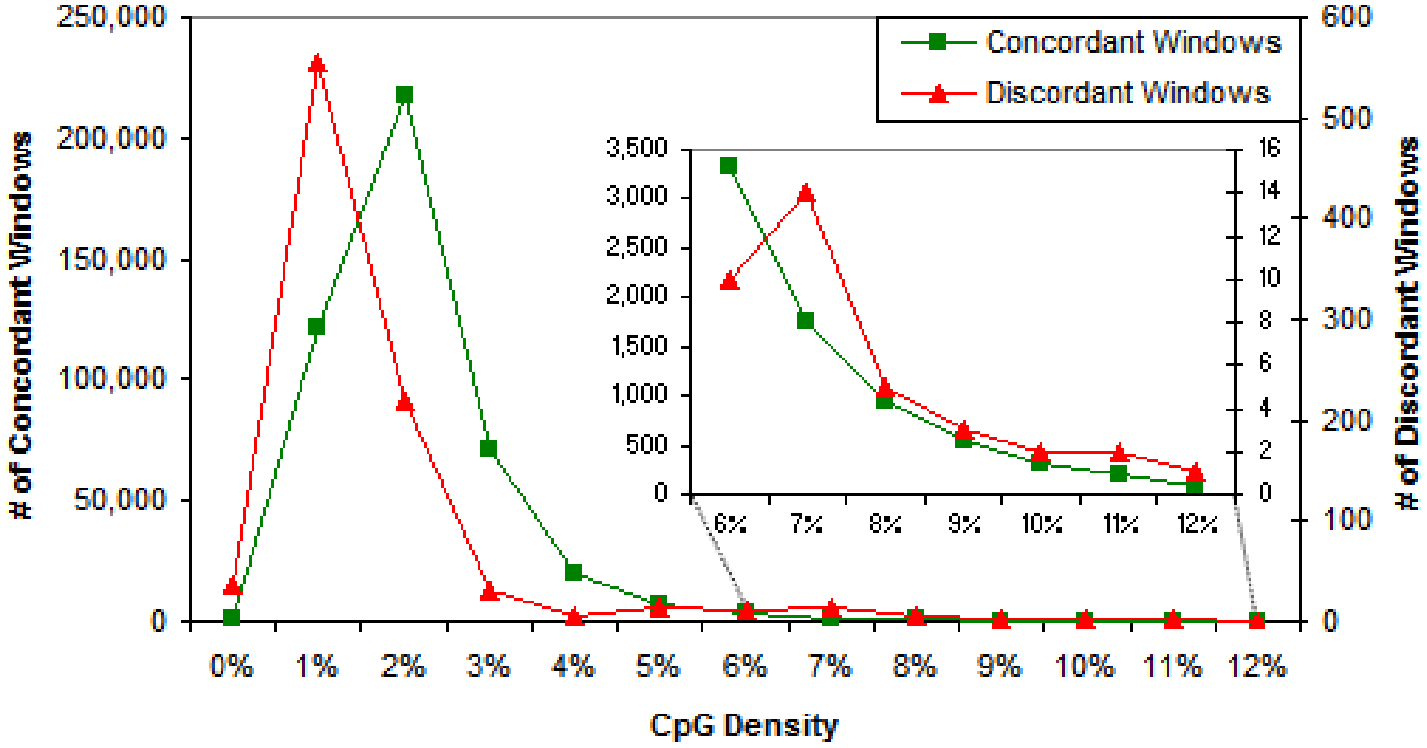


MeDIP-seq MBD-seq Comparison

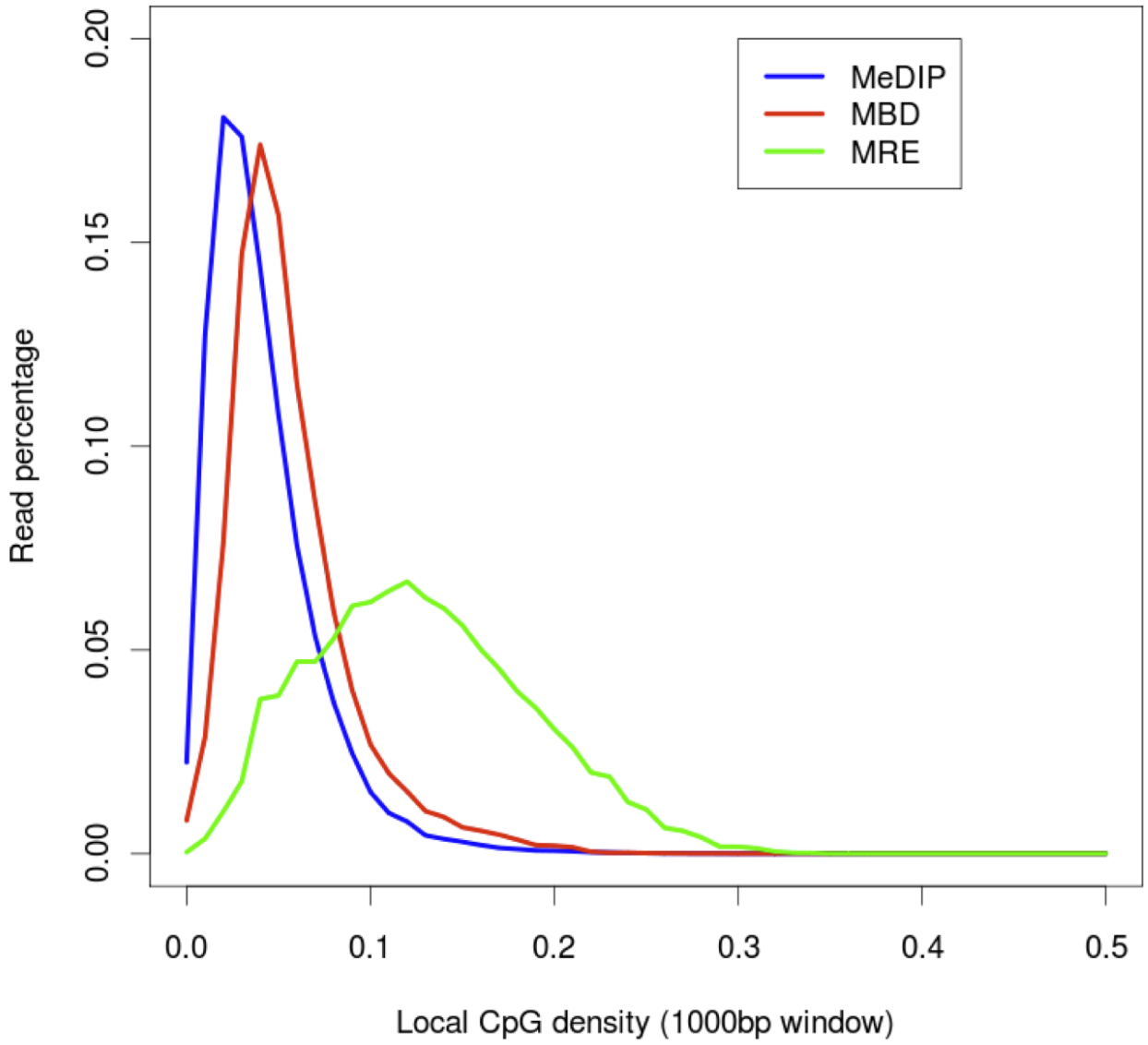
a

Minimum Read Depth	1000bp Windows			200bp Windows		
	Number of Windows	% Genome-wide CpGs	% Concordant	Number of Windows	% Genome-wide CpGs	% Concordant
2	1,189,545	61.82%	98.80%	2,136,710	37.96%	92.41%
5	446,096	32.65%	99.80%	753,329	17.72%	99.01%
10	162,661	15.07%	100.00%	273,767	7.74%	99.97%

b



MeDIP-seq MBD-seq Comparison

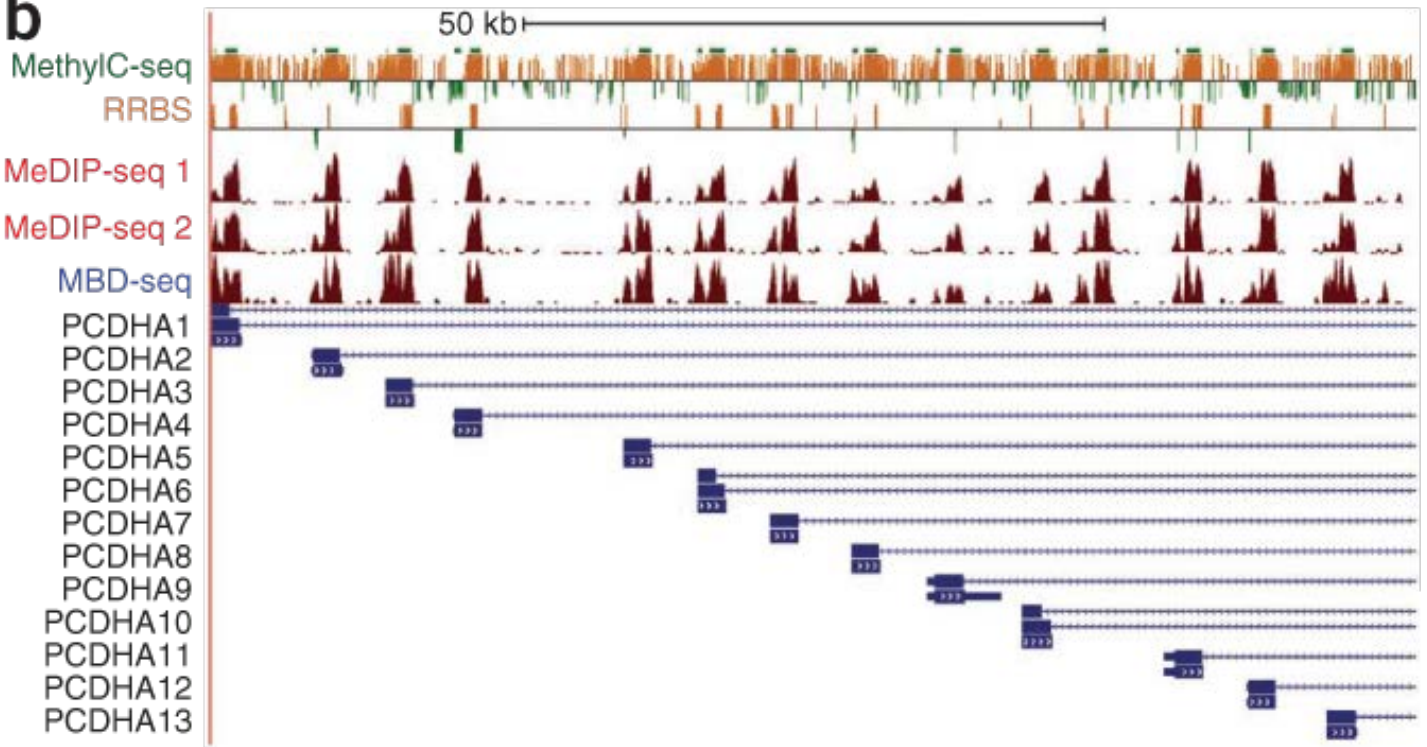


All Method Comparison

a

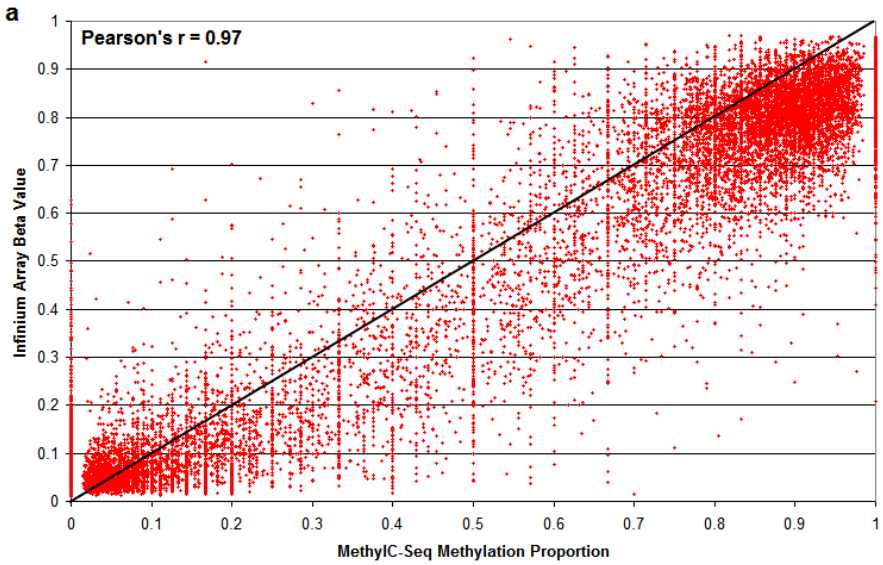
Methods	Minimum read depth of 5 199,438 windows (18.01% of genome-wide CpGs) Percent windows	Minimum read depth of 10 87,363 windows (9.39% of genome-wide CpGs) Percent windows
(MethylC, RRBS, MeDIP, MBD)	97.64	98.30
(MethylC, RRBS, MeDIP)(MBD)	0.07	0
(MethylC, RRBS, MBD)(MeDIP)	0.07	0
(MethylC, MeDIP, MBD)(RRBS)	1.98	1.60
(RRBS, MeDIP, MBD)(MethylC)	0.03	0.02
(MethylC, RRBS)(MeDIP, MBD)	0.20	0.07
(MethylC, MeDIP)(RRBS, MBD)	0.01	0
(MethylC, MBD)(RRBS, MeDIP)	0	0

b



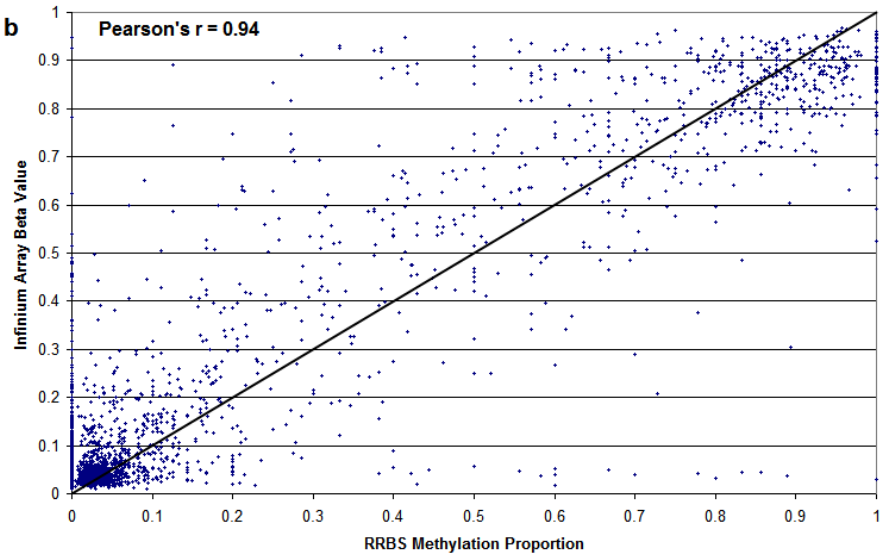
Sequencing methods compared to Infinium 27K Arrays

Pearson's r = 0.97



MethylC-seq

Pearson's r = 0.94



RRBS

- Methylation calls in 200bp windows:
  - Beta value >0.2 highly methylated
  - Beta value ≤0.2 weakly methylated

- MeDIP-Seq - 96.19% concordance
- MBD-Seq - 90.80% concordance

Methylation Comparison Conclusions

All four methods yield comparable results but differ in CpG coverage, resolution and cost

- **MethylC-Seq**: genome-wide single base resolution, but requires most sequencing.
- **RRBS**: primarily CGI coverage at single base resolution.
- **MBD**: genome-wide ~150bp resolution, but least expensive. No non-CpG detection.
- **MeDIP**: genome-wide ~150bp resolution. Can be integrated with assays of unmethylated sites

Biological importance of intermediate methylation levels

1. Imprinting

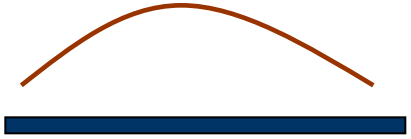
2. Non-imprinted monoallelic methylation

3. Cell type-specific methylation

4. Sites of inter-individual variation in methylation level

Integrative Method

1

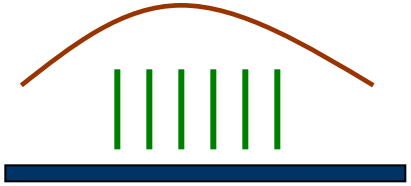


high MeDIP, no or low MRE
Methylated

2

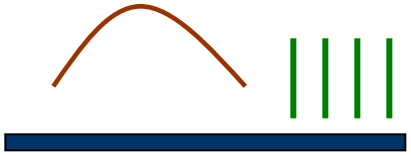
high MRE, no or low MeDIP
Unmethylated

3



high MRE and MeDIP
Intermediate Methylation

4



high MRE and MeDIP
Patch Methylation

Unmethylated CpGs

methylation-sensitive restriction digestion (**MRE**)

combine parallel digests, ligate adapters, size-select 100-300 bp

~20 million reads/sample

Methylated CpGs

methyl DNA immunoprecipitation (**MeDIP**)

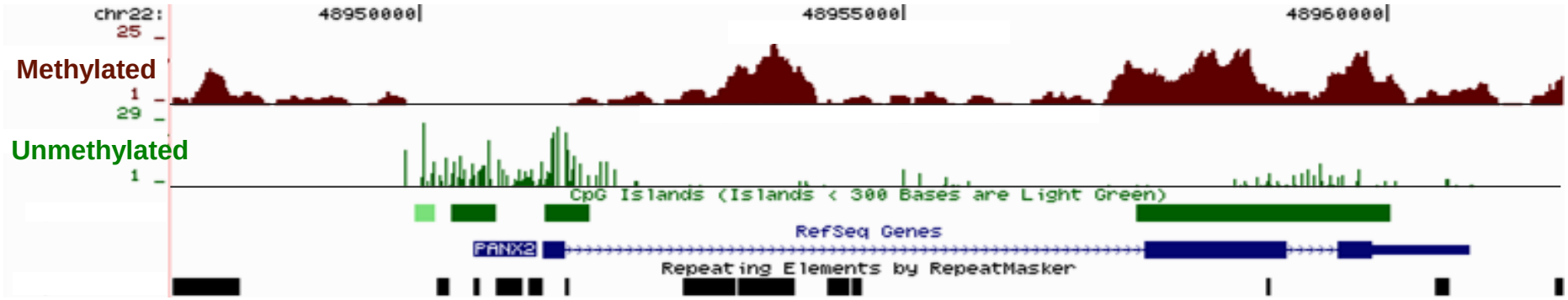
IP sonicated, adapter-ligated DNA, size-select 100-300 bp

~100 million reads/sample

Illumina library construction

IGAII sequencing

data visualization

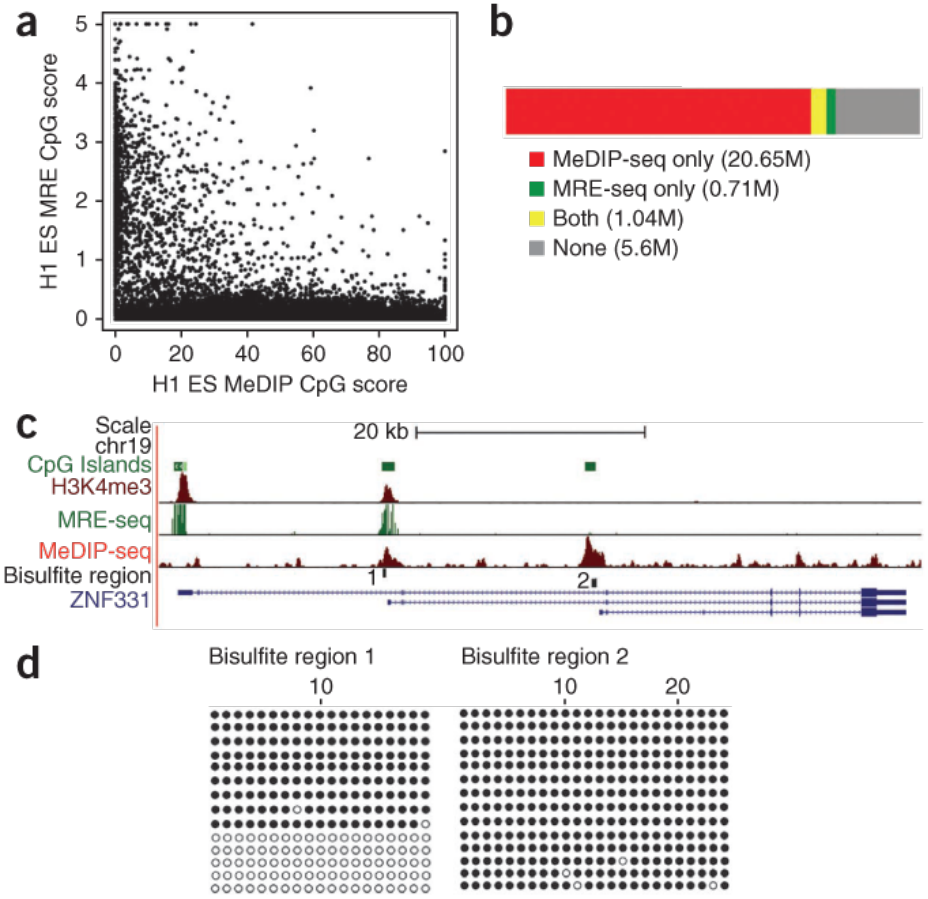


5' CpG islands are unmethylated

3' CpG island is partially methylated

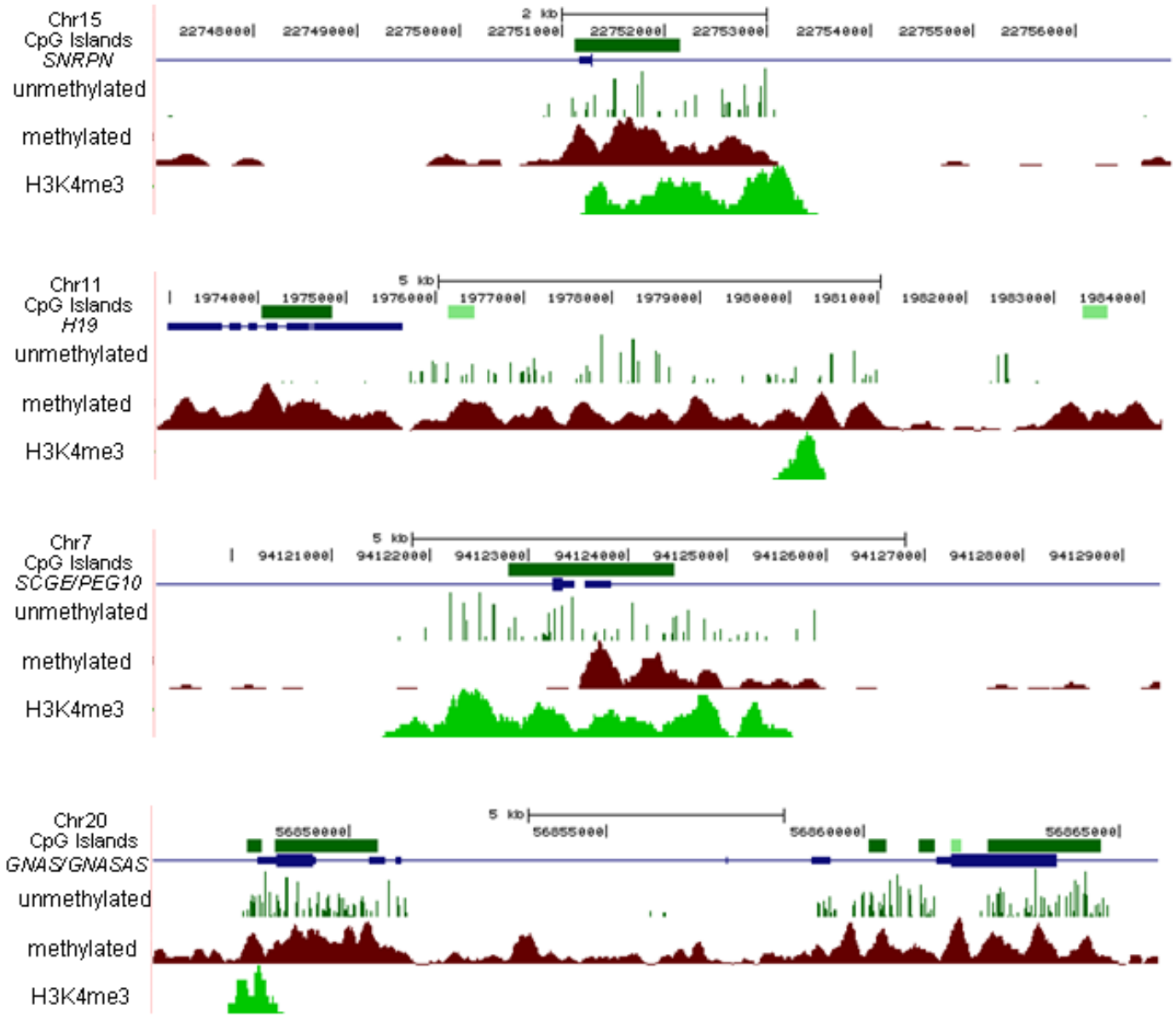
Catalog of intermediate methylation sites

- 992 CpG Islands with overlap between MeDIP-seq and MRE-seq signals



Ting Wang, Washington University

Intermediate methylation levels at imprinted genes



Rediscovered
16 of 19 known
imprinting DMRs

Using Genetic Variation to Detect Monoallelic Epigenomic and Transcription States

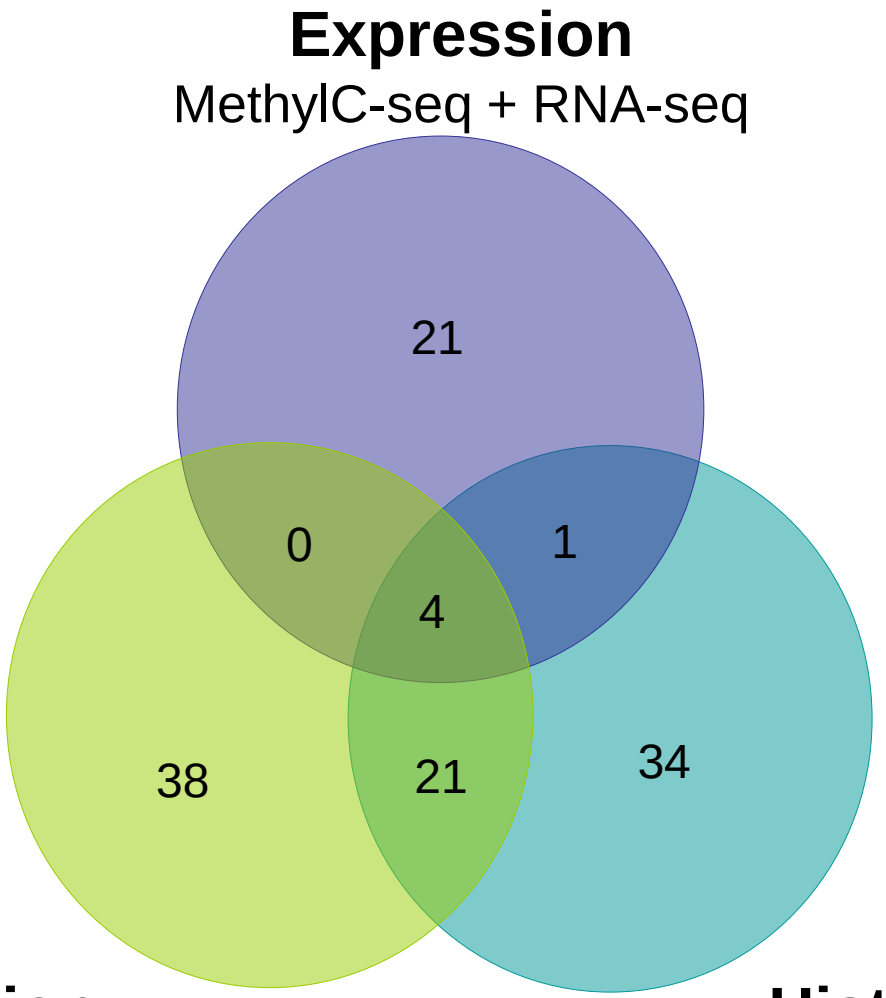
1. Monoallelic DNA methylation (MRE and MeDIP)

2. Monoallelic expression (MethylC-seq and RNA-seq)

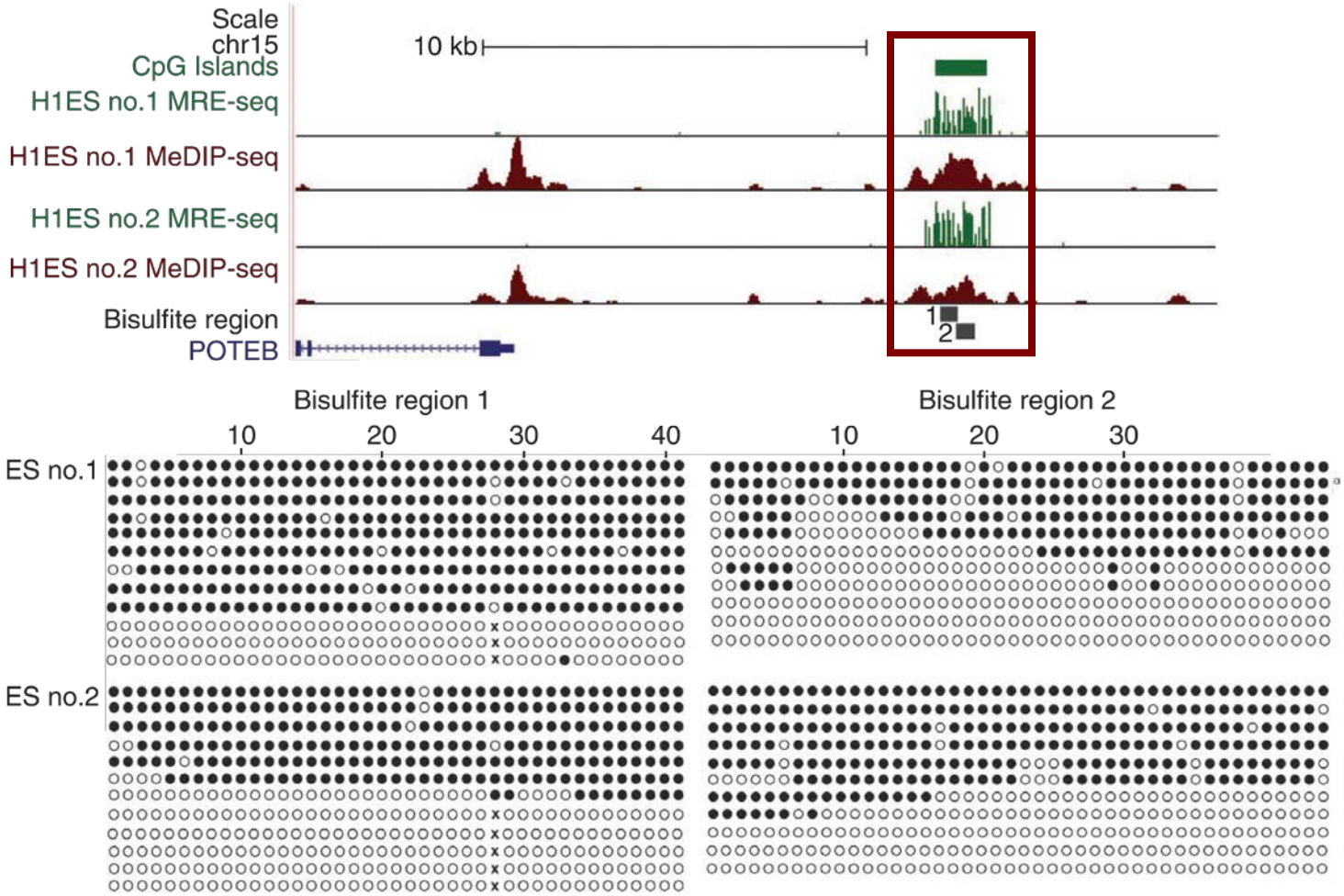
3. Monoallelic Histone H3K4me3 (MethylC-seq and Chip-seq)

Cristian Coarfa, BCM

Monoallelic Epigenomic Marks and Expression



Intermediate methylation levels in *POTEB*



Assay specific SNPs showing allele specific methylation

MeDIP Allele Count: G 9

MRE Allele Count: A 30

Integrative Method Conclusions

- Integrative method can identify regions of intermediate methylation
- Regions of intermediate methylation can point to imprinted or other types of differentially methylated regions
- Sequence reads can be used to identify epigenomic or expression differences between alleles

Acknowledgements

NIH Roadmap Epigenomics Project

EDACC: Cristian Coarfa, Yuanxin Xi, Wei Li, Robert A. Waterland, Aleksandar Milosavljevic

UCSF/GSC REMC: Raman Nagarajan, Chibo Hong, Sara Downey, Brett E. Johnson, Allen Delaney, Yongjun Zhao, Marco Marra, Martin Hirst, Joseph Costello

- **UCSC**: Tracy Ballinger, David Haussler
- **Washington University**: Xin Zhou, Kevin J Forsberg, Junchen Gu, Ting Wang
- **UCD**: Lorigail Echipare, Henriette O'Geen, Peggy J. Farnham

UCSD REMC: Ryan Lister, Mattia Pelizzola, Bing Ren, Joseph Ecker

- Cold Spring Harbor: Wen-Yu Chung, Michael Q. Zhang

Broad REMC: Hongcang Gu, Christoph Bock, Andreas Gnirke, Chuck Epstein, Brad Bernstein, Alexander Meissner