Epigenetics – Histone modifications

Biosciences 741: Genomics
Fall, 2013
Week 14

Table 5-1 Type of histone modifications in humans

Histone Type	Histone Modifications
H2A	H2AK5ac, H2AK9ac, H2AZ
H2B	H2BK120ac, H2BK12ac, H2BK20ac, H2BK5ac, H2BK5me1, UbH2B
H3	H3K14ac, H3K18ac, H3K23ac, H3K27ac, H3K27me1, H3K27me2, H3K27me3, H3K36ac, H3K36me1, H3K36me3, H3K4ac, H3K4me1, H3K4me2, H3K4me3, H3K79me1, H3K79me2, H3K79me3, H3K9ac, H3K9me1, H3K9me2, H3K9me3, H3R2me1, H3R2me2, H3ac
H4	H4K12ac, H4K16ac, H4K20me1, H4K20me3, H4K5ac, H4K8ac, H4K91ac, H4Kac, H4R3me2, H4ac

Epigenetics in Health and Disease by I. Kovalchuk and O. Kovalchuk (2012).

The inter-relationship of histone modifications and histone methylation in plants

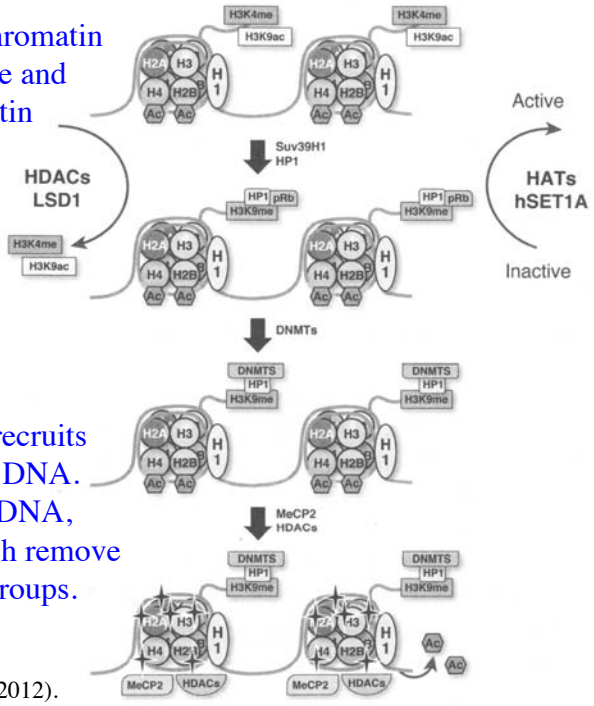
- In the *met1* mutant (a DNA methylase), the loss of CpG methylation leads to the loss of H3K9 methylation (showing that the former can cause the latter).
- On the other hand, the *KYP* mutant (Kryptonite; an H3K9 methylase) results in the loss of H3K9 methylation but does not alter the level of CpG methylation.
- The evidence in animals is less clear, and might work in both directions.

Chromatin marks define functional categories of genes

- Active (on) - H3K4me3, H3K9ac, phosphorylated RNA pol II.
- Permissive (stalled) - H3K4me3, H3K9ac, RNA pol II not phosphorylated.
- Bivalent (poised, inducible) - H3K4me3, H3K27me3.
- Repressive (off) - H3K9me3, H3K27me3.

Transcriptionally active chromatin is associated with H3K4me and H3K9ac. Inactive chromatin requires the removal of H3K4me and H3K9ac, followed by blocking this location with H3K9me.

HP1 binds H3K9me, and recruits DNMTs, which methylate DNA. MeCP2 binds methylated DNA, and recruits HDACs, which remove additional histone acetyl groups.



Epigenetics in Health and Disease
by I. Kovalchuk and O. Kovalchuk (2012).

Significance of histone modifications - me1 and ac usually stimulate transcription, but me2 and me3 effects are position-dependent.

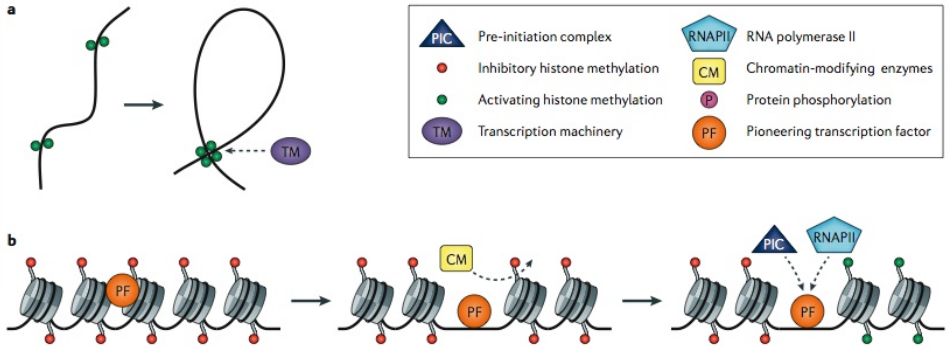
Table 5-4 Histone modifications and their effect on transcription in humans

Modification Type	Histone						
	H3K4	*H3K9*	*H3K14*	*H3K27*	*H3K79*	*H4K20*	*H2BK5*
me1	+	+		+	+	+	+
me2		-		-	+		
me3	+	-		-	+/-		-
ac		+	+				

+ represents activation of transcription; - represents repression of transcription. The data are collected from Barski et al. (2007).

Epigenetics in Health and Disease by I. Kovalchuk and O. Kovalchuk (2012).

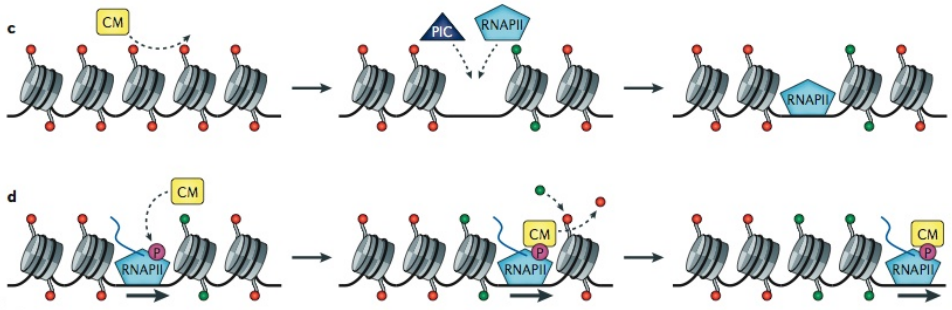
Top - histone methylation can influence transcription by bringing separate regions of chromatin together, forming loops.



Bottom - "pioneering factors" (PF, sequence-specific DNA binding proteins) can recruit DNA methylation and histone modification factors (CM), which make the chromatin more accessible for the Pre-initiation complex (PIC) and RNA polymerase II (RNAPII).

Greer & Shi (2012) Nat. Rev. Genet. 13, 343-357

Top - certain histone methylation patterns (such as H3K4 and H3K79 methylation) are characteristic of promoters and may be necessary for the binding of specific transcription factors (such as TFIID).



Bottom - the RNA polymerase II holoenzyme binds to H3K4me3. Moreover, the switch from initiation to extension is accomplished by phosphorylation of RNA pol II, which in turn switches its association with histone methylation/demethylation enzymes.

Greer & Shi (2012) Nat. Rev. Genet. 13, 343-357

Table 5-2 Histone acetyltransferases and deacetylases

	Histone H2A	Histone H2B	Histone H3	Histone H4
Acetyltransferases				
GCN5, PCAF			K9, K14, K18	
HAT1				K5, K12
CBP, P300	K5	K12, K15	K14, K18	K5, K8
TIP60/PLIP			K14	K5, K8, K12, K16
HBO1				K5, K8, K12
Deacetylases				*K5, K12*
SirT2-3				K16
RPD3/HDAC2/HDAC3			K9, K14, K16	K16

Table 5-2 Histone acetyltransferases and deacetylases

Specificity of histone methyl transferases in humans.

Table 1 | **Histone methyltransferases**

Histone and residue	*Homo sapiens*		
	me3	me2	me1
H3R2		CARM1(a); PRMT6(a)*; PRMT5(s); PRMT7(s)‡	CARM1; PRMT6*; PRMT5; PRMT7
H3K4	SETD1A; SETD1B; ASH1L; MLL; MLL2; MLL3; MLL4; SMYD3§; PRMD9	SETD1A; SETD1B; MLL; MLL2; MLL3; MLL4; SMYD3§	SETD1A; SETD1B; ASH1L§; MLL; MLL2; MLL3; MLL4; SETD7
H3R8		PRMT5(s)	PRMT5
H3K9	SUV39H1; SUV39H2; SETDB1; PRDM2§	SUV39H1; SUV39H2; SETDB1; G9a; EHMT1; PRDM2§	SETDB1; G9a; EHMT1; PRDM2§
H3R17		CARM1(a)	CARM1
H3R26		CARM1(a)	CARM1
H3K27	EZH2; EZH1	EZH2; EZH1	
H3K36	SETD2	NSD3; NSD2; NSD1; SMYD2§; SETD2	SETD2; NSD3; NSD2; NSD1;
H3K79	DOT1L	DOT1L	DOT1L
H4R3		PRMT1(a); PRMT6(a)*; PRMT5(s); PRMT7(s)‡	PRMT1; PRMT6*; PRMT5; PRMT7
H4K20	SUV420H1; SUV420H2	SUV420H1; SUV420H2	SETD8

Greer & Shi (2012)
Nat. Rev. Genet. 13, 343-357

Specificity of histone demethylase enzymes in humans.

Table 2 | **Histone demethylases**

Histone and residue	*Homo sapiens*		
	me3	me2	me1
H3R2			
H3K4	KDM2B; KDM5A; KDM5B; KDM5C; KDM5D; NO66	KDM1A; KDM1B; KDM5A; KDM5B; KDM5C; KDM5D; NO66	KDM1A; KDM1B; KDM5B; NO66
H3R8			
H3K9	KDM3B[§]; KDM4A; KDM4B; KDM4C; KDM4D	KDM3A; KDM3B[§]; KDM4A; KDM4B; KDM4C; KDM4D; PHF8; KDM1A; JHDM1D	KDM3A; KDM3B[§]; PHF8; JHDM1D
H3R17			
H3R26			
H3K27	KDM6A; KDM6B;	KDM6A; KDM6B; JHDM1D	JHDM1D
H3K36	NO66; KDM4A; KDM4B; KDM4C	KDM2A; NO66; KDM2B; KDM4A; KDM4B; KDM4C	KDM2A; KDM2B
H3K79			
H4R3			
H4K20			PHF8

Greer & Shi (2012)
Nat. Rev. Genet. 13, 343-357

Cancer is associated with decreases in histone methylation

Table 3 | **Global changes in histone methylation in various types of cancers**

Cancer type	Methyl mark	Consequence
Prostate cancer	↓H3K4me2	Higher recurrence
	↓H4K2me2	Higher recurrence
Lung cancer	↓H3K4me2	Poorer survival
Kidney cancer	↓H3K4me2	Poorer survival
Breast cancer	↓H3K4me2	Poorer survival
	↓H3K27me3	Poorer survival
	↓H4R3me2	Worse clinical outcomes
	↓H4K20me3	Worse clinical outcomes
Pancreatic cancer	↓H3K4me2	Poorer survival
	↓H3K9me2	Poorer survival
	↓H3K27me3	Poorer survival
Gastric adenocarcinoma	↑H3K9me3	Poorer survival
Ovarian cancer	↓H3K27me3	Poorer survival
Lymphomas	↓H4K20me3	Associated with
Colon adenocarcinomas	↓H4K20me3	Associated with

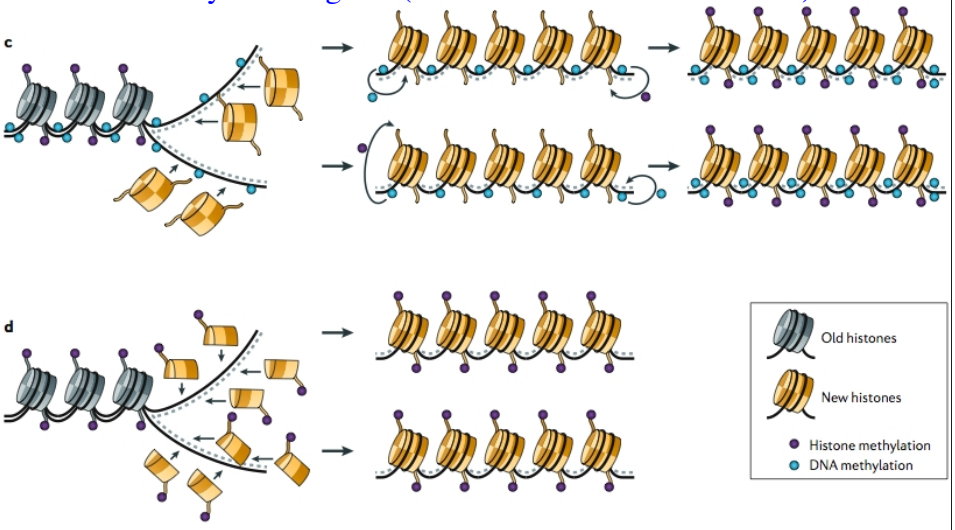
Greer & Shi (2012)
Nat. Rev. Genet. 13, 343-357

Chromosomal translocations that disrupt histone methyl transferases or histone demethylases are associated with cancer and cognitive defects.

Gene names (alternative names)	Methyl mark	Chromosomal location	Fused gene	Chromosomal translocation	Cancer type associated with translocation
KMT2A (MLL1, HRX, TRX1, ALL1)	Histone H3 lysine 4 (H3K4)	11q23	*AF4*, *ELL*, *AF9*, *ENL*, *AF6* and others	t(X;11)(q13;q23) t(1;11)(p32;q23) t(1;11)(q21;q23) t(2;11)(p21;q23) t(4;11)(q21;q23) t(6;11)(q27;q23) t(9;11)(p22;q23) t(10;11)(p11–13;q23) t(11,11)(q23;q25) t(11;14)(q23;q32) t(11;17)(q23;q12) t(11;17)(q23;q21) t(11;17)(q23;q25) t(11;19)(q23;p13)	Acute myeloid leukaemia, acute lymphoblastc leukaemia, mixed lineage leukaemia
KMT2D (MLL4, ALR)	H3K4	19q13.1	*HBXIP*	t(19;17)(q13;p11)	Hepatocellular carcinoma (HCC), hepatitis B virus related HCCs
KMT3B (NSD1, STO, SOTOS)	H3K36	5q35	*NUP98*	t(5;11)(q35;p15.5) t(5;2)(q35;p23)	Acute myeloid leukaemia, Sotos syndrome
NSD2 (WHS, TRX5, MMSET)	H3K36 dimethylation (H3K36me2)	4p16.3	*IGH*	t(4;14)(p16;q32)	Multiple myeloma tumours, lung cancers, Wolf–Hirschhorn syndrome
NSD3 (WHSC1L1)	H3K36me2	8p11.2	*NUP98*	t(8;11)(p11;p15) t(8;16)(p11;p13)	Acute myeloid leukaemia, myelodysplastic syndrome
KDM4C (JMJD2C, GASC1, JHDM3C)	H3K9me2, H3K9me3, H3K36me2, H3K36me3	9p24.1	*IGH*	t(9;14)(p24.1;q32)	Mucosa-associated lymphoid tissue lymphoma, chronic myeloid leukaemia
KDM5A (JARID1A, RBP2, RBBP2)	H3K4me2, H3K4me3	12p11	*NUP98*	t(11;21;12)(p15;p13;p13)	Acute myeloid leukaemia
JMJD1C (TRIP8)	H3K9	10q21.3		46,XY,inversion(10)(q11.1;q21.3)	Autism
HSPBAP1 (PASS1)	Unknown	3q21.1	*DIRC3*	t(2;3)(q35;q21)	Familial renal cell cancer

Greer & Shi (2012) Nat. Rev. Genet. 13, 343-357

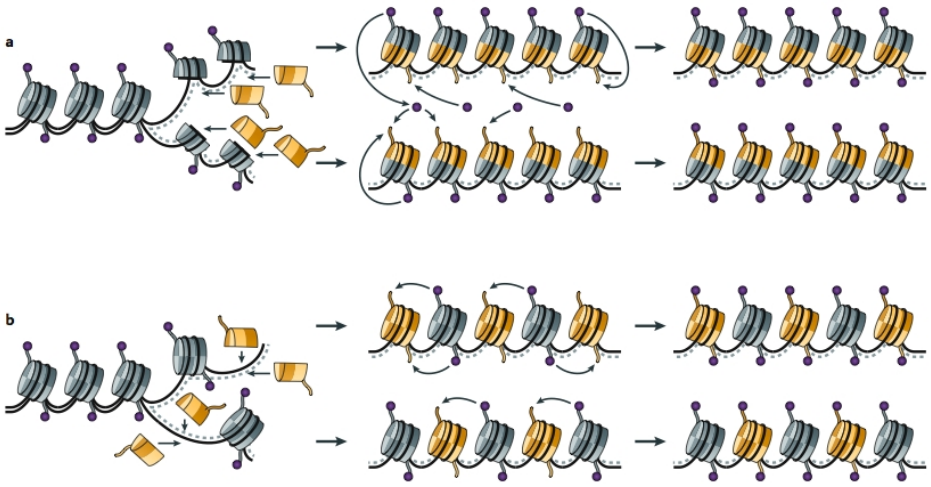
Top - another possibility is that chromatin is replicated by following DNA methylation signals (there is some evidence for this).



Bottom - a fourth possibility is that histones with pre-existing modifications (or variant histones) may be recruited to certain sites.

Greer & Shi (2012) Nat. Rev. Genet. 13, 343-357

Top - replication of chromatin might proceed through semiconservative assembly of half-nucleosomes.



Bottom - another possibility is fully conservative partitioning of intact nucleosomes between daughter chromatids.

Greer & Shi (2012) Nat. Rev. Genet. 13, 343-357

Histone variants

- Histone variants, such as H2A.Z, are distinct proteins that are encoded by separate genes. Their genes are evolutionarily related to the histone gene family.
- H2A.Z is mainly associated with the promoter region of active genes that lack DNA methylation. It is deposited by the ATP-dependent chromatin-remodeling complex SWR1. It destabilizes nucleosomes and helps to stimulate transcription.
- H2AX is another histone variant. It is associated with DNA strand breaks and helps to recruit DNA repair enzymes.
- H3.1 and H3.2 differ by a single amino acid and are deposited into chromatin during DNA replication. H3.3 is deposited independently of replication in quiescent, G1, and G2 cells, and is specifically associated with the promoters of active genes.

H2A.Z and H3.3 histone genes

- Most histones are synthesized and assembled into nucleosomes in parallel with DNA synthesis. However, the H2A.Z and H3.3 variants are synthesized from genes that are expressed throughout the cell cycle, and these histones therefore must be actively substituted for existing histone molecules within individual nucleosomes.
- H2A.Z typically occurs in nucleosomes as a heterodimer with H2A, which it likely destabilizes due to the divergent structure of its dimerization region.
- In *Arabidopsis*, H2A.Z antagonizes, and is antagonized by, DNA methylation.
- H3.3 differs from H3.1 and H3.2 at only four amino acids. Its role is unclear but may help to facilitate the insertion of H2A.Z.

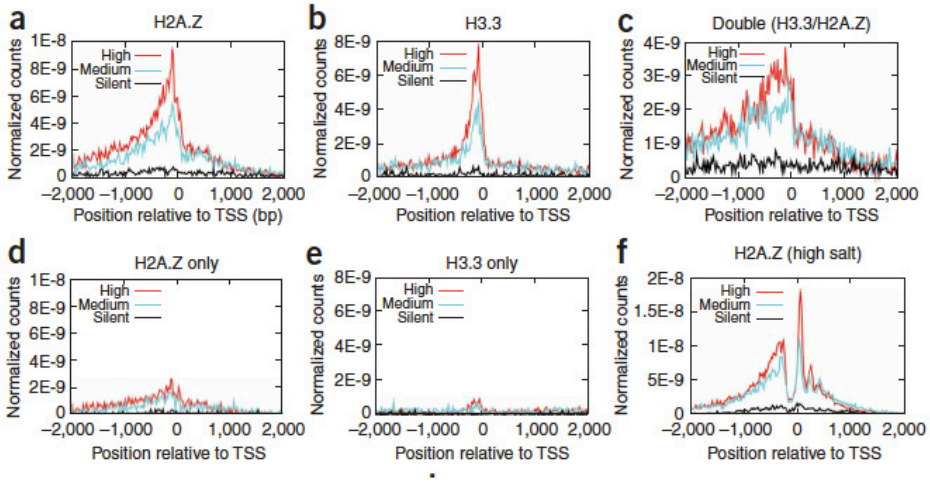
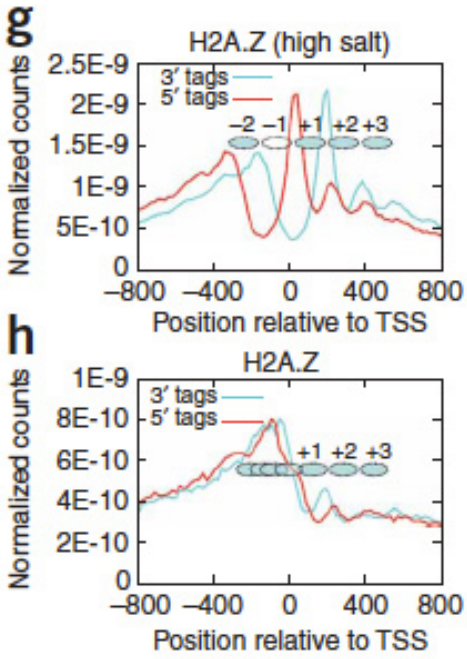


Figure 1 H3.3/H2A.Z NCPs mark 'nucleosome-free regions' of active promoters. Tags in non-overlapping 20-bp windows relative to the aligned transcription start sites (TSSs) were tallied in the gene set. The tag counts were normalized by the total numbers of bases. In a–f, island-filtered 5' tags were used and the profiles were further normalized by the total number of island-filtered tags in the library. TSSs for 1,000 highly active (red), intermediately active (cyan) and silent genes (black). (f) Profile of H2A.Z-containing NCPs isolated in high salt.

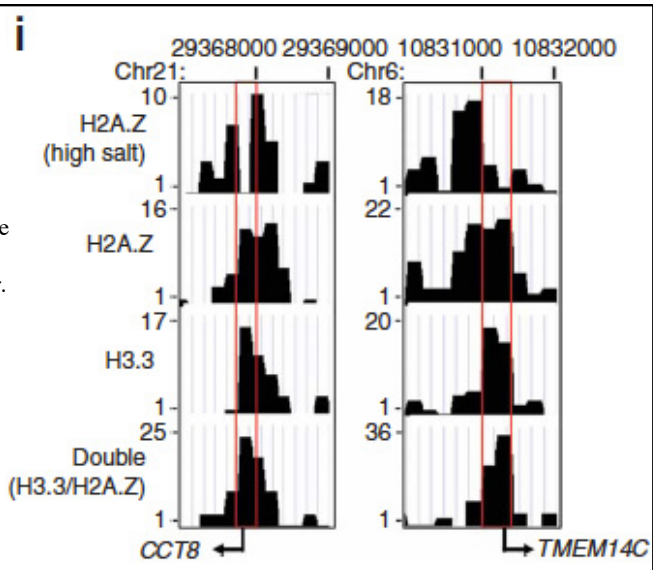
Jin et al. (2009) *Nat. Genet.* 41, 941-945.

(g,h) The H2A.Z nucleosome positioning near the TSSs at high (g) or low salt (h). The y axis shows the normalized counts of sequenced tags from the upper and lower strands of the DNA at each position, representing 5' and 3' boundaries of each nucleosome core particle (NCP). Open oval, depleted NCP; filled oval, phased NCP.

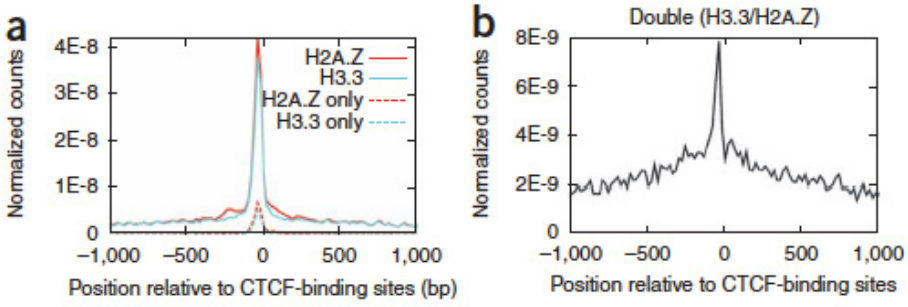


Jin et al. (2009) *Nat. Genet.* 41, 941-945.

(i) Two typical examples of histone variant patterns at high resolution at TSSs of two active genes, shown as custom tracks On the UCSC genome browser. Both active genes, CCT8 and TMEME14C, have high levels of H3.3/2A.Z NCPs at the TSS (lower three panels). The loss of these NCPs after exposure to high salt (top panels) is evident (red rectangles).

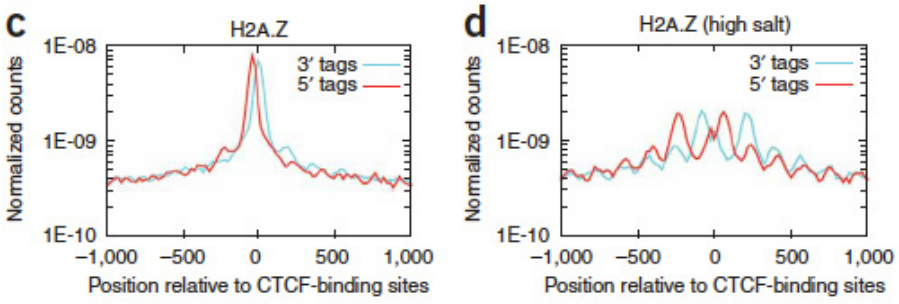


Jin et al. (2009) *Nat. Genet.* 41, 941-945.



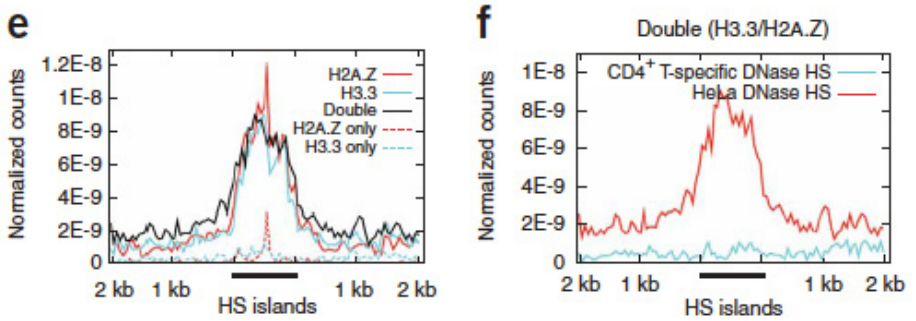
(a,b) Histone variants at intergenic CTCF-binding sites. The H2A.Z (as well as 'H2A.Z-only') and H3.3 (as well as 'H3.3-only') NCP levels were normalized by the total tag numbers of island-filtered tags in H2A.Z and H3.3 libraries, respectively, whereas the profile of H3.3/H2A.Z ('Double') NCP levels was Normalized by total island-filtered tags in the Double library.

Jin et al. (2009) *Nat. Genet.* 41, 941-945.



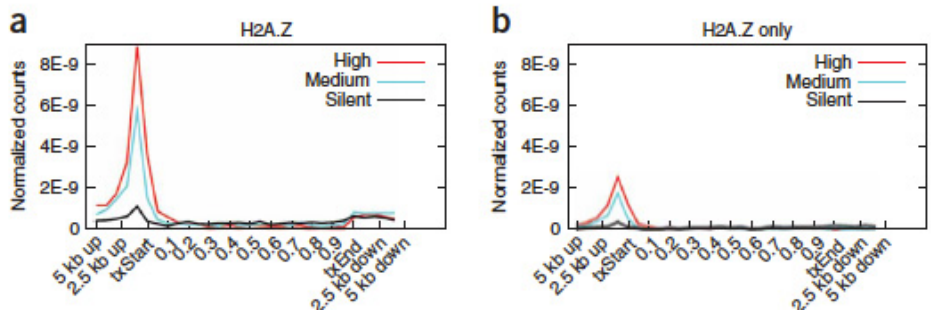
(c,d) Averaged H2A.Z nucleosome positioning near the CTCF-binding sites at low (c) or high (d) salt shown by the sequenced 5' (red) and 3' (cyan) tags, representing the 5' And 3' boundary of each NCP. Method was similar to that described in Figure 1g,h, except the Y axis is plotted on a logarithmic scale.

Jin et al. (2009) *Nat. Genet.* 41, 941-945.



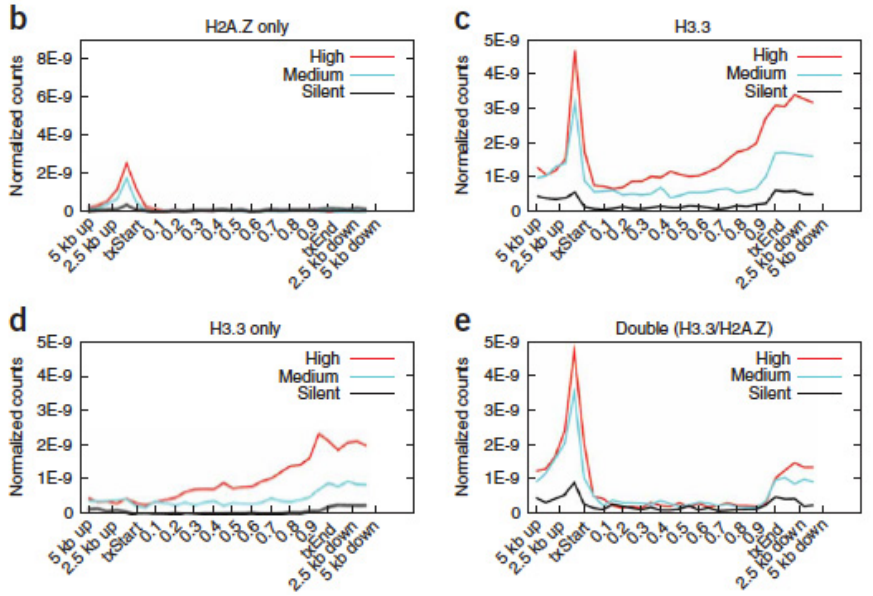
(e) Histone variants at ENCODE DNase I hypersensitive (HS) sites. All DNase I HS sites were aligned and normalized to the same length and were partitioned into 20 blocks. Island-filtered tags in each block were tallied and normalized by the total number of bases in each block. Outside the DNase I HS sites, island-filtered tags were tallied in 50-bp windows in the 2-kb upstream and downstream regions and normalized similarly. At the end, the profile was also normalized by the total number of island-filtered tags in each sample. (f) In HeLa cells, H3.3/H2A.Z NCPs are enriched only at HeLa DNase I hypersensitive sites (red) but not at sites (cyan) that are DNase I hypersensitive in CD4+ T cells but not in HeLa.

Jin et al. (2009) *Nat. Genet.* 41, 941-945.



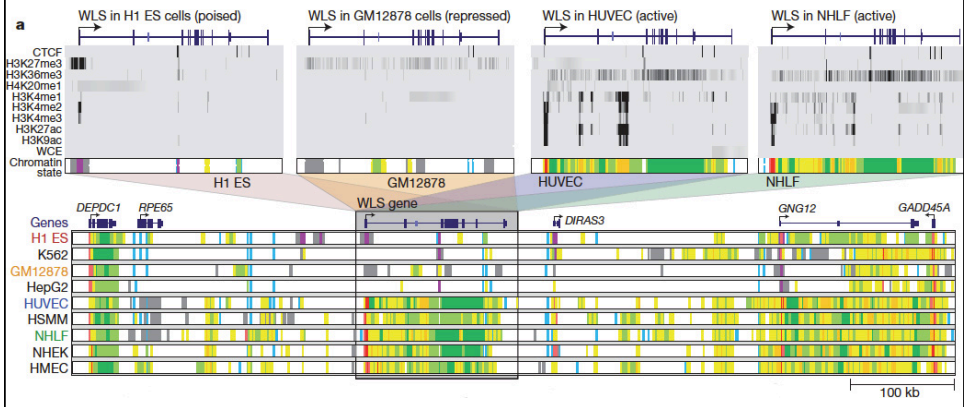
Profiles of histone variants indicated above each panel in and around gene bodies for 1,000 highly active (red), intermediately active (cyan) or silent (black) genes. For each gene, island-filtered tags were summed according to their shifted positions in 1-kb windows from 5 kb upstream of the TSS (txStart) to the txStart and from the TTS (txEnd) to 5 kb downstream of the TSS. Within the gene bodies, island-filtered tags were summed according to their shifted positions in windows equal to 5% of the gene length. All window tag counts were normalized by the total number of bases in the windows.

Jin et al. (2009) *Nat. Genet.* 41, 941-945.



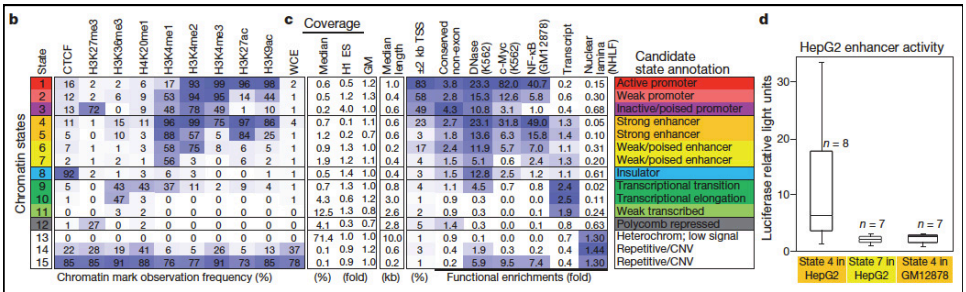
The profiles of H2A.Z and 'H2A.Z-only' (H3.3 and 'H3.3-only') were further normalized by the total tag numbers of island-filtered tags in the H2A.Z (H3.3) library, while the profile of H3.3/H2A.Z (Double) was further normalized by total island-filtered tags in the Double library.

Jin et al. (2009) *Nat. Genet.* 41, 941-945.



Profiles for nine chromatin marks (greyscale) are shown across the WLS gene in four cell types, and summarized in a single chromatin state annotation track for each (coloured according to b). WLS is poised in ESCs, repressed in GM12878 and transcribed in HUVEC and NHLF. Its TSS switches accordingly between poised (purple), repressed (grey) and active (red) promoter states; enhancer regions within the gene body become activated (orange, yellow); and its gene body changes from low signal (white) to transcribed (green).

Ernst et al. (2011) *Nature* 473, 43-49



b, Chromatin states learned jointly across cell types by a multivariate hidden Markov model. The table shows emission parameters learned de novo on the basis of genome-wide recurrent combinations of chromatin marks. Each entry denotes the frequency with which a given mark is found at genomic positions corresponding to the chromatin state. c, Genome coverage, functional enrichments and candidate annotations for each chromatin state. Blue shading indicates intensity, scaled by column. CNV, copy number variation; GM, GM12878. d, Box plots depicting enhancer activity for predicted regulatory elements. Sequences 250 bp long corresponding either to strong or weak/poised HepG2 enhancer elements or to GM12878-specific strong enhancer elements were inserted upstream of a luciferase gene and transfected into HepG2. Reporter activity was measured in relative light units. Robust activity is seen for strong enhancers in the matched cell type, but not for weak/poised enhancers or for strong enhancers specific to a different cell type. Boxes indicate 25th, 50th and 75th percentiles, and whiskers indicate 5th and 95th percentiles.

Ernst et al. (2011) *Nature* 473, 43-49

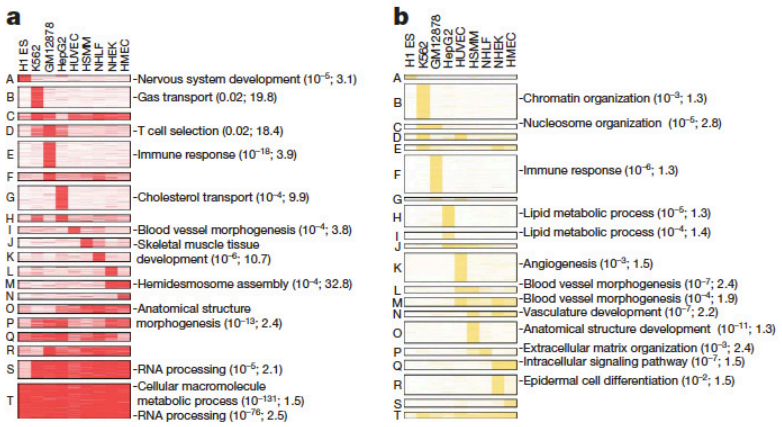


Figure 2 | Cell-type-specific promoter and enhancer states and associated functional enrichments. **a**, Clustering of genomic locations (rows) assigned to active promoter state 1 (red) across cell types (columns) reveals 20 common patterns of activity (A–T; Methods). For each cluster, enriched gene ontology terms are shown with hypergeometric *P* value and fold enrichment, based on the nearest TSS. For most clusters, several cell types show strong (dark red) or moderate (light red) activity. **b**, Analogous clustering and functional enrichments for strong enhancer state 4 (yellow). Enhancer states show greater cell type specificity, with most clusters active in only one cell type.

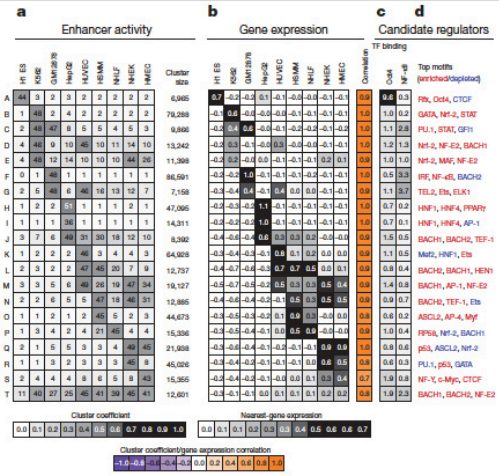
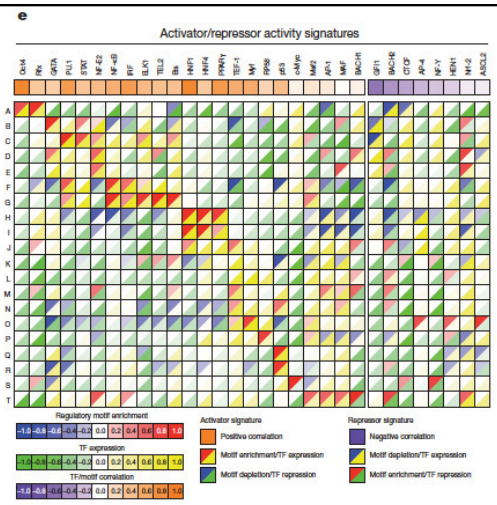


Figure 3 | Correlations in activity patterns link enhancers to gene targets and upstream regulators. **a**, Average enhancer activity across the cell types (columns) for each enhancer cluster (rows) defined in Fig. 2b (labelled A–T) and number of 200-bp windows in each cluster. **b**, Average messenger RNA expression of nearest gene across the cell types and correlation with enhancer activity profile from **a**. High correlations between enhancer activity and gene expression provide a means of linking enhancers to target genes. **c**, Enrichment for Oct4 binding in ESCs[24] and NF-κB binding in lymphoblastoid cells[14] for each cluster. TF, transcription factor. **d**, Strongly enriched (red) or depleted (blue) motifs for each cluster, from a catalogue of 323 consensus motifs. Rfx: Rfx family; Nrf-2: NFE2L2; STAT: STAT family; Ets: Ets family; Mef2: MEF2A and MYEF2;



Myf Myf family; NF-Y: NFYA, NFYB and NFYC. **e**, Predicted causal regulators for each cluster based on positive (activators) or negative (repressors) correlations between motif enrichment (top left triangles) and transcription factor expression (bottom right triangles). For example, the red–yellow combination indicates that Oct4 is a positive regulator of ESC-specific enhancers, as its motif-based predicted targets are enriched (red upper triangle) for enhancers active in ESCs (cluster A), and the Oct4 gene is expressed specifically in ESCs, resulting in a positive transcription factor expression correlation (yellow triangle). Overall correlations between motif enrichment and transcription factor expression across all clusters denote predicted activators (positive correlation, orange) and repressors (negative correlation, purple).

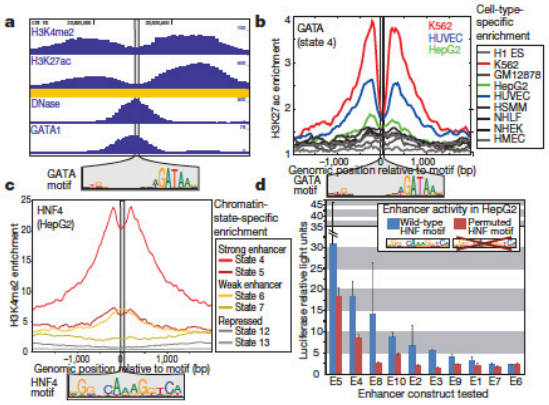


Figure 4 | Validation of regulatory predictions by nucleosome depletions and enhancer activity. **a**, Dips in chromatin intensity profiles in a K562-specific strong enhancer (orange) coincide with a predicted causal GATA motif instance (logo). The dips probably reflect nucleosome displacement associated with transcription factor binding, supported by DNase hypersensitivity[12] and GATA1 binding[25]. **b**, Superposition of H3K27ac signal across loci containing GATA motifs, centred on motif instances, shows dips in K562, as predicted. **c**, Superposition of H3K4me2 signal for HepG2 shows dips over HNF4 motifs in strong enhancer states, as predicted. **d**, HepG2-specific strong enhancers with predicted causal HNF motifs were tested in reporter assays. Constructs with permuted HNF motifs (red) led to significantly reduced luciferase activity in comparison with wild type (blue), with an average twofold reduction. Data shown are mean luciferase relative light units over three replicates and 95% confidence intervals.

Disease-related SNPs occur within enhancers (as defined by chromatin marks) that are activated in relevant cell types

a

Phenotype	Top cell type	Total no. SNPs from study	No. SNPs in enh. states 4 and 5	*P* value	FDR	H1 ES	K562	GM12878	HepG2	HUVEC	HSMM	NHLF	NHEK	HMEC
Erythrocyte phenotypes	K562	35	9	$<10^{-7}$	0.02	9	17	4	0	0	1	2	1	1
Blood lipids	HepG2	101	13	$<10^{-7}$	0.02	3	5	0	11	2	3	3	4	3
Rheumatoid arthritis	GM12878	29	7	2.0×10^{-7}	0.03	0	0	15	0	2	0	0	2	3
Primary biliary cirrhosis	GM12878	6	4	6.0×10^{-7}	0.03	0	11	41	0	0	0	0	8	8
Systemic lupus erythematosus	GM12878	18	6	9.0×10^{-7}	0.03	0	4	21	0	5	8	0	3	5
Lipoprotein cholesterol/triglycerides	HepG2	18	5	1.2×10^{-6}	0.03	17	8	0	24	3	6	4	3	3
Haematological traits	K562	39	7	1.7×10^{-6}	0.03	0	12	10	2	1	0	0	1	0
Haematological parameters	K562	28	6	2.2×10^{-6}	0.03	0	15	7	0	5	7	7	3	2
Colorectal cancer	HepG2	4	3	3.8×10^{-6}	0.03	0	0	0	66	0	12	0	12	12
Blood pressure	K562	9	4	5.0×10^{-6}	0.04	0	30	14	0	10	6	7	5	11

Discussion Topics

- Discuss the roles of H3K4, H3K9, and H3K27 modifications in the regulation of transcription (both positive and negative regulation).
- How (and why) does the expression of the standard histone genes (H1, H2A, H2B, H3, H4), as compared to H2 and H3 variants such as H2A.Z and H3.3, correlate with the cell cycle? With gene activation?
- Discuss the difference in chromatin marks between specific portions of genes such as promoters, enhancers, insulators, transcription start sites, and transcription termination sites.
- How did Ernst et al. use chromatin structure to identify enhancers vs. promoters? How did they link particular enhancers to particular genes? To particular transcription factors? How did they test these hypotheses? Identify relevant SNPs?