

# **Integrated analysis of sequencing data**

How to combine \*-seq data

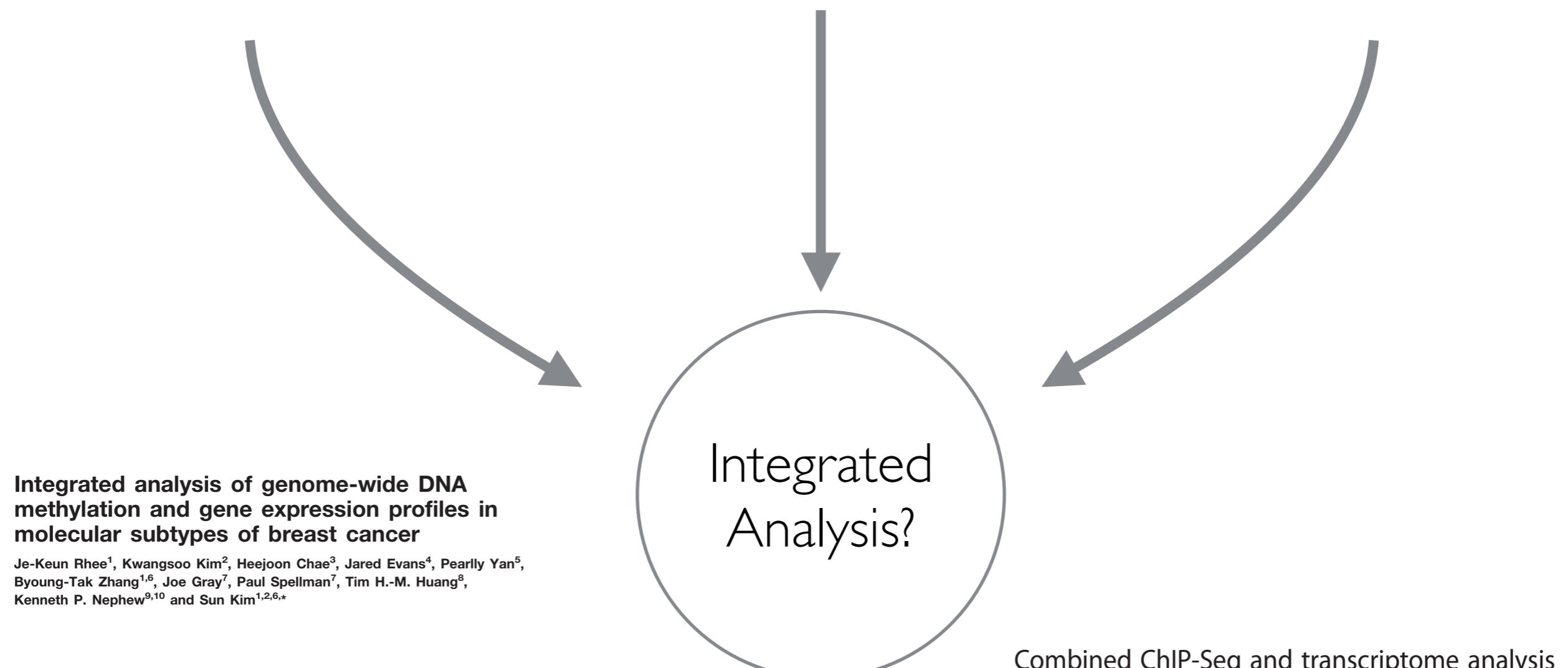
M. Defrance, M. Thomas-Chollier, C. Herrmann, D. Puthier, J. van Helden

\*ChIP-seq, RNA-seq, MeDIP-seq, ...

Transcription factor binding  
ChIP-seq

Expression quantification  
RNA-seq

DNA methylation  
MeDIP-seq



**Integrated analysis of genome-wide DNA methylation and gene expression profiles in molecular subtypes of breast cancer**

Je-Keun Rhee<sup>1</sup>, Kwangsoo Kim<sup>2</sup>, Heejoon Chae<sup>3</sup>, Jared Evans<sup>4</sup>, Pearly Yan<sup>5</sup>,  
Byoung-Tak Zhang<sup>1,6</sup>, Joe Gray<sup>7</sup>, Paul Spellman<sup>7</sup>, Tim H.-M. Huang<sup>8</sup>,  
Kenneth P. Nephew<sup>9,10</sup> and Sun Kim<sup>1,2,6,\*</sup>

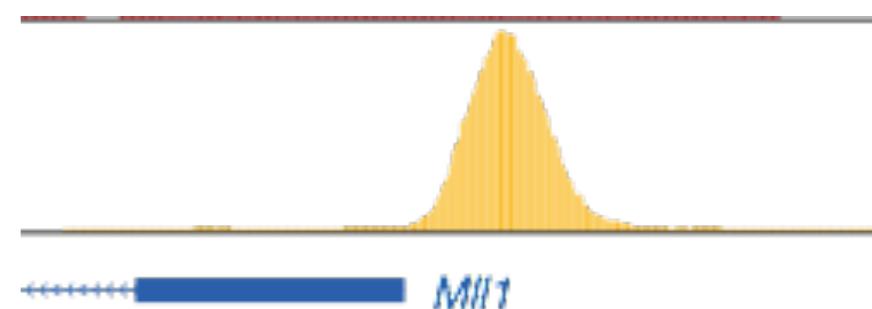
**Integrated analysis of DNA methylation and gene expression reveals specific signaling pathways associated with platinum resistance in ovarian cancer**

Meng Li<sup>1</sup>, Curt Balch<sup>1,2,3</sup>, John S Montgomery<sup>1</sup>, Mikyoung Jeong<sup>4</sup>,  
Jae Hoon Chung<sup>4</sup>, Pearly Yan<sup>5</sup>, Tim HM Huang<sup>5</sup>, Sun Kim<sup>\*6,7</sup> and  
Kenneth P Nephew<sup>\*1,2,3,8</sup>

Combined ChIP-Seq and transcriptome analysis identifies AP-1/JunD as a primary regulator of oxidative stress and IL-1 $\beta$  synthesis in macrophages

Richard P Hull<sup>1</sup>, Prashant K Srivastava<sup>1</sup>, Zelpha D'Souza<sup>1</sup>, Santosh S Atanur<sup>1</sup>, Fatima Mechta-Grigoriou<sup>2</sup>,  
Laurence Game<sup>1</sup>, Enrico Petretto<sup>1</sup>, H Terence Cook<sup>3</sup>, Timothy J Aitman<sup>1</sup> and Jacques Behmoaras<sup>3\*</sup>

Transcription factor binding  
ChIP-seq



OPTION 1

1 CONDITION

Enriched regions

OPTION 2

CONDITION 1  
CONDITION 2

Differentially enriched regions

**$\mu$ -array**

Commercial / custom  
Coding, non coding

Expression quantification  
**RNA-seq**

**RNA-seq**

Coding, non coding  
Alternative transcripts

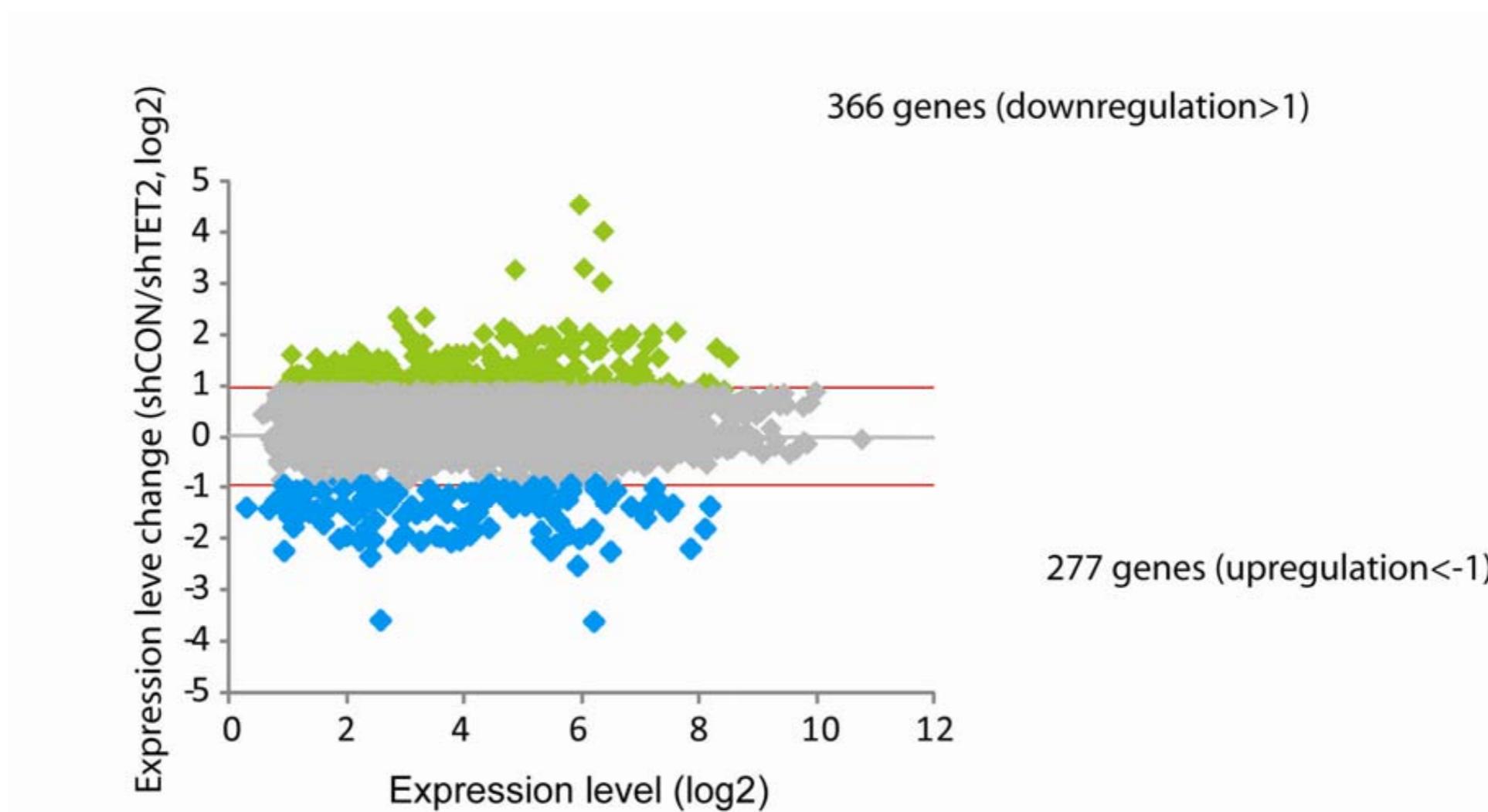
CONDITION 1  
CONDITION 2

Differentially expressed genes

## **Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks**

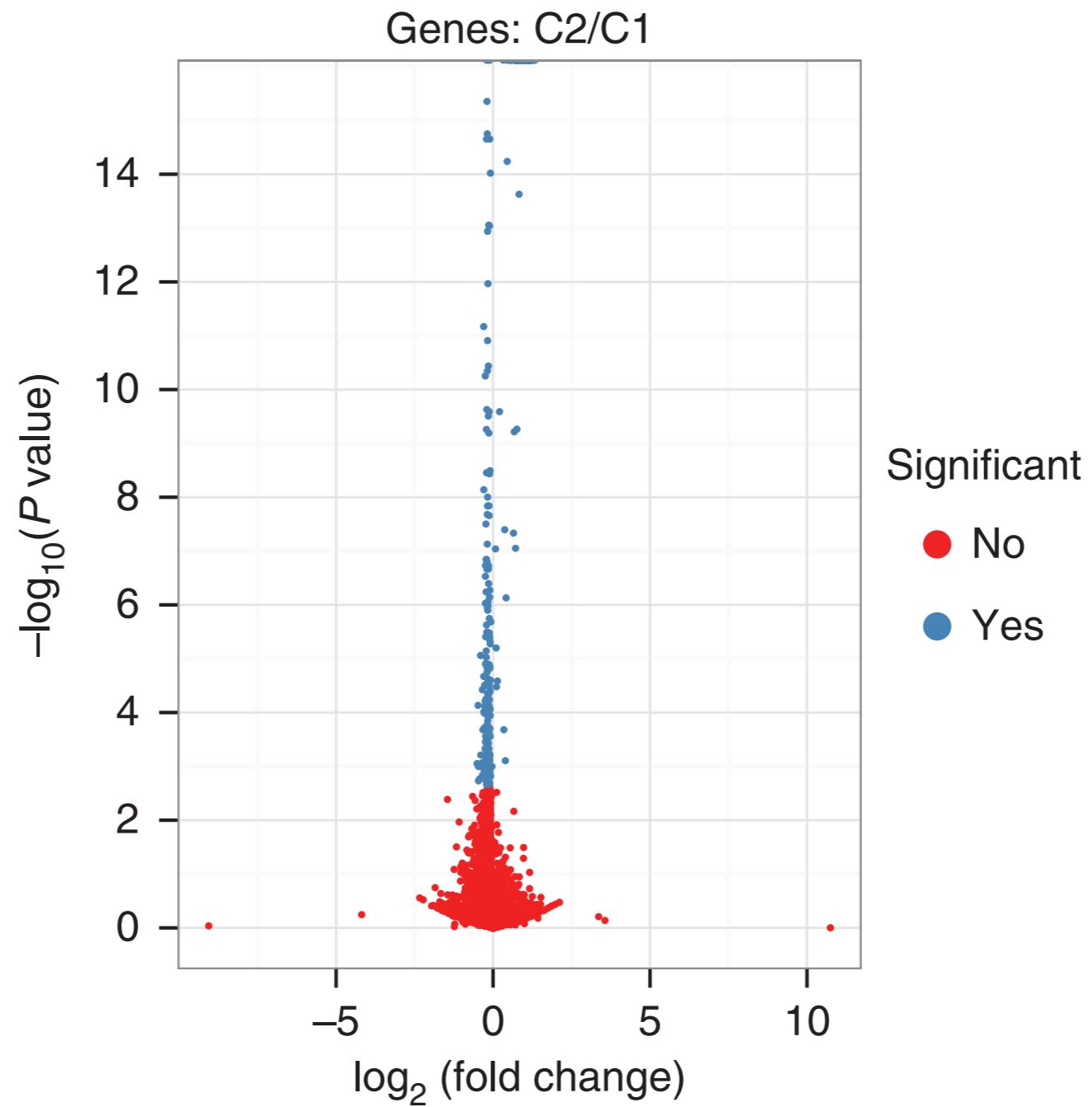
Cole Trapnell<sup>1,2</sup>, Adam Roberts<sup>3</sup>, Loyal Goff<sup>1,2,4</sup>, Geo Pertea<sup>5,6</sup>, Daehwan Kim<sup>5,7</sup>, David R Kelley<sup>1,2</sup>, Harold Pimentel<sup>3</sup>, Steven L Salzberg<sup>5,6</sup>, John L Rinn<sup>1,2</sup> & Lior Pachter<sup>3,8,9</sup>

# Differentially expressed genes?



2 condition x 1 sample

# Differentially expressed genes?



2 condition x n samples

## **Enrichment-based methods**

Methylated (alternatively, unmethylated) DNA fragments are enriched in a DNA library. The library composition is quantified by next-generation sequencing

DNA methylation-seq

## **Bisulphite sequencing**

DNA treatment with bisulphite specifically introduces mutations at unmethylated Cs. These mutations are mapped by next-generation sequencing

## **Enrichment-based methods**

Methylated (alternatively, unmethylated) DNA fragments are enriched in a DNA library. The library composition is quantified by next-generation sequencing

OPTION 1

1 CONDITION

Methylation level  
(CpG, region)

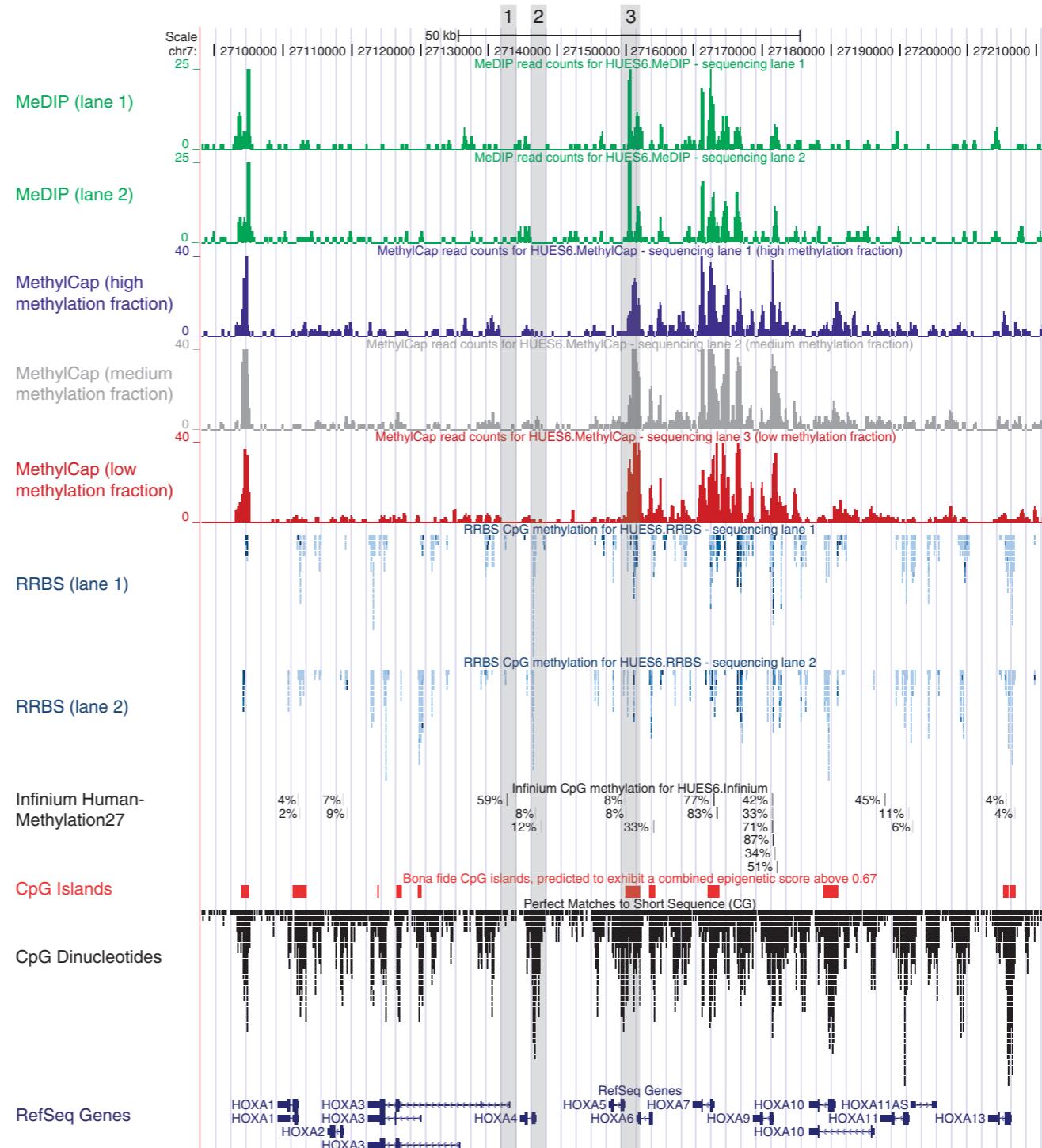
OPTION 2

CONDITION 1  
CONDITION 2

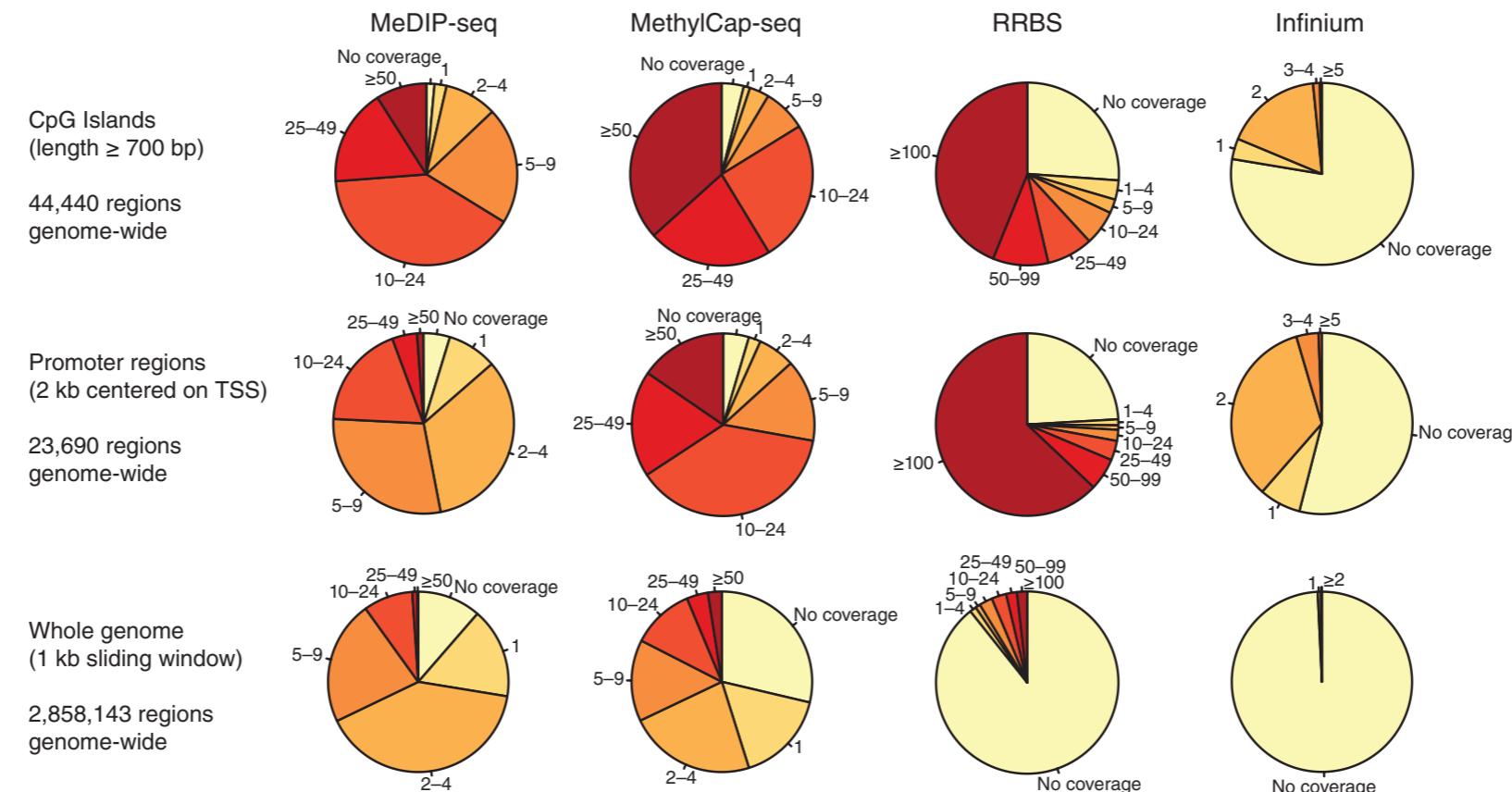
Differentially methylated regions

# DNA methylation mapping technologies

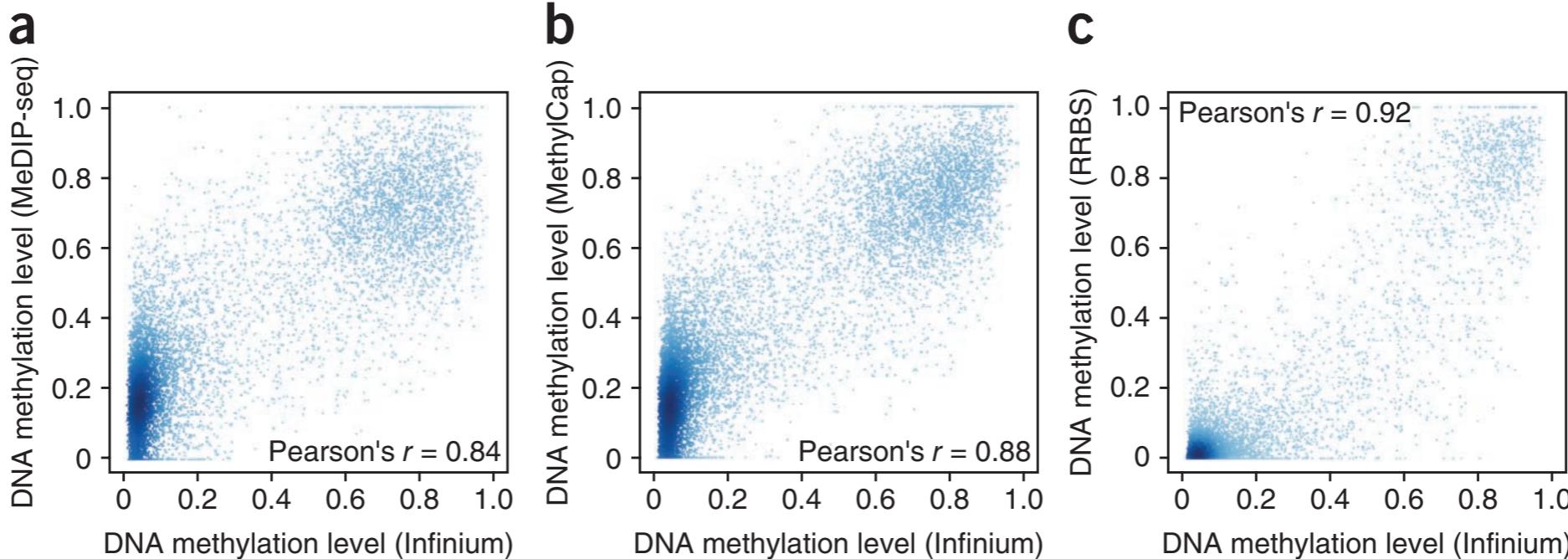
MethylCap-seq  
MeDIP-seq  
RRBS  
WG bisulfite-seq  
 $\mu$ -array (infinium)



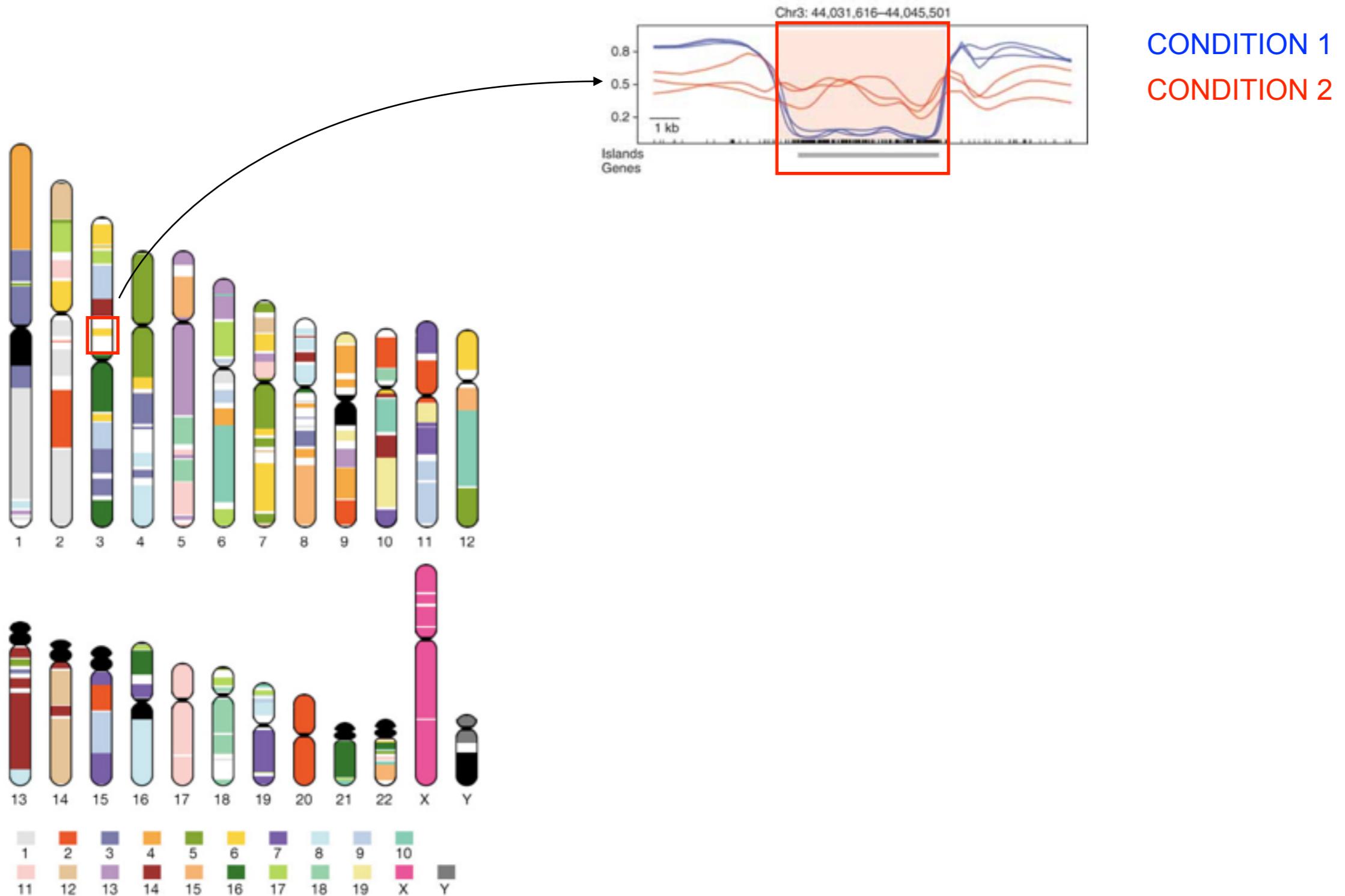
# DNA methylation mapping technologies: Genomic coverage



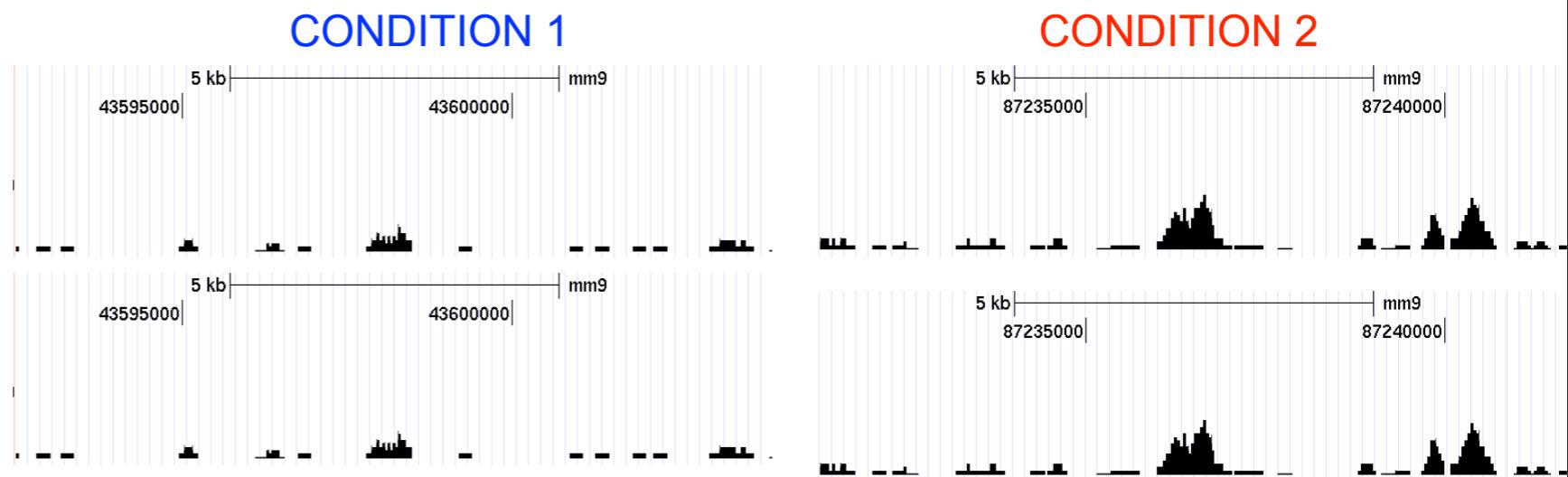
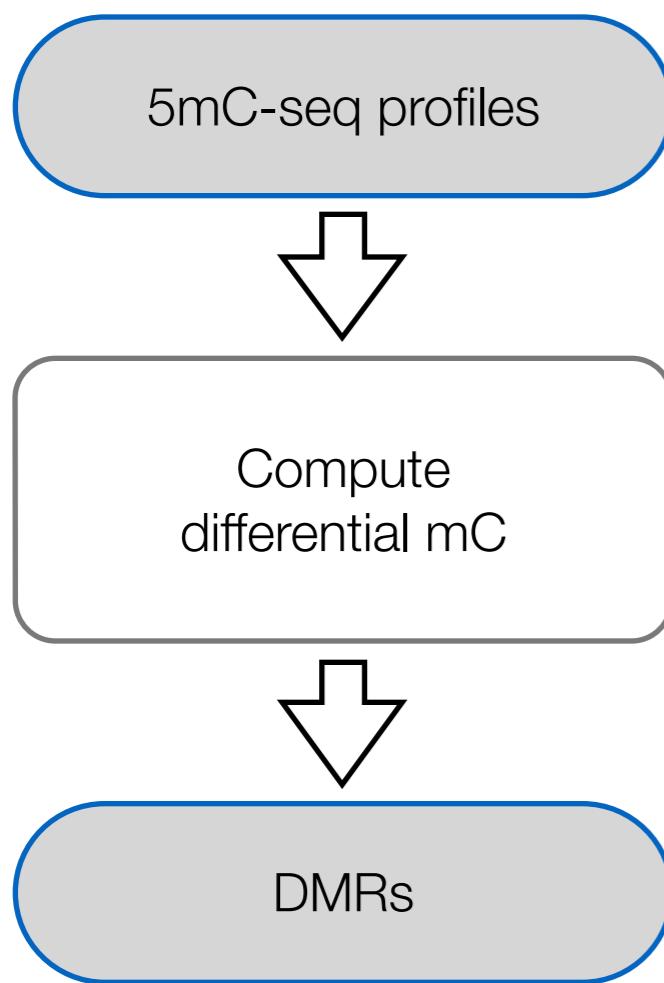
## Comparative quantification



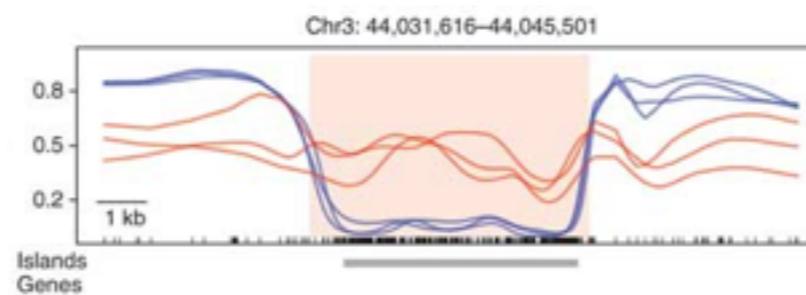
# Differentially Methylated Regions (DMRs)?



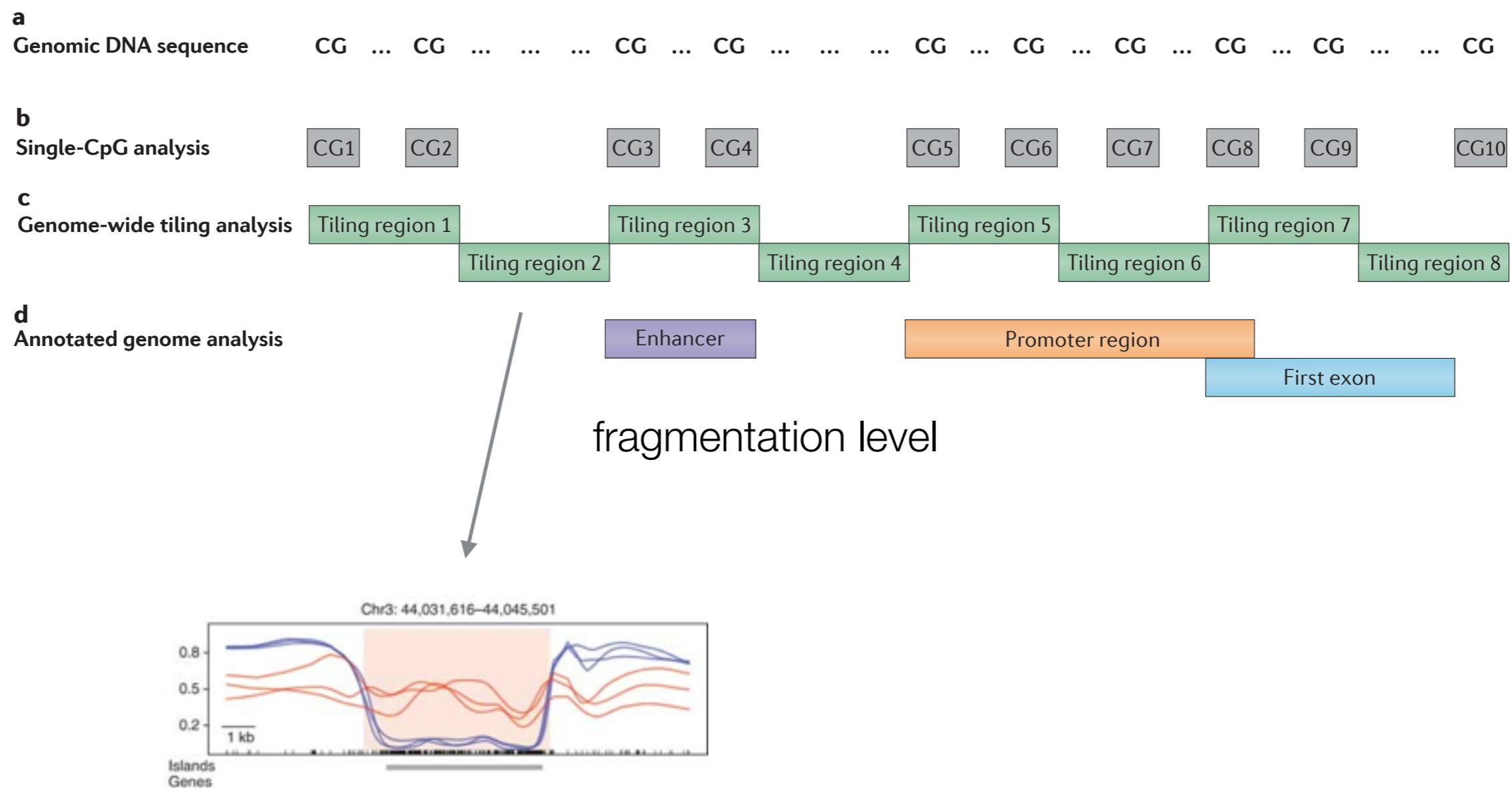
# Identifying differentially methylated regions (DMRs)



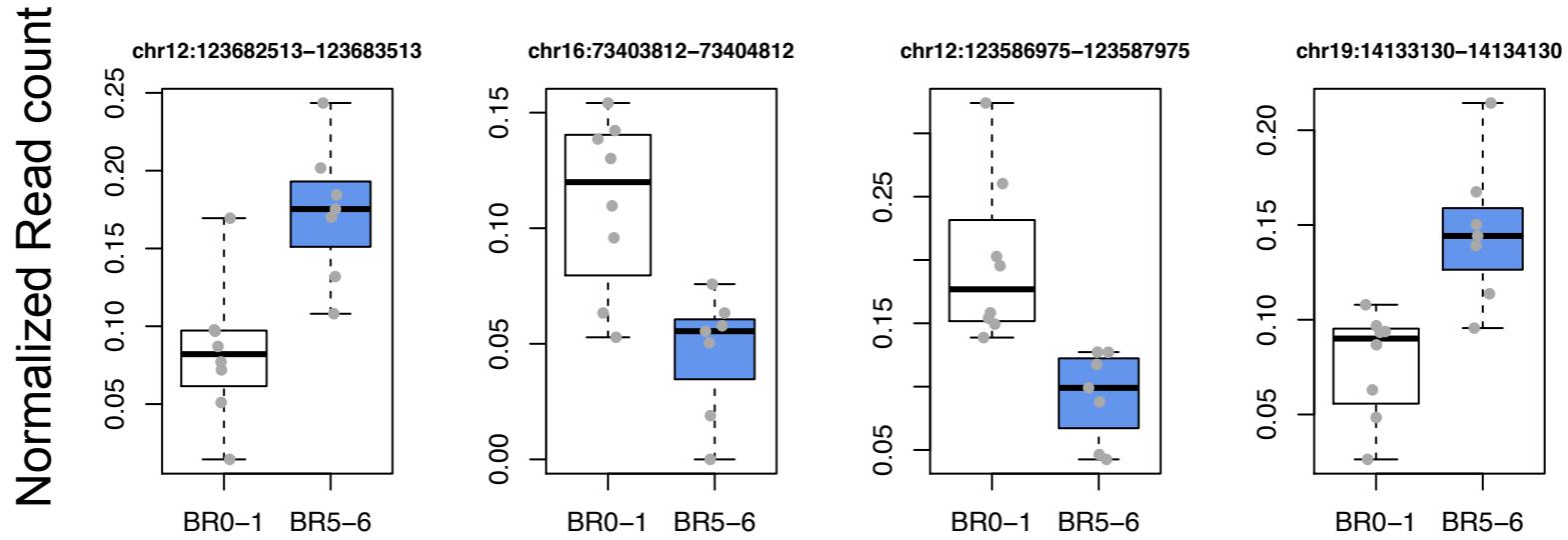
OPTION 1: fragment the genome (promoters, genes, ...)  
OPTION 2: sliding window approach



# Differentially methylated region, CpG or promoter?



# Differentially methylated regions



Region	Log ratio	Fold change	Pvalue	RefSeq.ID
chr12:123682513-123683513	-1.09	2.14	1.78E-03	
chr16:73403812-73404812	1.11	2.16	1.98E-03	
chr12:123586975-123587975	0.84	1.79	2.14E-03	NM_006312
chr19:14133130-14134130	-0.68	1.60	2.31E-03	NM_014921
chr5:134526390-134527390	0.88	1.84	3.21E-03	
chr14:66964536-66965536	0.83	1.78	3.47E-03	
chr4:17081181-17082181	0.70	1.63	3.49E-03	
chr6:21353381-21354381	1.16	2.23	3.53E-03	NM_013401
chr3:14347979-14348979	0.95	1.93	3.65E-03	NM_001134382
chr11:61421968-61422968	-0.93	1.91	3.85E-03	NM_032427
chr3:12995930-12996930	0.98	1.98	4.05E-03	NM_032251
chr11:95557069-95558069	0.95	1.93	4.83E-03	NM_007368
chr11:63865332-63866332	-1.12	2.17	4.90E-03	
chr13:113901339-113902339	-0.82	1.77	4.92E-03	NM_001077183
chr17:77586207-77587207	-1.22	2.34	5.43E-03	NM_004252
chr16:2059877-2060877	0.95	1.93	5.58E-03	NM_139057

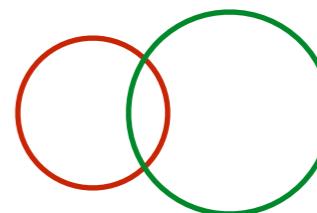
# Combining RNA-seq and ChIP-seq

Transcription factor binding  
ChIP-seq

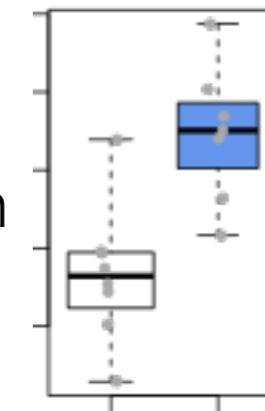
+

Expression quantification  
RNA-seq

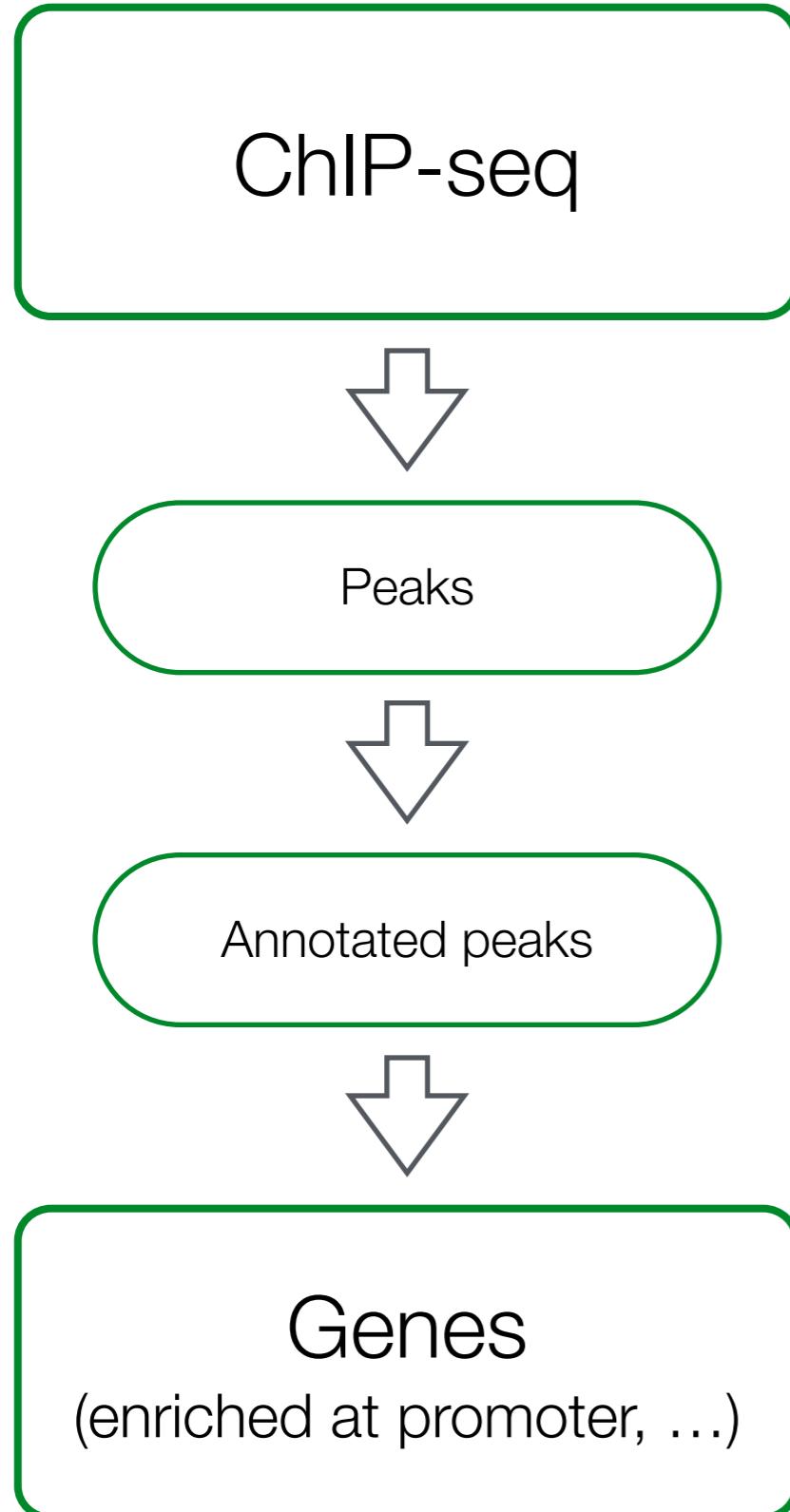
**OPTION 1:** focus on genes, on/off state



**OPTION 2:** quantitative analysis / correlation



## OPTION 1: focus on genes, on/off state



### PAVIS: a tool for Peak Annotation and Visualization

Weichun Huang<sup>1,§</sup>, Rasiah Loganathanraj<sup>2,3,§</sup>, Bryce Schroeder<sup>1,4,§</sup>, David Fargo<sup>2</sup> and Leping Li<sup>1,\*</sup>

ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data

Lihua J Zhu<sup>1,2\*</sup>, Claude Gazin<sup>3</sup>, Nathan D Lawson<sup>1,2</sup>, Hervé Pagès<sup>4</sup>, Simon M Lin<sup>5</sup>, David S Lapointe<sup>6</sup>, Michael R Green<sup>1,2</sup>

OPTION 1: focus on genes, on/off state

RNA-seq



Differential analysis

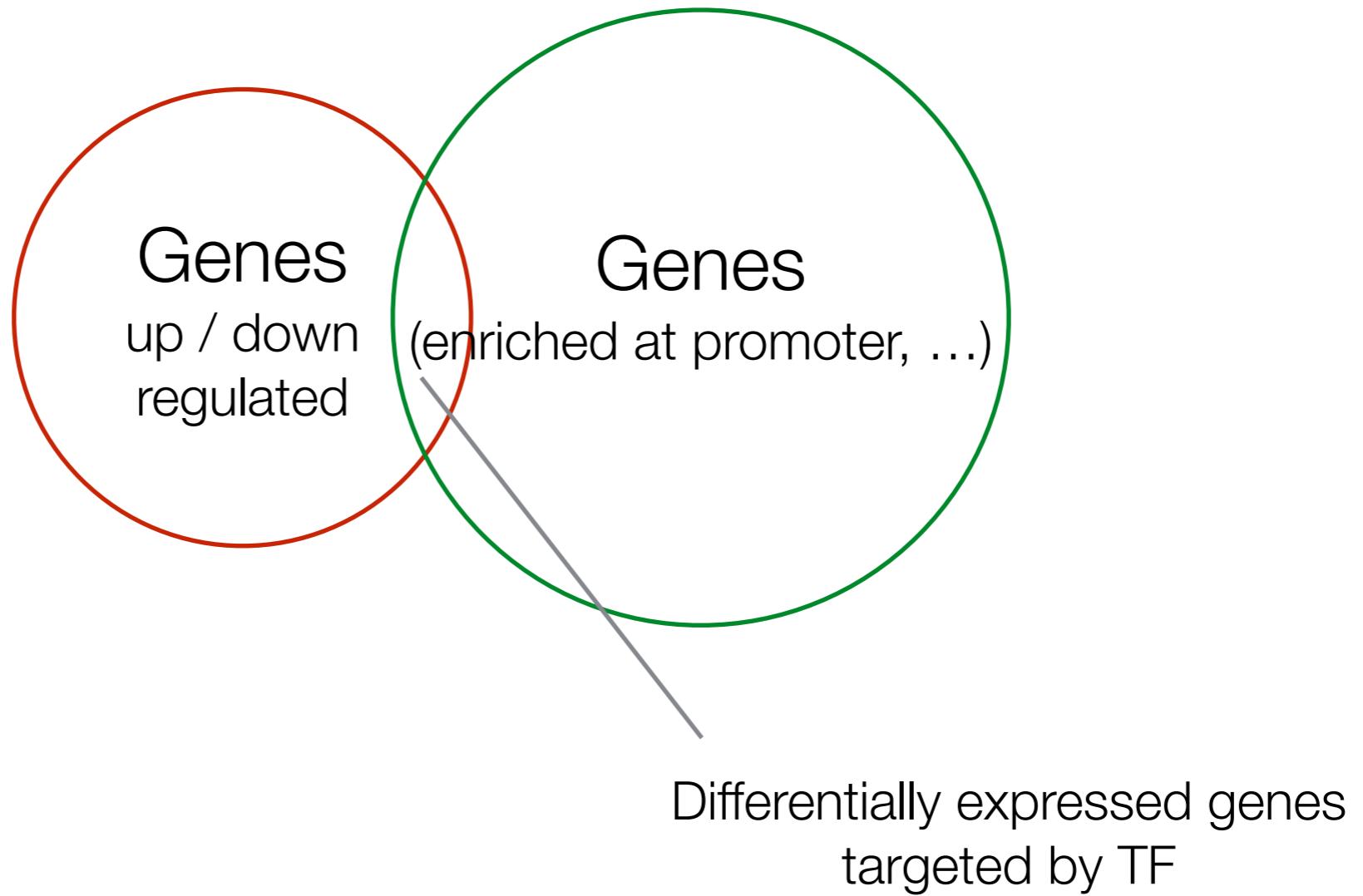


Genes  
up / down regulated

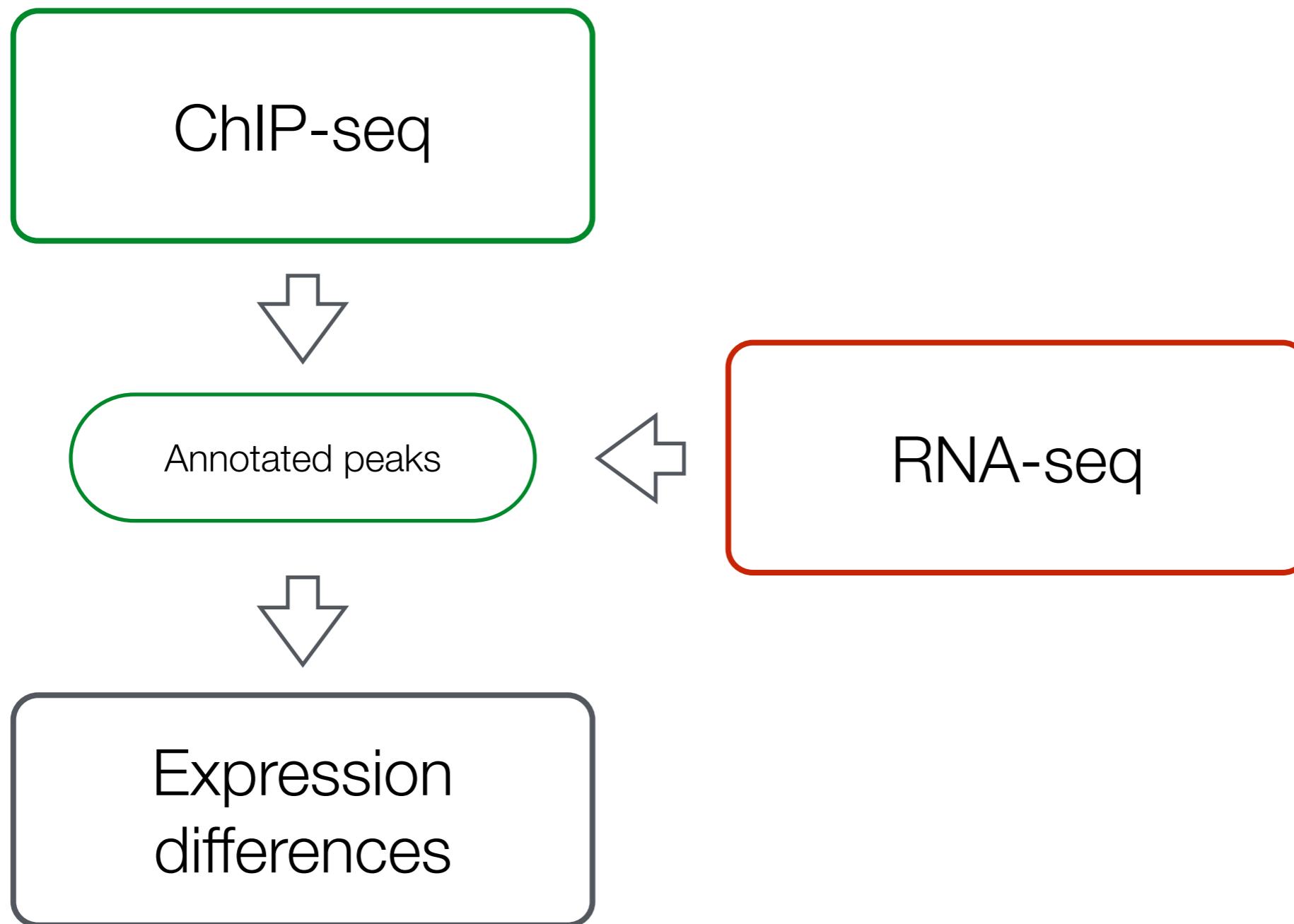
### **Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks**

Cole Trapnell<sup>1,2</sup>, Adam Roberts<sup>3</sup>, Loyal Goff<sup>1,2,4</sup>, Geo Pertea<sup>5,6</sup>, Daehwan Kim<sup>5,7</sup>, David R Kelley<sup>1,2</sup>, Harold Pimentel<sup>3</sup>, Steven L Salzberg<sup>5,6</sup>, John L Rinn<sup>1,2</sup> & Lior Pachter<sup>3,8,9</sup>

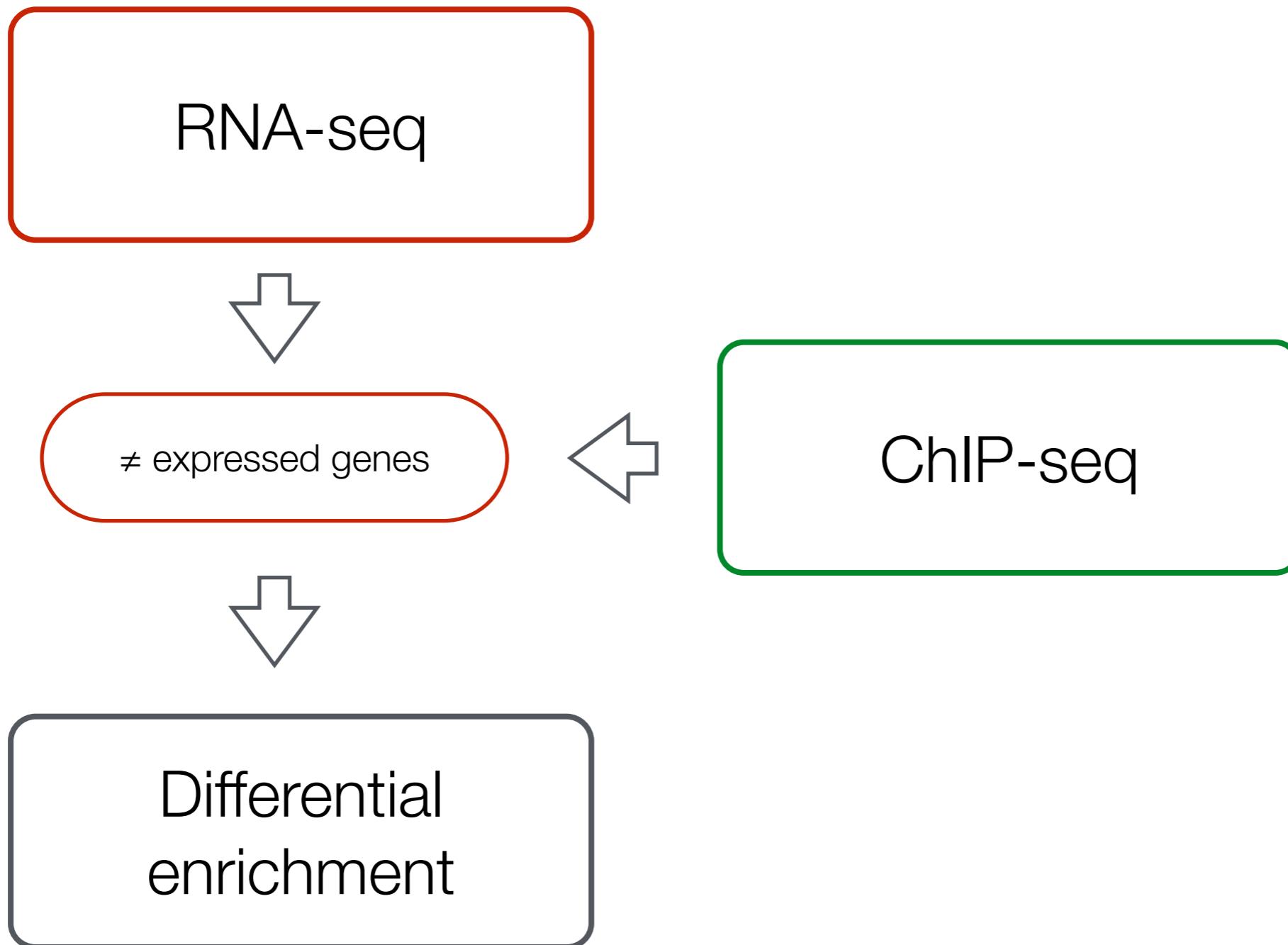
## OPTION 1: focus on genes, on/off state



## OPTION 2: quantitative analysis



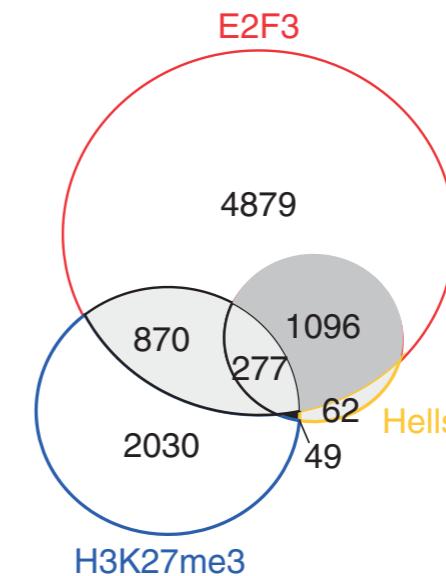
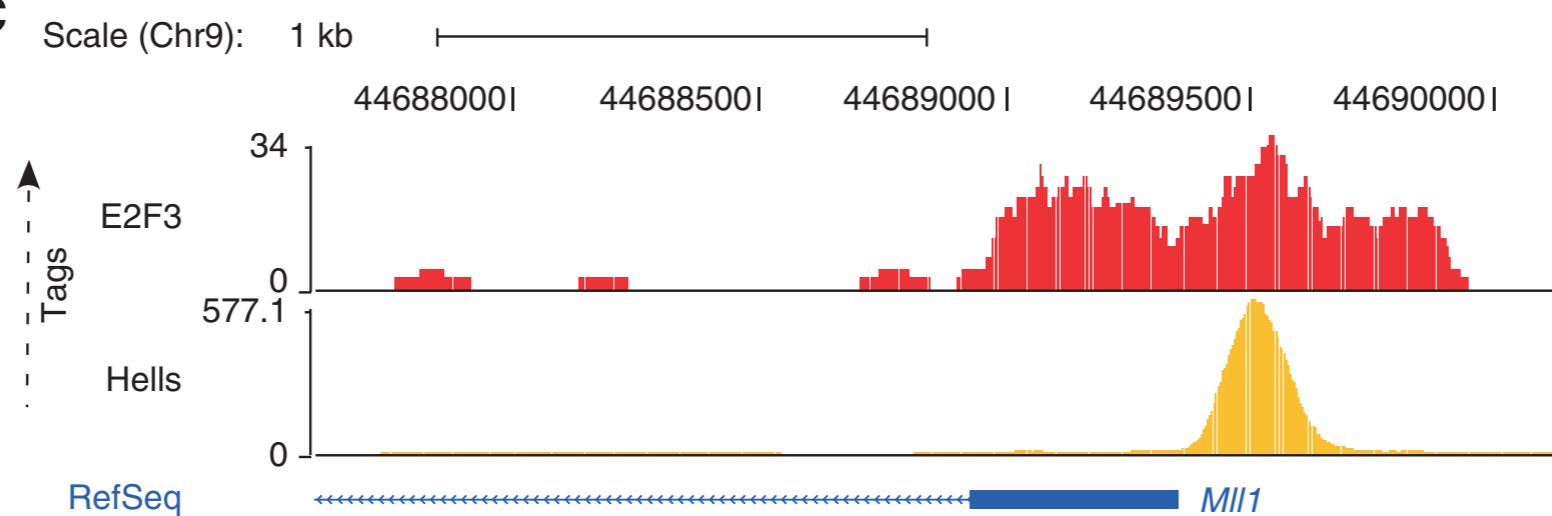
## OPTION 2: quantitative analysis



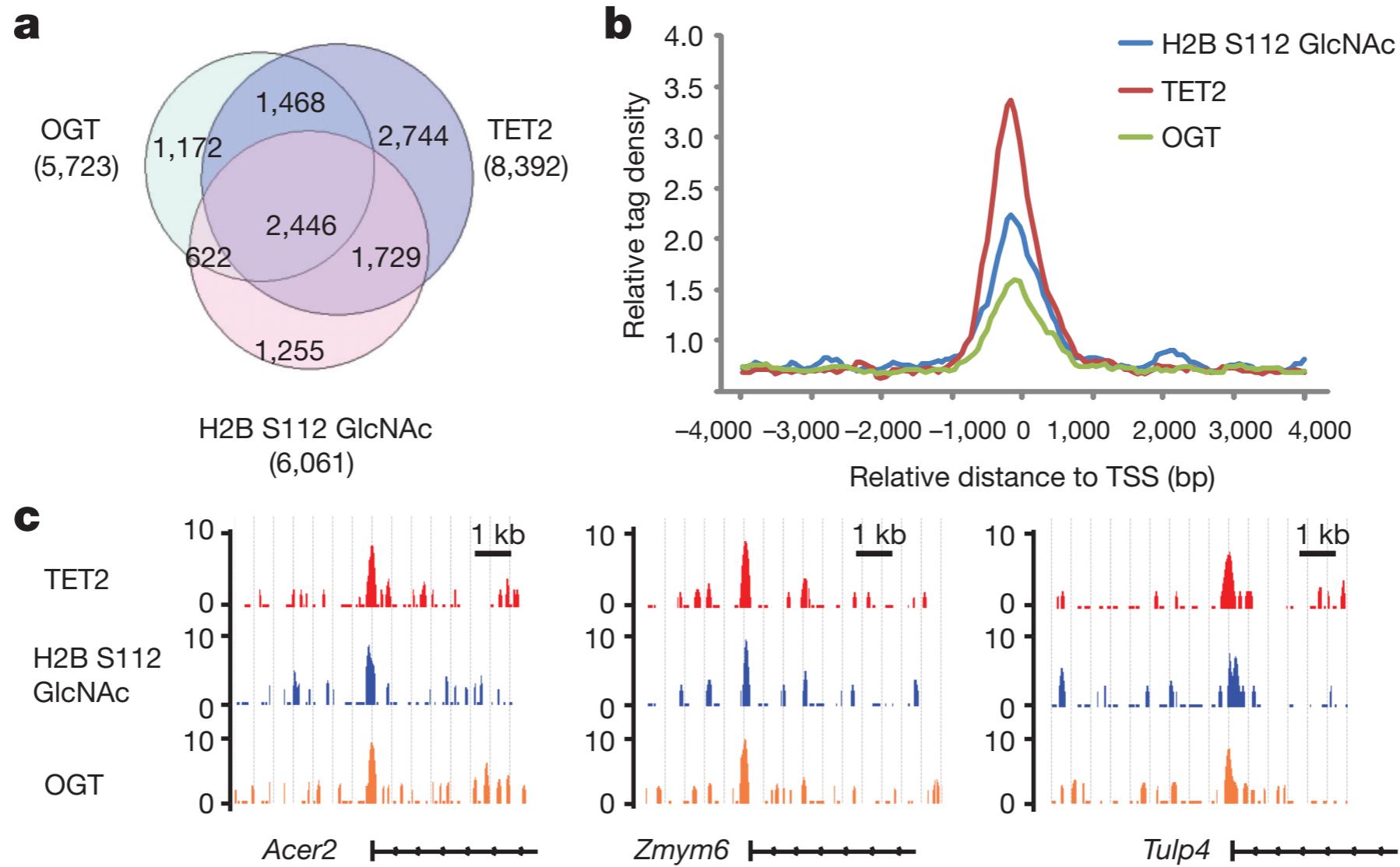
# Histone marks and TF binding

**A**

	Significant peaks	TSS within 1 kb of peaks
E2F3	8889	7122
Hells	2489	1484
H3K27me3	17 064	3226

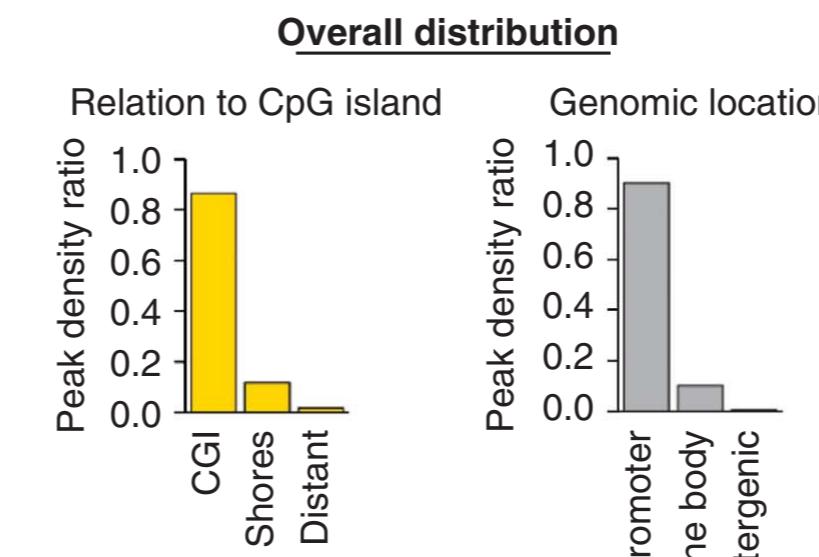
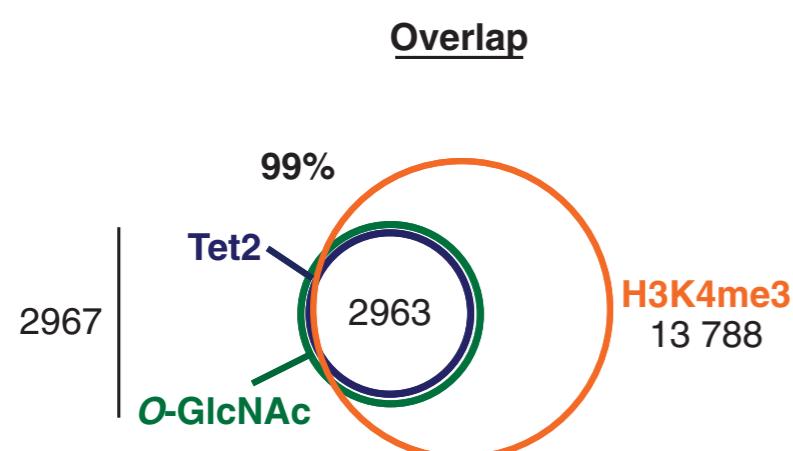
**B****C**

# Co-localization

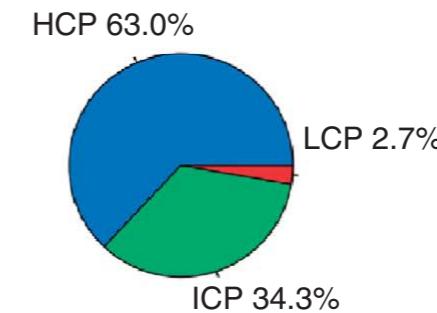
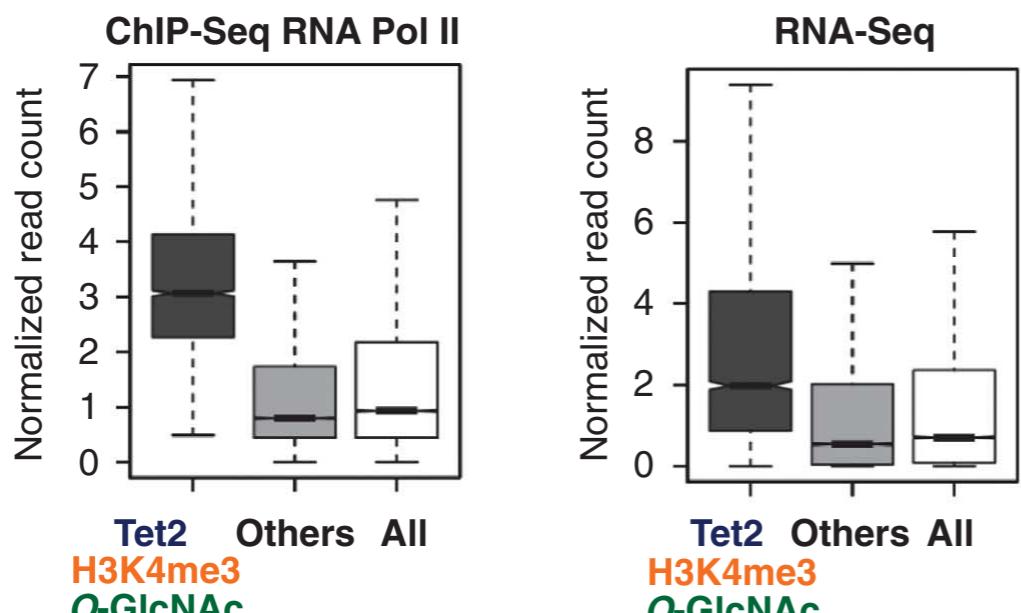
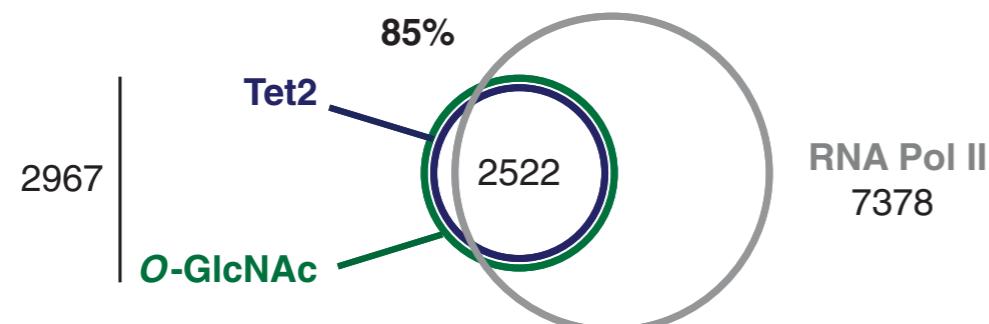


**A**

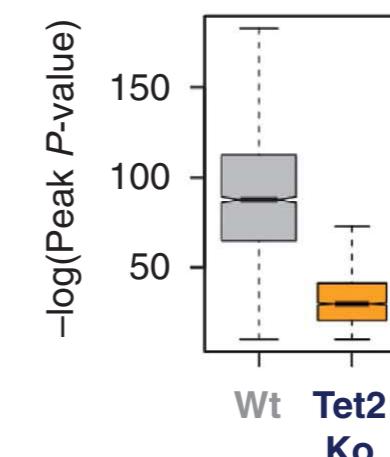
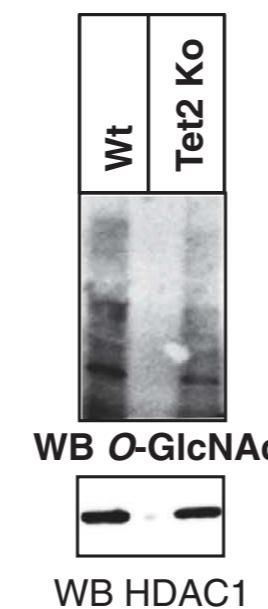
# RNA-seq ChIP-seq



**Promoter**  
**CpG density**

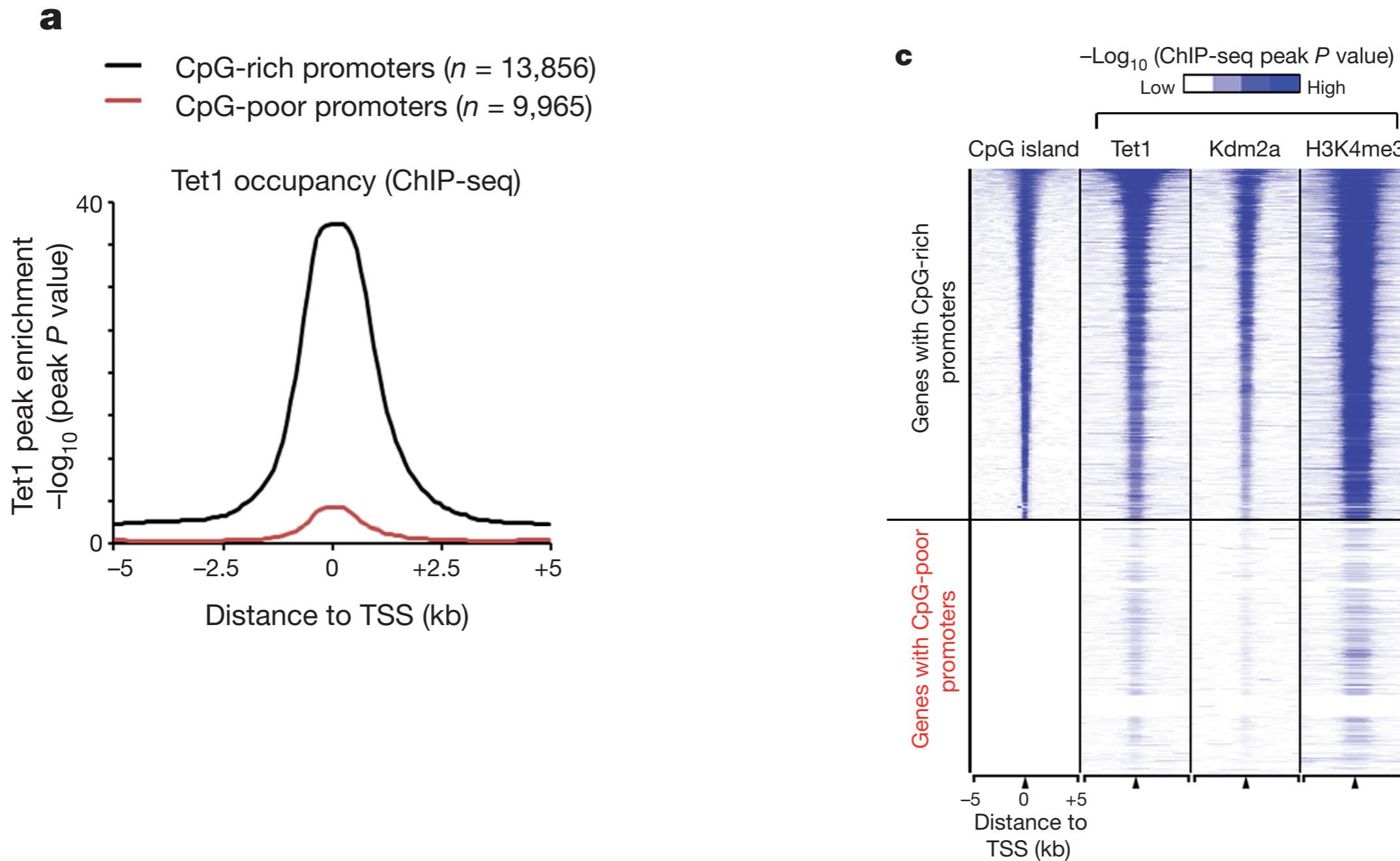
**B**

**C Tet2 knock-out:**  
**WB O-GlcNAc**



**D Tet2 knock-out:**  
**ChIP-Seq H3K4me3**

# Histone marks and TF binding



# Combining Histone marks and TF binding and expression

