Washington University in St.Louis
SCHOOL OF MEDICINE

Department of Genetics

Satellite Education Workshop (SW4)

DNA Methylomics

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

ABRF2012
March 17, 2012

ABRF

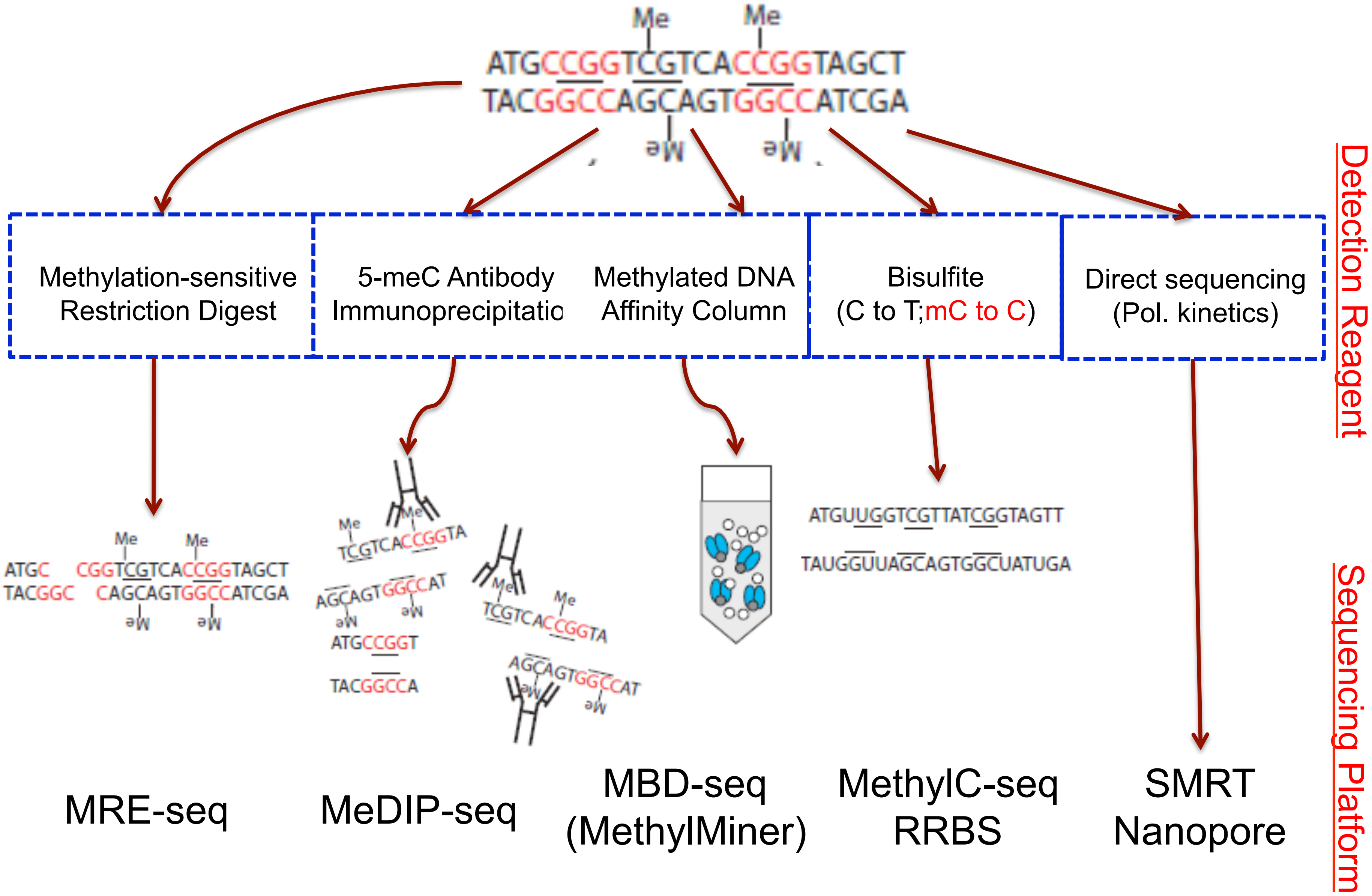
Part II

Analysis of DNA Methylomics Data

Outline

- Case 1: Comparison of popular methods
- Case 2: Function of Intragenic DNA methylation
- Integrative methods (combining MeDIP and MRE)
  - Case 3: Insights into monoallelic gene regulation and imprinting
  - Case 4: Predicting DNA methylation at single CpG resolution
  - (Case 5: Cancer methylomes)
- DNA methylomes/Epigenomes resources

Modern DNA Methylomics



Methylation-sensitive Restriction Digest

5-meC Antibody Immunoprecipitatio

Methylated DNA Affinity Column

Bisulfite (C to T;mC to C)

Direct sequencing (Pol. kinetics)

Detection Reagent

MRE-seq

MeDIP-seq

MBD-seq (MethylMiner)

MethylC-seq RRBS

SMRT Nanopore

Sequencing Platform

Case 1: Comparison of DNA methylation mapping technologies

ANALYSIS

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]

Quantitative comparison of genome-wide DNA methylation mapping technologies

Christoph Bock[1–4,6], Eleni M Tomazou[1–3,6], Arie B Brinkman[5], Fabian Müller[1–4], Femke Simmer[5], Hongcang Gu[1], Natalie Jäger[1–3], Andreas Gnirke[1], Hendrik G Stunnenberg[5] & Alexander Meissner[1–3]

Sequencing DNA methylomes

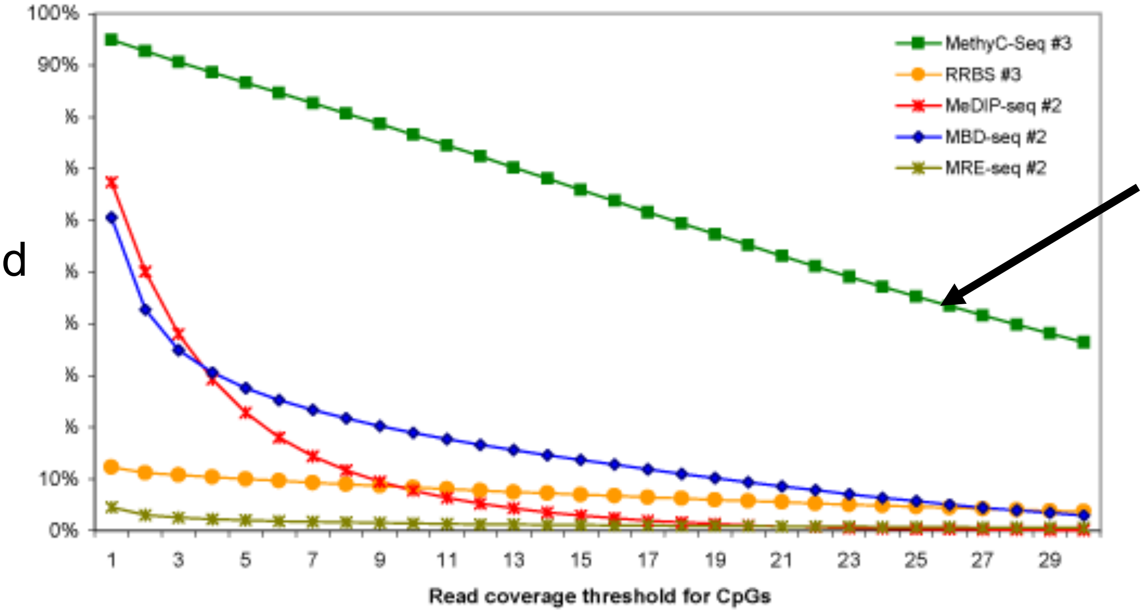
- Coverage
- Resolution
- Transposons
- Assign epigenetic state to genetic allele
- Mutation detection
- Copy number profile

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) 10 (now)
RRBS	0.2-1M	15 K	1	1
MRE-seq and MeDIP-seq	25 M	27 K	1 and 200	8 (2009) 1 (now)
Golden-Gate	1,500	800	1	NA
Infinium	27,500 480,000 (2012)	12,000	1	NA

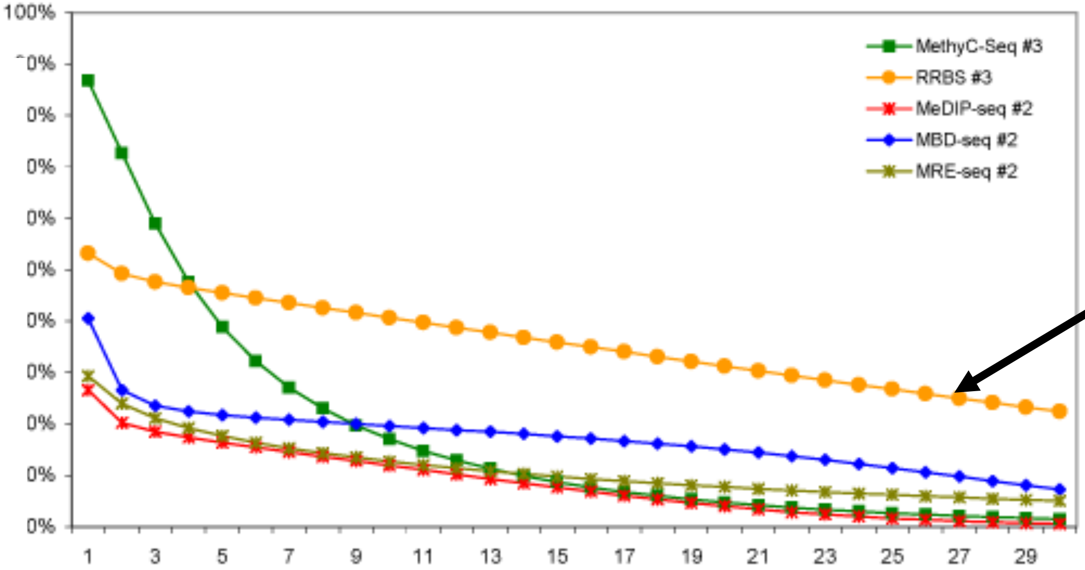
Genome-wide and CGI Coverage

% CpG covered
Genome-wide



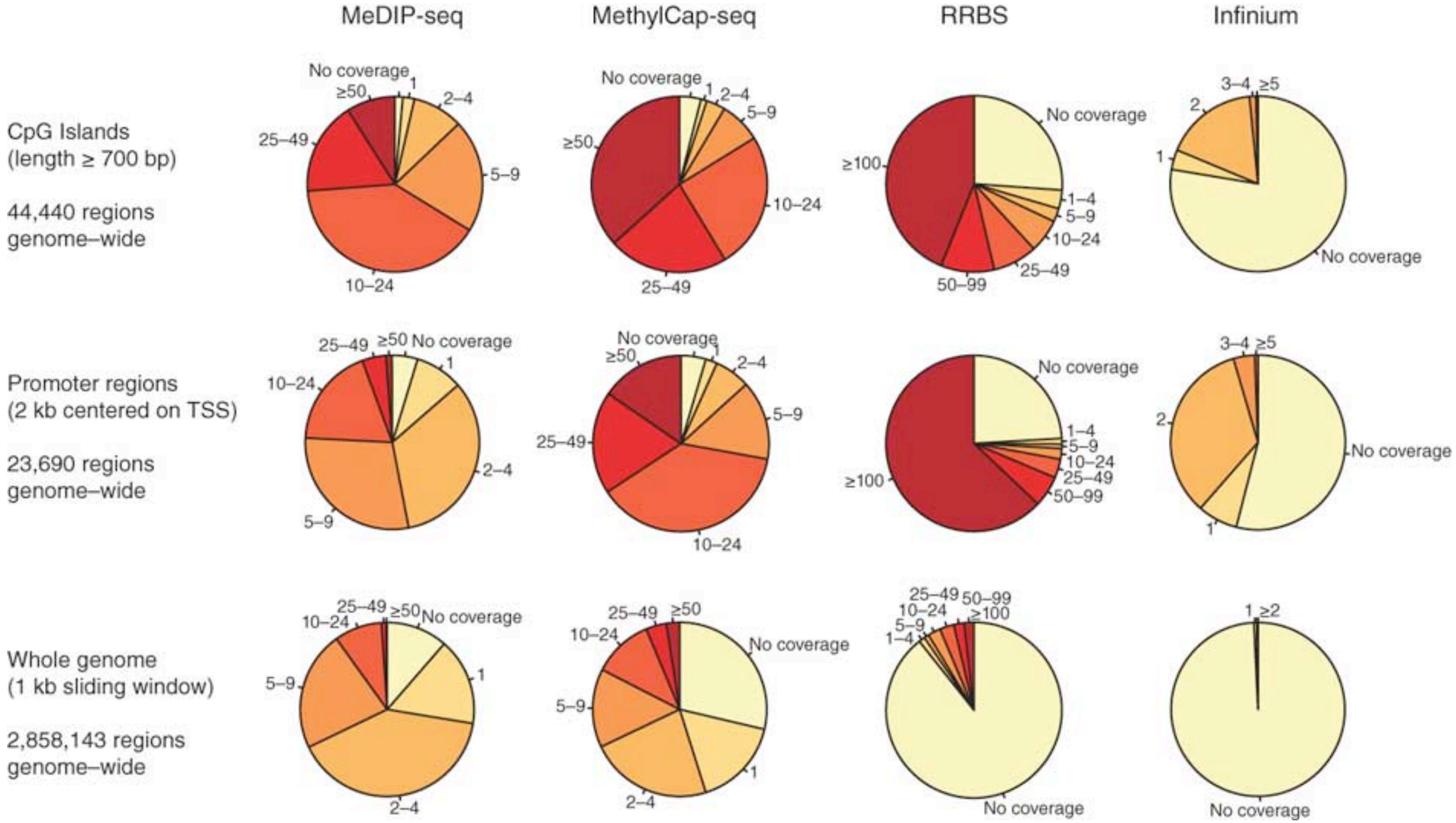
MethylC-seq

% CpG covered
In CpG islands

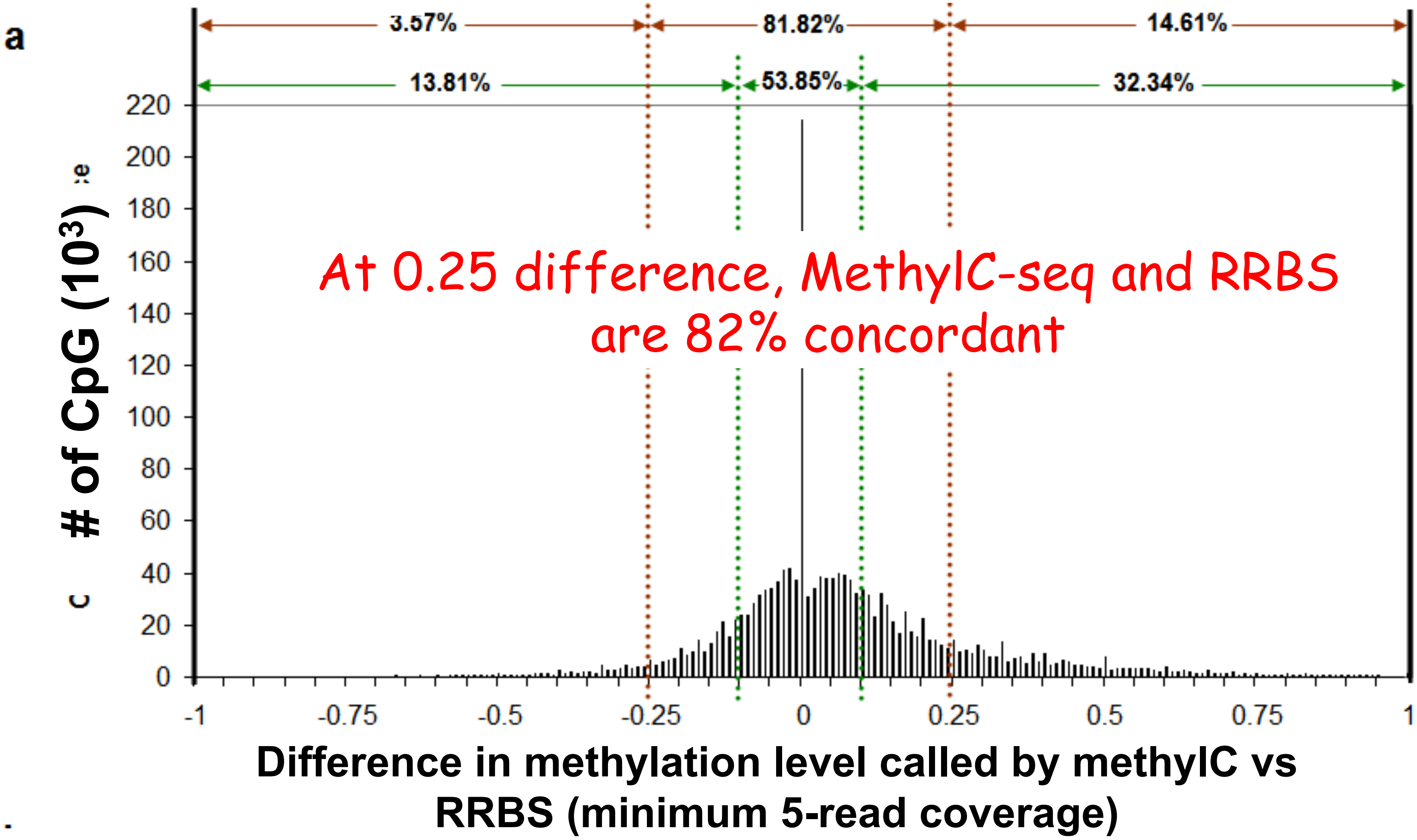


RRBS

Read coverage threshold


MeDIP-seq
MethylCap-seq
RRBS
Infinium
CpG Islands
(length ≥ 700 bp)
44,440 regions
genome–wide
No coverage
1
≥50
2–4
25–49
5–9
10–24
No coverage
1
2–4
5–9
≥50
10–24
25–49
No coverage
≥100
1–4
5–9
10–24
25–49
50–99
3–4
≥5
2
1
No coverage
Promoter regions
(2 kb centered on TSS)
23,690 regions
genome–wide
25–49
≥50
No coverage
1
10–24
2–4
5–9
No coverage
≥50
1
2–4
5–9
25–49
10–24
No coverage
1–4
5–9
10–24
25–49
≥100
50–99
3–4
≥5
2
No coverage
1
Whole genome
(1 kb sliding window)
2,858,143 regions
genome–wide
25–49
≥50
No coverage
10–24
1
5–9
2–4
25–49
≥50
10–24
No coverage
5–9
1
2–4
25–49
50–99
10–24
≥100
5–9
1–4
No coverage
1
≥2
No coverage

Comparison of MethylC-seq and RRBS

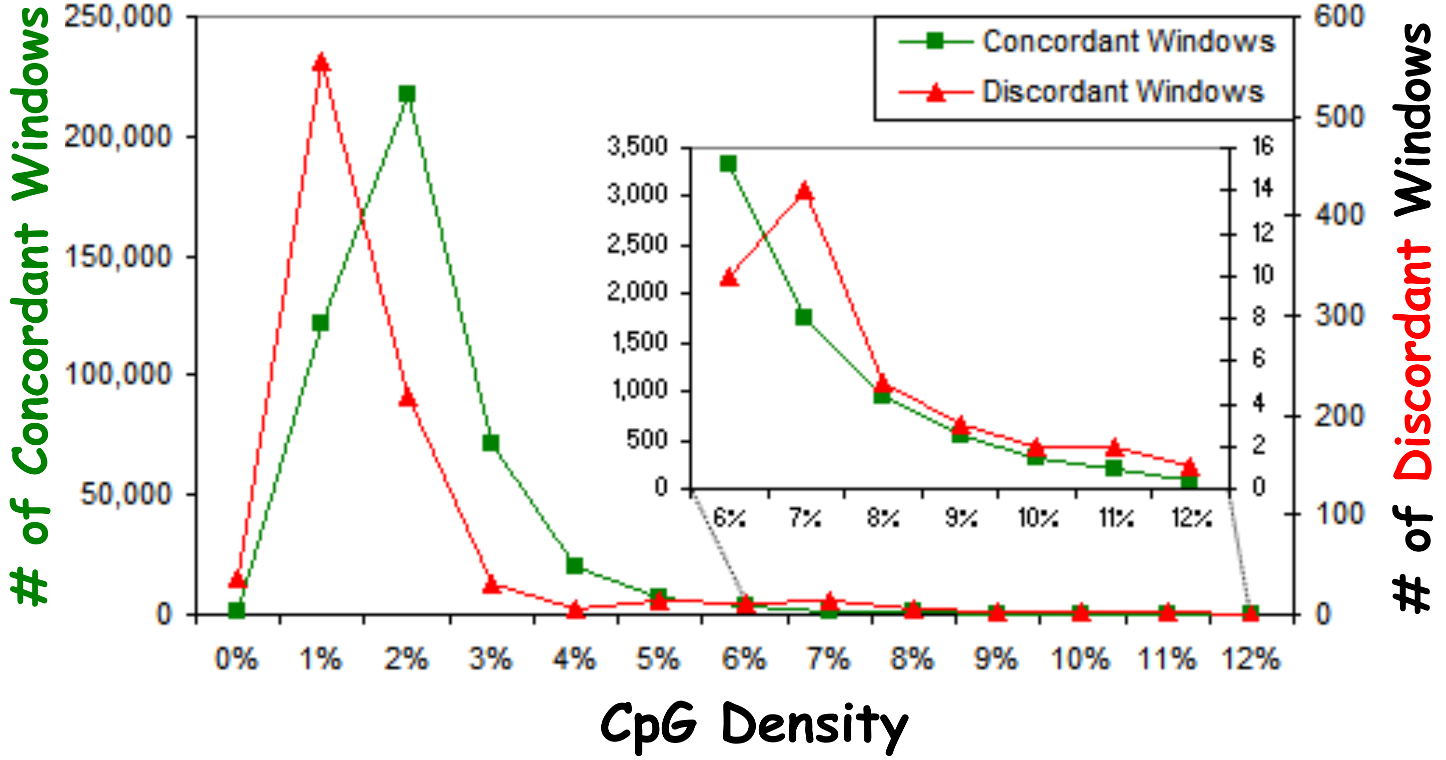


At 0.25 difference, MethylC-seq and RRBS are 82% concordant

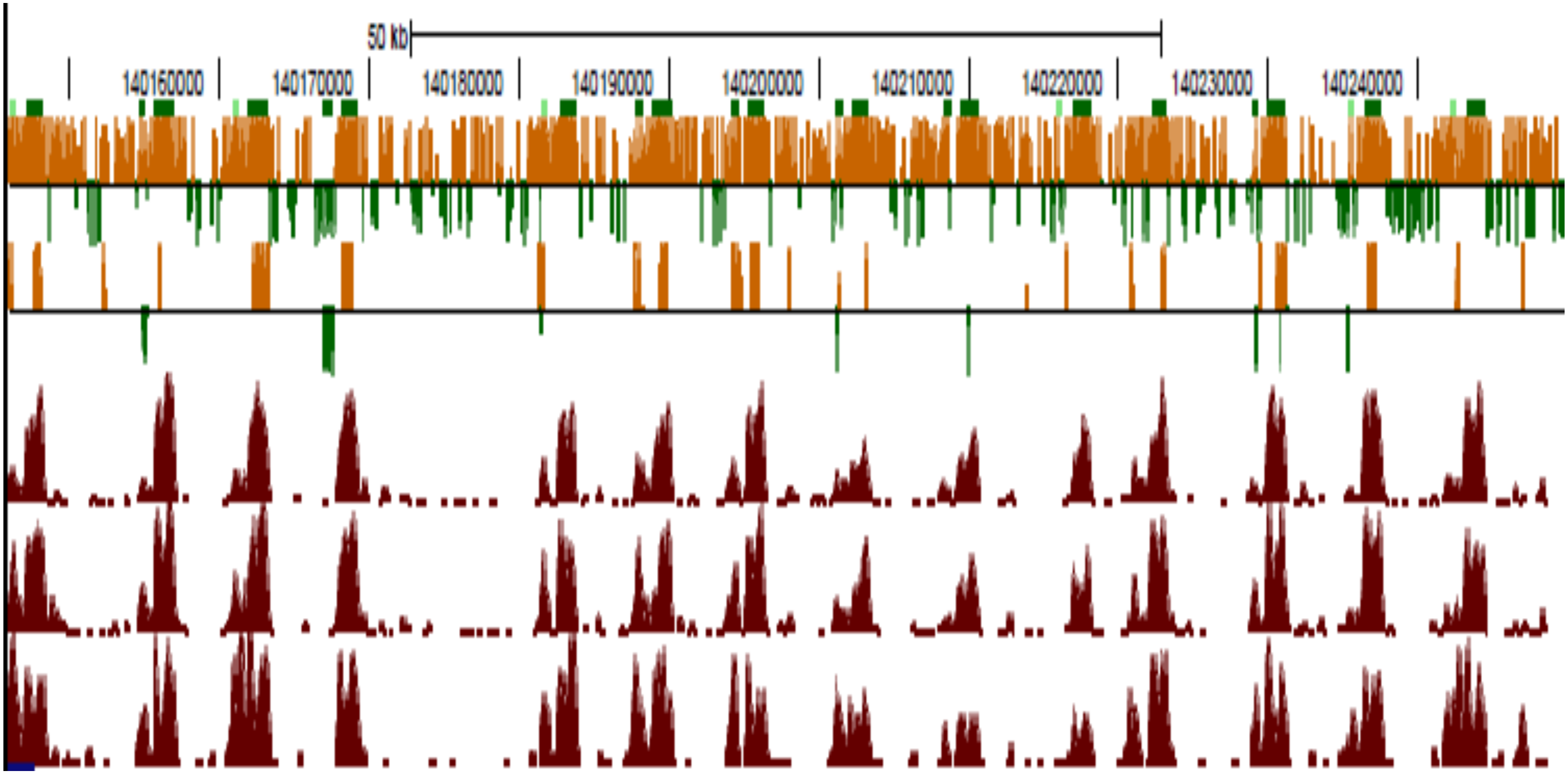
Comparison of MeDIP-seq and MBD-seq

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%



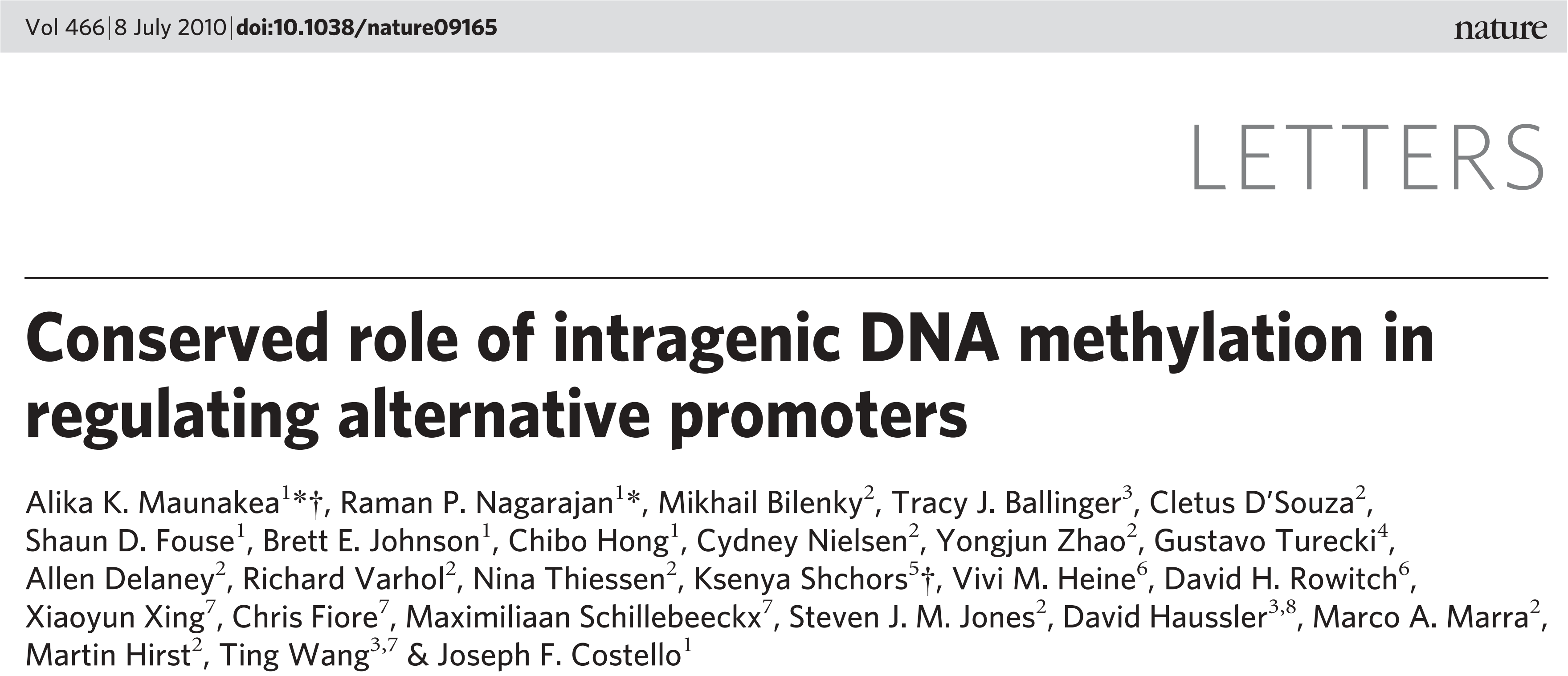
Comparison of
MethylC, RRBS, MeDIP, MeDIP, MBD



Methylome Methods Comparison

• Shotgun bisulfite	• Base resolution • Absolute quantitation • Higher cost/sample • mC not distinguished from hmC
• Enrichment	• 150bp resolution • Relative quantitation • Much lower cost/sample
• Integrative	• 1-150bp resolution • Detection of intermediate and allelic methylation states

Case 2: Functions of intragenic DNA methylation



Vol 466 | 8 July 2010 | doi:10.1038/nature09165

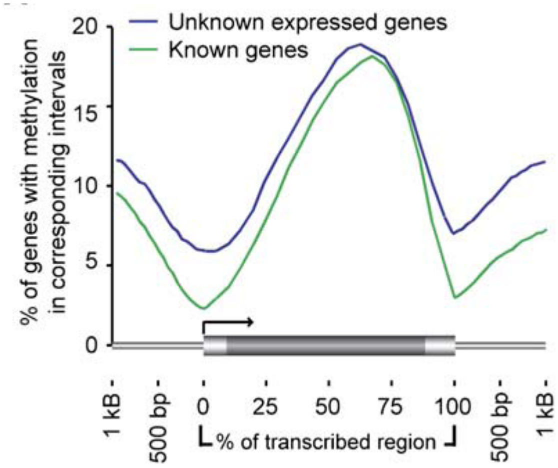
nature

LETTERS

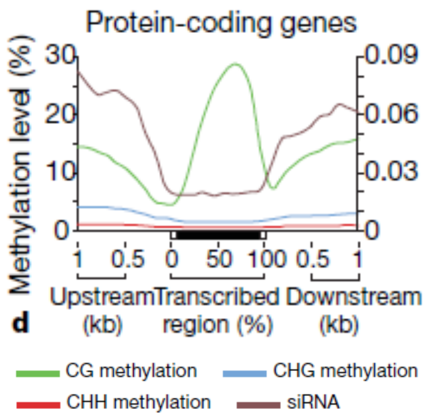
Conserved role of intragenic DNA methylation in regulating alternative promoters

Alika K. Maunakea[1]*†, Raman P. Nagarajan[1]*, Mikhail Bilenky[2], Tracy J. Ballinger[3], Cletus D'Souza[2], Shaun D. Fouse[1], Brett E. Johnson[1], Chibo Hong[1], Cydney Nielsen[2], Yongjun Zhao[2], Gustavo Turecki[4], Allen Delaney[2], Richard Varhol[2], Nina Thiessen[2], Ksenya Shchors[5]†, Vivi M. Heine[6], David H. Rowitch[6], Xiaoyun Xing[7], Chris Fiore[7], Maximiliaan Schillebeeckx[7], Steven J. M. Jones[2], David Haussler[3,8], Marco A. Marra[2], Martin Hirst[2], Ting Wang[3,7] & Joseph F. Costello[1]

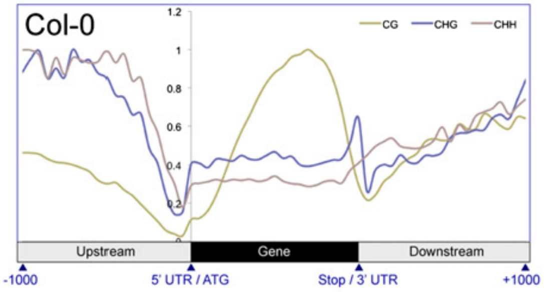
What is the function of gene body methylation?



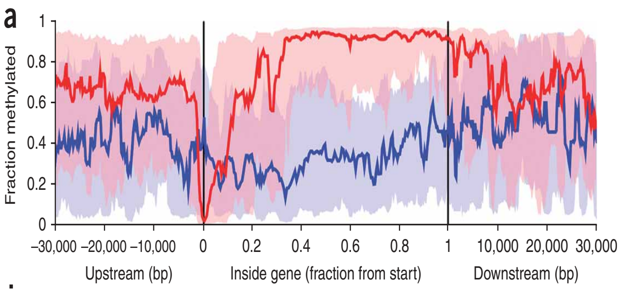
Zhang et al 2006
MeDIP-chip, Arabidopsis



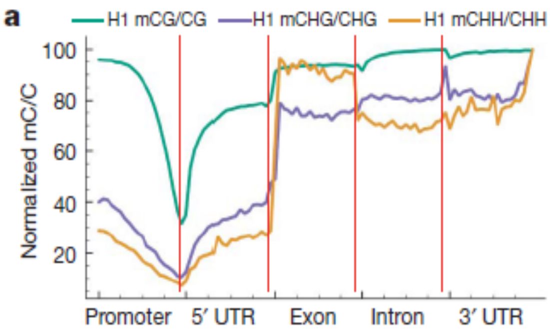
Cokus et al 2008,
BS-seq, Arabidopsis



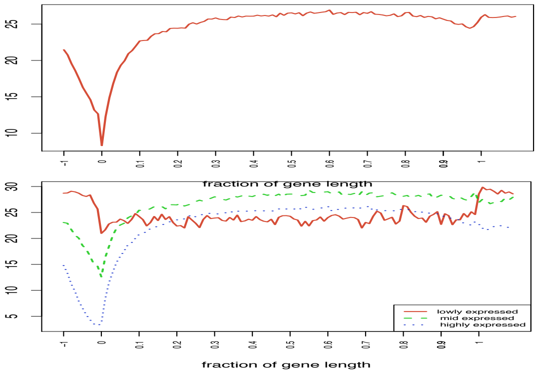
Lister et al 2008,
MethylC-seq, Arabidopsis



Ball et al 2009,
bisulfite padlock probes, human cell line

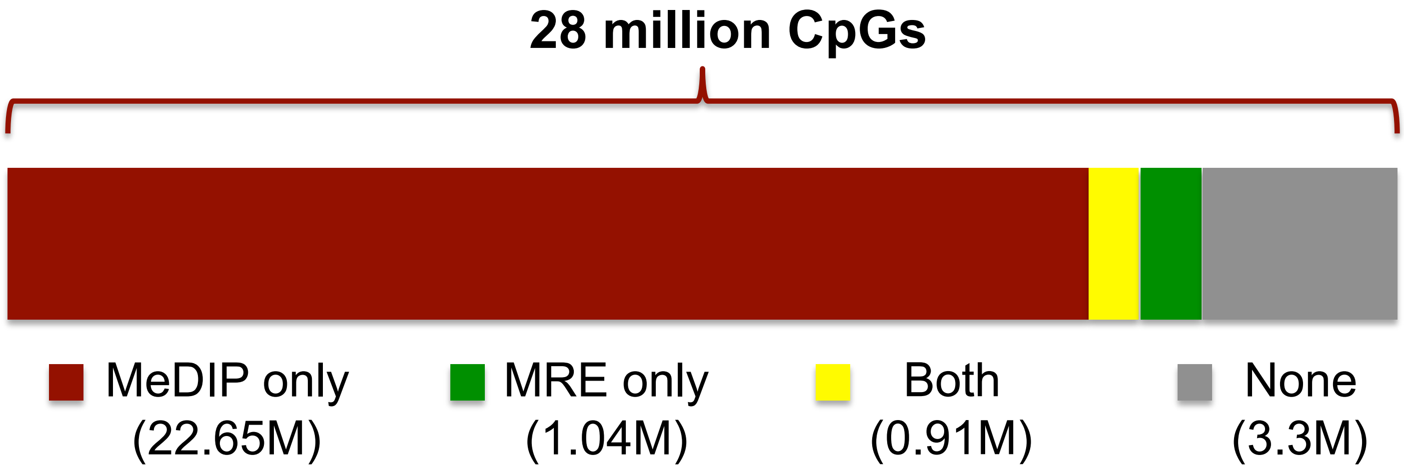


Lister et al 2009,
MethylC-seq, human ES cells



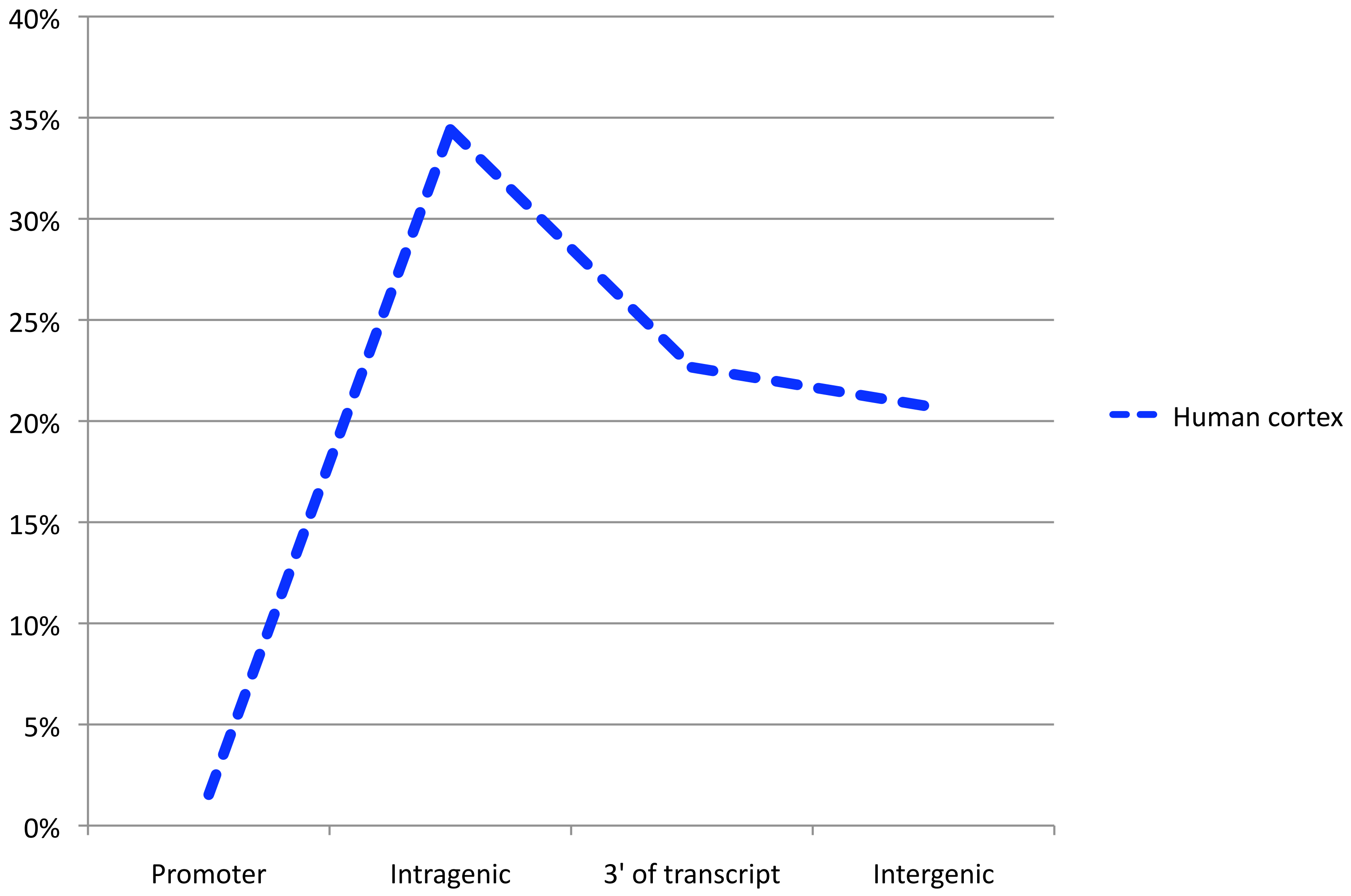
Maunakea et al 2010,
MeDIP+MRE, human brain

Genome-wide CpG site coverage

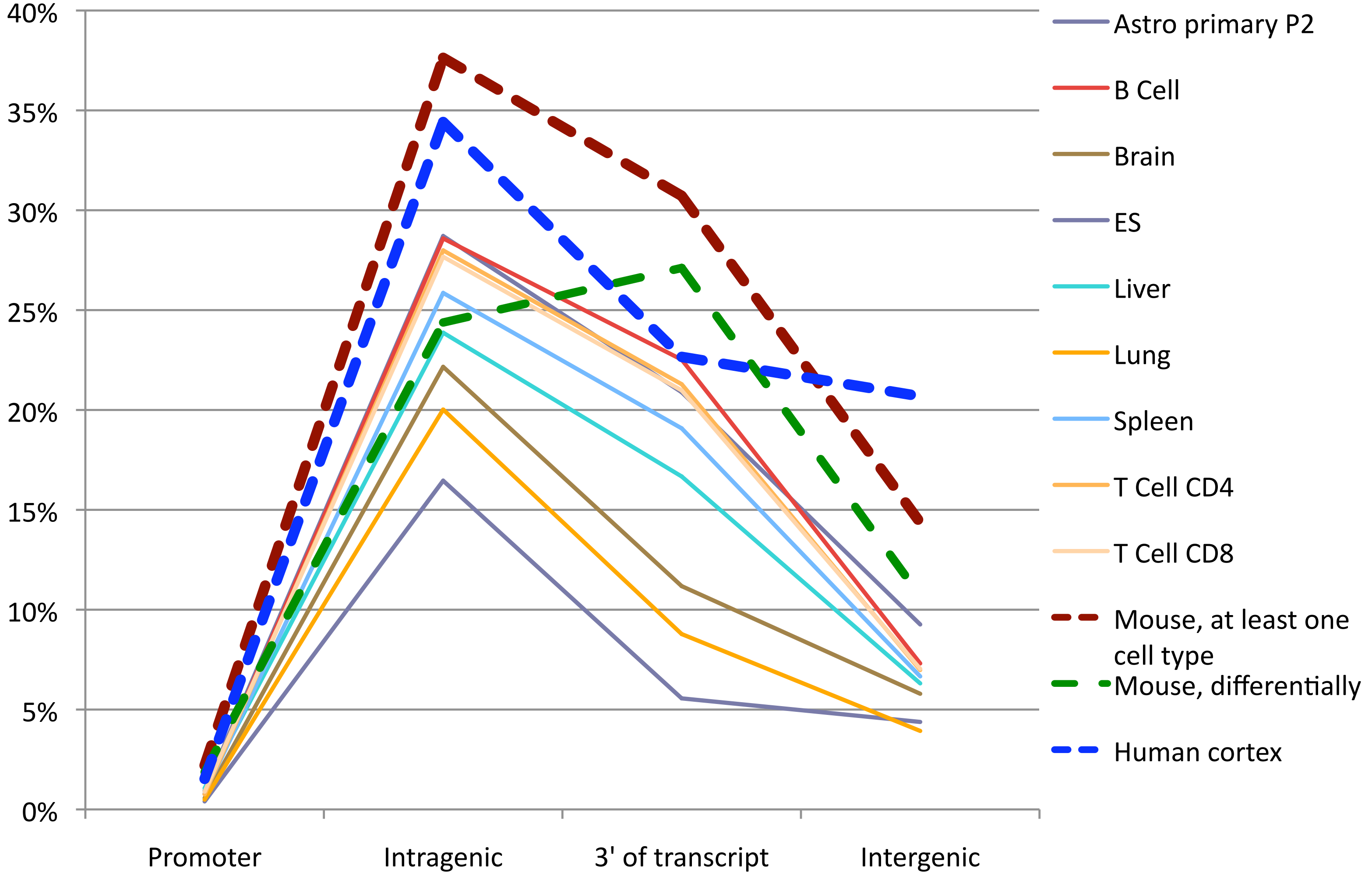


- Sample: human brain (frontal cortex)
- MeDIP-seq: 100 million reads
- MRE-seq: 30 million reads

Intragenic CpG island methylation is common; 5’ methylation is rare



Intragenic CpG island methylation is common; 5' methylation is rare

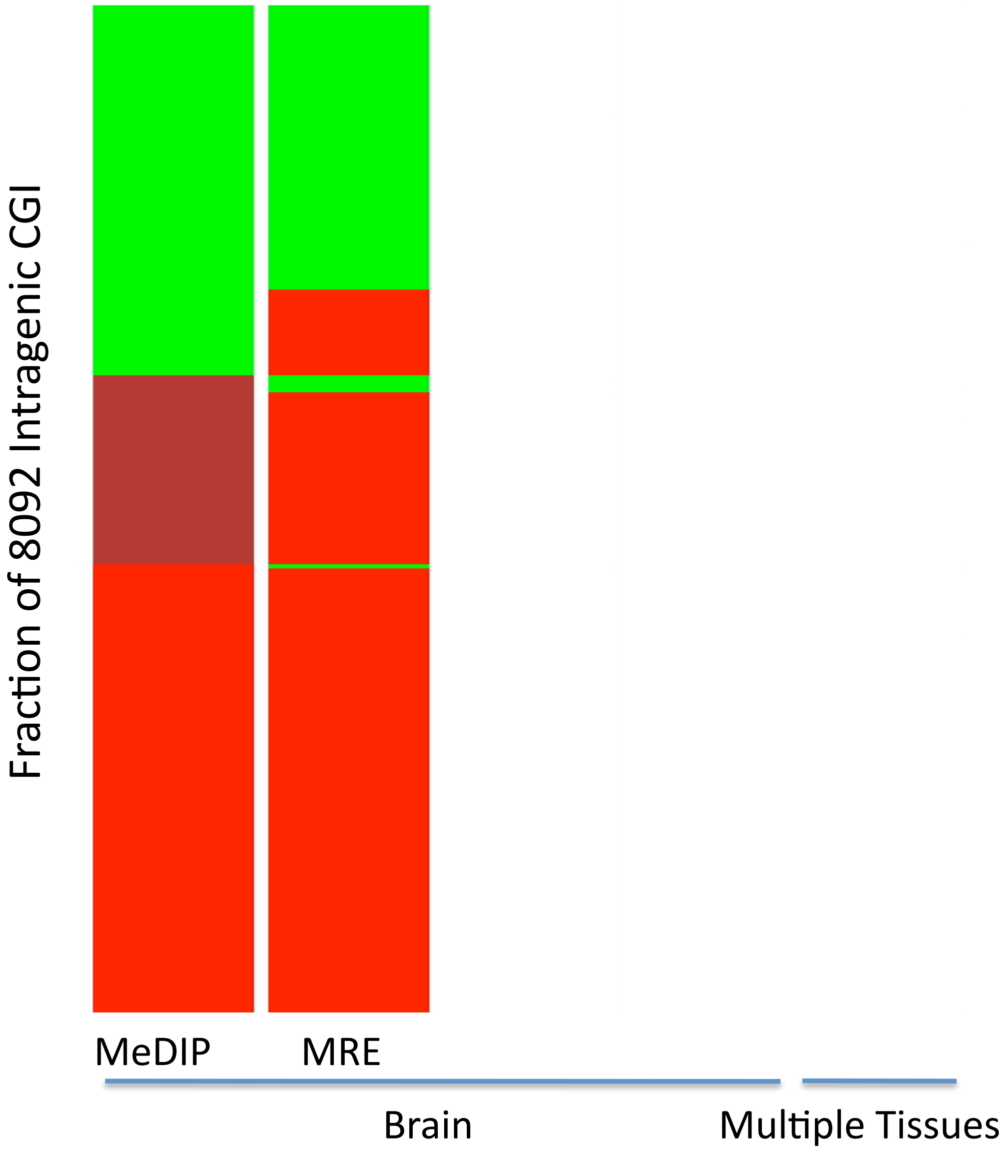


Mouse data from: Meissner, Mikkelsen et al, 2008

Methylation status of orthologous CpG islands is evolutionarily conserved

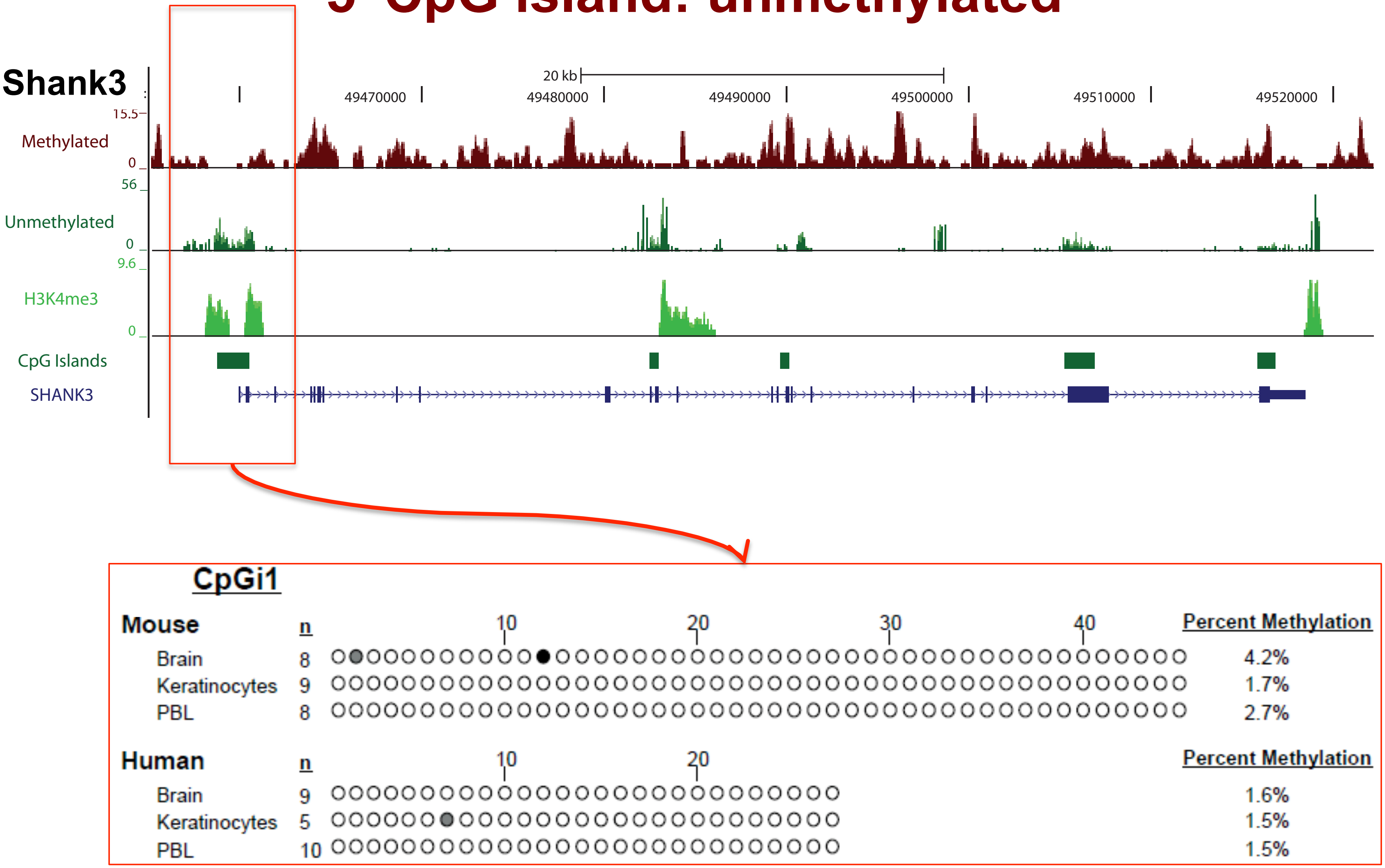
	Promoter CGIs (n=1245)		**Intragenic CGIs (n=502)**	
	Unmethylated in human	Methylated in human	Unmethylated in human	Methylated in human
Unmethylated in mouse	1234	3	293	66
Methylated in mouse	6	2	36	107
P-value	**< 0.0001**		**< 0.0001**	

Overlap of H3K4me3, CAGE and Intragenic CpG island

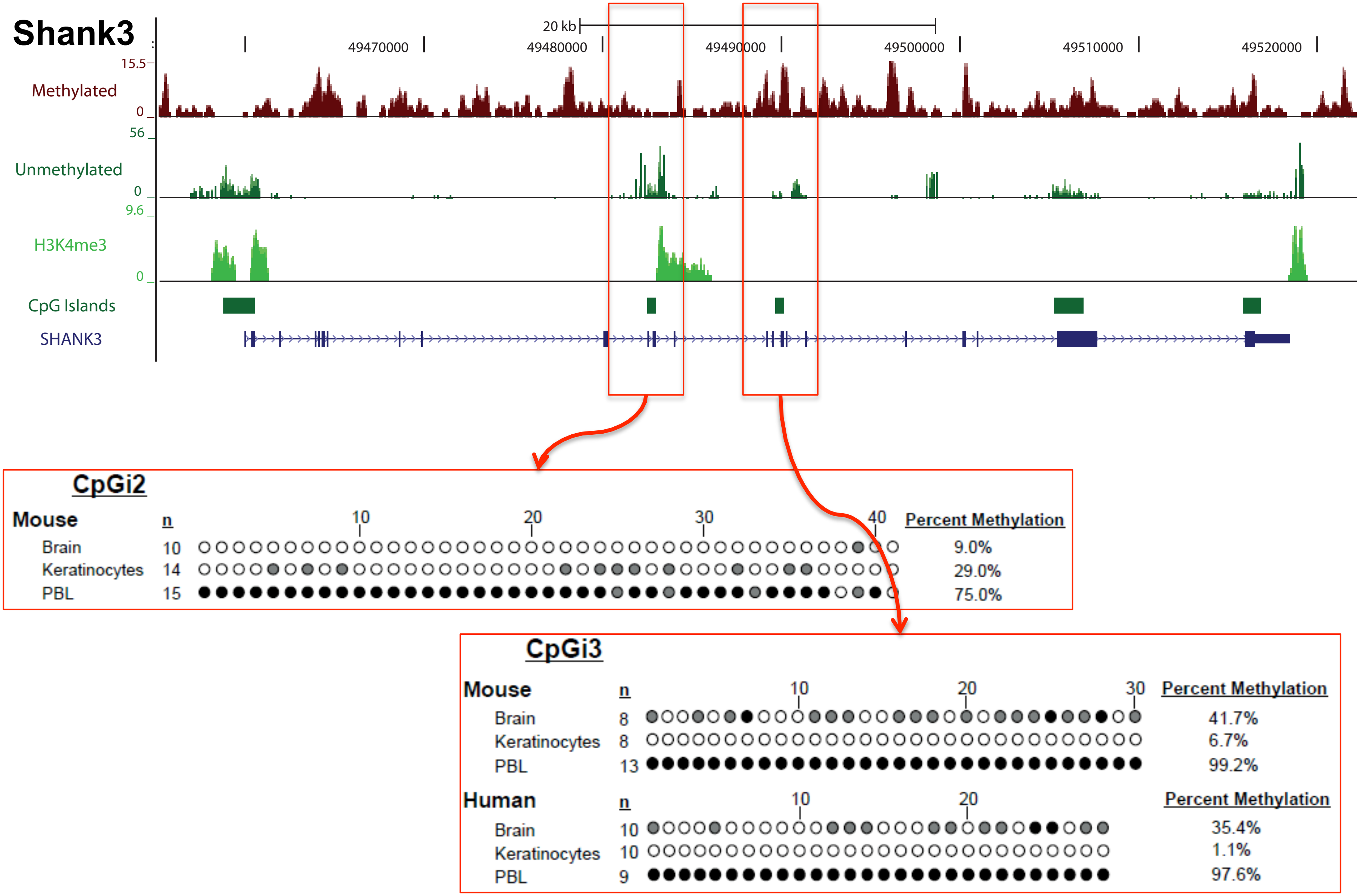


CAGE data from
Carninci et al
2005, 2008

5’ CpG island: unmethylated



Intragenic CpG island: tissue specific methylation



CpGi2

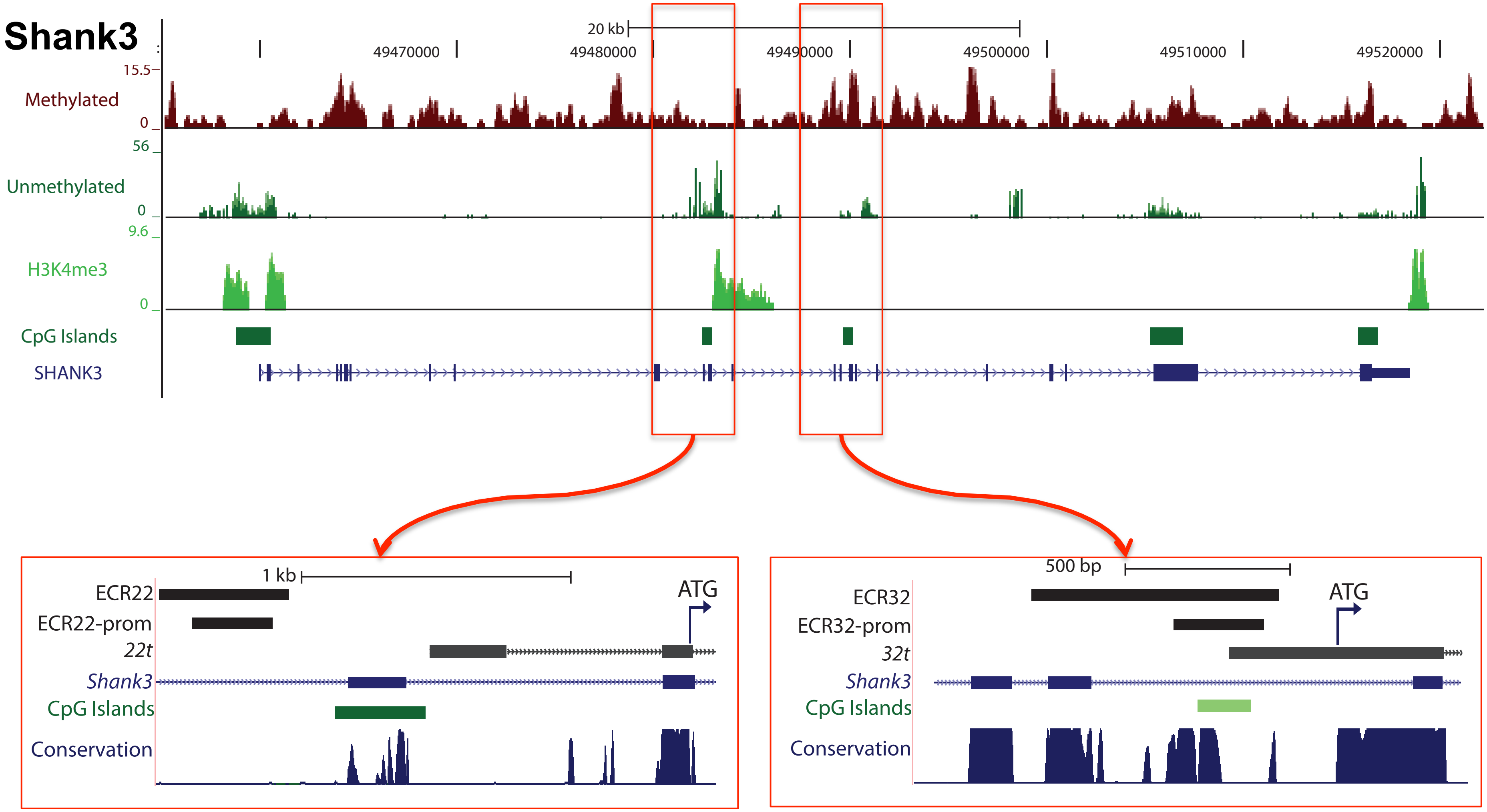
Mouse	n	Percent Methylation
Brain	10	9.0%
Keratinocytes	14	29.0%
PBL	15	75.0%

CpGi3

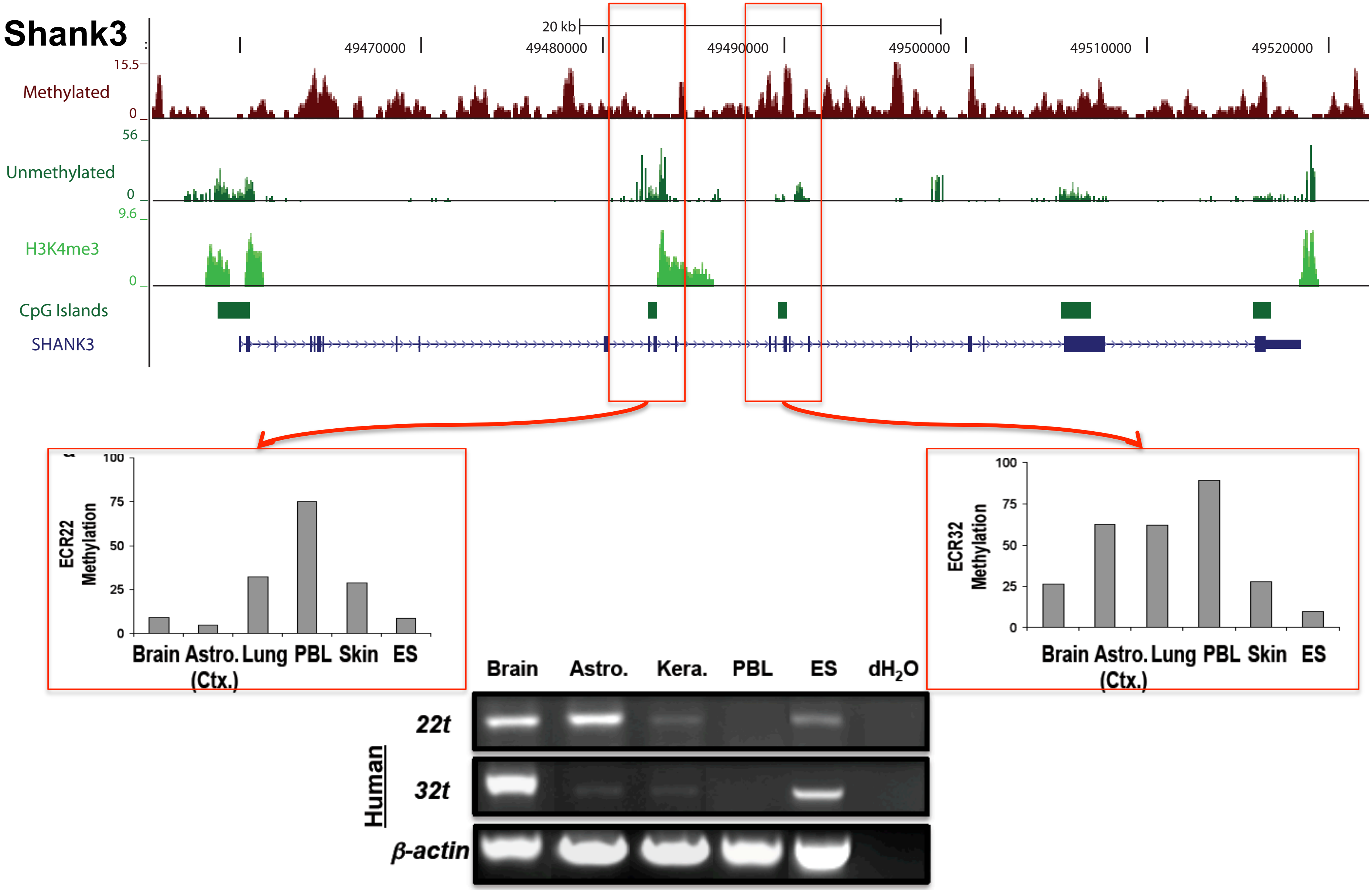
Mouse	n	Percent Methylation
Brain	8	41.7%
Keratinocytes	8	6.7%
PBL	13	99.2%

Human	n	Percent Methylation
Brain	10	35.4%
Keratinocytes	10	1.1%
PBL	9	97.6%

Intragenic CpG island initiates transcripts



Methylation level anti-correlates with transcript level



Initial observations from brain epigenomes

- Tissue-specific methylation is common in intragenic CpG islands, but rare in 5’ islands
- Genome sites of tissue-specific intragenic methylation overlap significantly with markers of TSS, and promoter-enriched histone modifications
  - Suggesting a major role for intragenic methylation in regulating cell context-specific alternative promoters in gene bodies

Case 3: Combining MeDIP and MRE to investigate intermediate methylation

- Integrative method identifies intermediate methylation states
- Genetic variation detected by epigenotyping
- Initial catalogue of novel imprinted gene candidates

Allele specific methylation

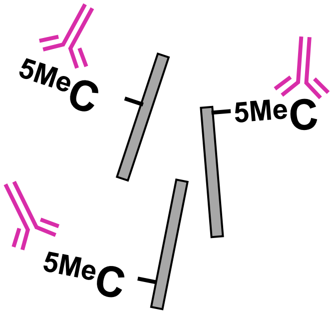
- Imprinting	- Parent-of-origin specific expression - Normal growth and brain development - 100 genes known, but many DMRs unknown (Chaofani, 2011, others)
- X-chromosome inactivation	- In Females, promoters vs gene bodies
- Monoallelic gene regulation	- Widespread on somatic chromosomes (Hellman and Chess, etc)

Integrative Method

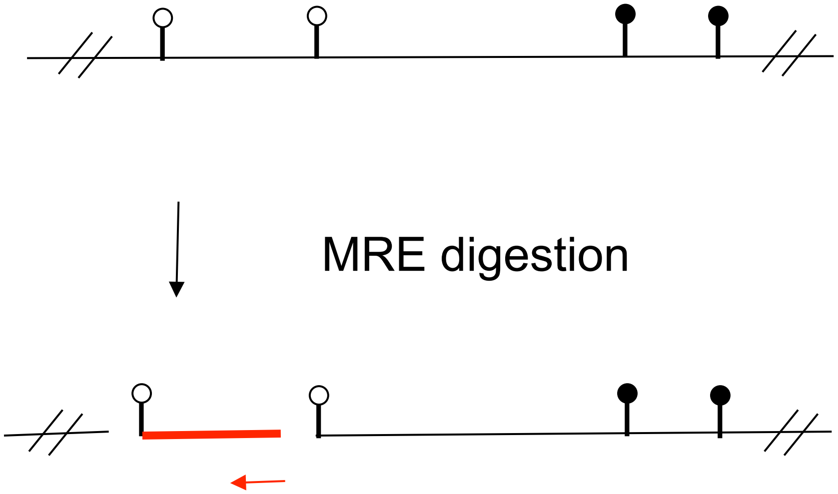
- Methyl-sensitive restriction enzyme – sequencing (MRE-seq)
  each read is a single unmethylated CpG site

- Methyl DNA immunoprecipitation - sequencing (MeDIP-seq)
  higher read density at methylated regions

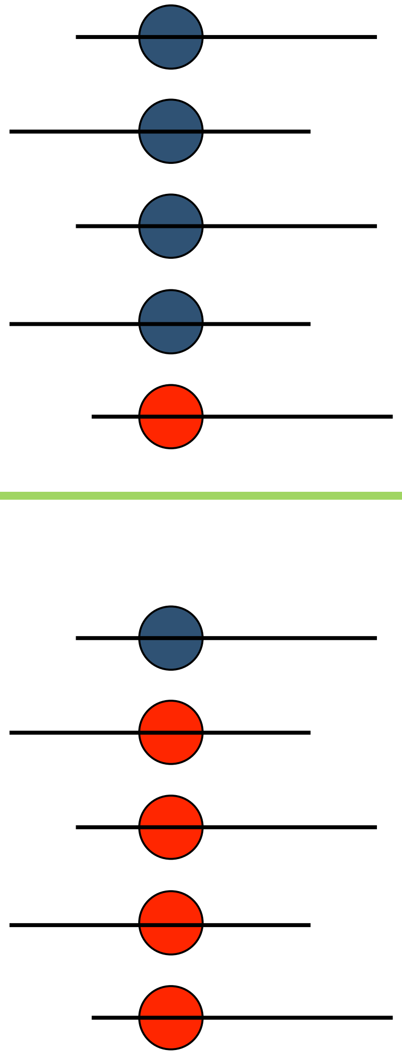
MeDIP-seq



MRE-seq



Allele Specific Methylation (ASM)



Unmethlyated CpGs
(MRE-Seq)

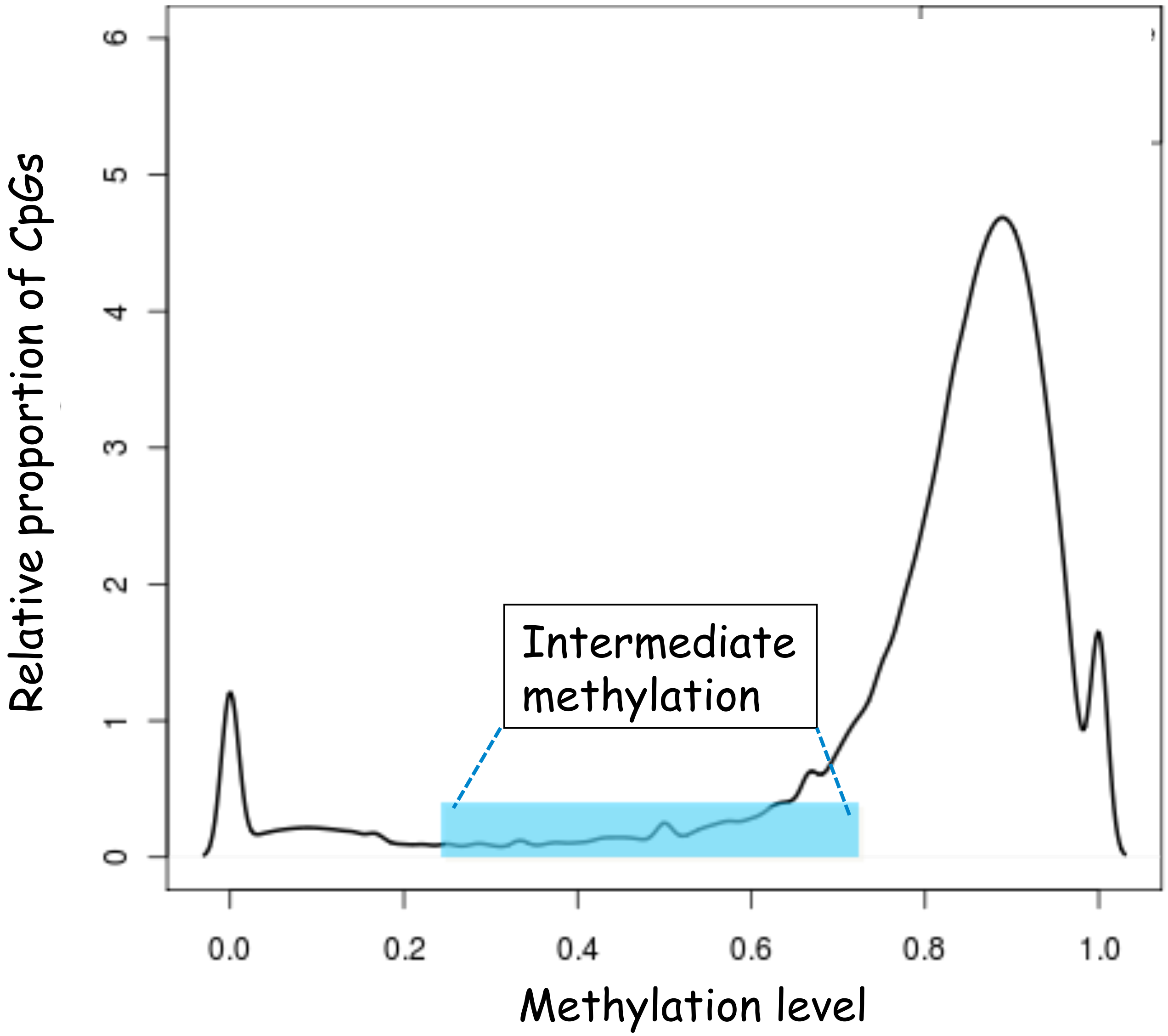
● Reference Allele

● Variant Allele

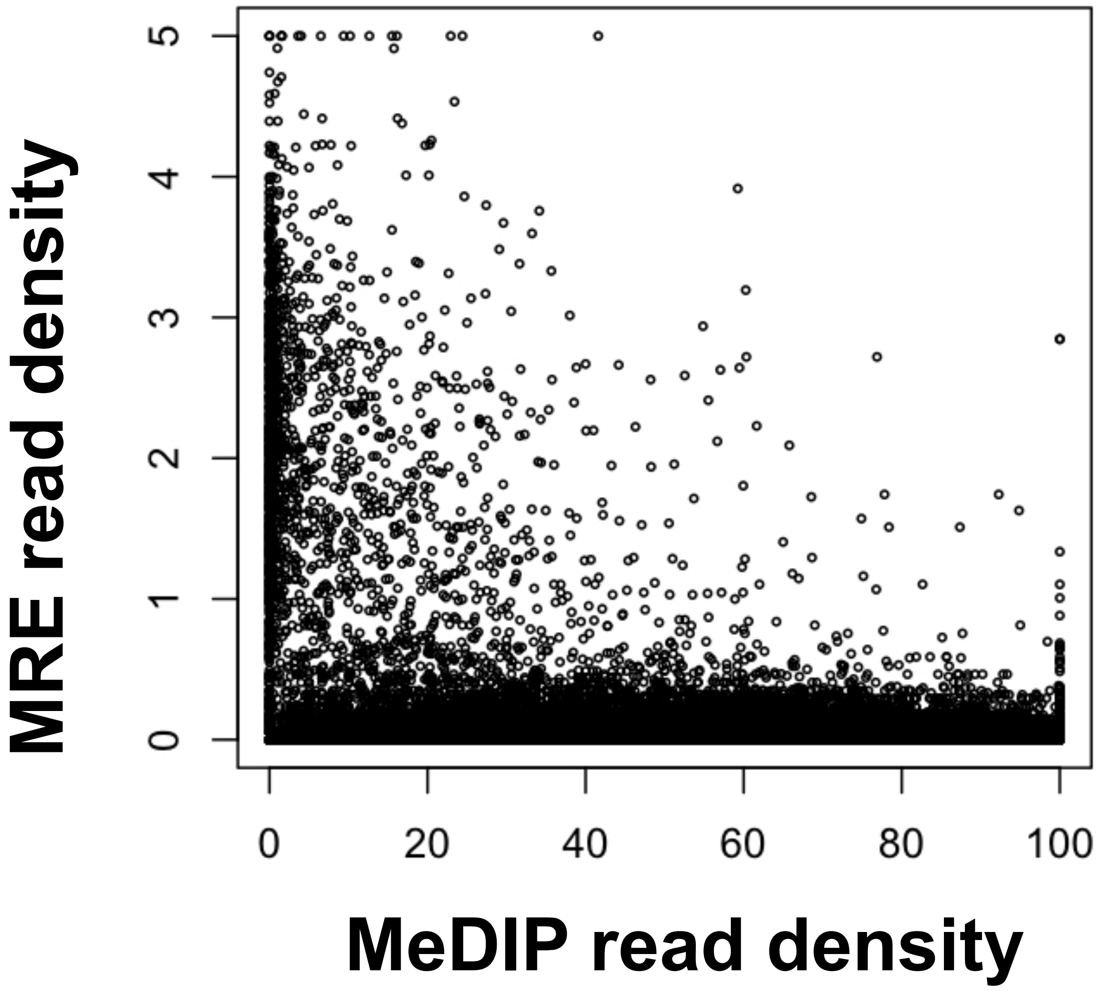
Methlyated CpGs
(MeDIP-Seq)

**Harris et al,
Nature Biotechnology 2010**

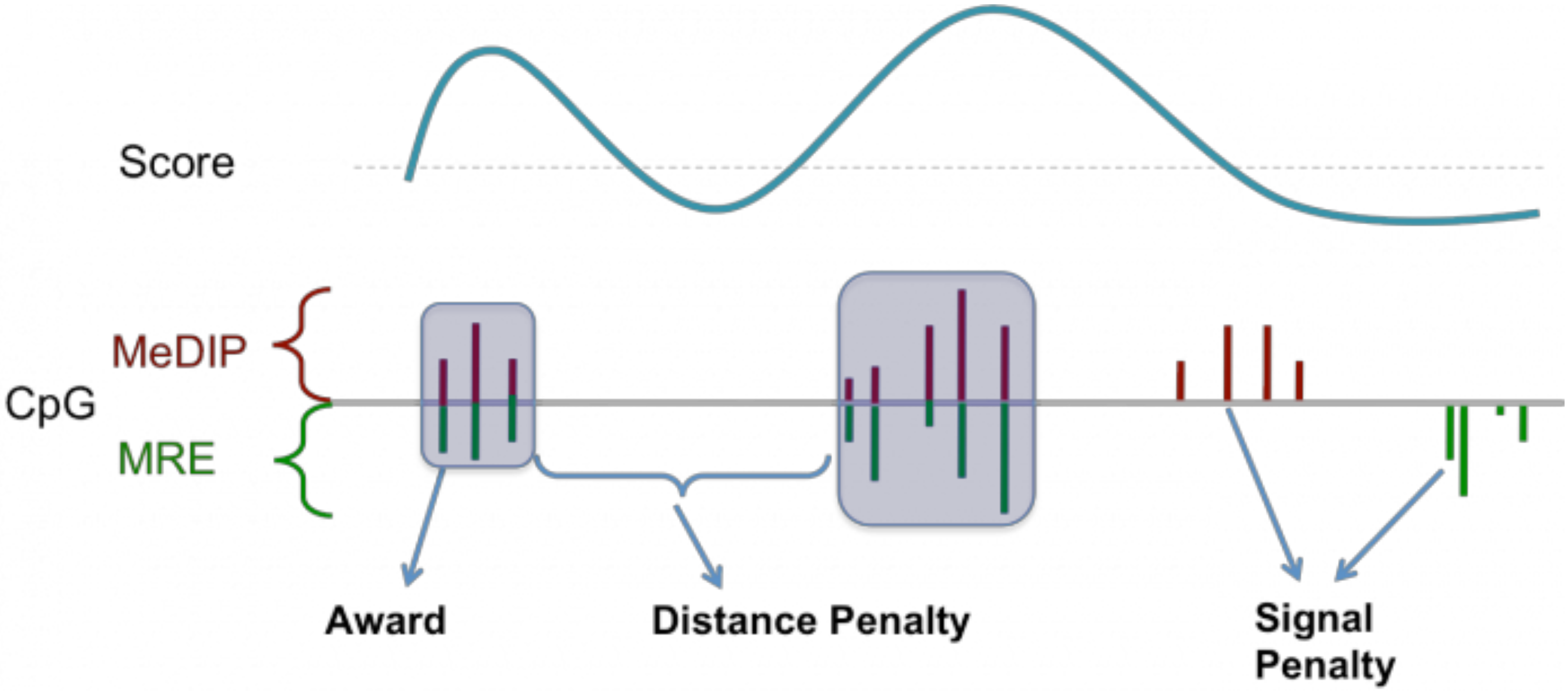
The genome is divided into methylated and unmethylated domains



Complementarity of MeDIP-seq and MRE-seq

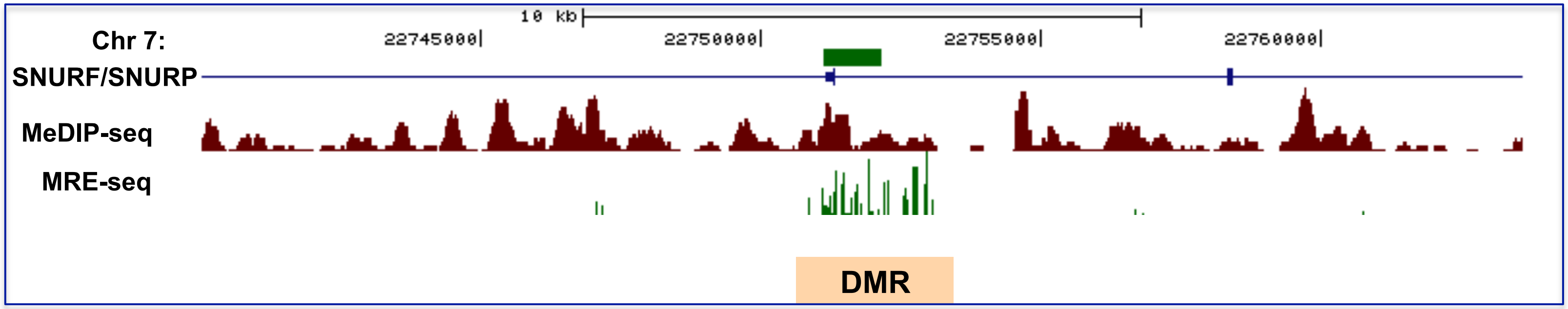


The “iMethylome” algorithm

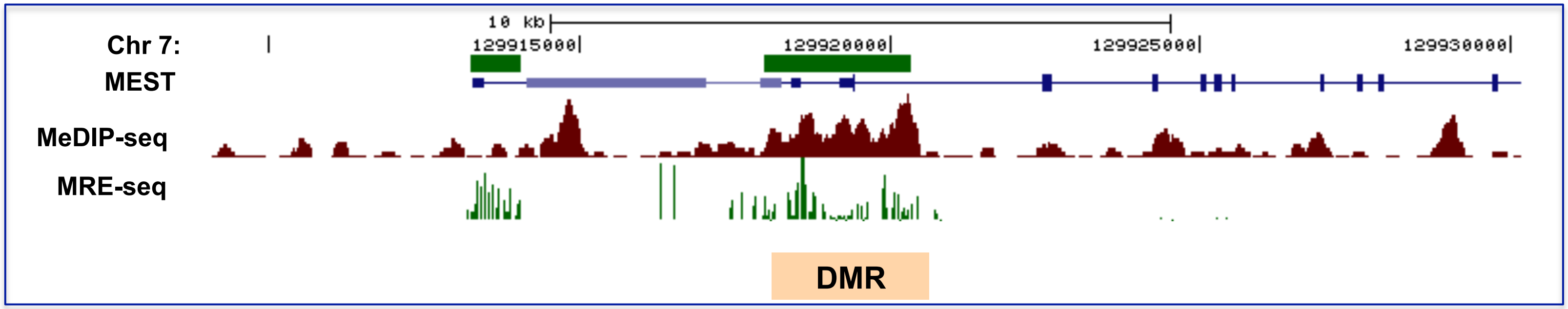


Allele-specific methylation at imprinted genes

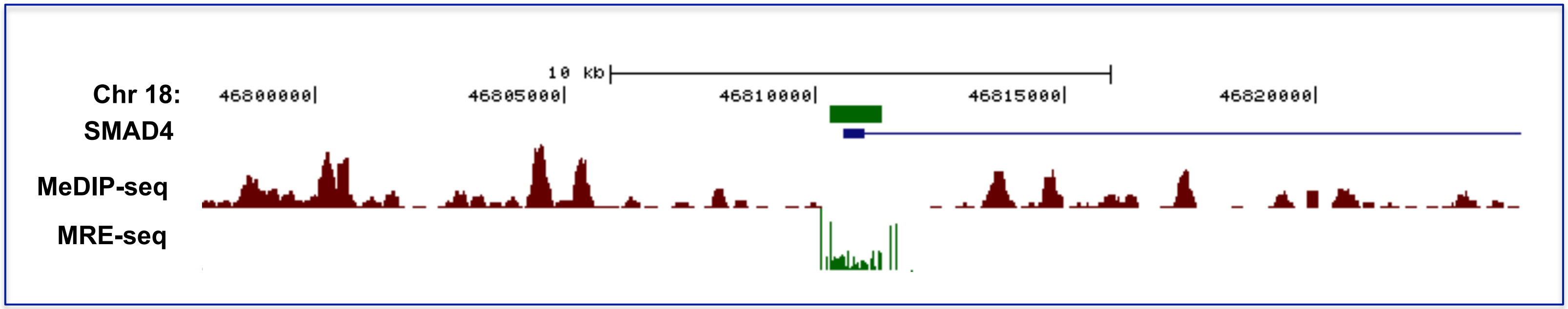
imprinted



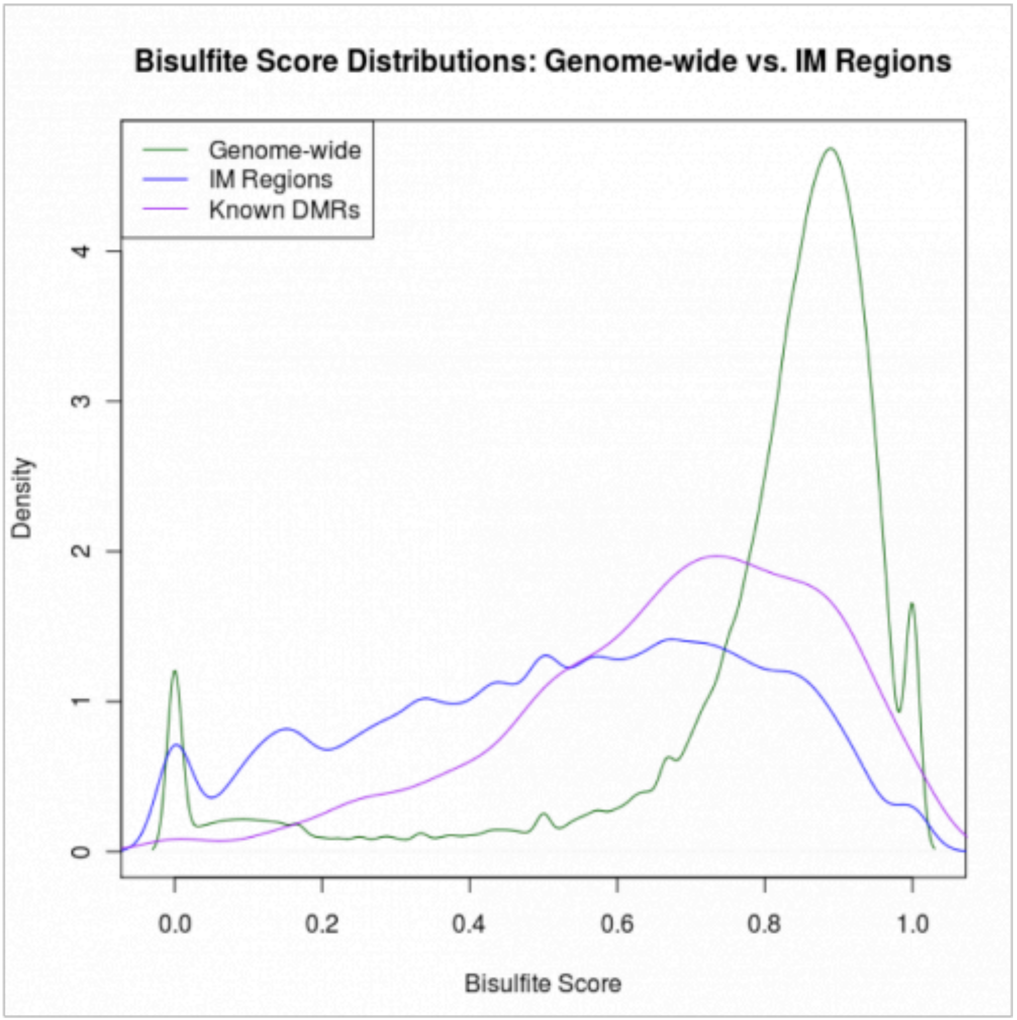
imprinted

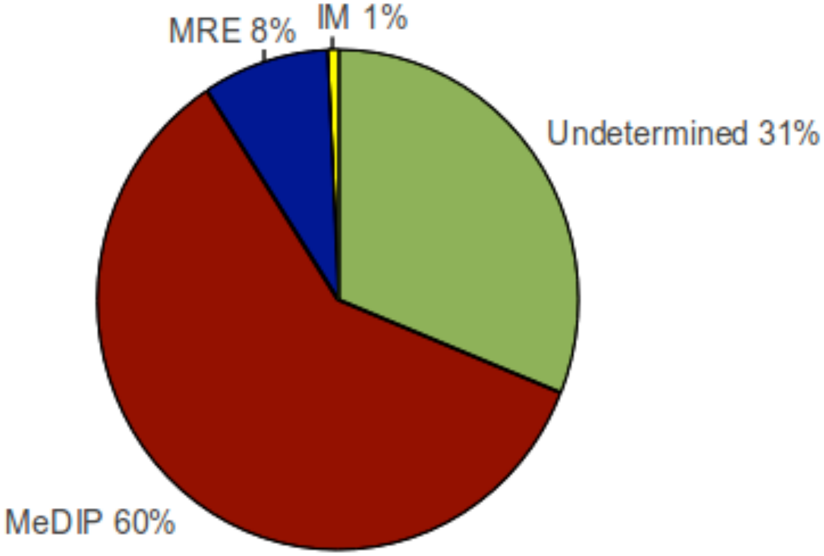


not imprinted



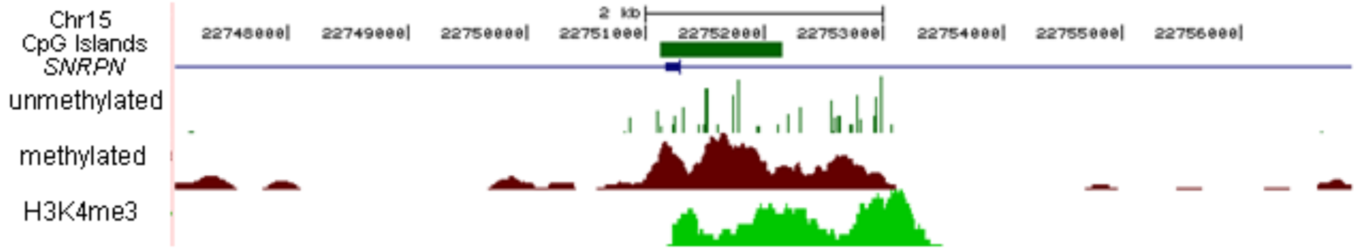
Enriching for Intermediate Methylation



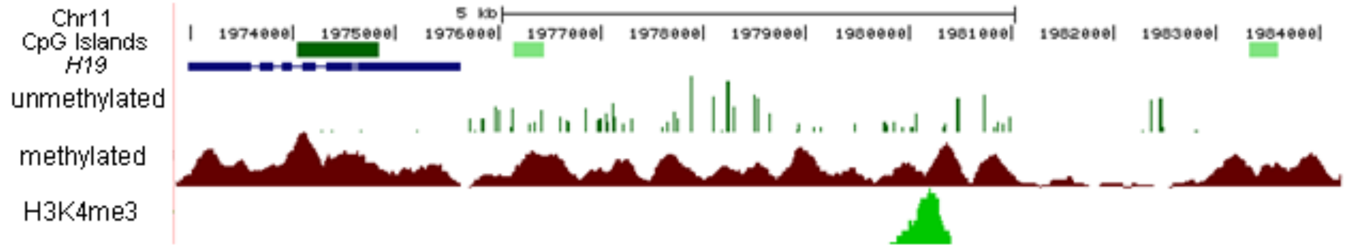


Intermediate methylation levels at imprinted genes

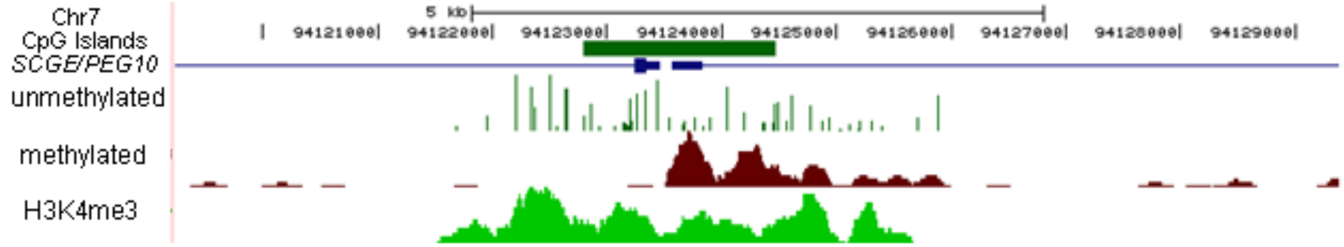
SNRPN



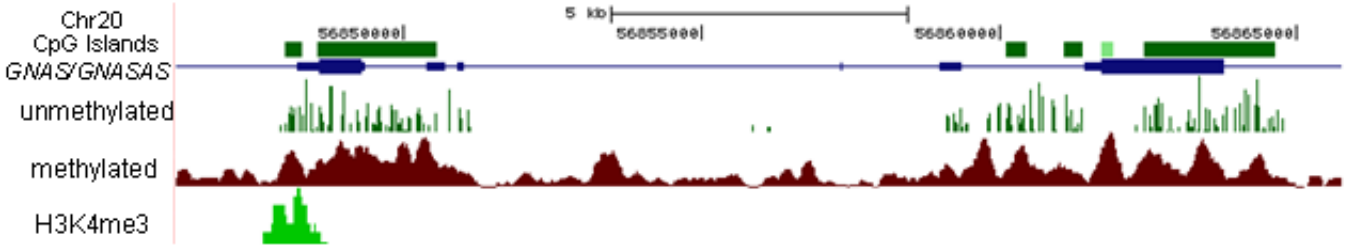
H19



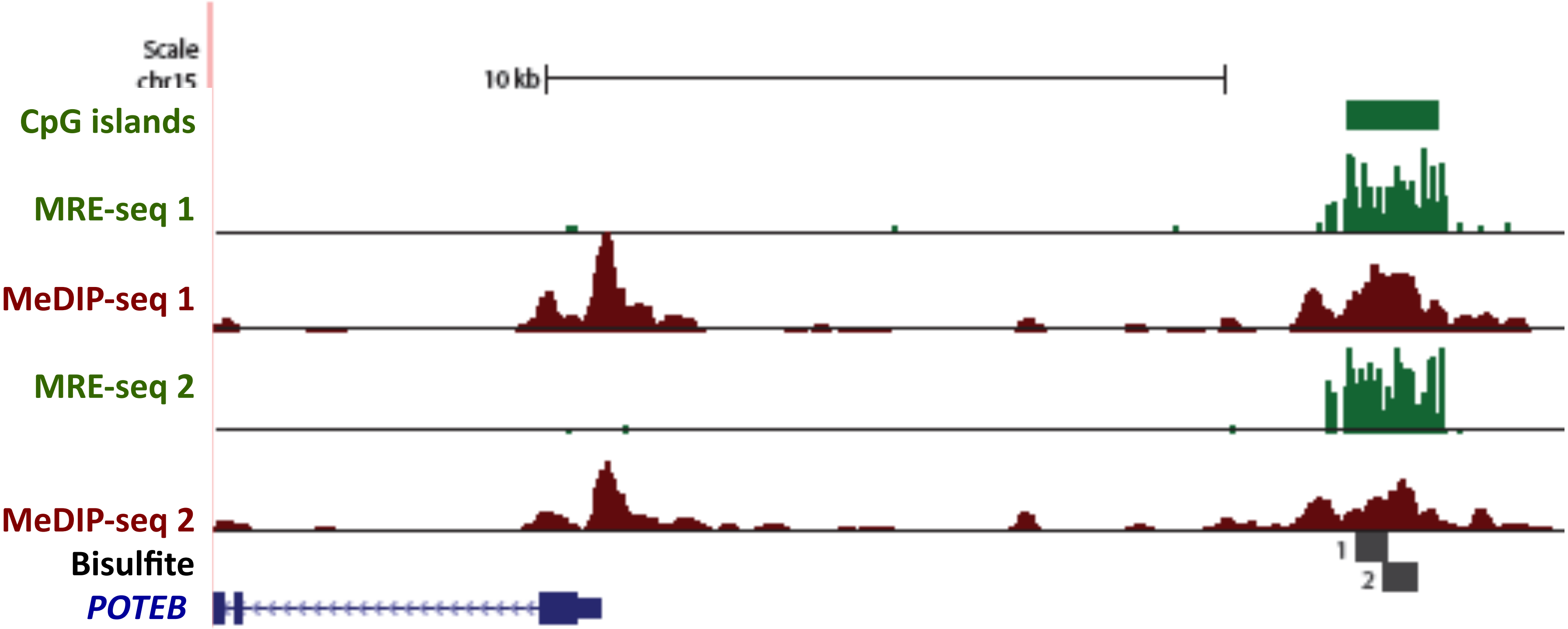
PEG10



GNAS

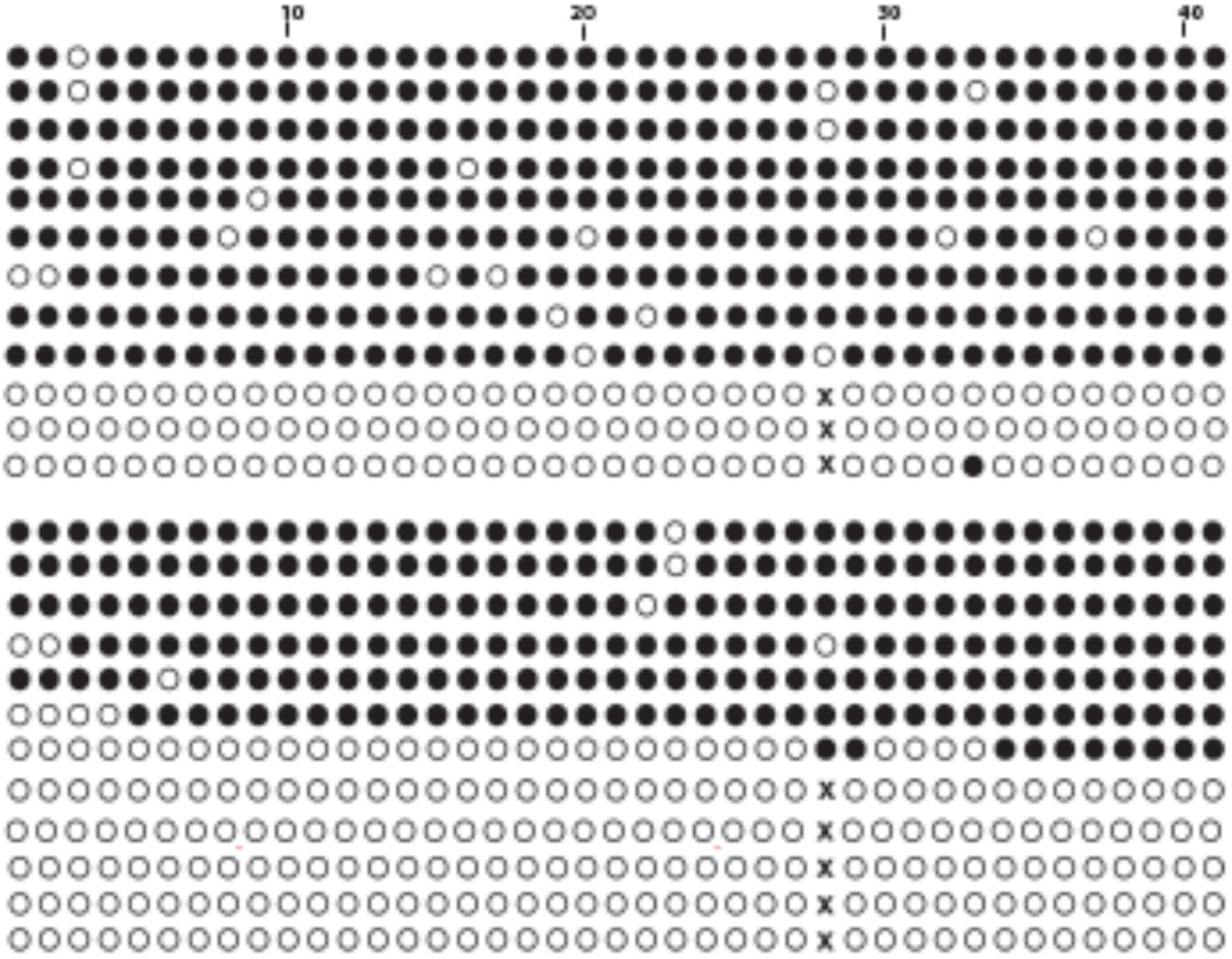


Intermediate methylation levels in *POTEB*

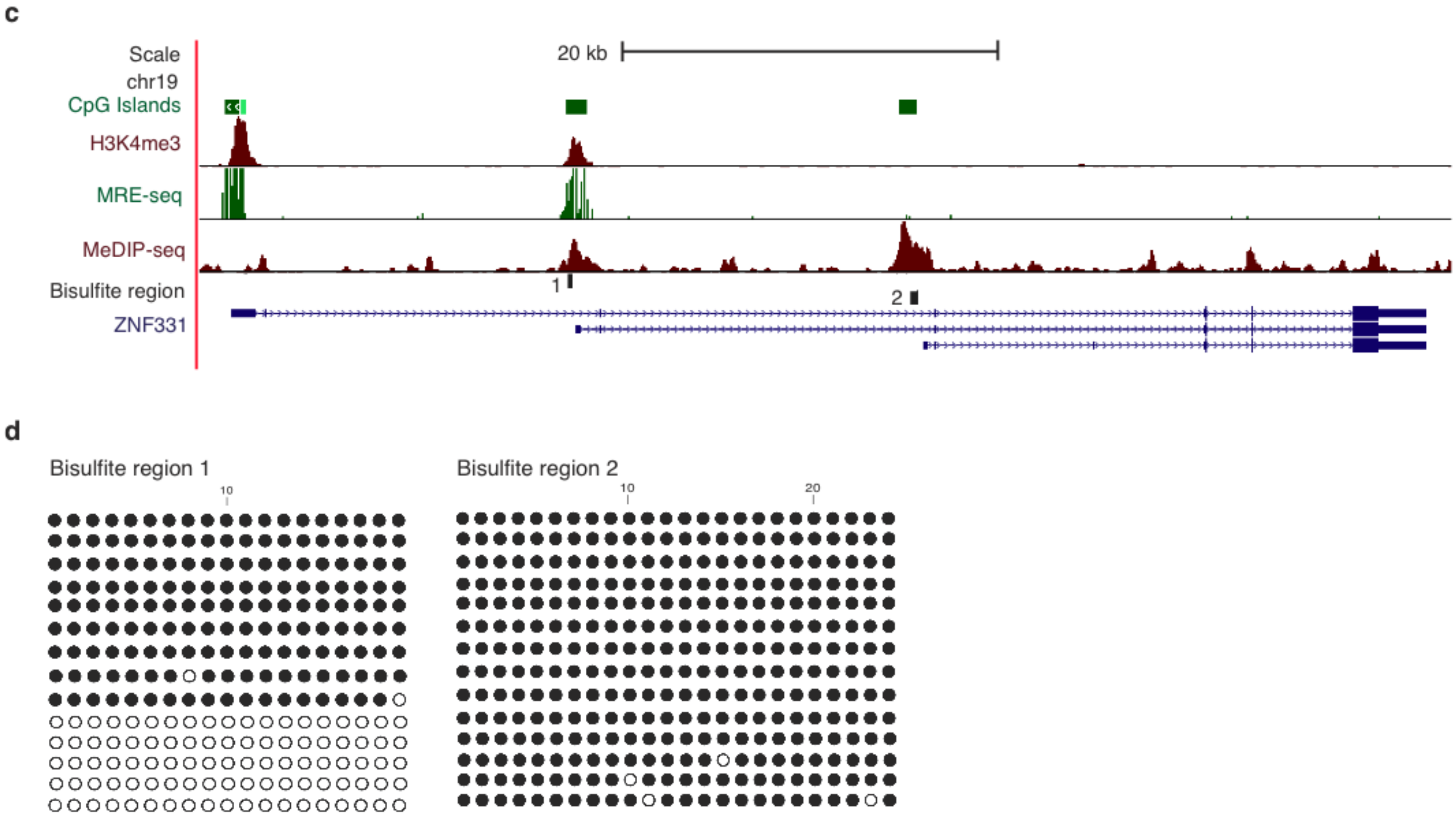


Location	Medip Allele	Count	MRE Allele	Count
chr15:19346666-19350003	G	9	A	30

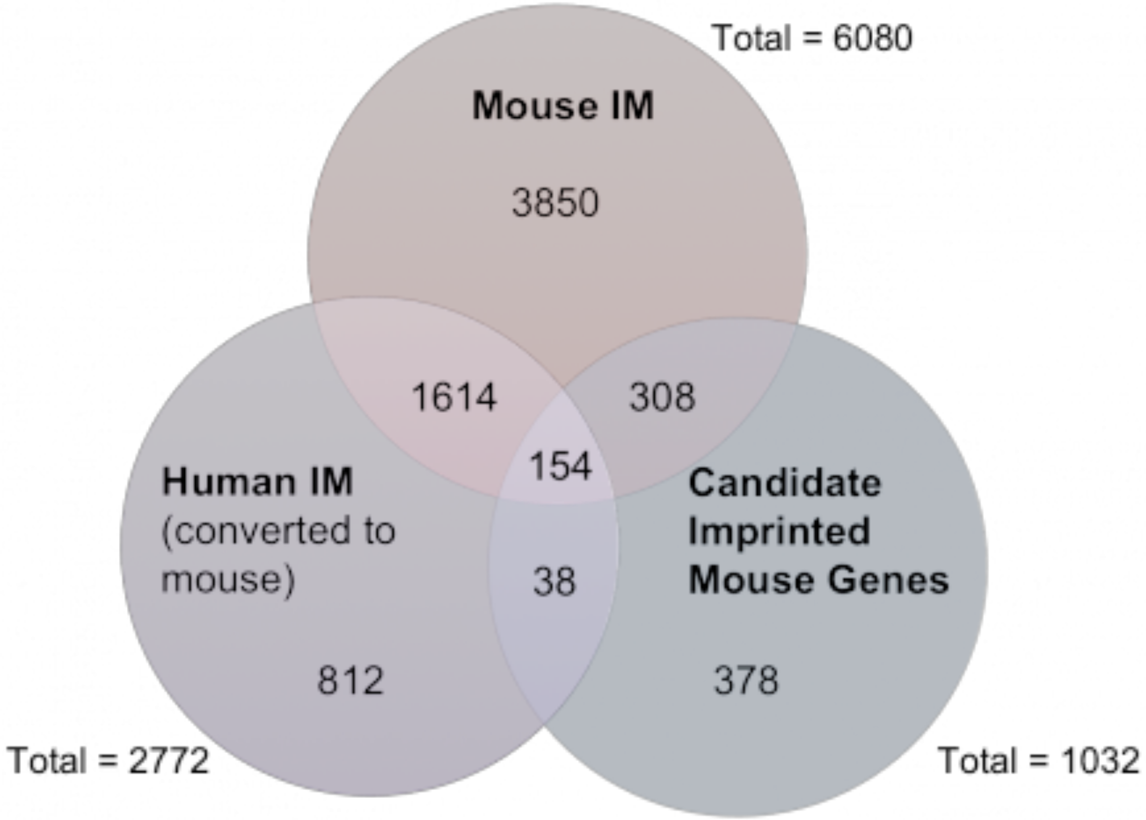
Validation of monoallelic DNA methylation in *POTEB*



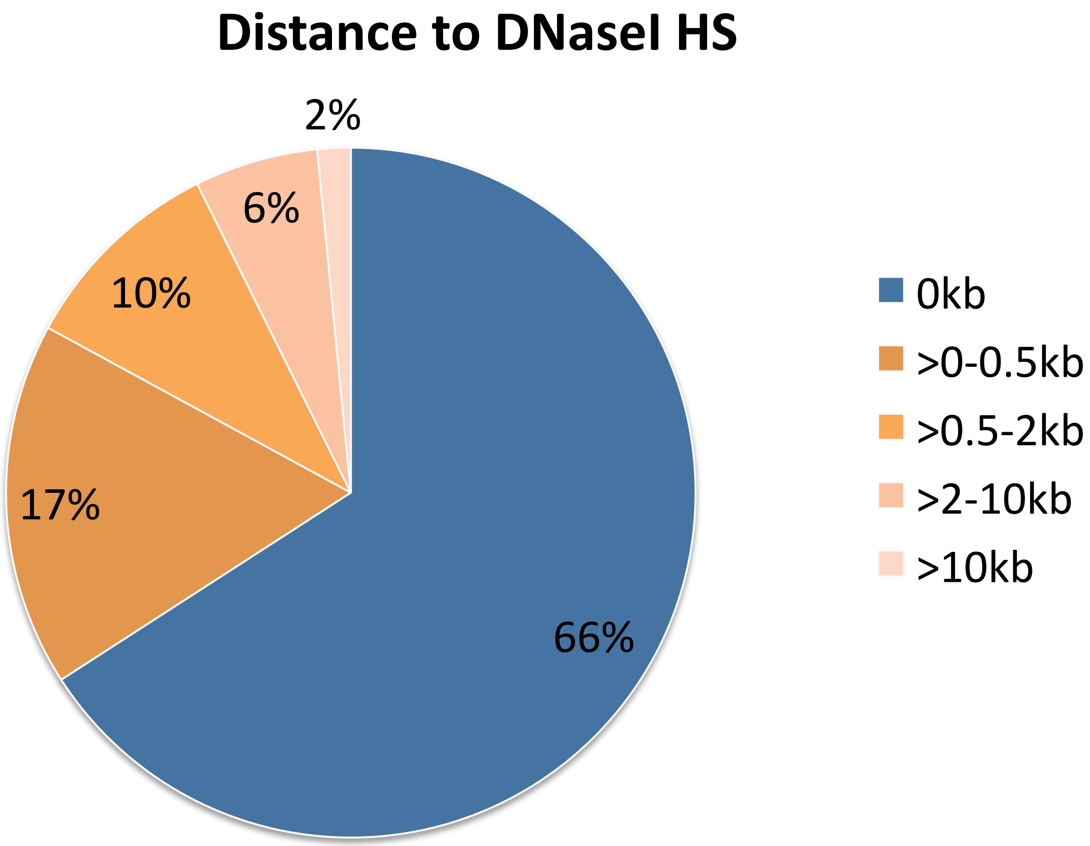
Putative Imprinted Gene



Conservation of IM



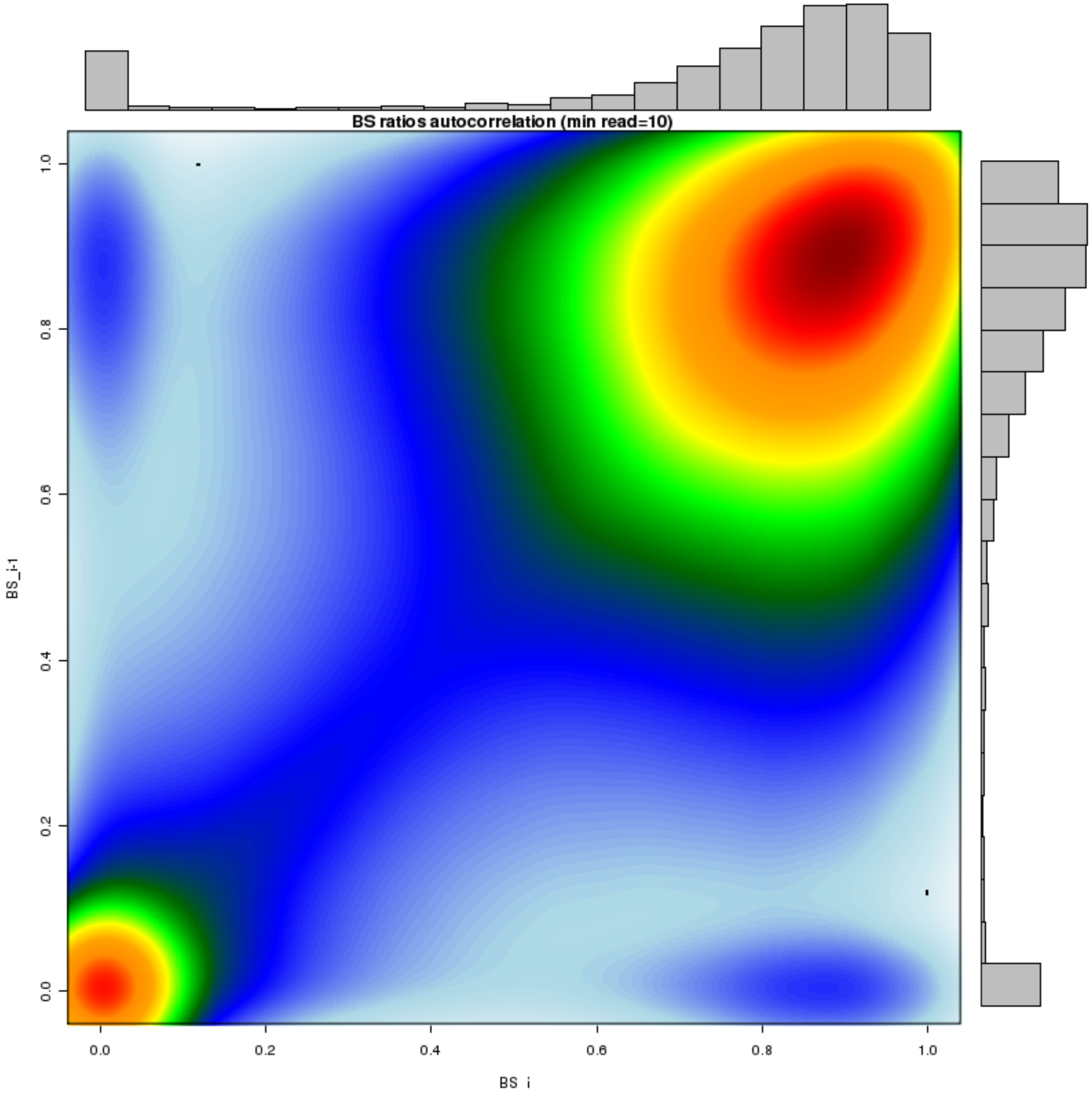
IMs are regulatory elements



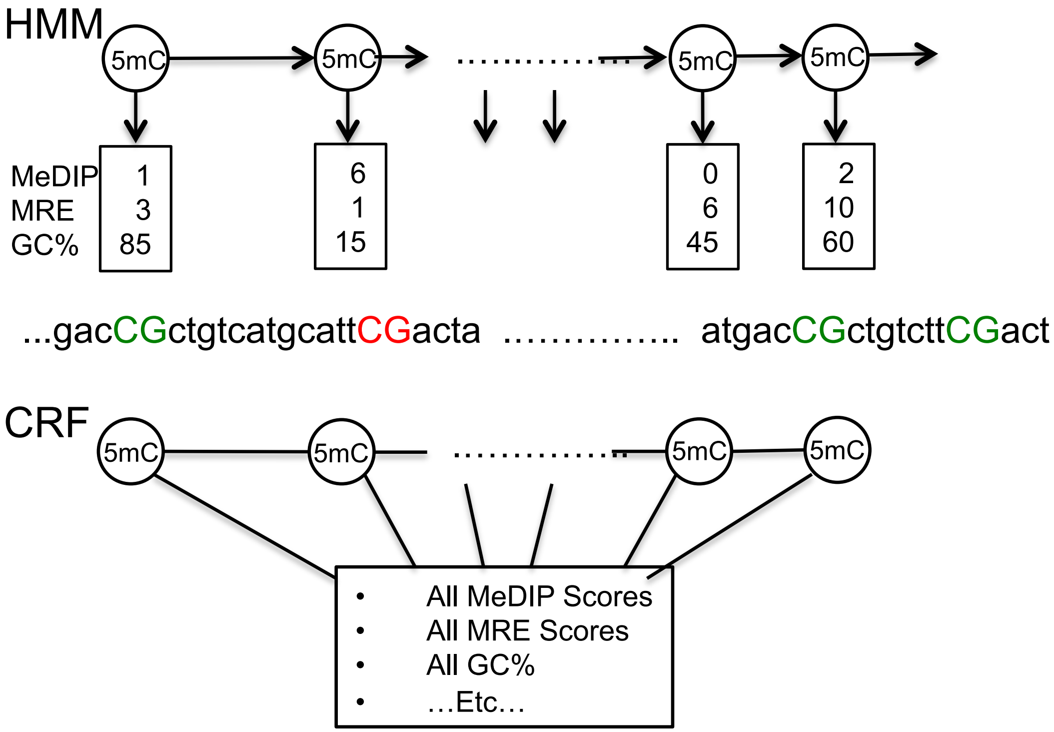
Case 4: Predicting single CpG methylation level with Conditional Random Field (methylCRF)

1. A novel statistical framework for integrative analysis of MeDIP and MRE data

DNA Methylation is locally correlated



A conditional random field model


HMM

5mC 5mC 5mC 5mC

MeDIP 1 6 0 2

MRE 3 1 6 10

GC% 85 15 45 60

...gacCGctgtcatgcattCGacta …………….. atgacCGctgtcttCGact

CRF

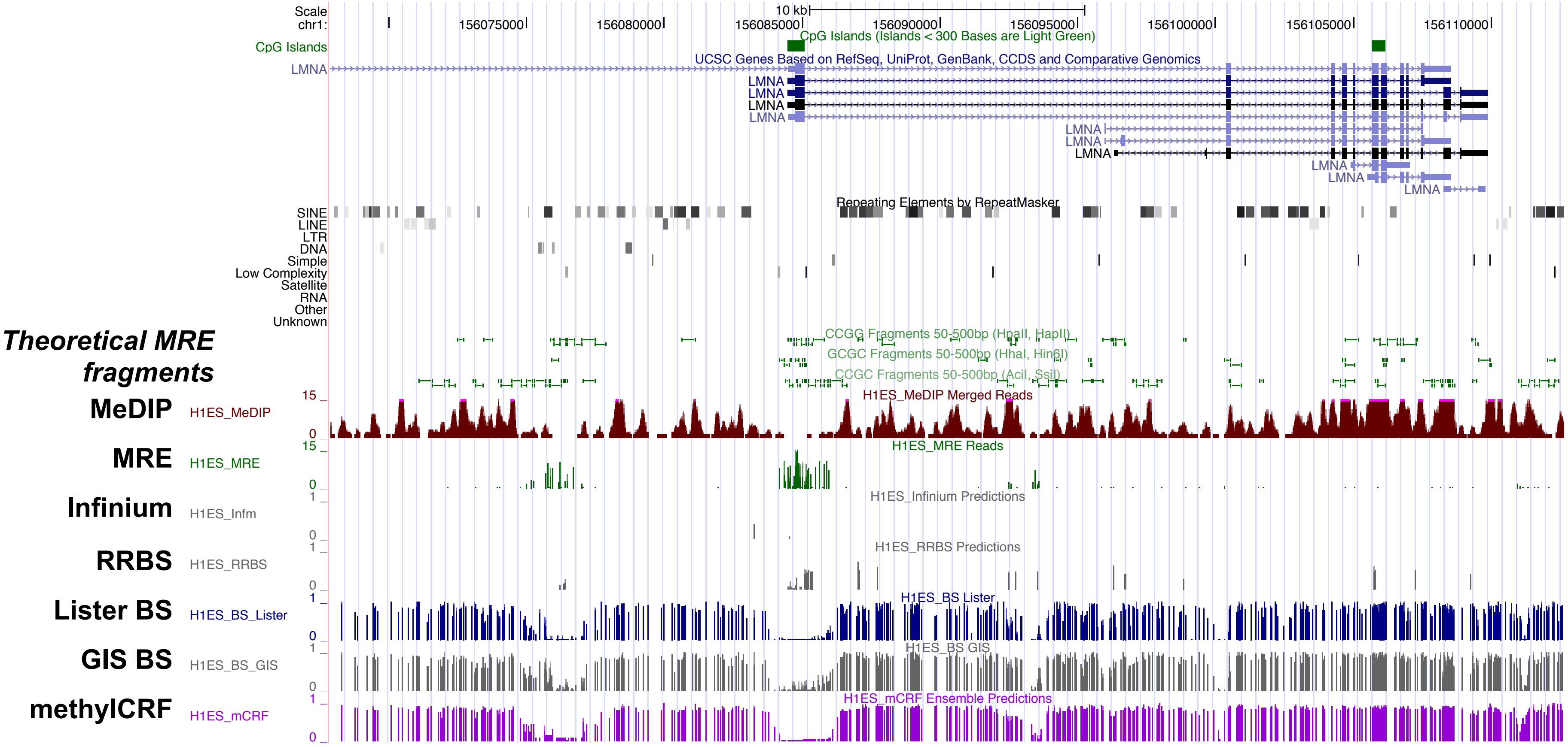
5mC 5mC 5mC 5mC

- All MeDIP Scores
- All MRE Scores
- All GC%
- …Etc…

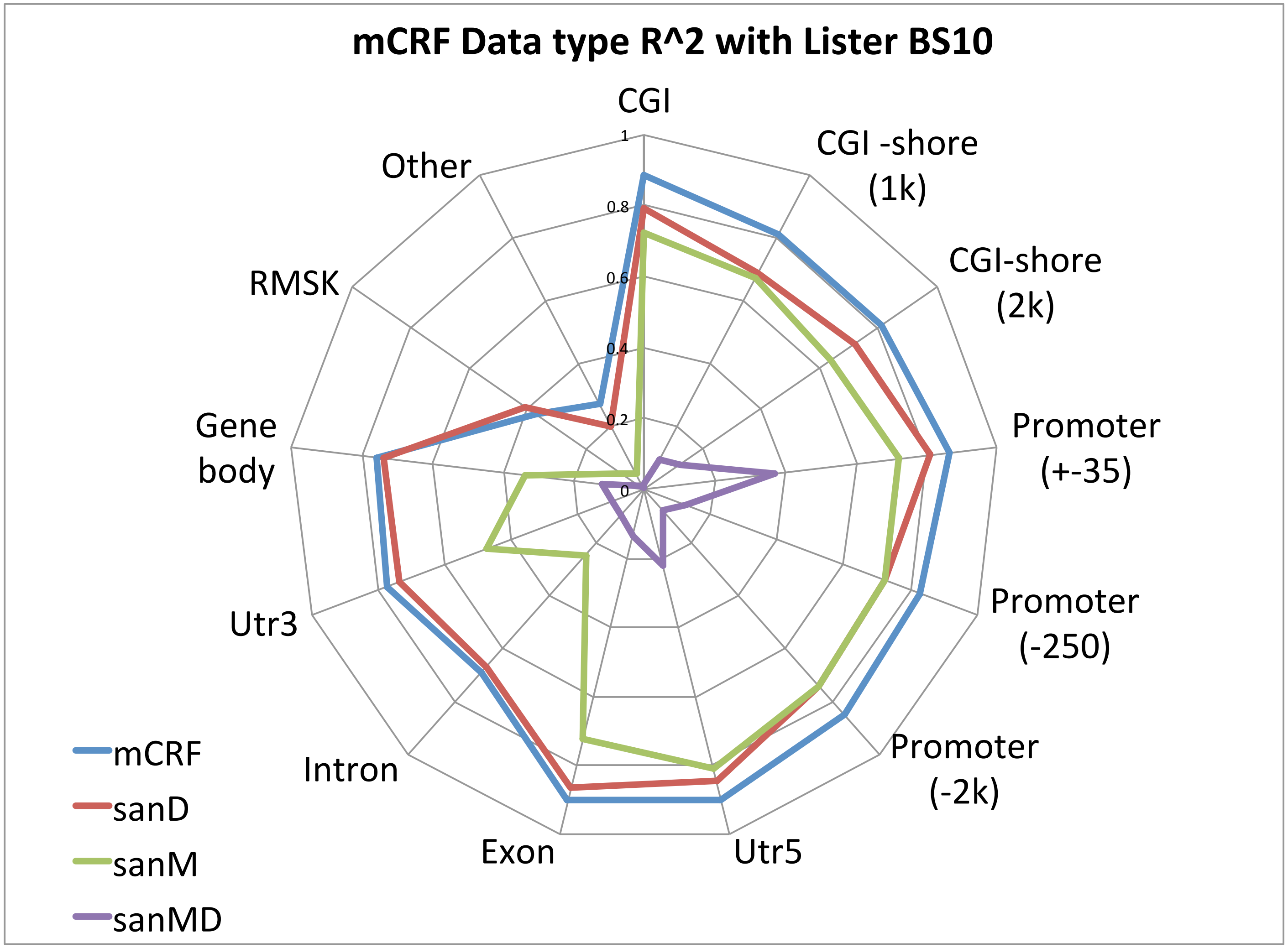
$$
\begin{aligned}
HMM &\Rightarrow P(5mC)P(MeDIP, MRE, \ldots \mid 5mC) \\
&= P(5mC, MeDIP, MRE, \ldots) \\
&= P(5mC \mid MeDIP, MRE, \ldots)P(MeDIP, MRE, \ldots)
\end{aligned}
$$

$$
CRF \Rightarrow P(5mC \mid MeDIP, MRE, \ldots)
$$

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



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



Sequence-based DNA Methylome/ Epigenome Resources

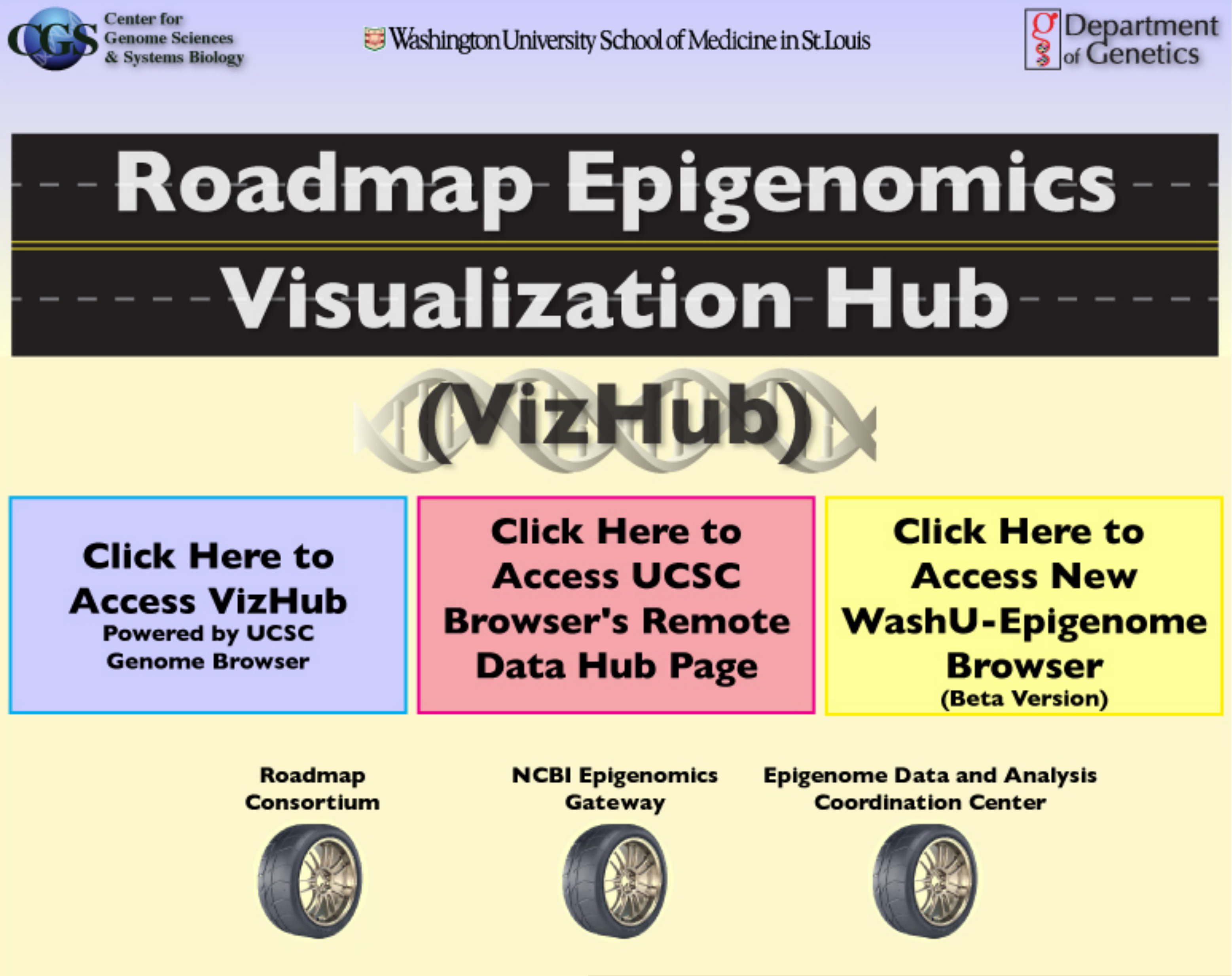
- http://VizHub.wustl.edu/
  - Data visualization hub for Roadmap Epigenomics Project
  - http://epigenomegateway.wustl.edu/
    - Next generation genome browser

- http://www.roadmapepigenomics.org/
  - Roadmap Consortium

- http://www.genboree.org/epigenomeatlas/index.rhtml
  - Data coordination center at Baylor

- http://www.ncbi.nlm.nih.gov/epigenomics
  - NCBI, where you can download data

Reference Epigenome Mapping Centers



CGS Center for Genome Sciences & Systems Biology

Washington University School of Medicine in St.Louis

Department of Genetics

Roadmap Epigenomics Visualization Hub (VizHub)

Click Here to Access VizHub
Powered by UCSC Genome Browser

Click Here to Access UCSC Browser's Remote Data Hub Page

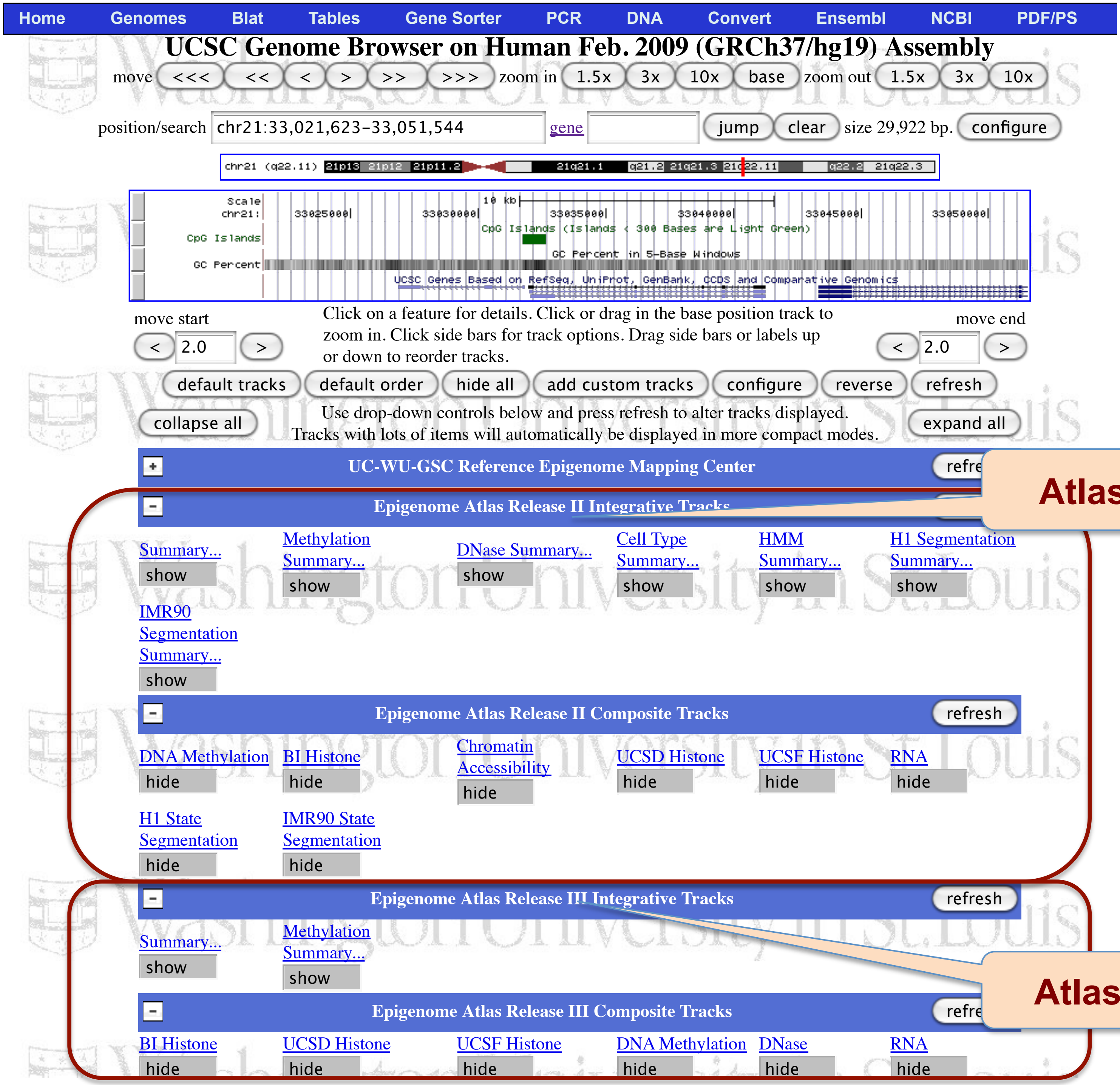
Click Here to Access New WashU-Epigenome Browser
(Beta Version)

Roadmap Consortium

NCBI Epigenomics Gateway

Epigenome Data and Analysis Coordination Center

http://VizHub.wustl.edu



UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

position/search chr21:33,021,623-33,051,544 gene size 29,922 bp.

move start 2.0

move end 2.0

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks.

Use drop-down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.

UC-WU-GSC Reference Epigenome Mapping Center

Epigenome Atlas Release II Integrative Tracks

Summary... show
Methylation Summary... show
DNase Summary... show
Cell Type Summary... show
HMM Summary... show
H1 Segmentation Summary... show
IMR90 Segmentation Summary... show

Epigenome Atlas Release II Composite Tracks

DNA Methylation hide
BI Histone hide
Chromatin Accessibility hide
UCSD Histone hide
UCSF Histone hide
RNA hide
H1 State Segmentation hide
IMR90 State Segmentation hide

Epigenome Atlas Release III Integrative Tracks

Summary... show
Methylation Summary... show

Epigenome Atlas Release III Composite Tracks

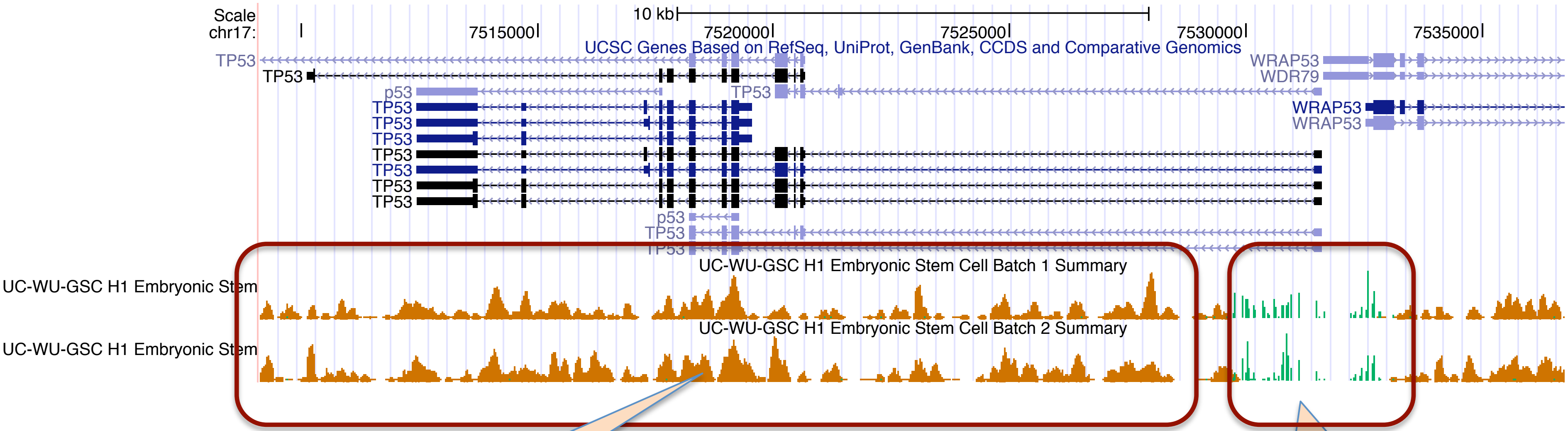
BI Histone hide
UCSD Histone hide
UCSF Histone hide
DNA Methylation hide
DNase hide
RNA hide

Atlas Release II

Atlas Release III

Integrative/summary tracks

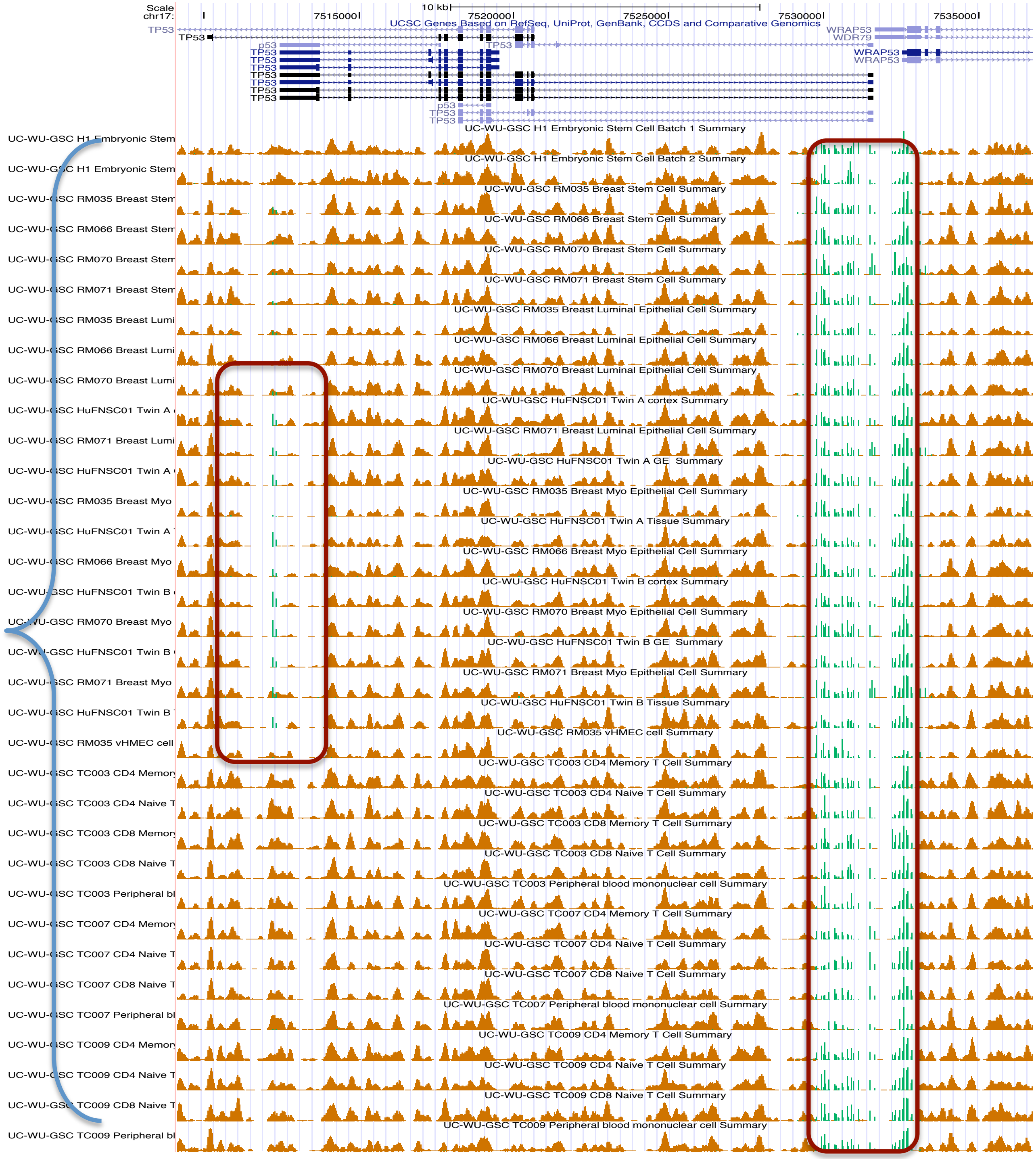
- New (well, a year old) UCSC- technology
- Overlay and summarize many data tracks
  - Different experiments for the same sample;
  - Same experiments for different samples;



MeDIP peaks, or methylated regions

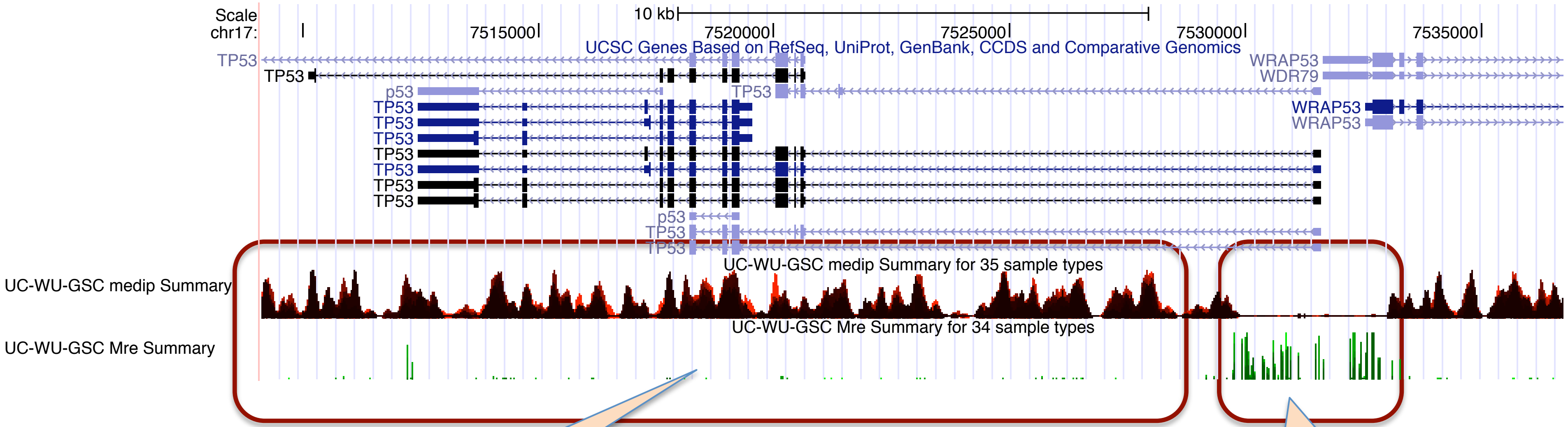
MRE peaks, or unmethylated regions

Combined MeDIP/MRE tracks of 34 samples



Integrative/summary tracks

- Can choose to integrate on samples or on epigenetic marks
- “Rainbow tracks”



MeDIP peaks, or methylated regions

MRE peaks, or unmethylated regions

Summary View of Epigenome Atlas 1

Methylation

H3K27me3

H3K36me3

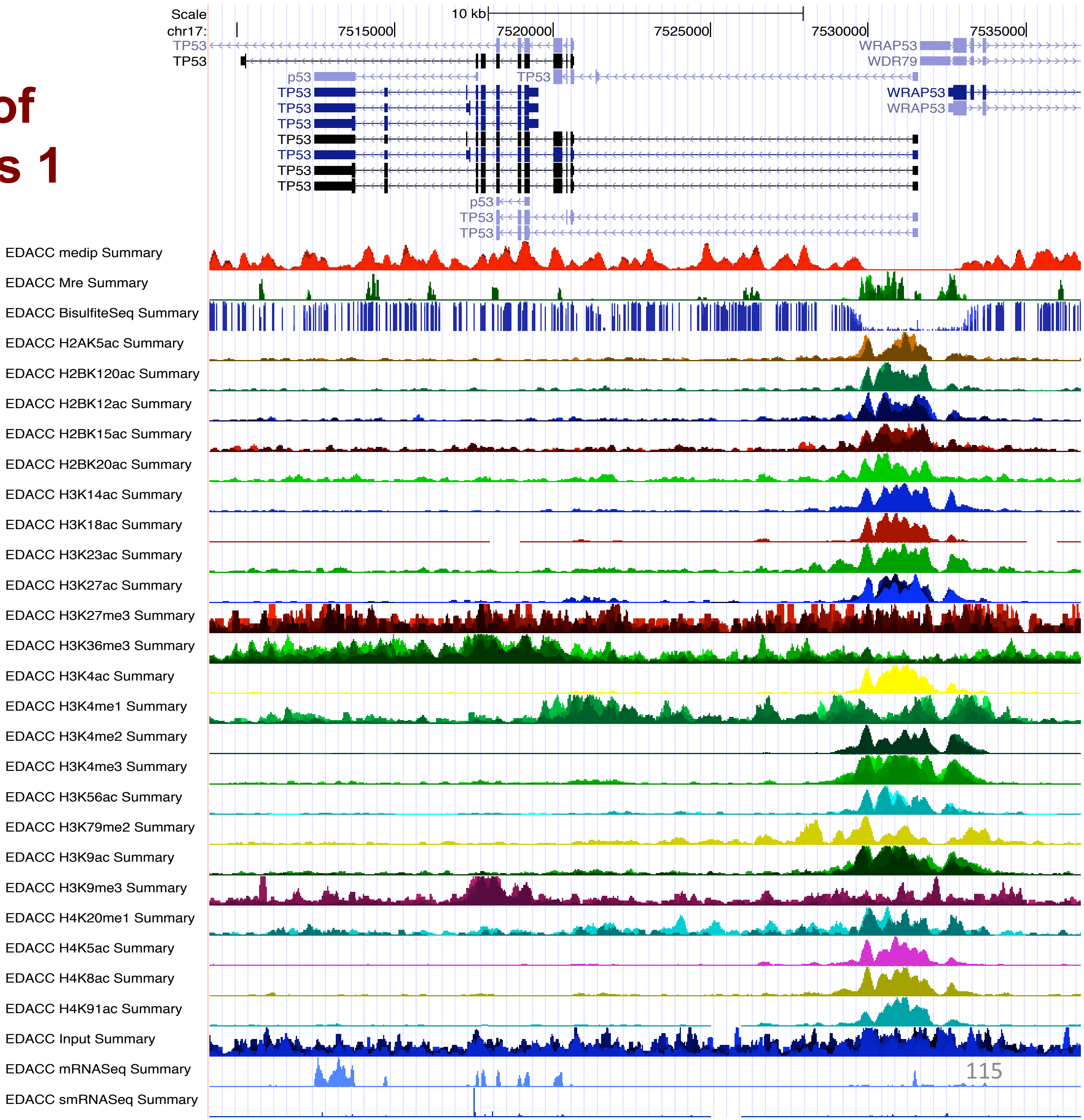
H3K4me1

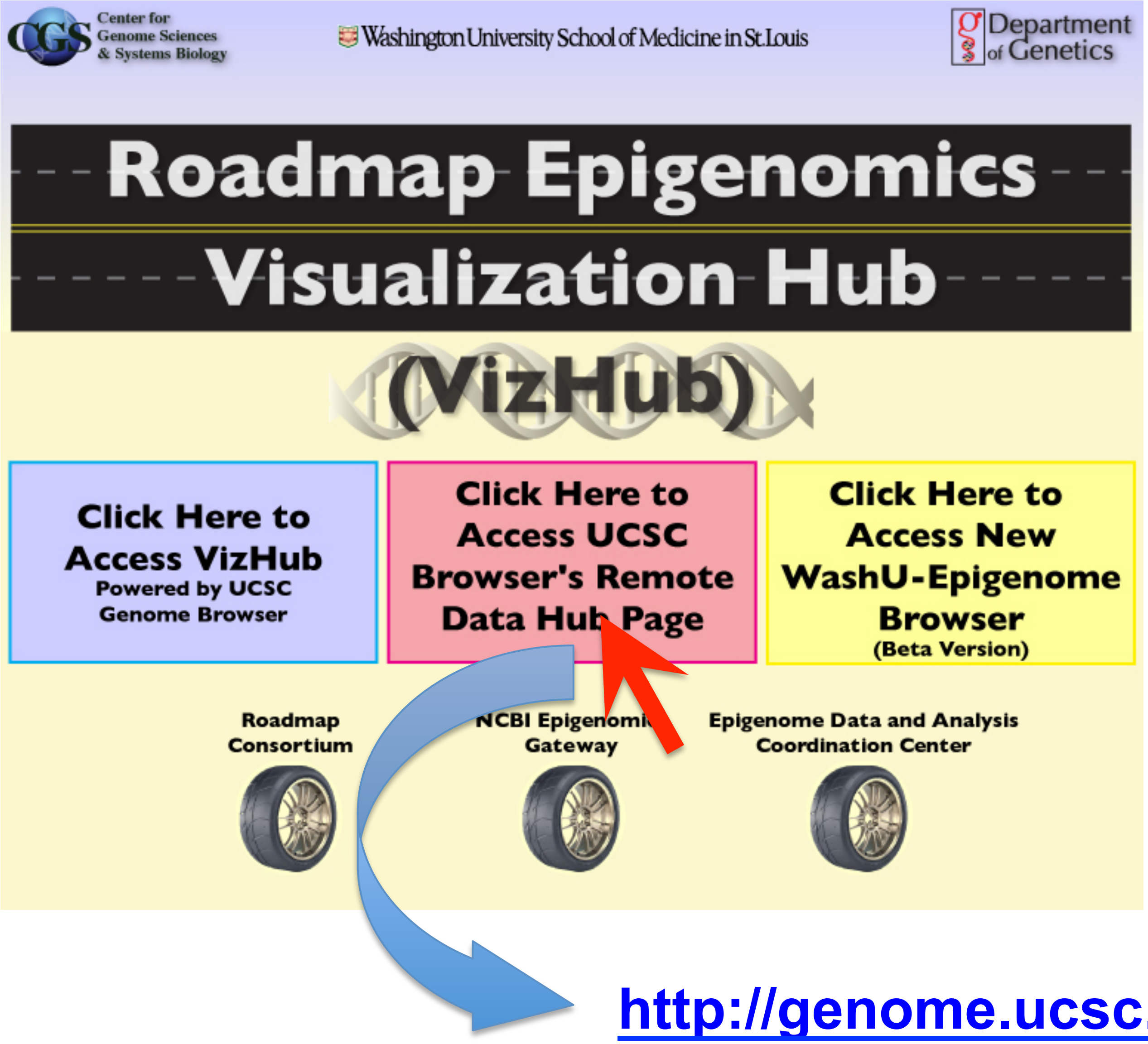
H3K4me3

H3K9me3

Input

RNA





Center for Genome Sciences & Systems Biology

Washington University School of Medicine in St.Louis

Department of Genetics

Roadmap Epigenomics Visualization Hub

(VizHub)

Click Here to Access VizHub
Powered by UCSC Genome Browser

Click Here to Access UCSC Browser's Remote Data Hub Page

Click Here to Access New WashU-Epigenome Browser
(Beta Version)

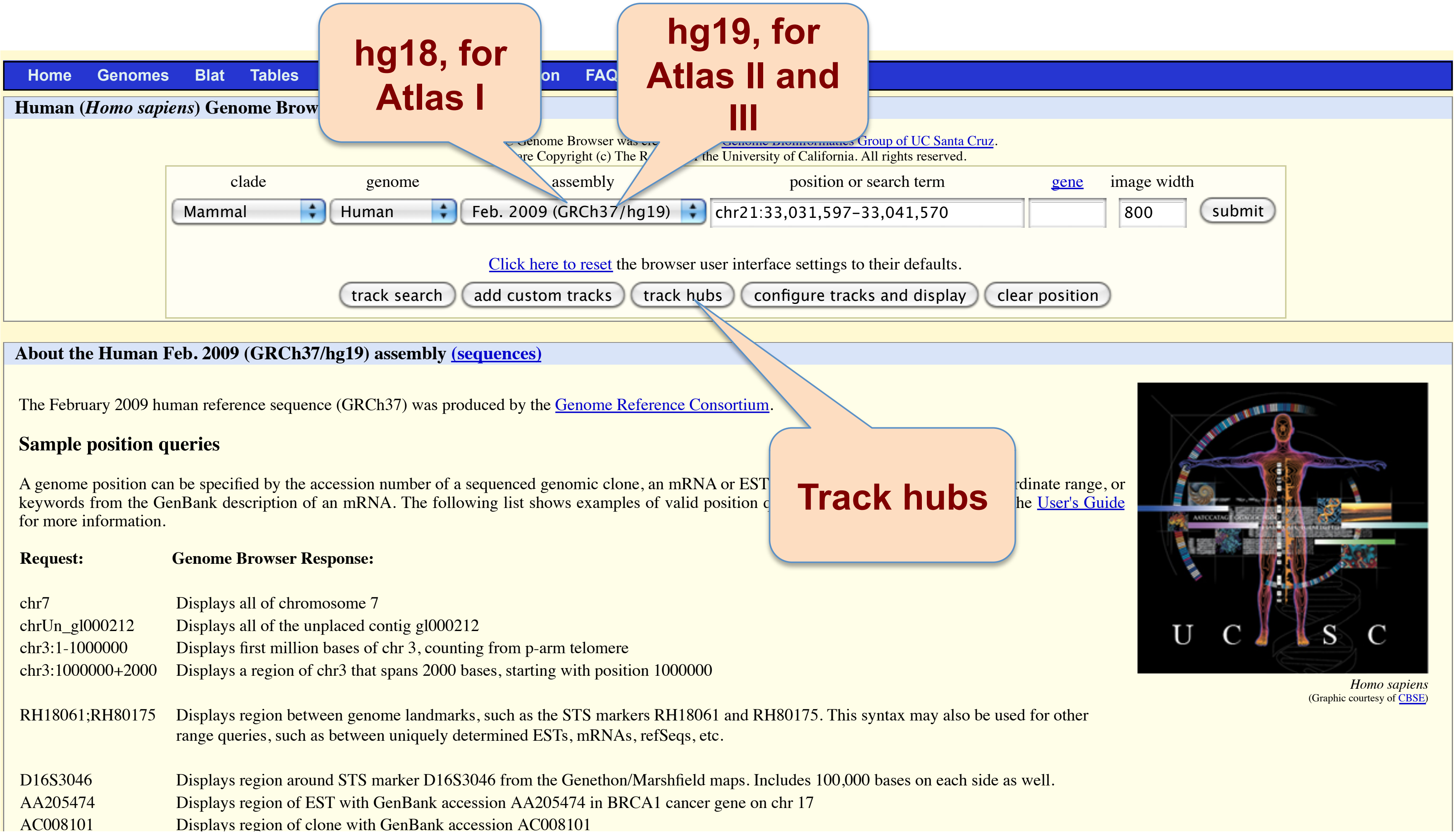
Roadmap Consortium

NCBI Epigenomi
Gateway

Epigenome Data and Analysis Coordination Center

http://genome.ucsc.edu/

Go to http://genome.ucsc.edu/, choose genomes



hg18, for Atlas I

hg19, for Atlas II and III

Track hubs

Human (*Homo sapiens*) Genome Brow

Genome Browser was cre… Genome Bioinformatics Group of UC Santa Cruz.
…re Copyright (c) The R… the University of California. All rights reserved.

clade	genome	assembly	position or search term	gene	image width
Mammal	Human	Feb. 2009 (GRCh37/hg19)	chr21:33,031,597-33,041,570		800

Click here to reset the browser user interface settings to their defaults.

About the Human Feb. 2009 (GRCh37/hg19) assembly (sequences)

The February 2009 human reference sequence (GRCh37) was produced by the Genome Reference Consortium.

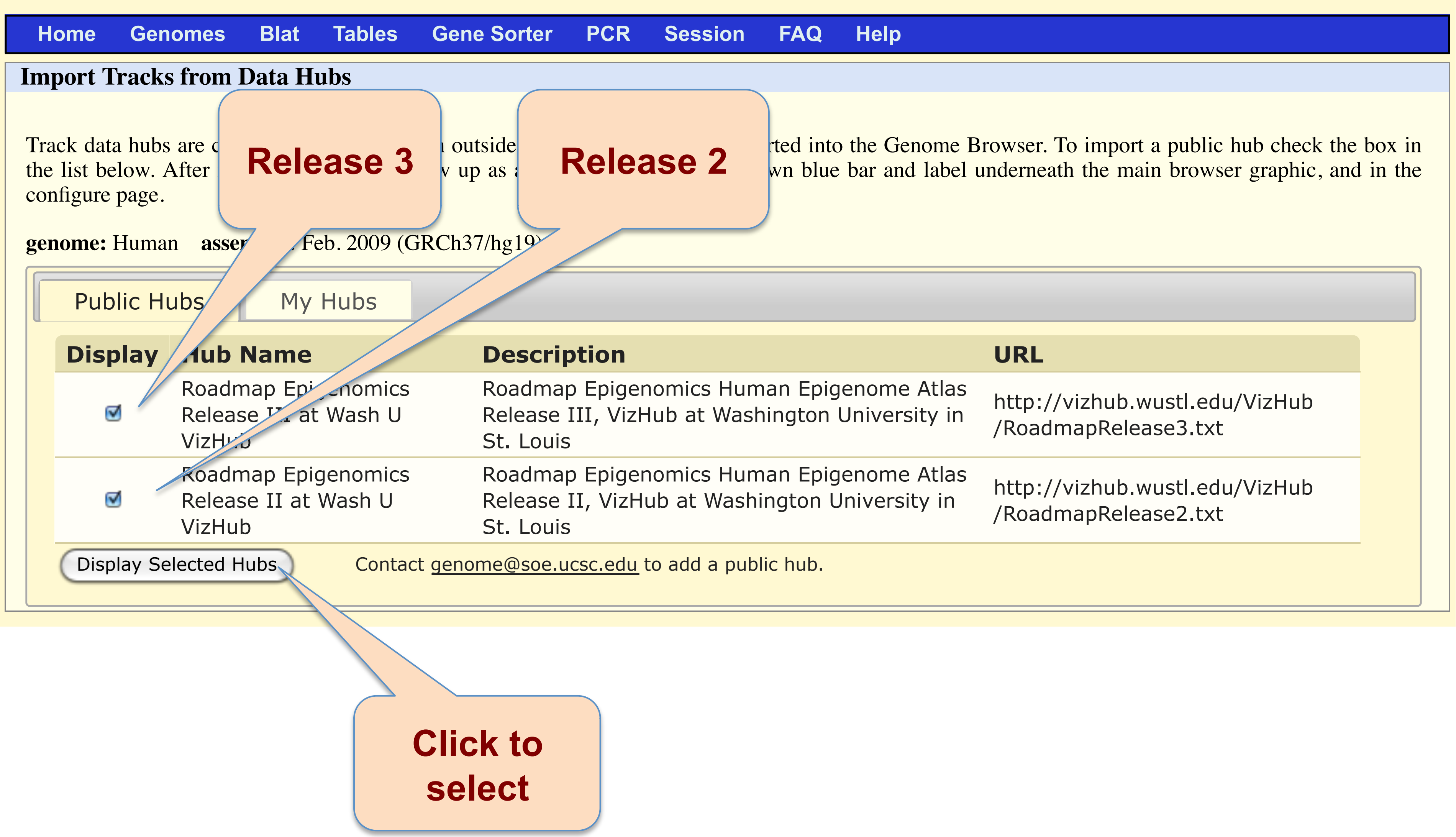
Sample position queries

A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST … rdinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position q… he User's Guide for more information.

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
chrUn_gl000212	Displays all of the unplaced contig gl000212
chr3:1-1000000	Displays first million bases of chr 3, counting from p-arm telomere
chr3:1000000+2000	Displays a region of chr3 that spans 2000 bases, starting with position 1000000
RH18061;RH80175	Displays region between genome landmarks, such as the STS markers RH18061 and RH80175. This syntax may also be used for other range queries, such as between uniquely determined ESTs, mRNAs, refSeqs, etc.
D16S3046	Displays region around STS marker D16S3046 from the Genethon/Marshfield maps. Includes 100,000 bases on each side as well.
AA205474	Displays region of EST with GenBank accession AA205474 in BRCA1 cancer gene on chr 17
AC008101	Displays region of clone with GenBank accession AC008101

Homo sapiens
(Graphic courtesy of CBSE)

Hub selection page



Import Tracks from Data Hubs

Track data hubs are c n outside rted into the Genome Browser. To import a public hub check the box in the list below. After v up as a wn blue bar and label underneath the main browser graphic, and in the configure page.

genome: Human **asser** Feb. 2009 (GRCh37/hg19)

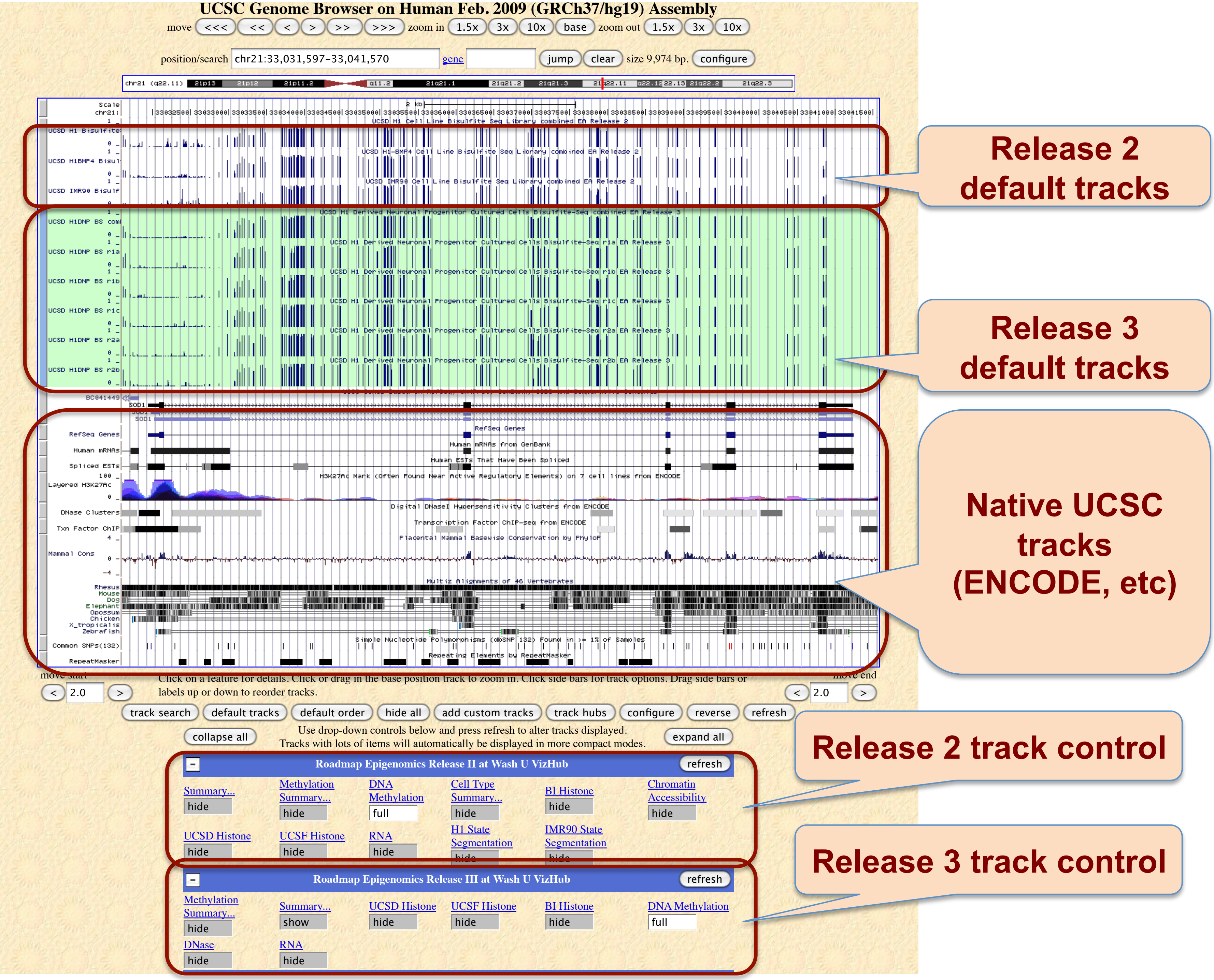
Display	Hub Name	Description	URL
☑	Roadmap Epi enomics Release I at Wash U VizH b	Roadmap Epigenomics Human Epigenome Atlas Release III, VizHub at Washington University in St. Louis	http://vizhub.wustl.edu/VizHub/RoadmapRelease3.txt
☑	Roadmap Epigenomics Release II at Wash U VizHub	Roadmap Epigenomics Human Epigenome Atlas Release II, VizHub at Washington University in St. Louis	http://vizhub.wustl.edu/VizHub/RoadmapRelease2.txt

Contact genome@soe.ucsc.edu to add a public hub.

Release 3

Release 2

Click to select



position/search chr21:33,031,597-33,041,570 size 9,974 bp.

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks.

Use drop-down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.

Roadmap Epigenomics Release II at Wash U VizHub

Summary...	Methylation Summary...	DNA Methylation	Cell Type Summary...	BI Histone	Chromatin Accessibility
hide	hide	full	hide	hide	hide
UCSD Histone	UCSF Histone	RNA	H1 State Segmentation	IMR90 State Segmentation	
hide	hide	hide	hide	hide	

Roadmap Epigenomics Release III at Wash U VizHub

Methylation Summary...	Summary...	UCSD Histone	UCSF Histone	BI Histone	DNA Methylation
hide	show	hide	hide	hide	full
DNase	RNA				
hide	hide				

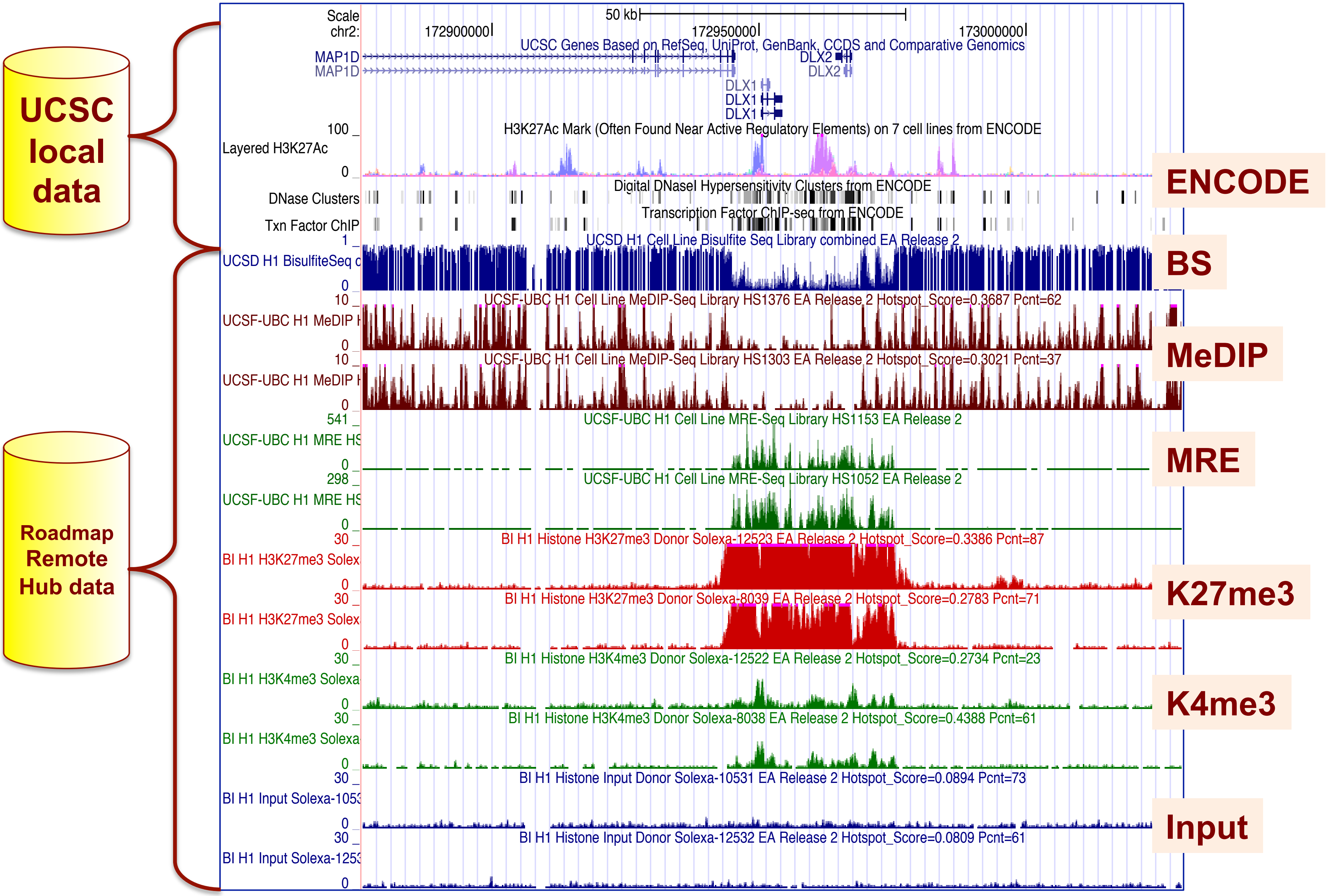
Release 2 default tracks

Release 3 default tracks

Native UCSC tracks (ENCODE, etc)

Release 2 track control

Release 3 track control



UCSC local data

Roadmap Remote Hub data

ENCODE

BS

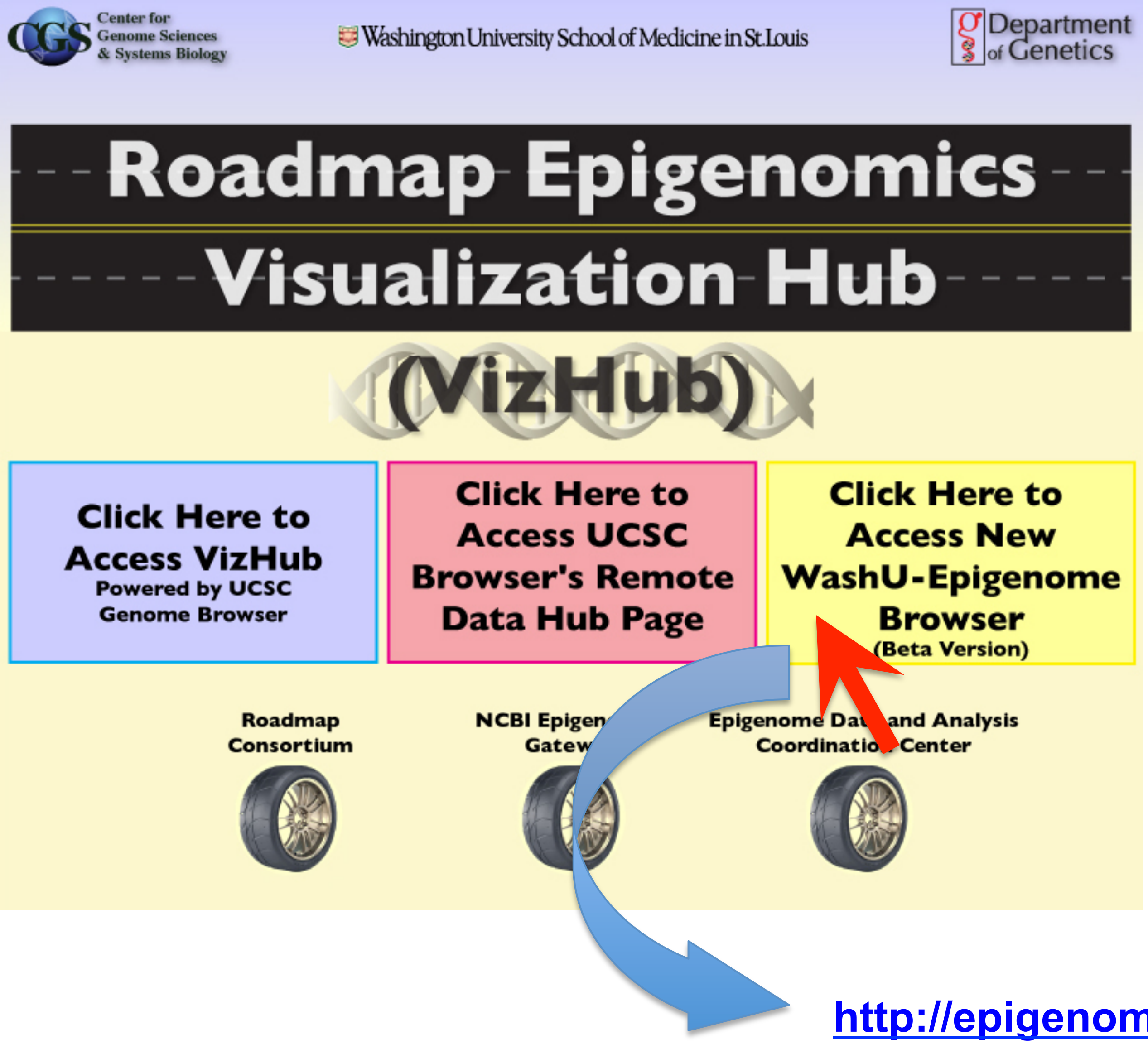
MeDIP

MRE

K27me3

K4me3

Input



Center for Genome Sciences & Systems Biology

Washington University School of Medicine in St.Louis

Department of Genetics

Roadmap Epigenomics Visualization Hub

(VizHub)

Click Here to Access VizHub
Powered by UCSC Genome Browser

Click Here to Access UCSC Browser's Remote Data Hub Page

Click Here to Access New WashU-Epigenome Browser
(Beta Version)

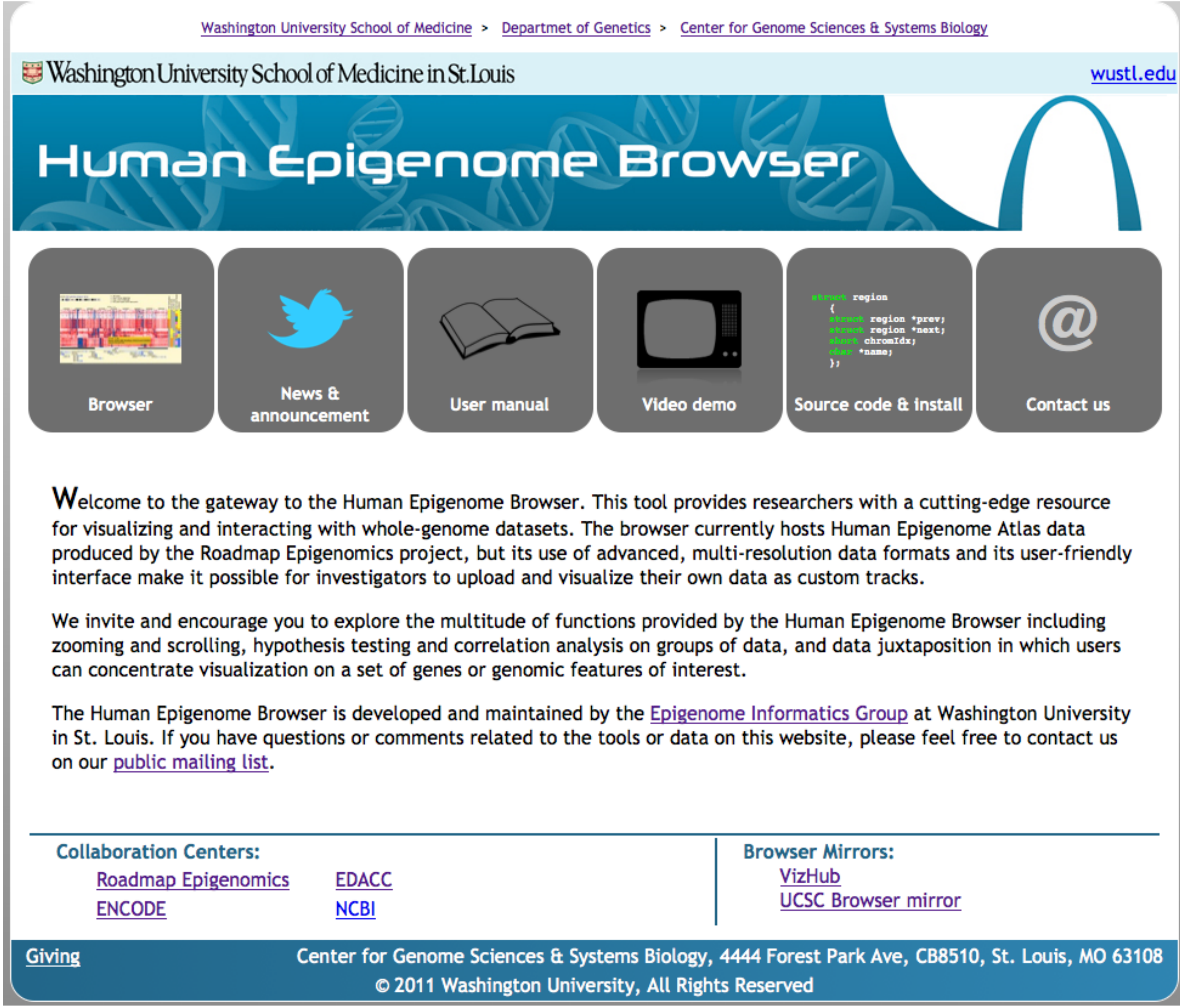
Roadmap Consortium

NCBI Epigen
Gatew

Epigenome Dat and Analysis
Coordinatio Center

http://epigenomegateway.wustl.edu/

(*Nature Methods*, in press)



Human Epigenome Browser

```
struct region
{
   struct region *prev;
   struct region *next;
   short chromIdx;
   char *name;
};
```

Welcome to the gateway to the Human Epigenome Browser. This tool provides researchers with a cutting-edge resource for visualizing and interacting with whole-genome datasets. The browser currently hosts Human Epigenome Atlas data produced by the Roadmap Epigenomics project, but its use of advanced, multi-resolution data formats and its user-friendly interface make it possible for investigators to upload and visualize their own data as custom tracks.

We invite and encourage you to explore the multitude of functions provided by the Human Epigenome Browser including zooming and scrolling, hypothesis testing and correlation analysis on groups of data, and data juxtaposition in which users can concentrate visualization on a set of genes or genomic features of interest.

The Human Epigenome Browser is developed and maintained by the Epigenome Informatics Group at Washington University in St. Louis. If you have questions or comments related to the tools or data on this website, please feel free to contact us on our public mailing list.

Collaboration Centers:

Roadmap Epigenomics	EDACC
ENCODE	NCBI

Browser Mirrors:

- VizHub
- UCSC Browser mirror

Giving

Center for Genome Sciences & Systems Biology, 4444 Forest Park Ave, CB8510, St. Louis, MO 63108

Summary

- Motivation
  - Larger datasets (whole genome, sequencing based)
  - More datasets (hundreds, thousands)
  - Rich metadata (clinical parameters, phenotypes)
  - Multi-dimensional datasets

- New way of browsing the genome
  - Hundreds of tracks in one view
  - Google-map style pan and zoom, drag and drop
  - Display epigenomic data alongside with their metadata
  - Heatmap, wiggle map, going from whole genome to single base
  - Viewing data on specific genomic features, genesets or pathways
  - Statistical analysis (comparing two or more samples, or groups of samples)
  - Support custom tracks
  - Support sessions
  - ENCODE data integrated

- Help page, mock data and video tutorial available
  - http://epigenomegateway.wustl.edu/

Acknowledgement

UCSF

Joseph Costello
- Ivan Smirnov
- Raman Nagarajan
- Chibo Hong
- Brett Johnson
- Shaun Fouse
- Rob Bell

Michael McManus

Thea Tlsty

UC Santa Cruz

David Haussler
Jim Kent
Tracy Ballinger
Donna Karolchik
Bob Kuhn
Ann Zweig
Brian Raney
Kate Rosenbloom

GSC Canada

Marco Marra
Martin Hirst
Steven Jones
Cydney Nielsen
Yongjun Zhao
Gordon Robertson

USC

Peggy Farnham

Washington University

Xin Zhou
Mingchao Xie
Brett Maricque
Eric Martin
Brian Koebbe

EDACC

Aleks Milosavljevic
Alan Harris
Christian Coarfa
Rob Waterland

Other REMCs

John Stamatoyannopoulos
Richard Sandstrom
Alex Meissner
Brad Bernstein
Bing Ren
Joe Ecker
Noam Shoresh
Bob Thurman

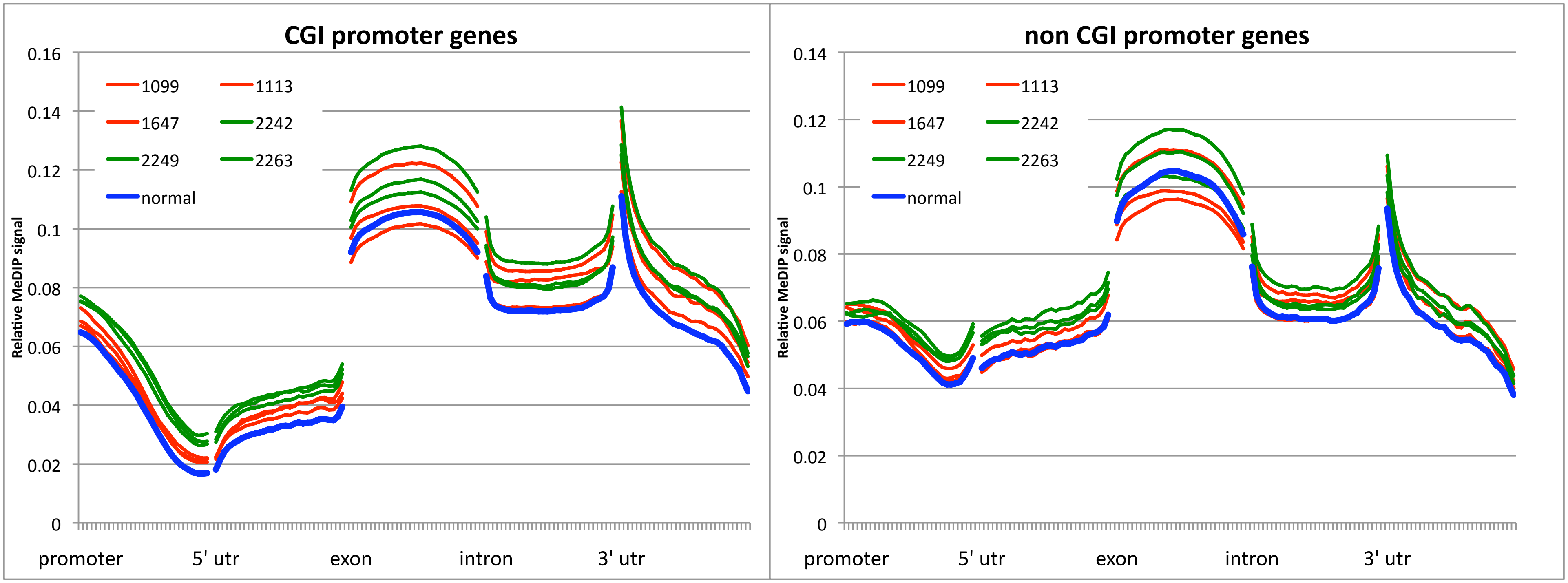
NCBI

Greg Schuler
Tanya Barrett

Case 5: Cancer Methylomes

- Comparing endometrial cancers

Methylation level across gene structure



- Increased DNA methylation over genic regions;
- More methylation in promoters in type I than in type II;

MLH1 Promoter