Epigenetics and High Throughput Biology

Xinghua Lu
Dept Biomedical Informatics

Genetics and Epigenetics

- Genetics
  - The genome sequence and its information
- Epigenetics
  - Greek, epi=above, upon
  - The study of inheritable changes in gene function and phenotypes caused by mechanisms other than changes in the underlying DNA sequences.
  - DNA methylation,
  - Histone modification
  - Chromatin modification

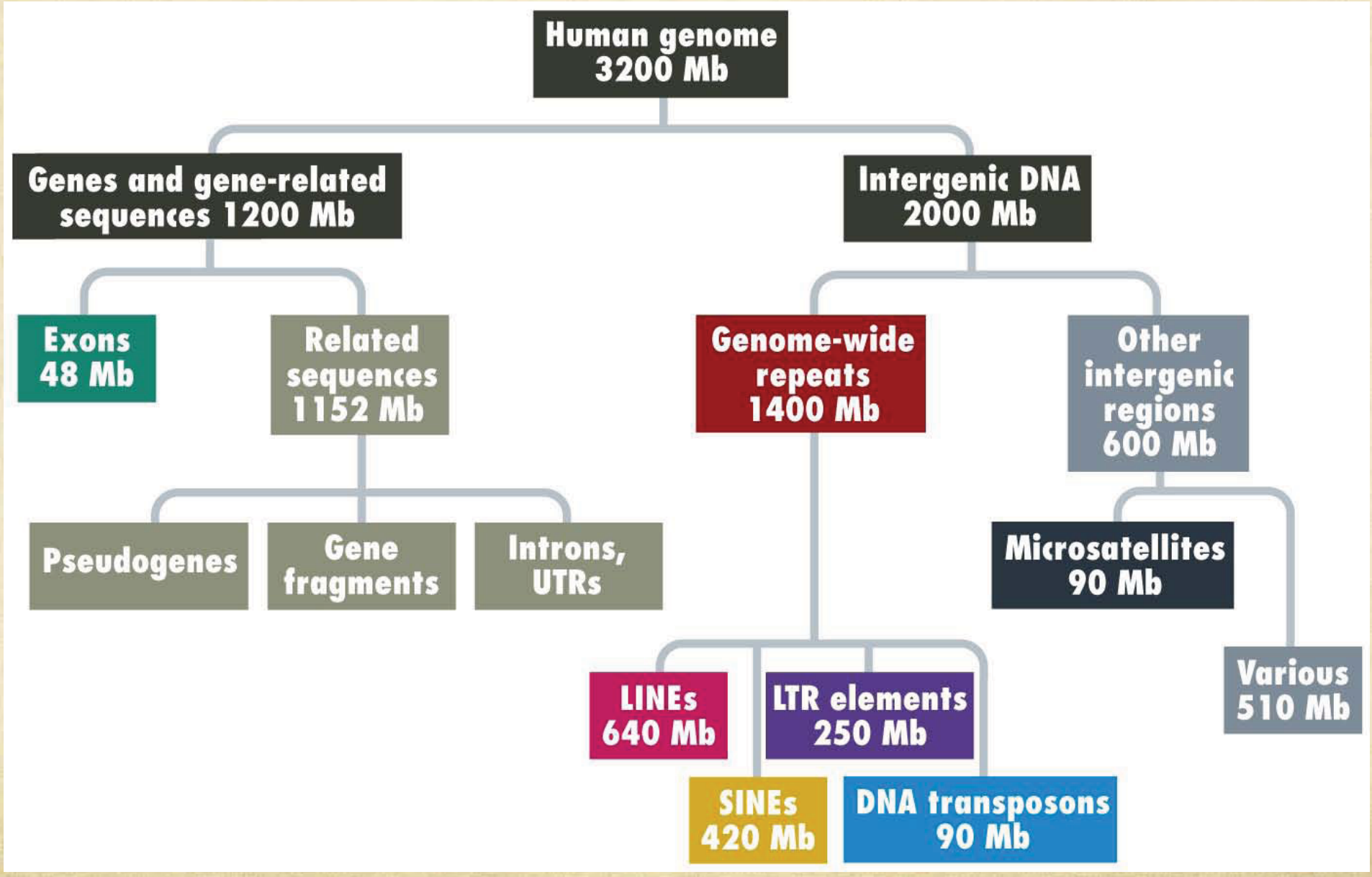
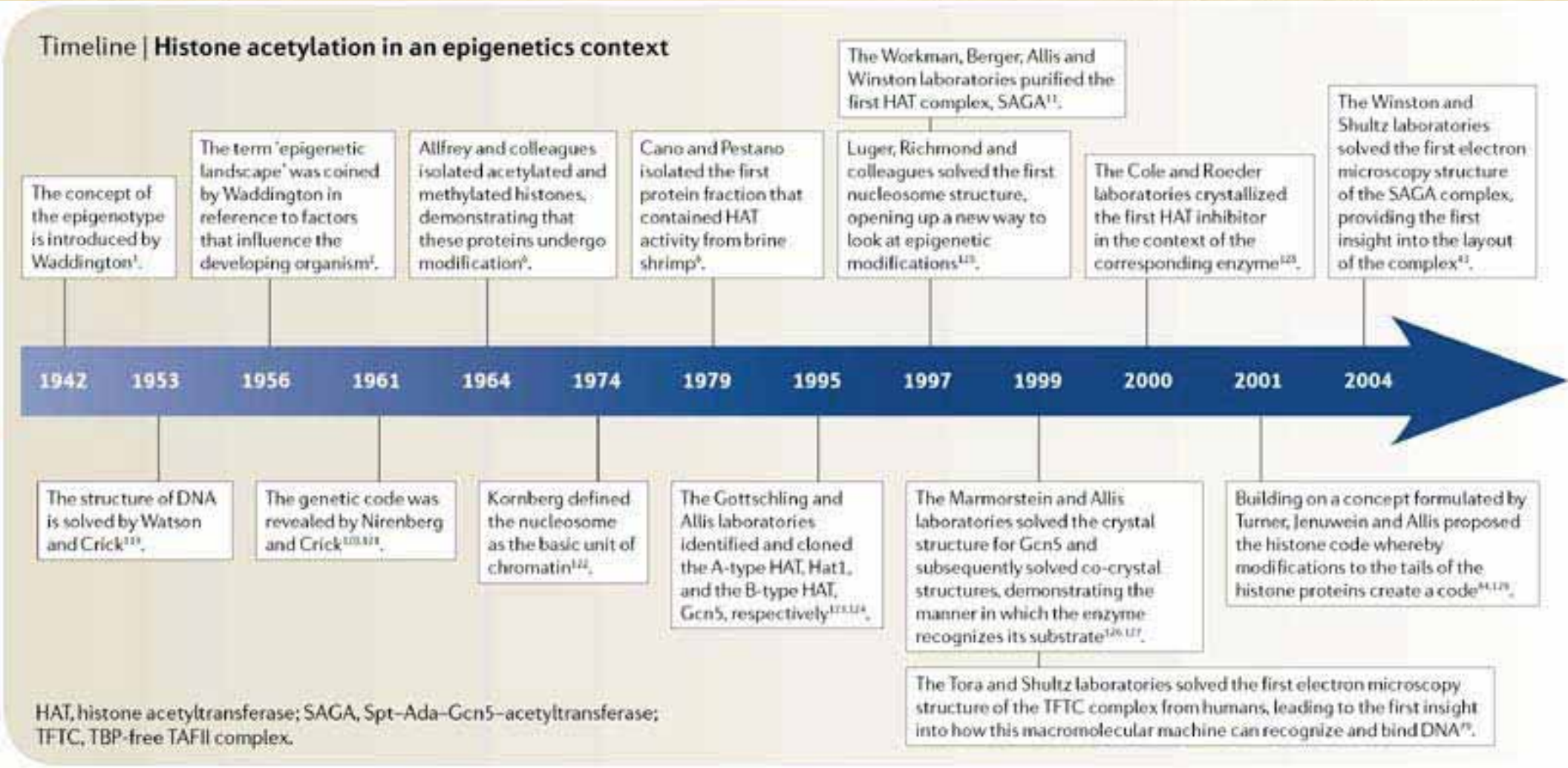
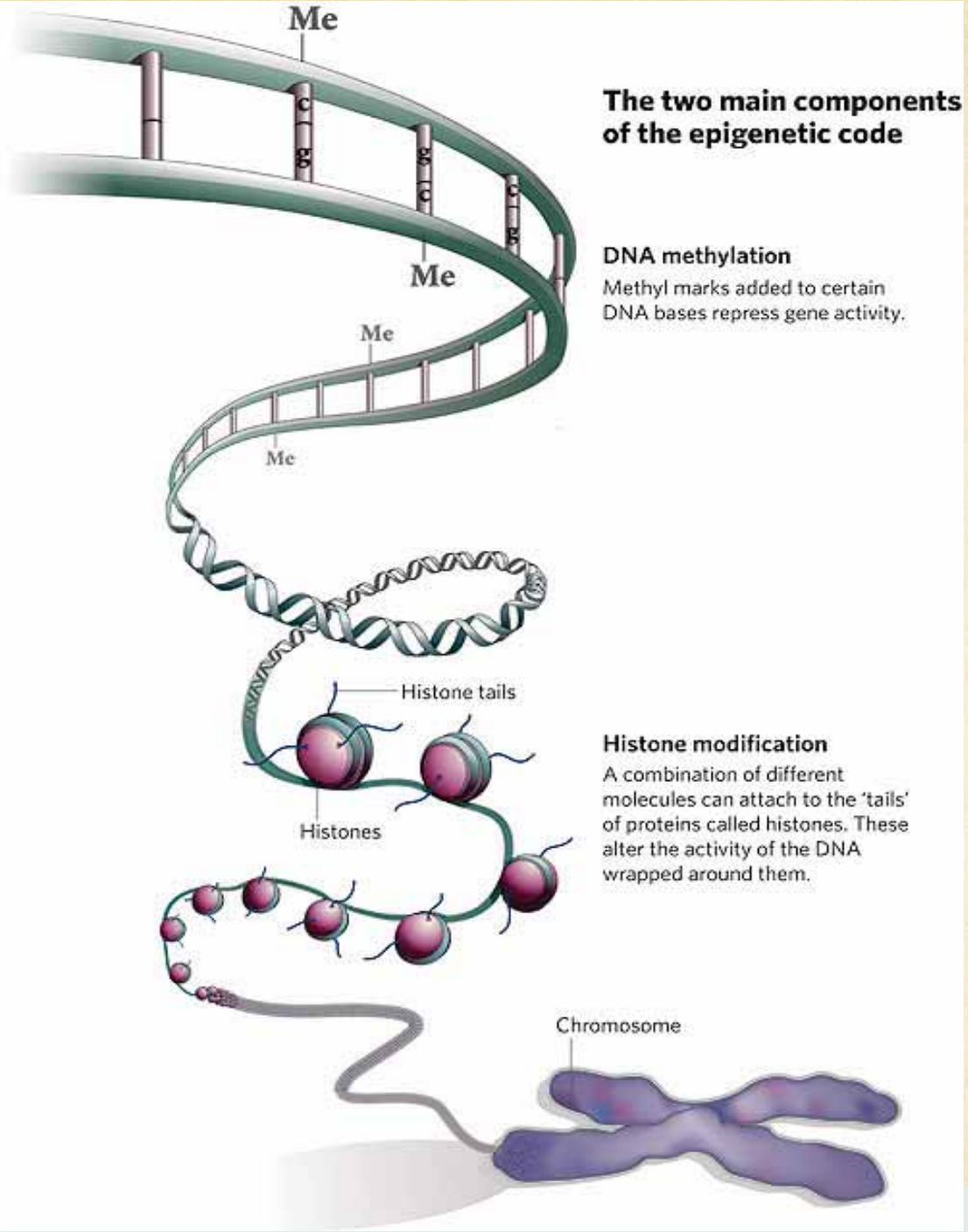


Figure 7.13 *Genomes 3* (© Garland Science 2007)

A bit of History

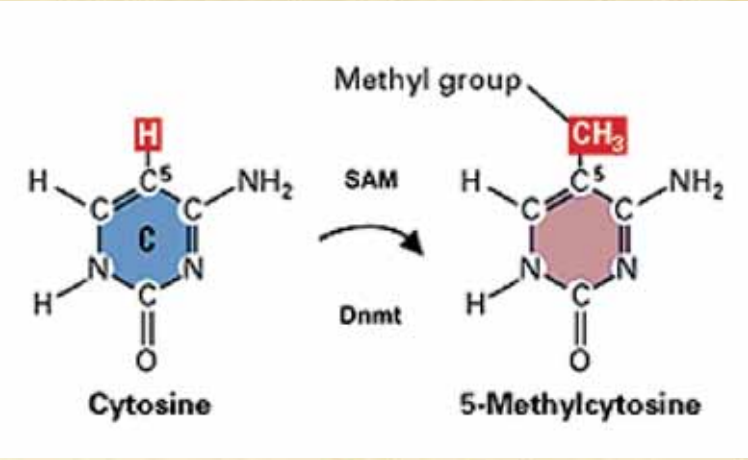


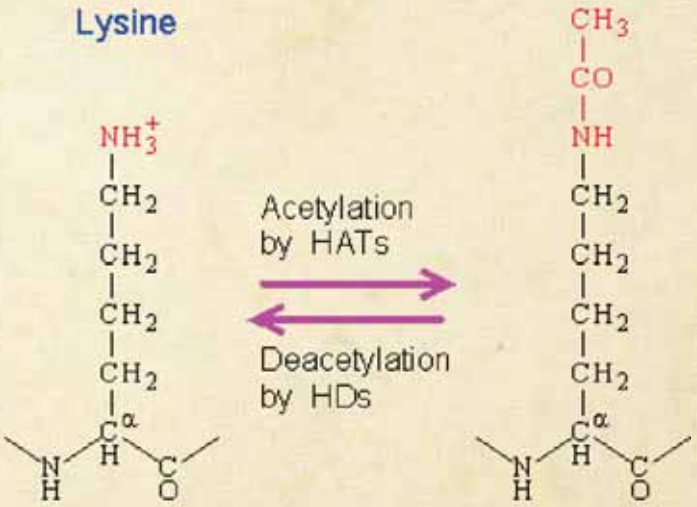


The two main components of the epigenetic code

DNA methylation
Methyl marks added to certain DNA bases repress gene activity.

Histone modification
A combination of different molecules can attach to the 'tails' of proteins called histones. These alter the activity of the DNA wrapped around them.





REVIEWS

APPLICATIONS OF NEXT-GENERATION SEQUENCING

Charting histone modifications and the functional organization of mammalian genomes

Vicky W. Zhou[*‡§||¶], *Alon Goren*[*‡§¶] *and Bradley E. Bernstein*[*‡§]

http://www.nature.com/nrg/journal/v12/n1/pdf/nrg2905.pdf

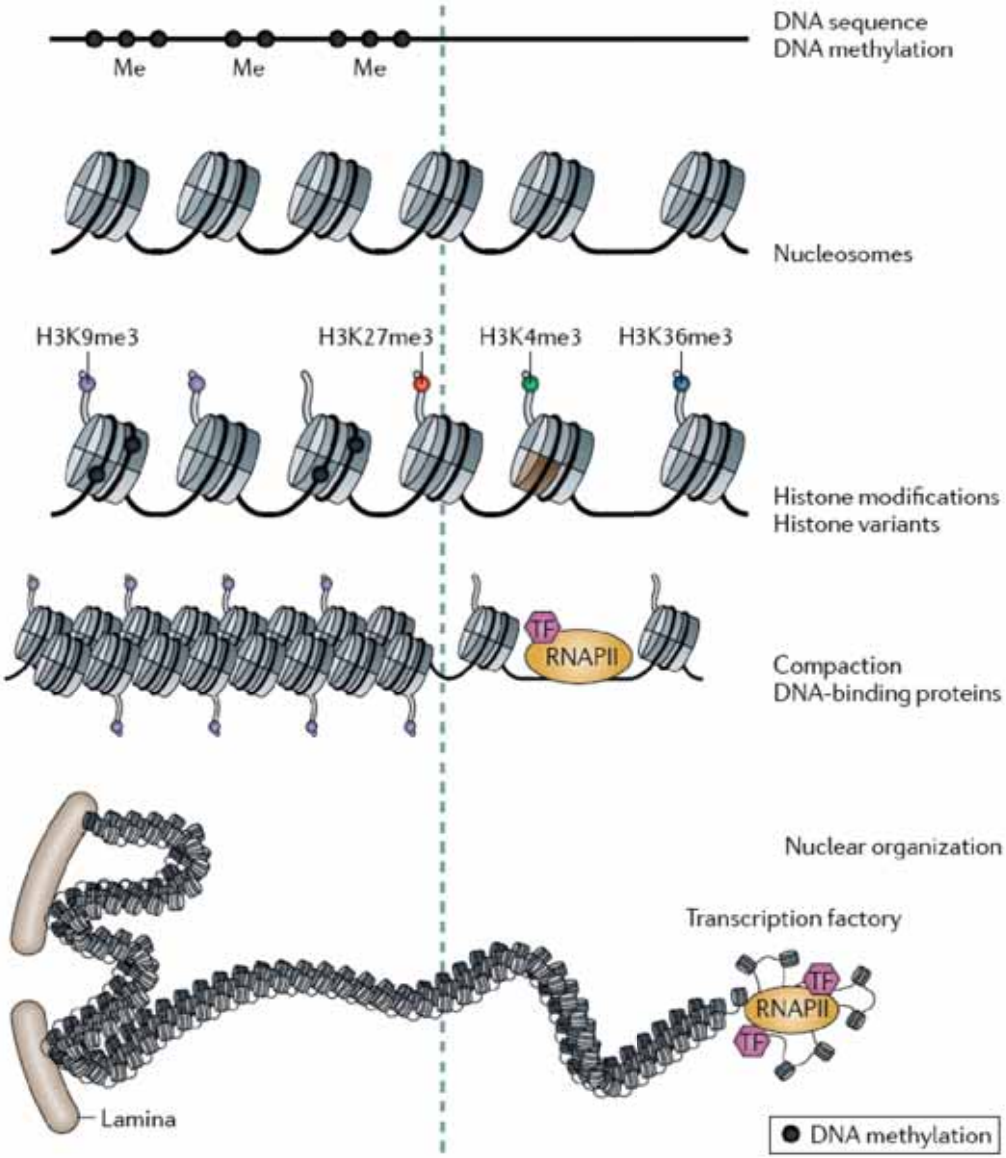
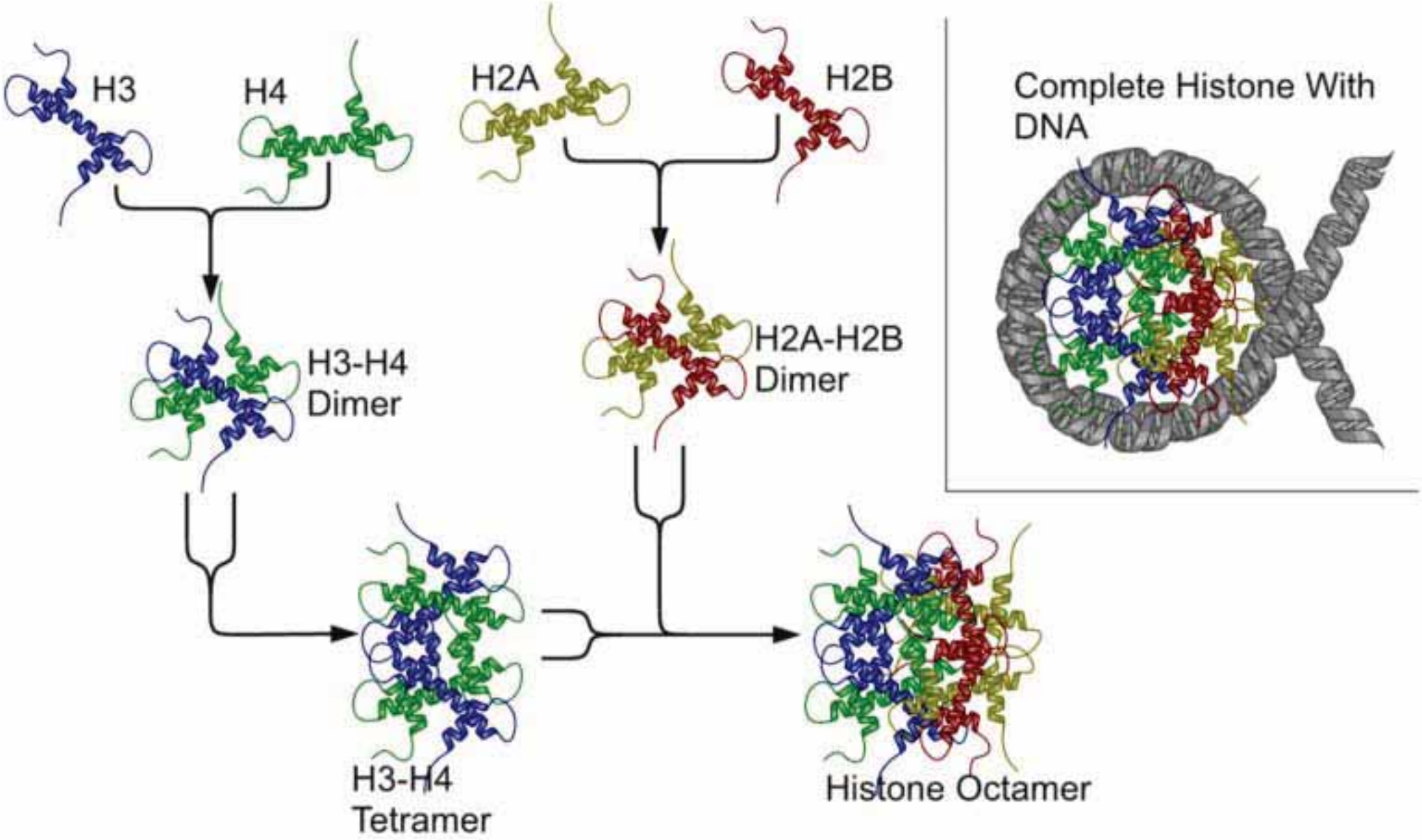
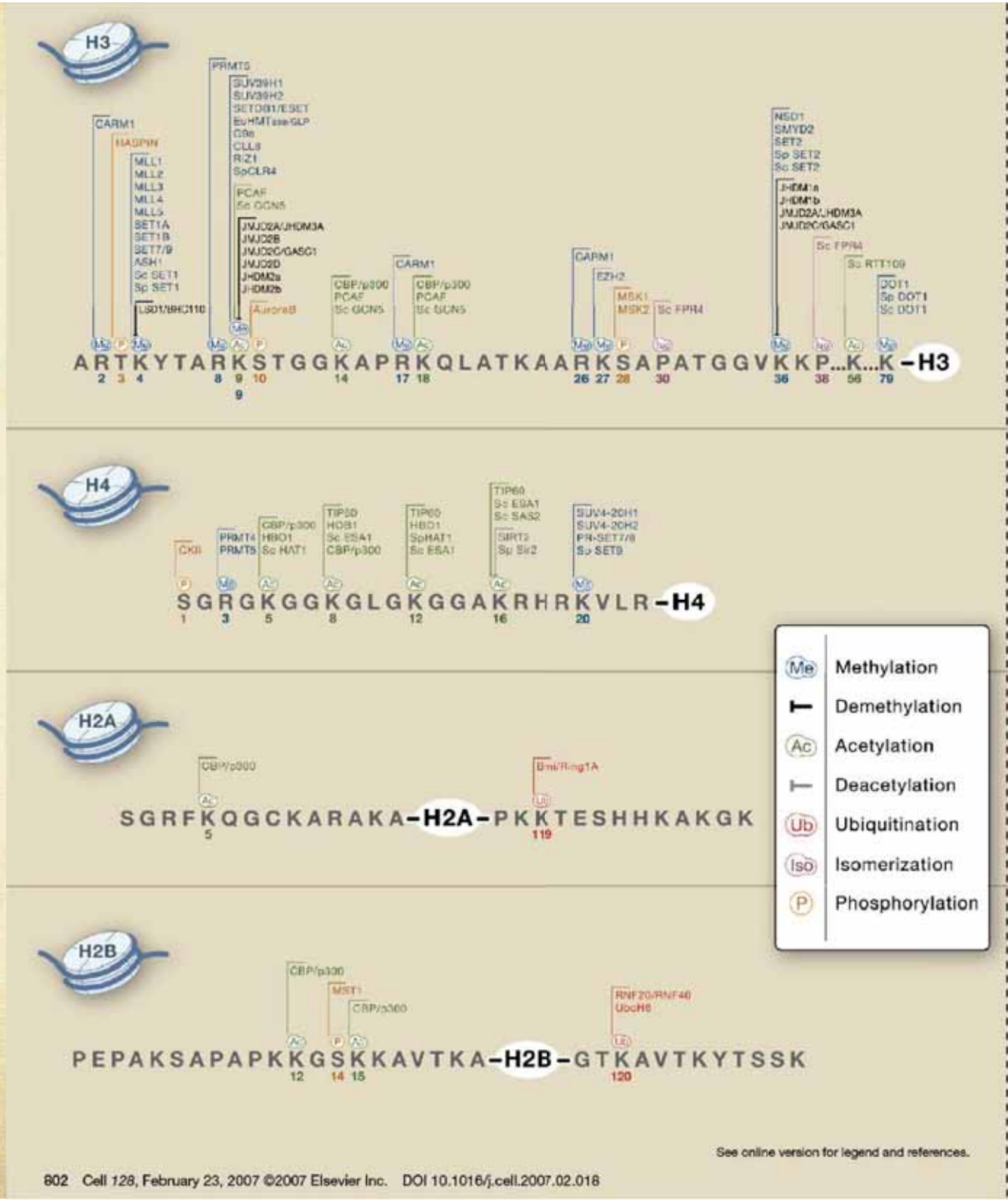


Figure 1 | **Layers of chromatin organization in the mammalian cell nucleus.** Broadly, features at different levels of chromatin organization are generally associated with inactive (off) or active (on) transcription. From the top, genomic DNA is methylated (Me) on cytosine bases in specific contexts and is packaged into nucleosomes, which vary in histone composition and histone modifications (for example, histone H3 lysine 9 trimethylation (H3K9me3)); these features constitute the primary layer of chromatin structure. Here, different histone modifications are indicated by coloured dots and histone variants such as H2A.Z are brown. DNA in chromatin may remain accessible to DNA-binding proteins such as transcription factors (TFs) and RNA polymerase II (RNAPII) or may be further compacted. Chromatin can also organize into higher-order structures such as nuclear lamina-associated domains and transcription factories. Each layer of organization reflects aspects of gene and genome regulation.


H3
H4
H2A
H2B
Complete Histone With DNA
H3-H4 Dimer
H2A-H2B Dimer
H3-H4 Tetramer
Histone Octamer

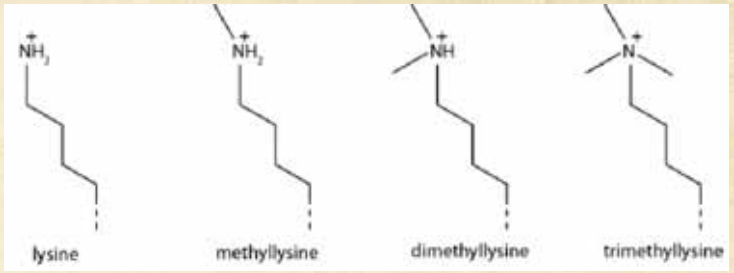
Enzymes Involved in Histone modifications



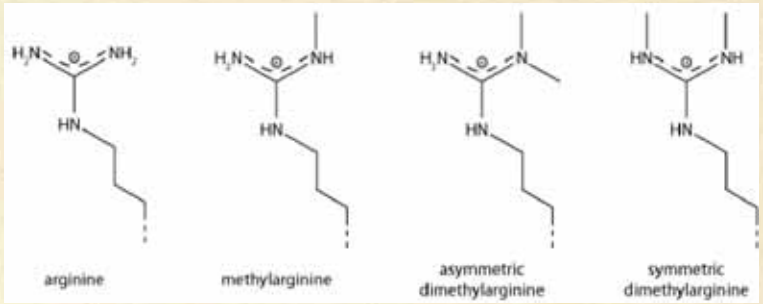
See online version for legend and references.

Histone Modifications

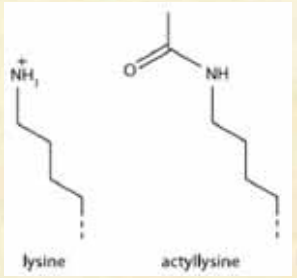
- Lysine methylation



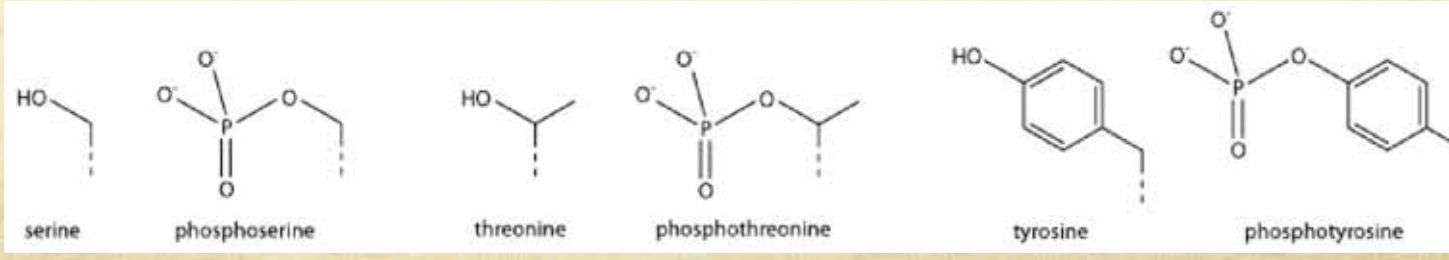
- Arginine methylation



- Lysine acetylation



- Phosphorylation



Techniques of Detecting Histone Modification

- Chip-seq
  - Different Antibodies recognizing distinct modifications
- FAIRE
- SONO-seq
- DNase-seq

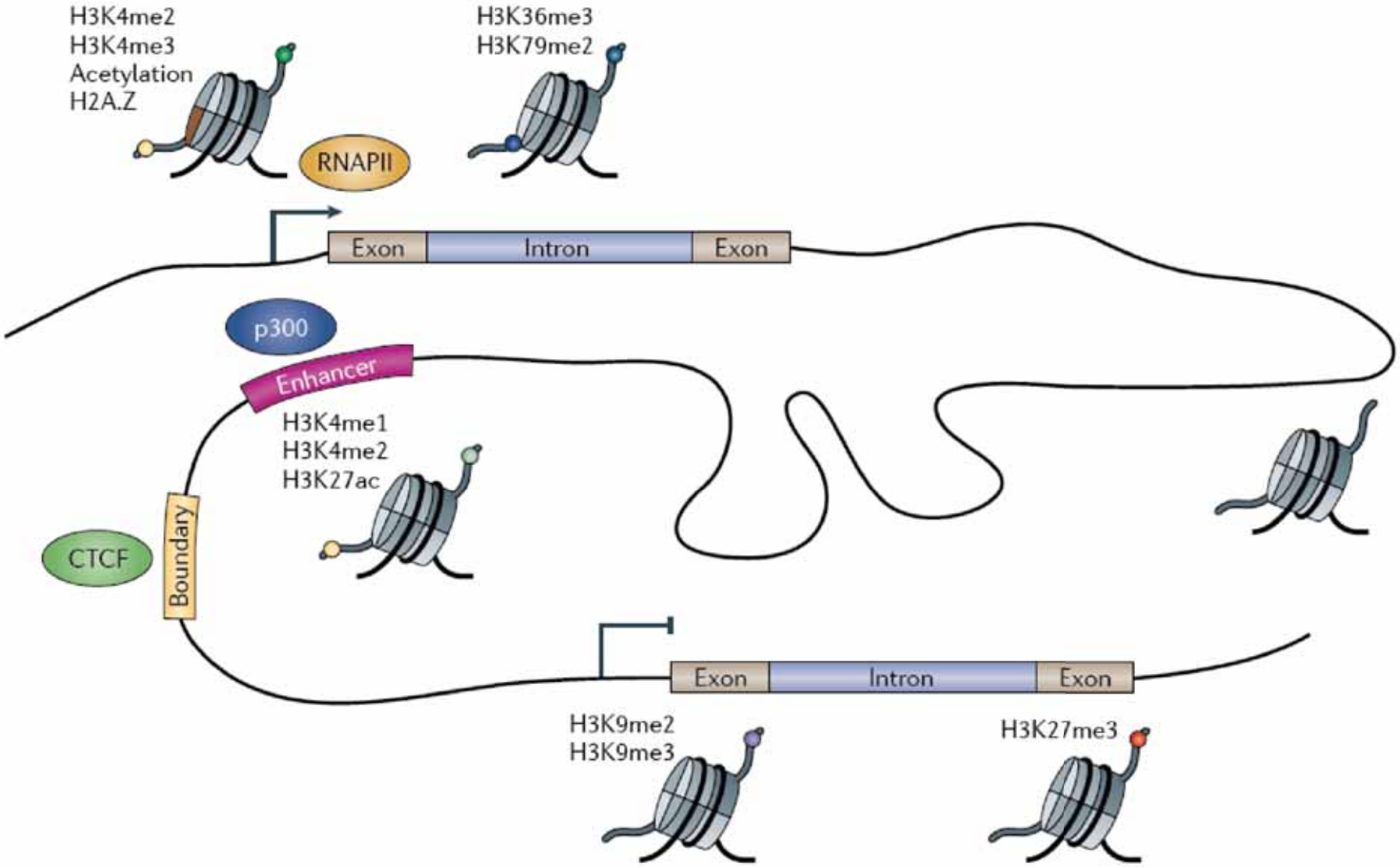


Figure 2 | **Histone modifications demarcate functional elements in mammalian genomes.** Promoters, gene bodies, an enhancer and a boundary element are indicated on a schematic genomic region. Active promoters are commonly marked by histone H3 lysine 4 dimethylation (H3K4me2), H3K4me3, acetylation (ac), and H2A.Z. Transcribed regions are enriched for H3K36me3 and H3K79me2. Repressed genes may be located in large domains of H3K9me2 and/or H3K9me3 or H3K27me3. Enhancers are relatively enriched for H3K4me1, H3K4me2, H3K27ac and the histone acetyltransferase p300. CTCF binds many sites that may function as boundary elements, insulators or structural scaffolds. These various features of chromatin help organize the DNA and distinguish functional elements in the large expanse of the genome. RNAPII, RNA polymerase II.

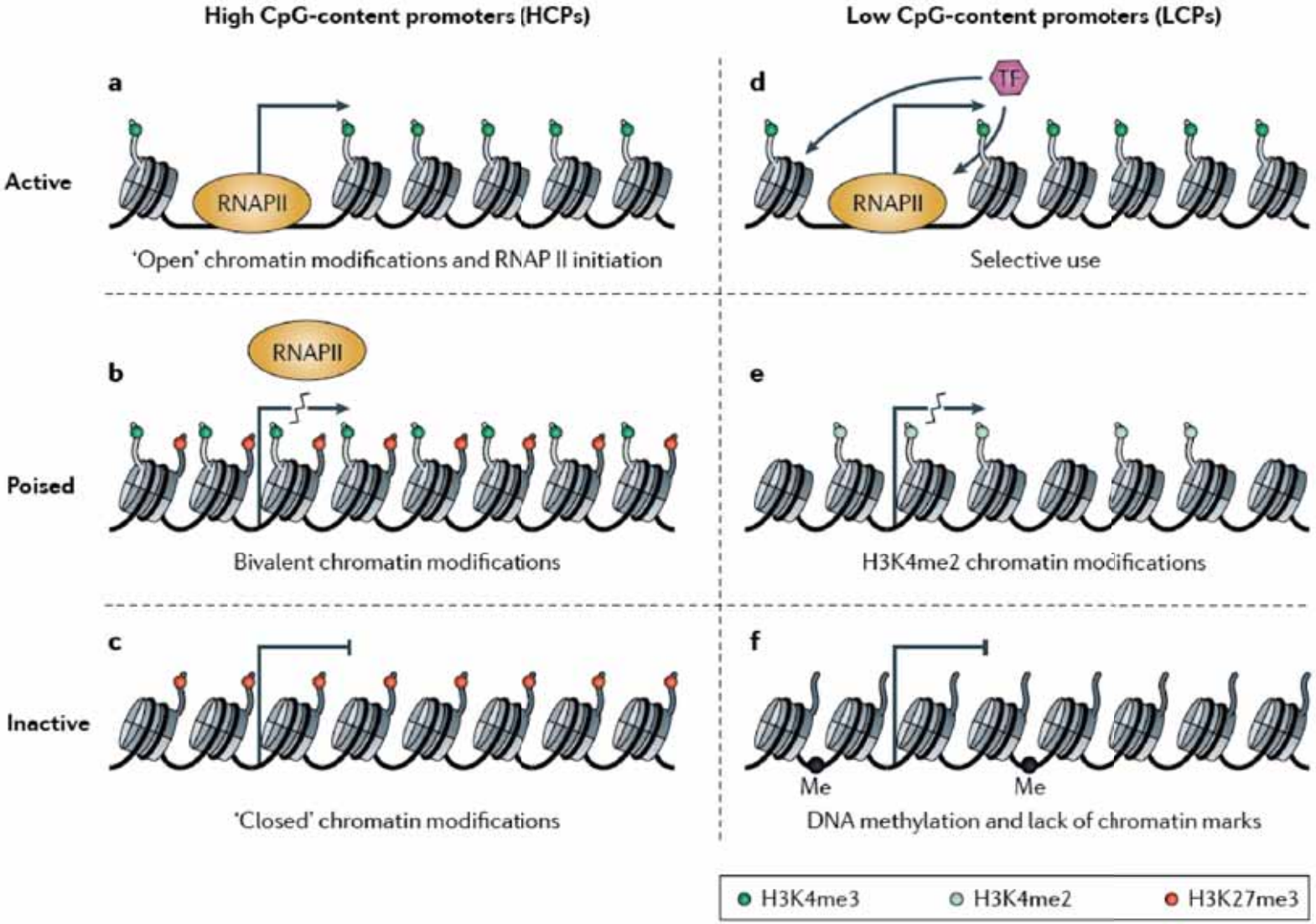


Figure 3 | **Chromatin patterns and regulation by promoter class.** Promoters can be classified according to their CpG content. High CpG-content promoters (HCPs) and low CpG-content promoters (LCPs) are subject to distinct chromatin patterns and regulation. **a** | HCPs have characteristics of accessible or 'active' chromatin by default. Active HCPs (for example, housekeeping gene promoters) are enriched for histone H3 lysine 4 trimethylation (H3K4me3) and subject to RNA polymerase II (RNAPII) initiation. They may be subject to additional regulation at the transition to elongation. **b** | Poised HCPs (for example, developmental regulator gene promoters in embryonic stem cells) are marked by the bivalent combination of *H3K4me3* and *H3K27me3*. They may be subject to *RNAPII* initiation, but tend not to elongate or make productive mRNA. **c** | Inactive HCPs carry 'repressive' chromatin modifications such as H3K27me3 and are relatively inaccessible to RNAPII. Unlike HCP chromatin, LCP chromatin seems to be selectively activated (for example, by specific transcription factors (TFs)). **d** | Active LCPs are enriched for H3K4me3 and transcribed. **e** | Poised LCPs may be marked by H3K4me2 without H3K4me3. **f** | Inactive LCPs typically lack chromatin marks but may be DNA methylated (Me).

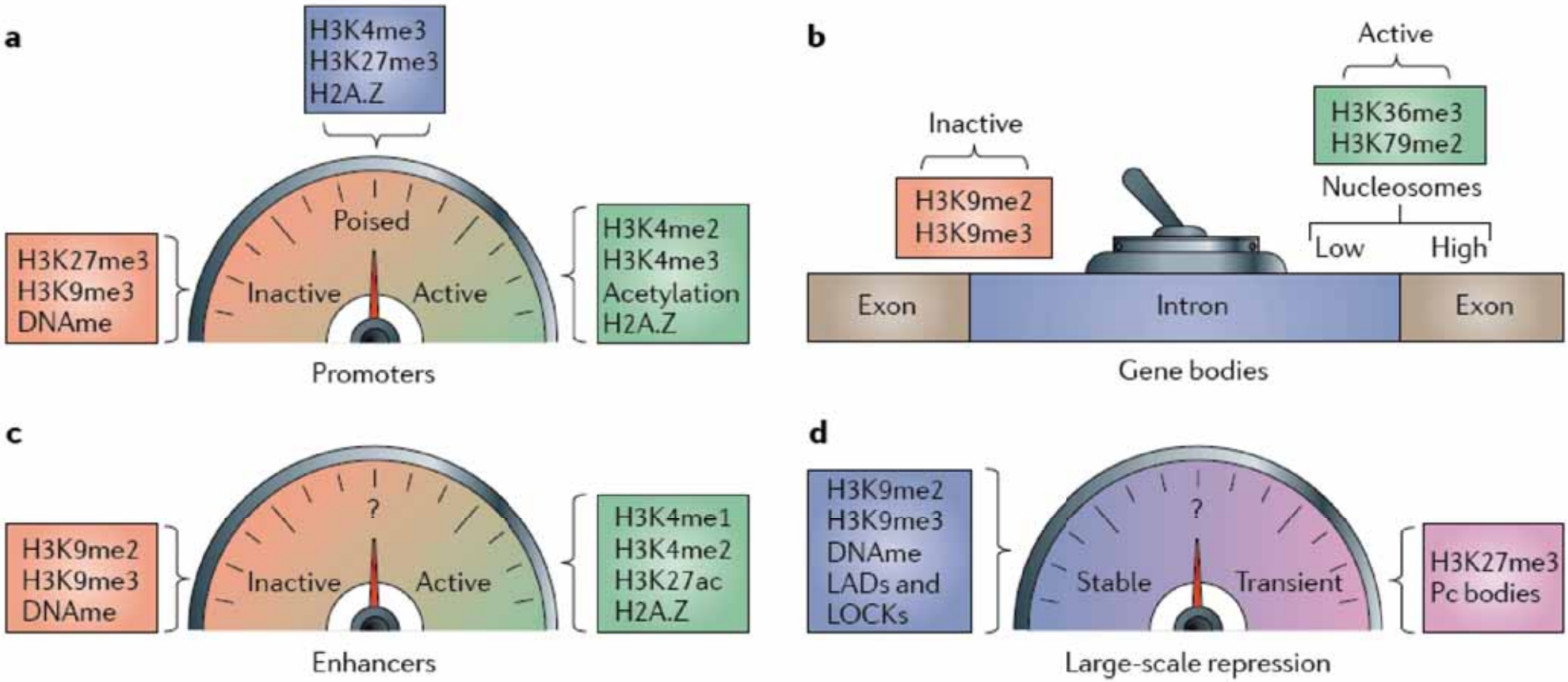


Figure 4 | **'Dashboard' of histone modifications for fine-tuning genomic elements.** In addition to enabling annotation, histone modifications may serve as 'dials' or 'switches' for cell type specificity. **a** | At promoters, they can contribute to fine-tuning of expression levels — from active to poised to inactive — and perhaps even intermediate levels. **b** | At gene bodies, they discriminate between active and inactive conformations. In addition, exons in active genes have higher nucleosome occupancy and thus more histone H3 lysine 36 trimethylation (H3K36me3) and H3K79me2-modified histones than introns. **c** | At distal sites, histone marks correlate with levels of enhancer activity. **d** | On a global scale, they may confer repression of varying stabilities and be associated with different genomic features. For example, lamina-associated domains (LADs) in the case of stable repression and Polycomb (Pc) bodies in the case of context-specific repression. DNAme, DNA methylation; LOCK, large organized chromatin K modification.

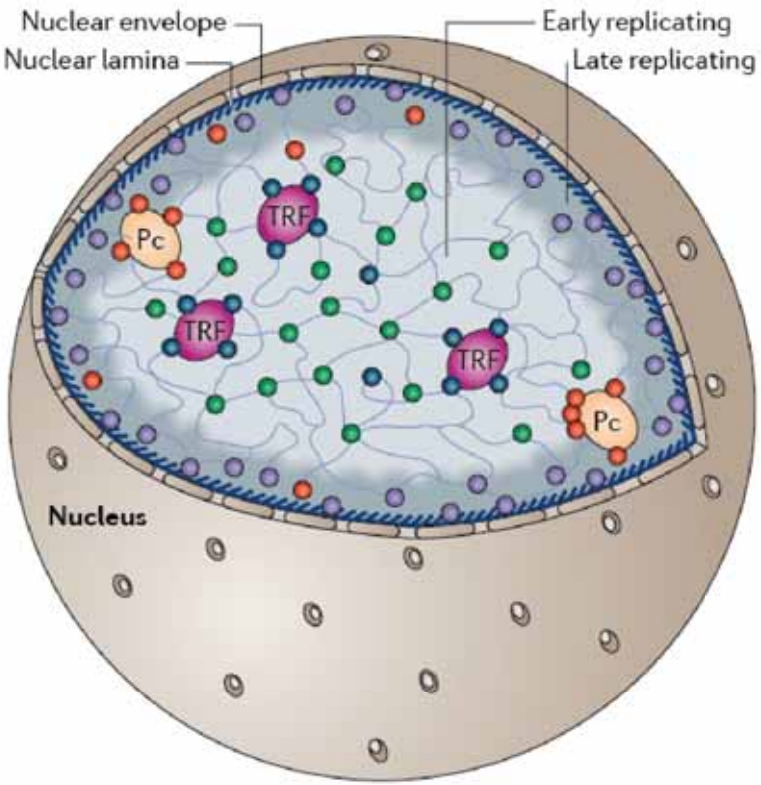


Figure 5 | **Histone modification signatures associated with features in the mammalian cell nucleus.** Signature histone modifications correlate with various nuclear features, although the relationships might be indirect. Chromatin with modifications generally associated with active transcription (green dots) often replicates early, whereas chromatin with generally repressive modifications (purple dots) replicates late. Regions enriched for some sets of active modifications (blue dots) may converge into transcription factories (TRFs). Blocks of histone H3 lysine 27 trimethylation (H3K27me3; red dots) may form Polycomb bodies (Pc) and diffuse domains marked by H3K9me2 or H3K9me3 (purple dots) may contact the nuclear lamina.

Leading Edge
Review

Cell

The Role of Chromatin during Transcription

Bing Li,[1] Michael Carey,[1,2] and Jerry L. Workman[1,*]
[1] Stowers Medical Research Institute, 1000 East 50th Street, Kansas City, MO 64110, USA
[2] Department of Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles, 10833 LeConte Avenue, Los Angeles, CA 90095, USA
*Correspondence: jlw@stowers-institute.org
DOI 10.1016/j.cell.2007.01.015

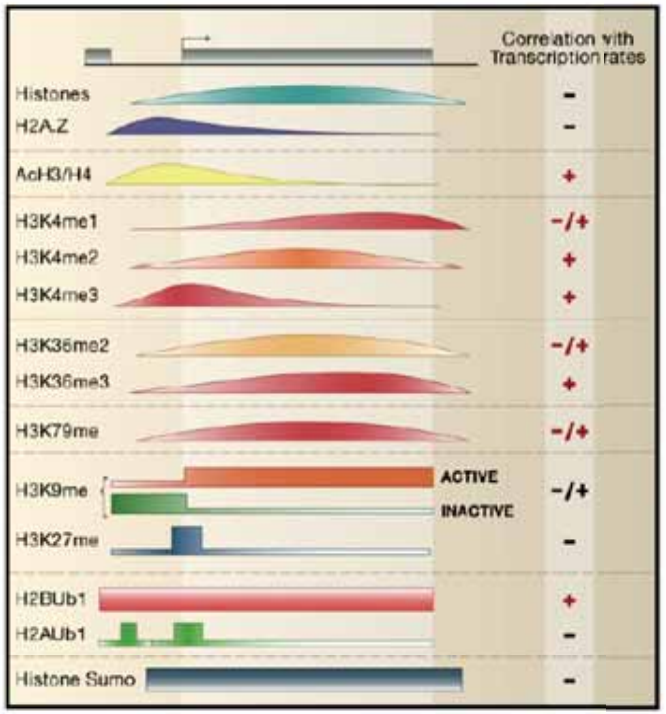


Figure 1. Genome-Wide Distribution Pattern of Histone Modifications from a Transcription Perspective

The distribution of histones and their modifications are mapped on an arbitrary gene relative to its promoter (5′ IGR), ORF, and 3′ IGR (original references were reviewed in Shilatifard, 2006; Workman, 2006). The curves represent the patterns that are determined via genome-wide approaches. The squares indicate that the data are based on only a few case studies. With the exception of the data on K9 and K27 methylation, most of the data are based on yeast genes.

Table 1. Histone Modifications Associated with Transcription

Modifications	Position		Enzymes				Recognition Module(s)[a]	Functions in Transcription
			S. cerevisiae	*S. pombe*	*Drosophila*	Mammals		
Methylation	H3	K4	Set1	Set1	Trx, Ash1	MLL, ALL-1, Set9/7, ALR-1/2, ALR, Set1	PHD, Chromo, WD-40	Activation
		K9	n/a	Clr4	Su(var)3-9, Ash1	Suv39h, G9a, Eu-HMTase I, ESET, SETBD1	Chromo (HP1)	Repression, activation
		K27				E(Z)	Ezh2, G9a	Repression
		K36	Set2			HYPB, Smyd2, NSD1	Chromo(Eaf3), JMJD	Recruiting the Rpd3S to repress internal initiation
		K79	Dot1			Dot1L	Tudor	Activation
	H4	K20		Set9	PR-Set7, Ash1	PR-Set7, SET8	Tudor	Silencing
Arg Methylation	H3	R2				CARM1		Activation
		R17				CARM1		Activation
		R26				CARM1		Activation
	H4	R3				PRMT1	(p300)	Activation
Phosphorylation	H3	S10	Snf1				(Gcn5)	Activation
Ubiquitination	H2B	K120/123	Rad6, Bre1	Rad6		UbcH6, RNF20/40	(COMPASS)	Activation
	H2A	K119				hPRC1L		Repression
Acetylation	H3	K56					(Swi/Snf)	Activation
	H4	K16	Sas2, NuA4		dMOF	hMOF	Bromodomain	Activation
	Htz1	K14	NuA4, SAGA					Activation

[a] The proteins that are indicated within the parentheses are shown to recognize the corresponding modifications but specific domains have yet to be determined.

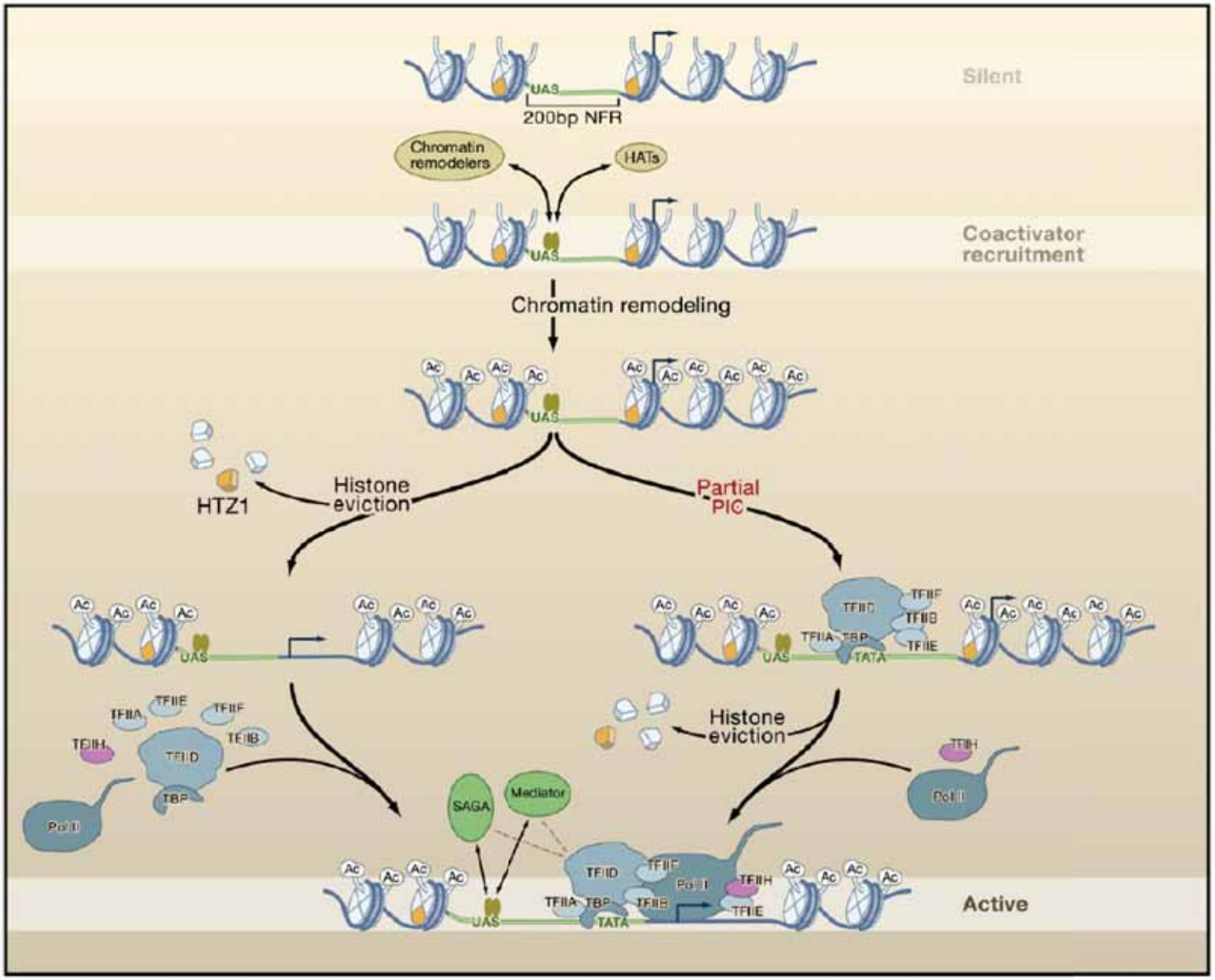


Figure 2. Models of Chromatin Regulation during Transcription Initiation

At the silent promoter, Htz1-containing nucleosomes flank a 200 bp NFR on both sides. Upon targeting to the upstream-activation sequence (UAS), activators recruit various coactivators (such as Swi/Snf or SAGA). This recruitment further increases the binding of activators, particularly for those bound within nucleosomal regions. More importantly, histones are acetylated at promoter-proximal regions, and these nucleosomes become much more mobile. In one model (left), a combination of acetylation and chromatin remodeling directly results in the loss of Htz1-containing nucleosome, thereby exposing the entire core promoter to the GTFs and Pol II. SAGA and mediator then facilitate PIC formation through direct interactions. In the other model (right), which represents the remodeled state, partial PICs could be assembled at the core promoter without loss of Htz1. It is the binding of Pol II and TFIIH that leads to the displacement of Htz1-containing nucleosomes and the full assembly of PIC.

Cell

Resource

Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification

Minjia Tan,[1,6] Hao Luo,[1,6] Sangkyu Lee,[1,6] Fulai Jin,[2] Jeong Soo Yang,[1] Emilie Montellier,[3] Thierry Buchou,[3] Zhongyi Cheng,[1] Sophie Rousseaux,[3] Nisha Rajagopal,[2] Zhike Lu,[1] Zhen Ye,[2] Qin Zhu,[4] Joanna Wysocka,[5] Yang Ye,[4] Saadi Khochbin,[3] Bing Ren,[2] and Yingming Zhao[1,*]

[1]Ben May Department of Cancer Research, The University of Chicago, Chicago, IL 60637, *USA*

[2]Ludwig Institute for Cancer Research and Department of Cellular and Molecular Medicine, University of California San Diego School of Medicine, 9500 Gilman Drive, La Jolla, CA 92093, USA

[3]INSERM, U823; Université Joseph Fourier - Grenoble 1; Institut Albert Bonniot, Faculté de Médecine, Domaine de la Merci, 38706 La Tronche Cedex, France

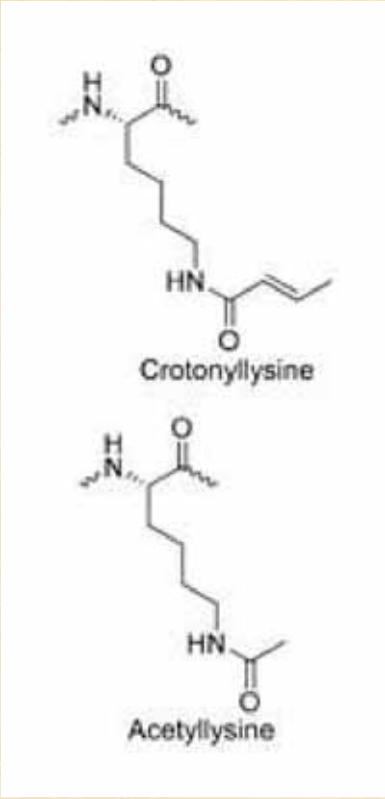
[4]Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zu Chong Zhi Road, Shanghai 201203, P.R. China

[5]Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA 94305, USA

[6]These authors contributed equally to this work

*Correspondence: yingming.zhao@uchicago.edu

DOI 10.1016/j.cell.2011.08.008



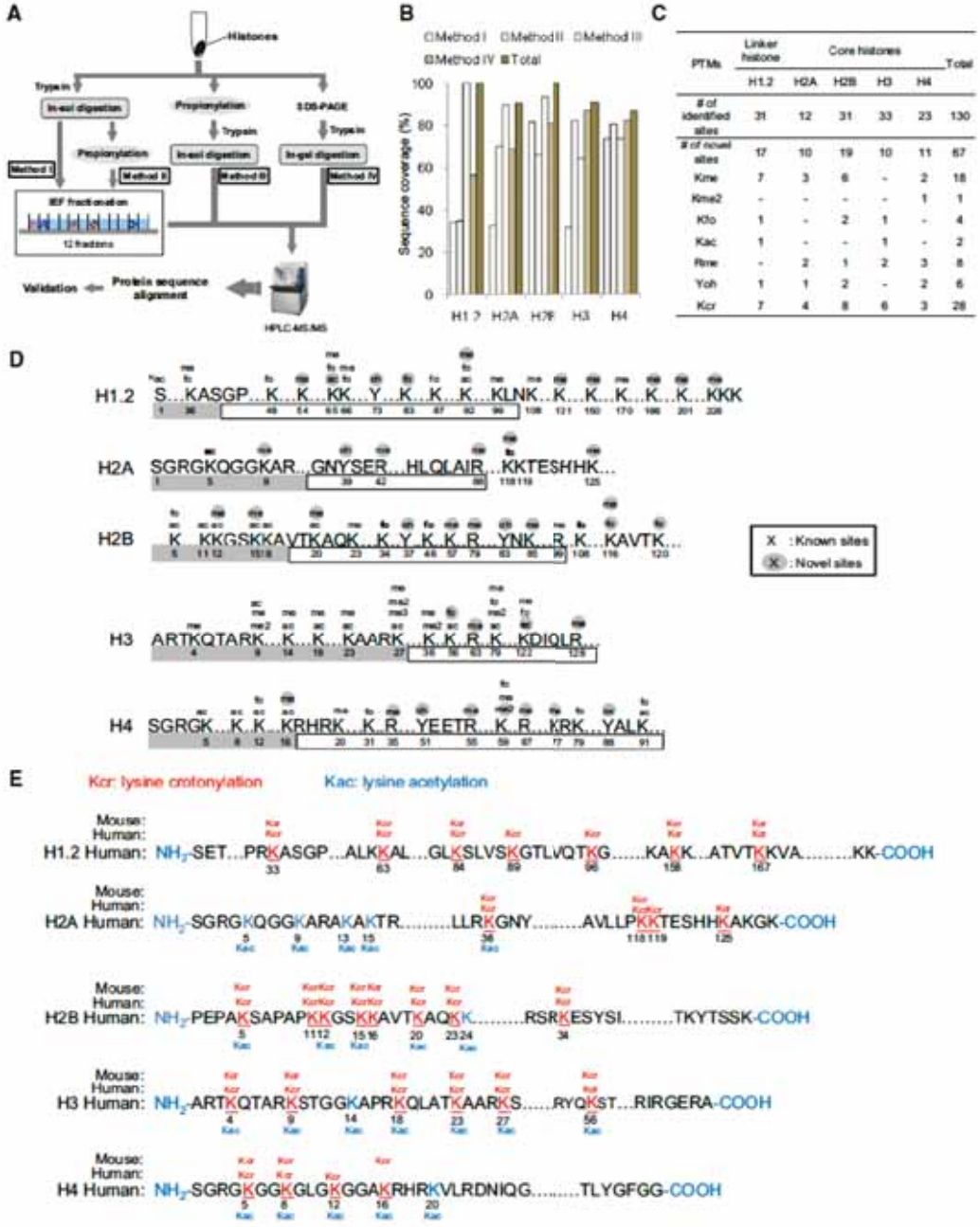


Figure 1. Experimental Strategy and Results for Identified Histone PTM Sites

(A) Schematic diagram of the experimental design for comprehensive mapping of PTM sites in linker and core histones from HeLa cells. Histone extracts were in-solution trypsicly digested without chemical propionylation (Method I), chemically propionylated after in-solution tryptic digestion (Method II), chemically propionylated before in-solution tryptic digestion (Method III), and in-gel digested after SDS-PAGE gel separation. Samples from Methods I and II were further subjected to IEF fractionation to generate 12 fractions.

(B) Peptide sequence coverage of linker and core histones in each of the four methods s shown.

(C) A table summarizing all the PTM sites identified by this study. Abbreviations: me, monomethylation; me2, dimethylation; me3, trimethylation; fo, formylation; ac, acetylation; oh, hydroxylation; and cr, crotonylation.

(D) A diagram showing sites of histone PTMs other than Kcr identified in this study. Amino acid residue number is indicated below its sequence. Gray and blank boxes indicate N-terminal and globular core domains, respectively.

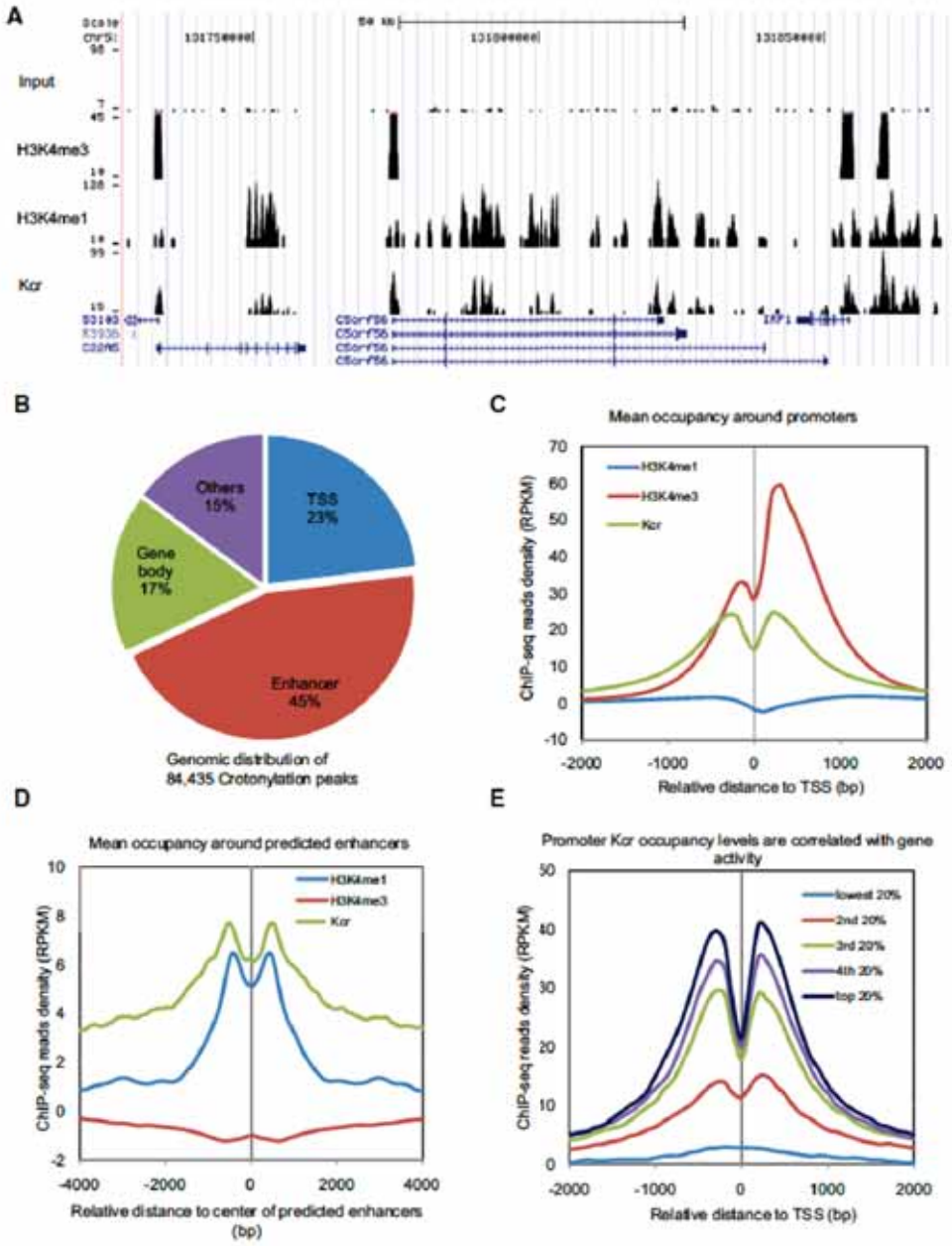


Figure 5. Enrichment of Histone Kcr on Active Chromatin

(A) ChIP-seq snapshots of input, H3K4me3, H3K4me1 and Kcr in IMR90 cells.

(B) Pie chart showing the genomic distribution of all histone Kcr peaks with annotated genomic regions. TSS is defined as regions +/– 2.5kb around known transcription starting sites in RefSeq database. Enhancers are promoter distal regions associated with H3K4me1 as predictive mark.

(C) Curves showing the mean reads density of indicated histone modification around all known TSS. Reads densities are calculated within a 100bp sliding window and normalized by subtracting reads density in the control input ChIP-seq data. RPKM is calculated as the number of reads which map per kilobase of genomic region per million mapped reads.

(D) Average normalized read densities of H3K4me3, H3K4me1 and histone Kcr around predicted enhancers are plotted.

(E) All RefSeq genes are divided into 5 groups based on their expression level calculated from mRNA-seq data, and the average normalized read densities of Kcr around each group of TSS are plotted. See also Figure S3.

LETTER

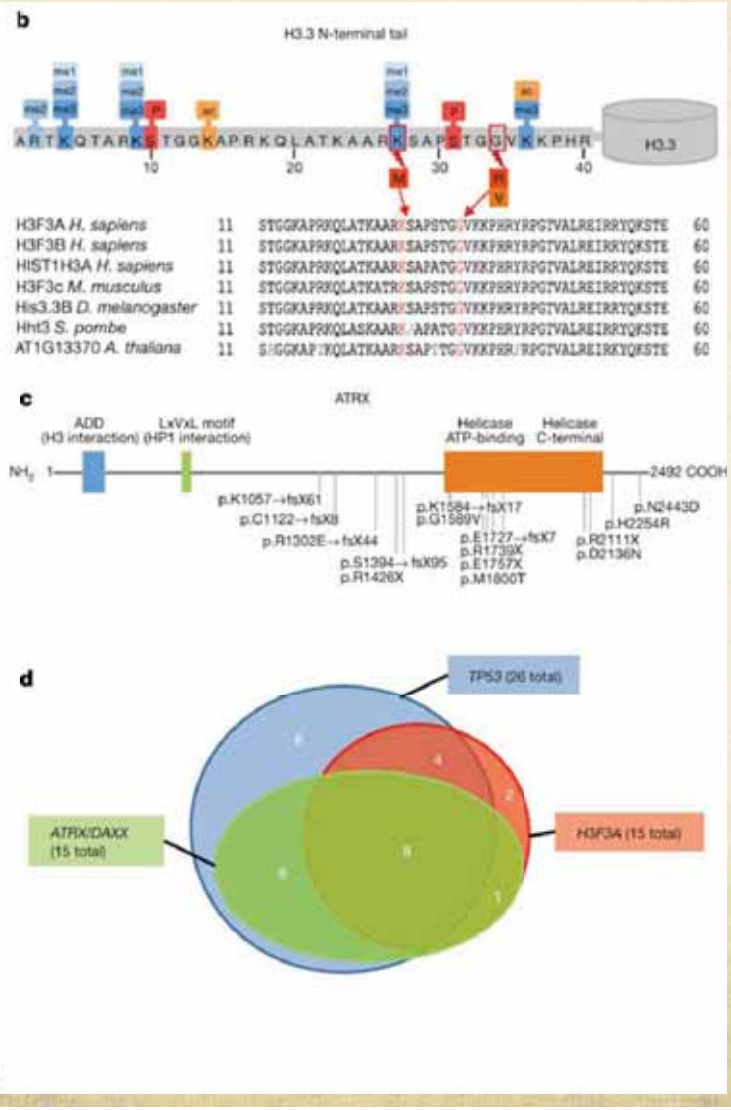
doi:10.1038/nature10833

Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma

a

Sample ID	H3F3A	ATRX/DAXX	TP53	IDH1	NF1	PDGFRA
PGBM1	K27M	C1122fs	P152fs;R306X	WT	WT	WT
PGBM2	K27M	WT	R213X*	WT	WT	K385I
PGBM3	K27M	WT	N131del	WT	WT	WT
PGBM4	K27M	K1057fs*	G262fs*	WT	WT	WT
PGBM5	K27M	WT	WT	WT	Y2264fs	WT
PGBM6	K27M	M1800T*	WT	WT	F1247fs;V2230del	WT
PGBM8	K27M	WT	R273C	WT	WT	WT
PGBM9	K27M	WT	R273P*	WT	WT	WT
PGBM10	K27M	WT	WT	WT	T990fs	WT
PGBM11	G34R	S1394fs*	Y163C*	WT	WT	WT
PGBM12	G34R	E1727fs	R342X;R175H	WT	WT	Y849D*
PGBM13	G34R	R1739X*	T256fs*	WT	WT	WT
PGBM14	G34R	E1757X*	R273C;R248Q	WT	WT	WT
PGBM15	G34R	H2254R*	S51delins	WT	WT	WT
PGBM16	G34V	R2111X*	R342X*	WT	WT	K385M
PGBM17	WT	G1589V*	Y220C	R132H	WT	WT
PGBM18	WT	R1426X*	R273C;R196X	R132H	C622X;L1489fs	WT
PGBM19	WT	K1584fs†	R267Q;T230I	WT	R440X	WT
PGBM20	WT	N2443D*	R248W*	WT	WT	WT
PGBM21	WT	R238X (DAXX)	R267W;P152L	WT	R1947X	WT
PGBM22	WT	R1302fs;K1584fs‡	R337C;R175H	WT	R2616X;R461X‡	WT
PGBM23	WT	WT	I254S	R132H	WT	WT
PGBM24	WT	WT	R196X*	WT	WT	WT
PGBM25	WT	WT	R342X*	WT	G1526fs*	WT
PGBM26	WT	WT	R175H*	WT	Y2264fs*	WT
PGBM27	WT	WT	I251L*	WT	WT	WT
PGBM28	WT	WT	R273H	WT	T676fs	WT
PGBM29	WT	WT	V10G	R132H	WT	WT
PGBM30	WT	WT	G245S*	WT	WT	WT
PGBM31	WT	WT	WT	WT	WT	WT
PGBM32	WT	WT	WT	WT	T1627S	WT
PGBM33	WT	WT	WT	WT	Splicing	WT
PGBM34	WT	WT	WT	WT	WT	D842_I853delinsV
PGBM35	WT	WT	WT	WT	WT	WT
...	...	...	...	...	...	...
PGBM49	WT	WT	WT	WT	WT	WT

*Homozygous mutations.
†Sample PGBM19 additionally has a DAXX mutation C629Sfs, whereas PGBM21 has no ATRX mutation but has the DAXX mutation shown.
‡Sample PGBM22 has a third ATRX mutation, p.D2136N, and a third NF1 mutation, p.A887T.



DNA Methylation

- Modification of C at CpG dinucleotide
- Methylation of CpG island is often associated with suppression of expression
- Chromosome imprinting
  - Only one allele from a certain parental line is expressed
  - Inactivation of X-chromosome

Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals

Rudolf Jaenisch[1] & Adrian Bird[2]

doi:10.1038/ng1089

Table 1 • Phenotypes of mouse mutants of epigenetic regulatory factors

Protein	Function	Mutant phenotype	References
methyltransferases			
Dnmt1	maintenance of methylation	embryonic lethal, loss of imprinting and X-linked gene expression, ES cells viable	20,21
Dnmt1o	oocyte-specific isoform	loss of maternal imprints	23
Dnmt2	non-CpG methylation in *Drosophila*	no phenotype	28
Dnmt3a, Dnmt3b	*de novo* methyltransferases, establishment of methylation	embryonic lethal, ICF syndrome	24,134
Dnmt3L	no catalytic activity, colocalizes with Dnmt3a and Dnmt3b	abnormal maternal imprinting	26,27
methyl binding proteins			
MeCP2	methyl binding proteins, recruit HDACs	RTT	80,81,84
MBD1	methyl binding proteins, recruit HDACs		
MBD2	methyl binding proteins, recruit HDACs	behavior abnormalities	76
MBD3	methyl binding proteins, recruit HDACs	lethal	76
MBD4	repair enzyme	increased mutation frequency	180
histone-modifying proteins			
HDAC1	histone deacetylase	embryonic lethal	181
Suvar39	Lys9 methylation in histone H3	embryonic lethal, chromosomal instability, increased tumor risk	182

Principles and challenges of genome-wide DNA methylation analysis

Peter W. Laird

Abstract | Methylation of cytosine bases in DNA provides a layer of epigenetic control in many eukaryotes that has important implications for normal biology and disease. Therefore, profiling DNA methylation across the genome is vital to understanding the influence of epigenetics. There has been a revolution in DNA methylation analysis technology over the past decade: analyses that previously were restricted to specific loci can now be performed on a genome-scale and entire methylomes can be characterized at single-base-pair resolution. However, there is such a diversity of DNA methylation profiling techniques that it can be challenging to select one. This Review discusses the different approaches and their relative merits and introduces considerations for data analysis.

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.

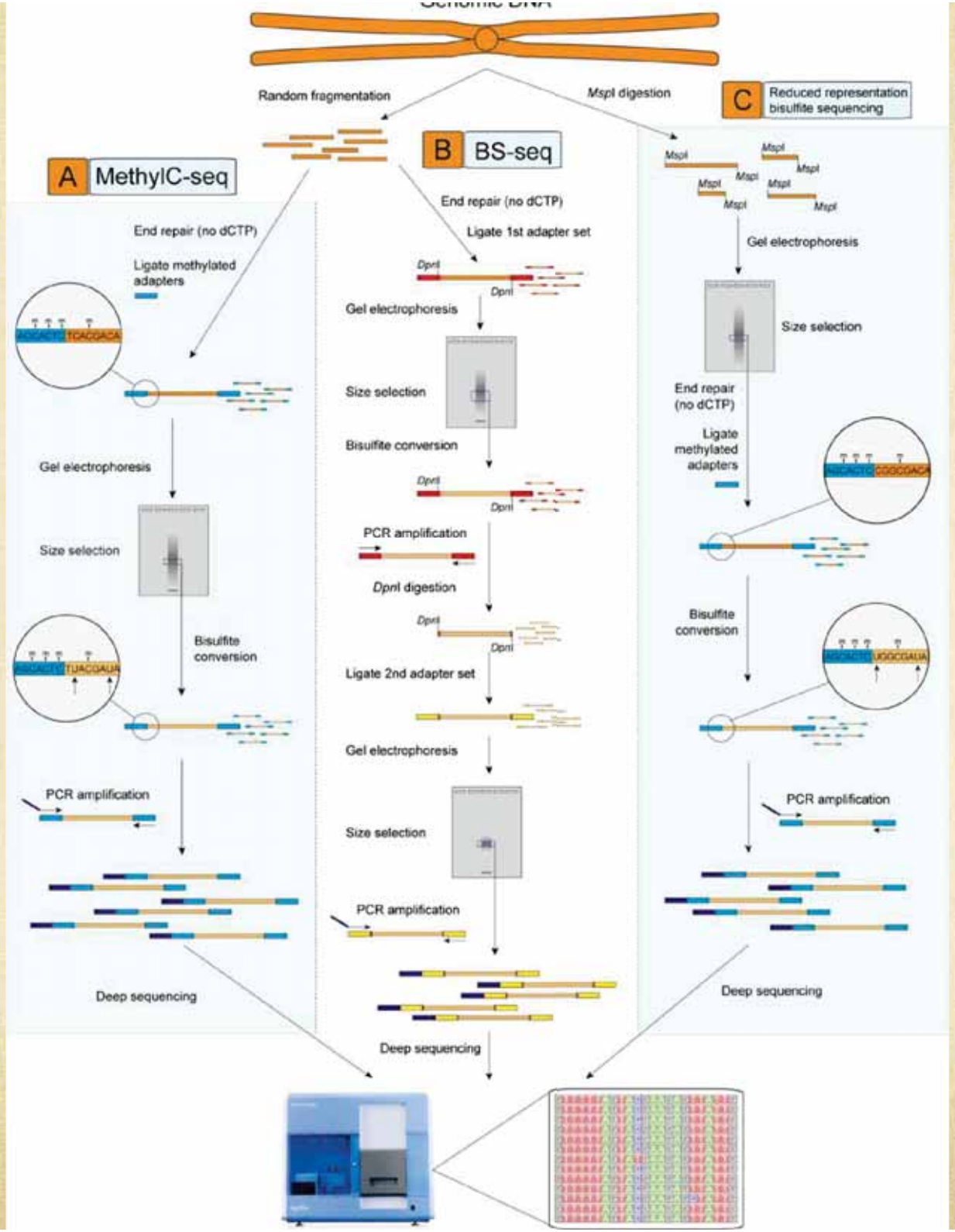

Genomic DNA
Random fragmentation
MspI digestion
C Reduced representation bisulfite sequencing
A MethylC-seq
B BS-seq
End repair (no dCTP)
Ligate methylated adapters
Gel electrophoresis
Size selection
Bisulfite conversion
PCR amplification
Deep sequencing
End repair (no dCTP)
Ligate 1st adapter set
DpnI
Gel electrophoresis
Size selection
Bisulfite conversion
PCR amplification
DpnI digestion
Ligate 2nd adapter set
Gel electrophoresis
Size selection
PCR amplification
Deep sequencing
MspI
Gel electrophoresis
Size selection
End repair (no dCTP)
Ligate methylated adapters
Bisulfite conversion
PCR amplification
Deep sequencing

Figure 1. Techniques for genome-wide sequencing of cytosine methylation sites. Three techniques used recently to generate bisulfite (BS) sequencing libraries compatible with next-generation sequencing are depicted. (*A*) MethylC-seq (Lister et al. 2008). Double-stranded universal adapter sequences in which all cytosines are methylated are ligated to fragmented genomic DNA. Sodium bisulfite treatment converts unmethylated cytosines to thymine, after which library yield enrichment by PCR with primers complementary to the universal adapter sequences produces the final library that can be sequenced. (*B*) BS-seq (Cokus et al. 2008). Ligation of a first set of double-stranded adaptors that contained methylated adenine bases within DpnI restriction sites close to the site of ligation with genomic DNA. After BS conversion, PCR is performed using primers complementary to the converted adapter sequences, yielding double-stranded DNA that is digested with DpnI to remove only the first adapter set. Sequencing adapters are subsequently ligated to the double-stranded BS-converted genomic DNA fragments, and PCR with primers complementary to the adapters performed to yield a sequencing library. (*C*) Reduced representation BS sequencing (RRBS) (Meissner et al. 2008). Genomic DNA is first digested by the methylation-insensitive MspI restriction enzyme, which cleaves the phosphodiester bond upstream of the CpG dinuclotide in its CCGG recognition element. Digested DNA is then separated by gel electrophoresis, and one or more specific size fractions are selected. The size-selected DNA is then end repaired, ligated to double-stranded methylated sequencing adapters (as described above for MethylC-seq), BS converted, and amplified by PCR with primers complementary to the adapter sequences.

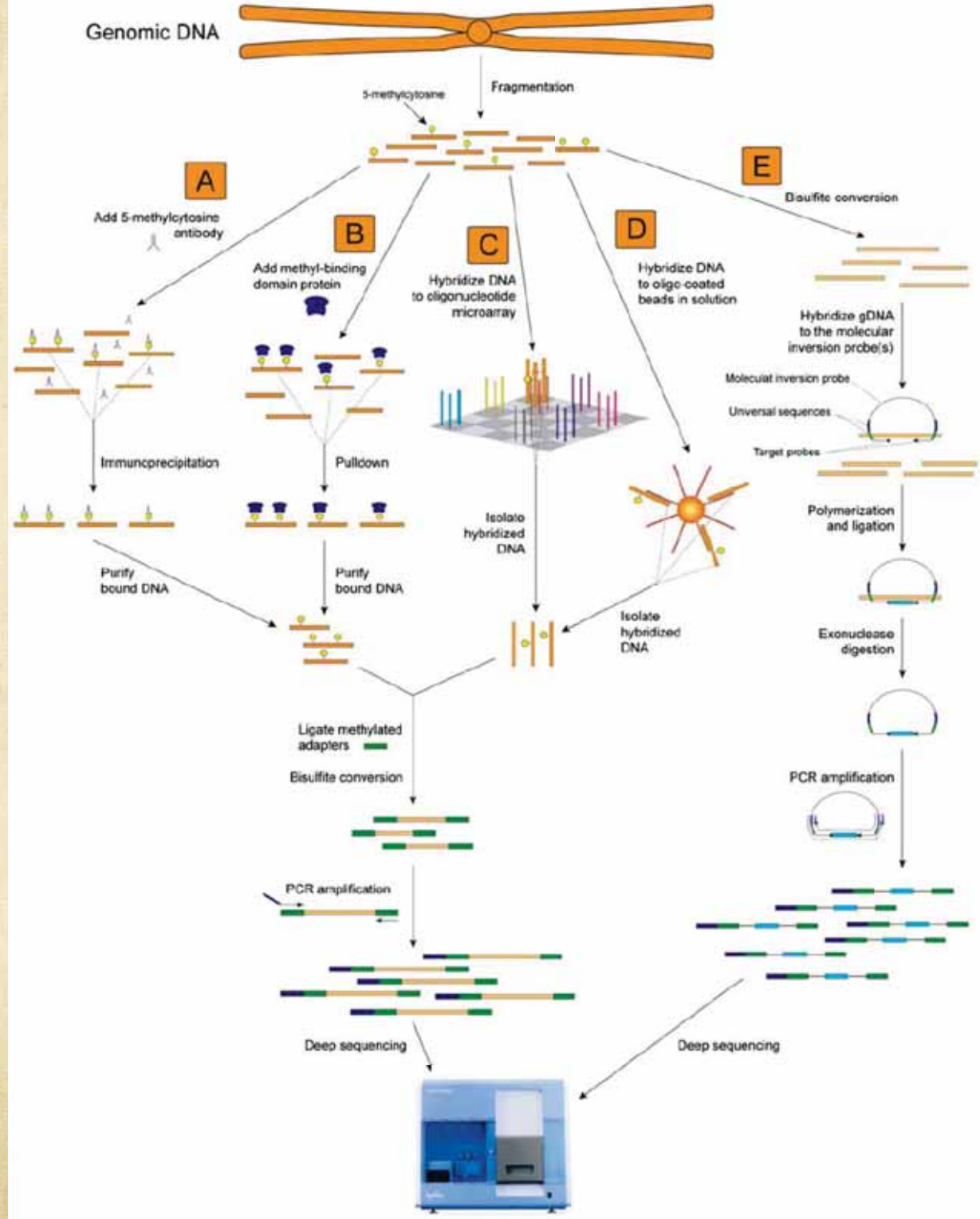

Genomic DNA
5-methylcytosine
Fragmentation
A
Add 5-methylcytosine antibody
B
Add methyl-binding domain protein
C
Hybridize DNA to oligonucleotide microarray
D
Hybridize DNA to oligo-coated beads in solution
E
Bisulfite conversion
Immunoprecipitation
Pulldown
Isolate hybridized DNA
Isolate hybridized DNA
Purify bound DNA
Purify bound DNA
Hybridize gDNA to the molecular inversion probe(s)
Molecular inversion probe
Universal sequences
Target probes
Polymerization and ligation
Exonuclease digestion
PCR amplification
Ligate methylated adapters
Bisulfite conversion
PCR amplification
Deep sequencing
Deep sequencing

Figure 2. Techniques for enrichment of methylated or target regions prior to BS sequencing. Five approaches that may be used to reduce the complexity of a sample before BS conversion and next-generation sequencing are depicted, targeting methylated regions or select target sequences. (*A*) MeDIP. Methylated fragments of genomic DNA are immunoprecipitated with an anti-5-methylcytosine antibody. Purified, immunoprecipitated DNA is ligated to double-stranded universal adapter sequences in which all cytosines are methylated. Sodium bisulfite treatment converts unmethylated cytosines to thymine, after which library yield enrichment by PCR with primers complementary to the universal adapter sequences produces the final library that can be sequenced. (*B*) MBD. Methylated fragments of genomic DNA are isolated from a complex mix of fragmented genomic DNA with a methyl binding domain protein, after which adapter ligation, BS conversion, and PCR enrichment are performed as in *A*. (*C*) Microarray capture. Target sequences within a complex mix of fragmented genomic DNA are captured by hybridization to specific oligonucleotides on the surface of a microarray. Following isolation of the hybridized genomic DNA, adapter ligation, BS conversion, and PCR enrichment are performed as in *A*. (*D*) Capture in solution. Specific target regions within a mix of fragmented genomic DNA are captured by hybridization to specific oligonucleotides attached to beads in solution. Following isolation of the hybridized genomic DNA, adapter ligation, BS conversion, and PCR enrichment are performed as in *A*. (*E*) Molecular inversion probe capture. Fragmented genomic DNA is BS converted, after which molecular inversion probes are added that are designed to hybridize to specific target sequences after conversion. Polymerization primed by the 3′ end of the molecular inversion probe followed by ligation generates a circular molecule that contains the target sequence and is not digested by subsequent exonuclease treatment. PCR using primers that hybridize to the ends of the molecular inversion probes allows amplification of the target region, to which double-stranded universal adapter sequences are ligated to produce a library that is sufficient for next-generation sequencing.

Taking the measure of the methylome

Stephan Beck

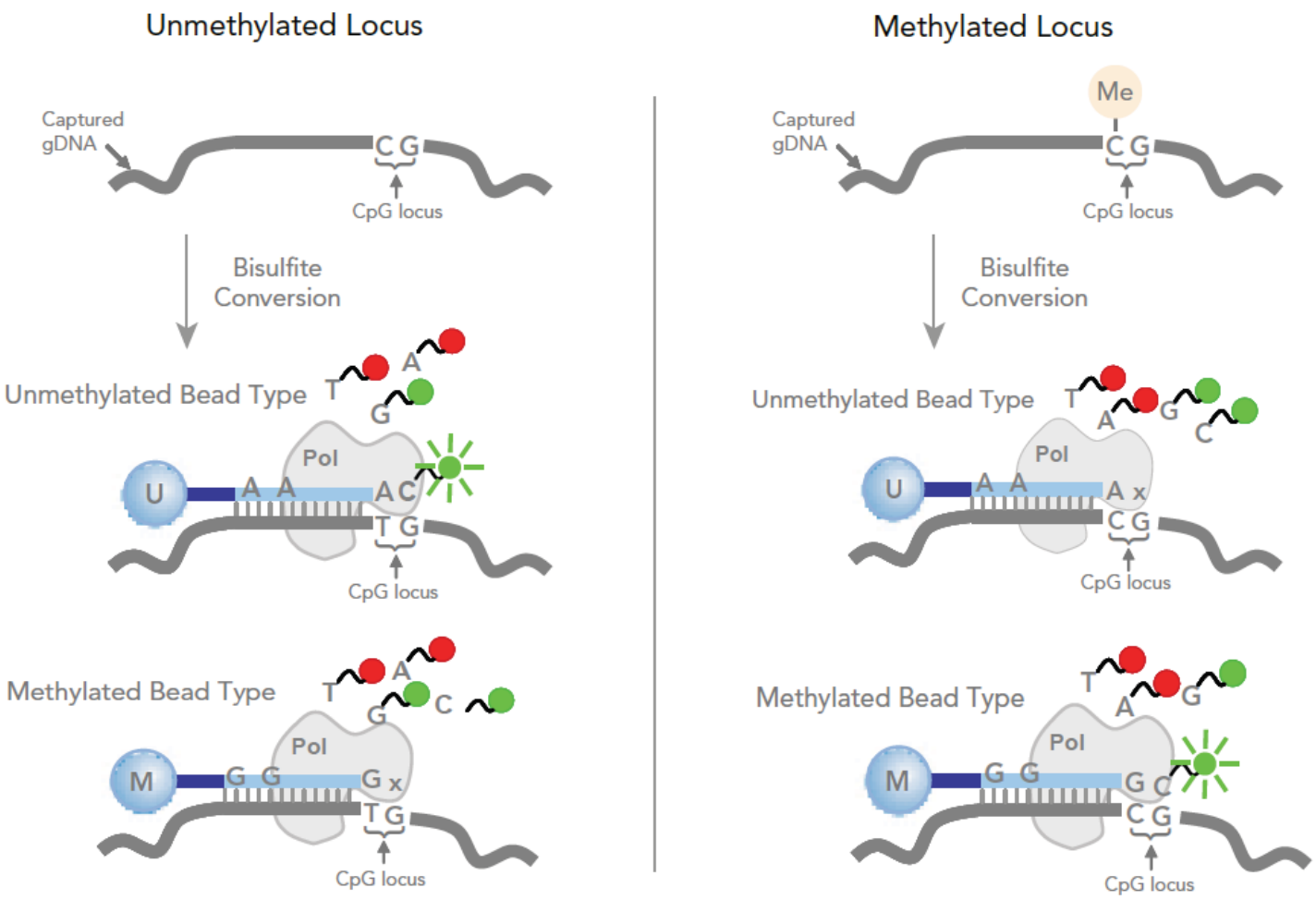
Table 1 Key metrics of the technology comparison

	MethylC-seq	MeDIP-seq	MethylCap-seq	MBD-seq	RRBS	Infinium-27K[a]
Genomic DNA	5 μg	0.3–5 μg	1 μg	3 μg	0.03–0.05 μg	0.5–1 μg
Readout	Sequence	Sequence	Sequence	Sequence	Sequence	Array
Assay	Bisulfite conversion	Capture with monoclonal antibody	Capture with MBD of MeCP2	Capture with MBD of MBD2	Bisulfite conversion	Bisulfite conversion
Resolution	1 bp	100–1,000 bp	100–1,000 bp	100–1,000 bp	1 bp	1 bp
Theoretical coverage	Whole-genome (~100%)	Whole-genome (~100%)	Whole-genome (~100%)	Whole-genome (~100%)	Genome-wide (~10%)	Genome-wide (~0.1%)
Actual coverage (1 read threshold)	~95%[1]	~67%[1]	~67%[2]	~61%[1]	~12%[1]	(~0.1%)
Actual coverage (5 reads threshold)	~87%[1]	~23%[1]	~28%[2]	~28%[1]	~10%[1]	(~0.1%)
Actual coverage (10 reads threshold)	~76%[1]	~9%[1]	~14%[2]	~20%[1]	~9%[1]	(~0.1%)
Cost	~$100 K[b]	~$2 K	~$3 K	~$2 K	~$2 K	~$0.2 K
Concordance (6-, 5-way)	NA	NA	NA	NA	NA	NA
Concordance (4-way)[1]	~99%	~99%		~99%	~99%	
Concordance (3-way)[1]	~100%	~100%		~100%		
Concordance (2-way)[1]	~96%					~96%
Concordance (2-way)[1]		~96%				~96%
Concordance (2-way)[2]		~84%				~84%
Concordance (2-way)[2]			~88%			~88%
Concordance (2-way)[1]				~91%		~91%
Concordance (2-way)[1]					~97%	~97%
Concordance (2-way)[2]					~92%	~92%
Conclusion	Gold standard but issue with hmC	Good all-rounder	Good all-rounder	Good all-rounder	Good for CpG islands	Good for promoters

Where appropriate, numbers are rounded or shown as a range. As sequencing costs are falling rapidly, the estimates shown are approximate and based on the assumption of ~$1K per lane on an Illumina Genome Analyser (see refs. 1 and 2 for details on the models used). To determine the maximum achievable genome coverage, MethylC and RRBS were subjected to saturation sequencing, entailing 2 lanes of sequencing for RRBS.

[a]For the MethylC data reported by Lister *et al.*[7] in 2009, the exact number of lanes could not be determined and the estimated 100 lanes (resulting in estimated costs of $100K), are likely to represent the upper limit. The current costs for a MethylC methylome is closer to $20K. MeDIP and MBD were subjected to 2 lanes and MethylCap to 3 lanes of sequencing. [b]A 450K upgrade of the current 27K Infinium array has been announced for later this year.

Figure 4: Infinium Assay for Methylation



The Infinium Assay for Methylation detects methylation status at individual CpG loci by typing bisulfite-converted DNA. Unmethylated C is converted to T (left), whereas methylation protects C from conversion (right). A pair of bead-bound probes are used to detect the presence of T or C by hybridization followed by single-base extention with a labeled nucleotide.

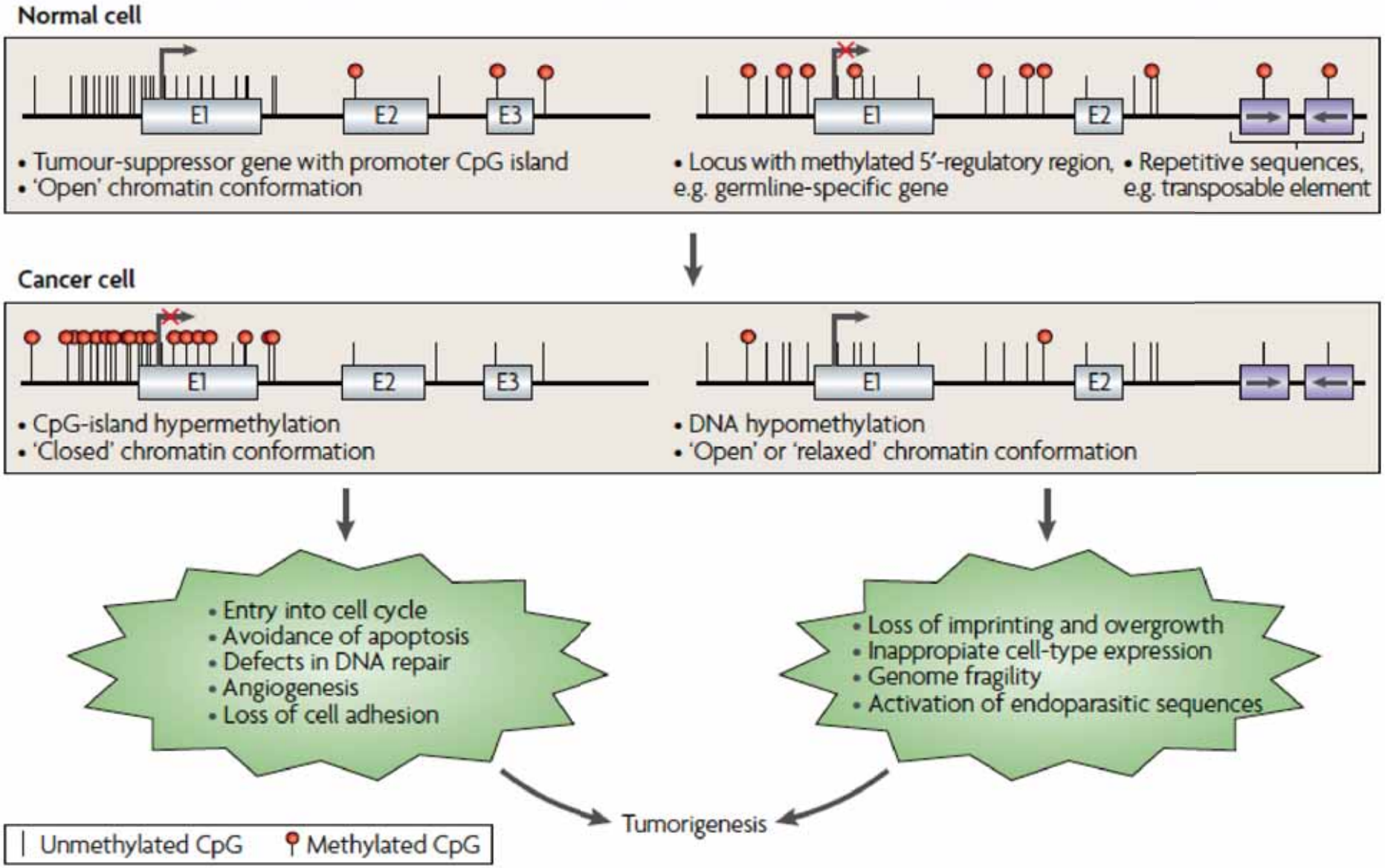


Figure 1 | **Altered DNA-methylation patterns in tumorigenesis.** The hypermethylation of CpG islands of tumour-suppressor genes is a common alteration in cancer cells, and leads to the transcriptional inactivation of these genes and the loss of their normal cellular functions. This contributes to many of the hallmarks of cancer cells. At the same time, the genome of the cancer cell undergoes global hypomethylation at repetitive sequences, and tissue-specific and imprinted genes can also show loss of DNA methylation. In some cases, this hypomethylation is known to contribute to cancer cell phenotypes, causing changes such as loss of imprinting, and might also contribute to the genomic instability that characterizes tumours. E, exon.

Cancer Cell
Article

Identification of a CpG Island Methylator Phenotype that Defines a Distinct Subgroup of Glioma

Houtan Noushmehr,[1,13] Daniel J. Weisenberger,[1,13] Kristin Diefes,[2,13] Heidi S. Phillips,[3] Kanan Pujara,[3] Benjamin P. Berman,[1] Fei Pan,[1] Christopher E. Pelloski,[4] Erik P. Sulman,[4] Krishna P. Bhat,[2] Roel G.W. Verhaak,[5,6] Katherine A. Hoadley,[7,8] D. Neil Hayes,[7,8] Charles M. Perou,[7,8] Heather K. Schmidt,[9] Li Ding,[9] Richard K. Wilson,[9] David Van Den Berg,[1] Hui Shen,[1] Henrik Bengtsson,[10] Pierre Neuvial,[10] Leslie M. Cope,[11] Jonathan Buckley,[1,12] James G. Herman,[11] Stephen B. Baylin,[11] Peter W. Laird,[1,14,*] Kenneth Aldape,[2,14] and The Cancer Genome Atlas Research Network

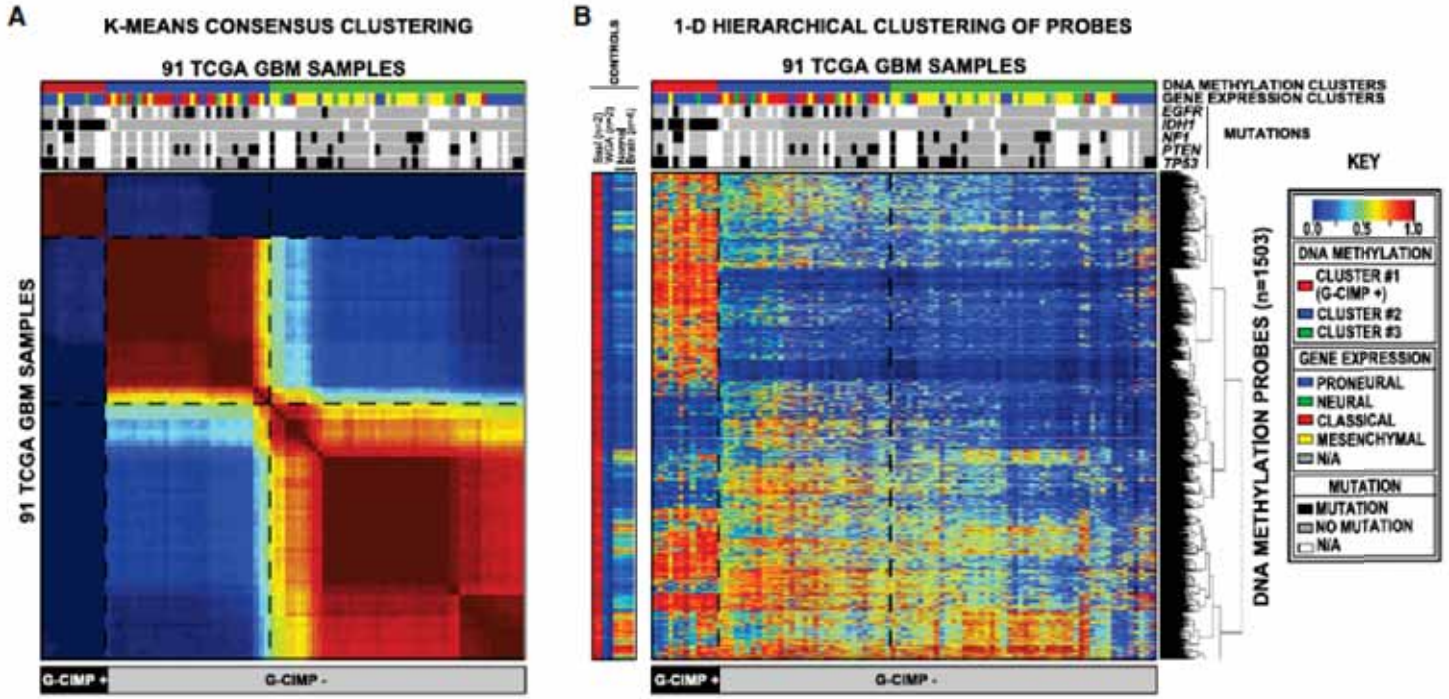


Figure 1. Clustering of TCGA GBM Tumors and Control Samples Identifies a CpG Island Methylator Phenotype

Unsupervised consensus clustering was performed with the 1503 Infinium DNA methylation probes whose DNA methylation beta values varied the most across the 91 TCGA GBM samples. DNA methylation clusters are distinguished with a color code at the top of the panel: red, consensus cluster 1 (n = 12 tumors); blue, consensus cluster 2 (n = 31 tumors); and green, consensus cluster 3 (n = 48 samples). Each sample within each DNA Methylation cluster are color labeled as described in the key for its gene expression cluster membership (proneural, neural, classical, and mesenchymal). The somatic mutation status of five genes (*EGFR*, *IDH1*, *NF1*, *PTEN*, and *TP53*) are indicated by the black squares, the gray squares indicate the absence of mutations in the sample, and the white squares indicate that the gene was not screened in the specific sample. G-CIMP-positive samples are labeled at the bottom of the matrix.

(A) Consensus matrix produced by k-means clustering ($K = 3$). The samples are listed in the same order on the x and y axes. Consensus index values range from 0 to 1, with 0 being highly dissimilar and 1 being highly similar.

(B) One-dimensional hierarchical clustering of the same 1503 most variant probes, with retention of the same sample order as in (A). Each row represents a probe; each column represents a sample. The level of DNA methylation (beta value) for each probe, in each sample, is represented with a color scale as shown in the legend; white indicates missing data. M.SssI-treated DNA (n = 2), WGA-DNA (n = 2), and normal brain (n = 4) samples are included in the heatmap but did not contribute to the unsupervised clustering. The probes in the eight control samples are listed in the same order as the y axis of the GBM sample heatmap. See also Figure S1 and Table S1.