

Genome-Wide Analysis of the Chromatin Composition of Histone H2A and H3 Variants in Mouse Embryonic Stem Cells

Masashi Yukawa^{1,9}, Tomohiko Akiyama^{1,9}, Vedran Franke^{2,9}, Nathan Mise³, Takayuki Isagawa⁴, Yutaka Suzuki⁵, Masataka G. Suzuki¹, Kristian Vlahovicek², Kuniya Abe³, Hiroyuki Aburatani⁴, Fugaku Aoki^{1*}

1 Department of Integrated Biosciences, Graduate school of Frontier Sciences, The University of Tokyo, Kashiwa, Japan, **2** Bioinformatics Group, Division of Biology, Faculty of Science, Zagreb University, Zagreb, Croatia, **3** Mammalian Cellular Dynamics Team, Bioresource Center, RIKEN Tsukuba Institute, Tsukuba, Japan, **4** Genome Science Division, Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan, **5** Department of Medical Genome Science, Graduate school of Frontier Sciences, The University of Tokyo, Kashiwa, Japan

Abstract

Genome-wide distribution of the majority of H2A and H3 variants (H2A, H2AX, H2AZ, macroH2A, H3.1, H3.2 and H3.3) was simultaneously investigated in mouse embryonic stem cells by chromatin immunoprecipitation sequencing. Around the transcription start site, histone variant distribution differed between genes possessing promoters of high and low CpG density, regardless of their expression levels. In the intergenic regions, regulatory elements were enriched in H2A.Z and H3.3, whereas repeat elements were abundant in H2A and macroH2A, and H3.1, respectively. Analysis of H2A and H3 variant combinations composing nucleosomes revealed that the H2A.Z and H3.3 combinations were present at a higher frequency throughout the genome than the other combinations, suggesting that H2A.Z and H3.3 associate preferentially with each other to comprise the nucleosomes independently of genome region. Finally, we found that chromatin was unstable only in regions where it was enriched in both H2A.Z and H3.3, but strongly quantified stable in regions in which only H3.3 was abundant. Therefore, histone variant composition is an important determinant of chromatin structure, which is associated with specific genomic functions.

Citation: Yukawa M, Akiyama T, Franke V, Mise N, Isagawa T, et al. (2014) Genome-Wide Analysis of the Chromatin Composition of Histone H2A and H3 Variants in Mouse Embryonic Stem Cells. PLoS ONE 9(3): e92689. doi:10.1371/journal.pone.0092689

Editor: Brian P. Chadwick, Florida State University, United States of America

Received: November 18, 2013; **Accepted:** February 24, 2014; **Published:** March 21, 2014

Copyright: © 2014 Yukawa et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported in part by a Grant-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan (#20062002; URL: <http://www.jsps.go.jp/english/index.html>). No additional funding was received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: aokif@k.u-tokyo.ac.jp

These authors contributed equally to this work.

Introduction

The eukaryotic genome consists of genes and intergenic regions, which include enhancers, insulators, repeat elements, *etc.* These genomic components play roles in various genomic functions, such as gene expression, DNA replication and maintenance of genome stability. Emerging evidence from epigenetic studies revealed that various genome regions contain specific modifications on their nuclear core histones, which are involved in functional regulation [1,2]. Furthermore, recent studies showed that histone variants also play important roles in the regulation of genomic function [3–8]. Nuclear core histones H2A and H3 exist as several variants: H2A has the canonical H2A, H2A.X, H2A.Z, macroH2A (mH2A) and H2A.Bbd; H3 has H3.1, H3.2, H3.3, CENP-A and H3t [9]. These variants have been suggested to have specific functions in the regulation of gene expression and genome stability.

Of the H2A variants, canonical H2A is the most abundant in chromatin [10]. Although its function is unclear, recent studies suggest that canonical H2A is involved in gene silencing [11].

H2A.X is a highly conserved H2A variant. It is involved in the DNA damage repair pathway [12,13]. H2A.X is present at higher levels in mouse embryonic stem (mES) cells than in neuronal precursor cells, which are differentiated from mES cells [14]. H2A.X is preferentially incorporated into the nucleus of totipotent one-cell stage embryos [15]. These results suggest that H2A.X is involved in regulation of the undifferentiated state and cell proliferation. H2A.Z is a highly conserved variant, as is H2A.X. It is enriched in promoters and is associated with transcriptional regulation [16]. H2A.Z enrichment in promoters is negatively correlated with CpG methylation in plant and mammalian cells [17,18]. mH2A is a vertebrate-specific histone H2A variant and contains a large C-terminal domain, which is termed a macro domain [9,19,20]. mH2A is implicated in transcriptional repression during X chromosome inactivation [21]. Recent studies have shown that mH2A enrichment in gene bodies is negatively correlated with gene expression [6,22].

Of the H3 variants, H3.3 is the evolutionarily conserved replacement H3 [23]. H3.3 seems to be involved in the activation of gene expression. It is modified by active marks (*i.e.*, H3K4me3,

H3K36me3) and acetylation of various N-terminal lysines [24,25], and is enriched in the promoters and gene bodies of actively expressed genes [7]. It is also enriched in enhancers [7,26]. H3.1 and H3.2 have similar amino acid sequences; they differ in only a single residue [27]. Therefore, no antibody that can distinguish them is available, which has hampered both the clarification of their distribution in the genome and their function. However, mass spectrometry analysis revealed differences in their N-terminal modifications. H3.1 is preferentially acetylated and dimethylated on lysines 14 and 9, respectively, while H3.2 is preferentially subjected to trimethylation on lysine 27 [24,25]. Recently, genome-wide analyses of H3.2 in *Arabidopsis thaliana* showed that H3.2 enrichment in genes is negatively correlated with their expression levels, and H3.2 was deposited abundantly in regions with methylated CpG and H3K9me2, as well as H3K27me3 [28,29]. These results suggest that H3.2 is involved in the formation of heterochromatin structure. However, because *Arabidopsis thaliana* does not possess H3.1, the difference in the function of H3.1 and H3.2 remains to be elucidated.

Although several studies have assessed the genome-wide distributions of histone H2A and H3 variants, most examined only a single variant [6,22,30,31]. Because these studies used different species and cell types and were conducted using different protocols, it is difficult to integrate their data to gain a comprehensive understanding of the H2A and H3 variant compositions of various genome regions. Furthermore, few studies have investigated the combination of H2A and H3 variants. One report showed that nucleosomes containing H3.3 and H2A.Z were less stable than those containing H3.3 and H2A [32]. Therefore, the function of a nucleosome could vary according to changes in its H2A and H3 variant composition. Thus, conducting a genome-wide analysis of H2A and H3 variant combinations will provide an understanding of the mechanism underlying the regulation of various genomic functions. Although only a single report has shown that nucleosomes containing H2A.Z and H3.3 are enriched at promoters, enhancers, and insulators [26], other combinations of H2A and H3 variants have not been investigated.

In the present study, we examined the distribution of the majority of genomic H2A and H3 variants in mES cells using chromatin immunoprecipitation (ChIP) followed by high-throughput sequencing. We found that the enrichment of histone variants in gene regions (promoter and gene body) differed depending on gene expression level. Further, the enrichment was markedly different around the transcription start sites (TSSs) of genes with high- and low-CpG promoters. Furthermore, specific H2A and H3 variant compositions were observed in various regulatory and repeat elements. Also, H2A.Z and H3.3 associated preferentially with each other throughout the genome. Taken together, our results suggest that specific H2A and H3 variant compositions regulate the functions of various genome regions.

Materials and Methods

ZHBTc4 mouse ES cell culture

Feeder-free ZHBTc4 ES cells [33] were maintained on gelatin-coated dishes with ES medium, which consisted of Glasgow Minimal Essential Medium (GMEM; Sigma-Aldrich, G5154) supplemented with 2 mM GlutaMAX I (Invitrogen, cat#35050-061), 110 μ M 2-mercaptoethanol (Invitrogen, cat#21985-023), 1 \times MEM non-essential amino acids solution (NEAA; Invitrogen, cat#11140-050), 25 U/mL of penicillin-streptomycin (Invitrogen, cat#15140-122), 11 μ M sodium pyruvate (Invitrogen, cat#11360-070), 14% (v/v) KnockOut Serum Replacement (KSR; Invitrogen, cat#10828-028), 1% (v/v) fetal bovine serum

(FBS; Invitrogen, cat#16141-079), and 1,000 U/mL leukemia inhibitory factor (LIF; Millipore, cat#ESG1107). Blasticidin (1 mg/mL; Funakoshi, cat#KK-400) was added to the medium to maintain the pluripotency of the cells. To induce differentiation of the cells, the medium was replaced with ES medium without blasticidin, supplemented with tetracycline (Tc; Sigma, cat#T7660), and cultured for 72 h [33].

Plasmid construction and production of transgenic cell lines with Flag-tagged histone variants

A pCAGGS vector was used to generate transgenic ES cell lines constitutively expressing Flag-histone variants (H2A, H2A.X, H2A.Z, mH2A, H3.1, H3.2, and H3.3). A DNA fragment encoding a histone variant tagged with Flag at its N-terminal region was inserted into the EcoRI site of the pCAGGS vector. The puromycin N-acetyl-transferase (*Pac*) gene obtained by SalI digestion of pCre-Pac was inserted into the SalI site of the vector. The vector coding a Flag-histone variant was transfected into ZHBTc4 cells by lipofection, and cells were selected in ES medium using 100 μ g/mL of puromycin. A single colony of cells for each variant was picked up, cultured and subjected to ChIP-sequencing analysis. The difference in the genome distribution of a Flag-histone variant among the cell lines was assessed using the Flag-H2A.X cell lines. ChIP-quantitative polymerase chain reaction (qPCR) analysis of several genes showed no significant difference among the four cell lines examined (Figure S1).

Flag-tagged protein expression did not seem to affect ES cell properties, as no difference in the morphology of wild-type cells and those expressing Flag-tagged variants was observed (Figure S2). Furthermore, cells expressing Flag-tagged protein differentiated into trophectoderm like cells as well as the wild-type ones. Cluster analysis using gene expression profiles showed that the gene expression patterns of Flag-tagged-protein-expressing cells were not significantly altered from that of the wild-type cells. All types of undifferentiated cells expressing Flag-H2A and H3 variants were classified into the same cluster as the undifferentiated wild-type cells, while all differentiated cells clustered separately, regardless of the time after induction of differentiation (Figure S3).

RNA isolation and RNA sequencing

Following the TruSeq RNA Sample Preparation Guide, total RNA was isolated from ZHBTc4 ES cells without polyA selection at 0 and 72 h after the induction of differentiation. RNA was fragmented by heat treatment and then subjected to first and second strand cDNA synthesis, amplified, and used for library construction. The library thus prepared was applied to a Genome Analyzer IIx (Illumina), following the manufacturer's protocols, and high-throughput sequencing was carried out. Sequences (35 bp) were aligned against the cDNA sets from Ensemble GRC38 using Bowtie. The number of mapped reads per kilobase of exon per million reads (RPKM) for each gene was calculated using the following formula: $\text{RPKM} = (\text{gene read number} / (\text{total read number} \times \text{gene length})) \times 10^9$.

ChIP and sample preparation for sequencing

ES cells were cultured in 10-cm plates to ~80% confluence (1.2×10^7 cells per plate). One plate was used per ChIP aliquot. Formaldehyde solution (50 mM HEPES-KOH [pH 7.5], 100 mM NaCl, 1 mM EDTA [pH 8.0], 0.5 mM EGTA [pH 8.0], 11% formaldehyde) was directly added to culture media at a final concentration of 1% for fixation. After 10-min incubation at room temperature, a glycine solution was added to the plates at a final concentration of 125 mM, and the plates were

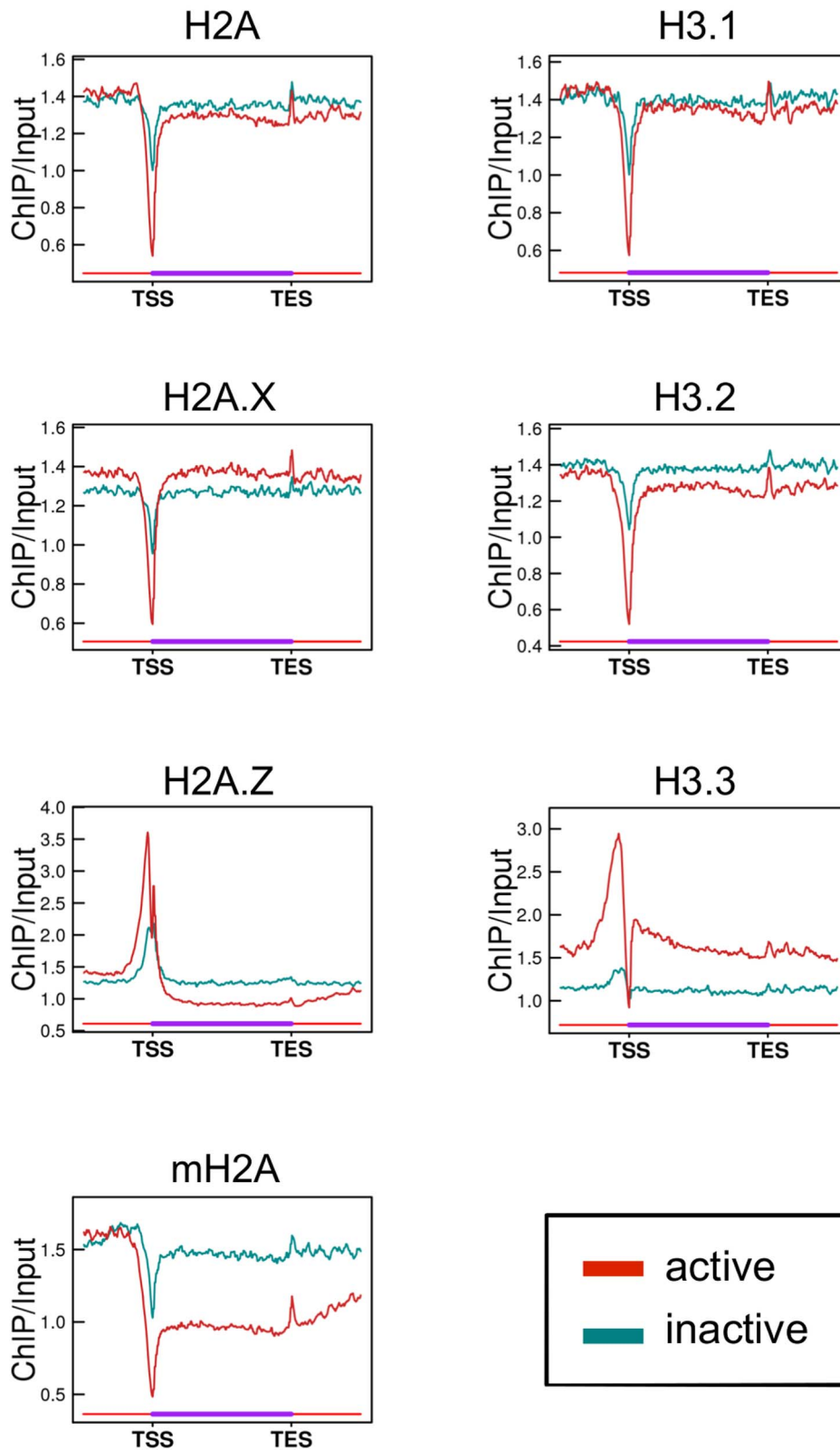


Figure 1. Enrichment of histone variants in active and inactive genes. Line graphs show the average enrichment (ChIP/input) of each histone variant in active (red) and inactive (dark cyan) genes. TSS and TES are the transcription start and end sites, respectively. The length of genes was normalized. In this analysis, genes of lengths <4 kb were excluded, as overlapping TSS and TES vicinities after length normalization would result in inaccuracies in the data of the gene bodies. In addition, genes whose distance from neighboring genes was <4 kb were also excluded to avoid overlapping of the flanking regions.

doi:10.1371/journal.pone.0092689.g001

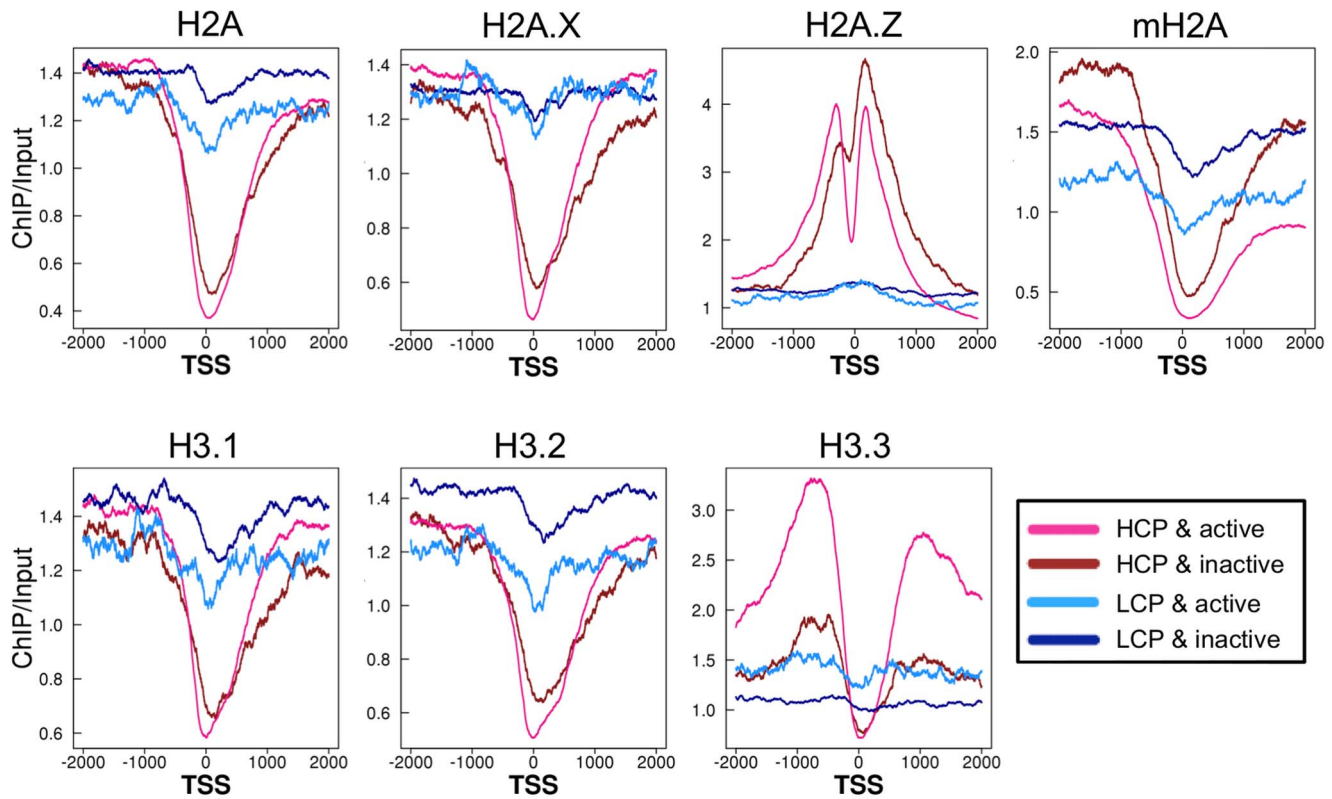


Figure 2. Enrichment of histone variants around the TSS with promoters of high and low CpG density. Line graphs show average enrichment of histone variants around the TSS of HCP & active, HCP & inactive, LCP & active and LCP & inactive genes. doi:10.1371/journal.pone.0092689.g002

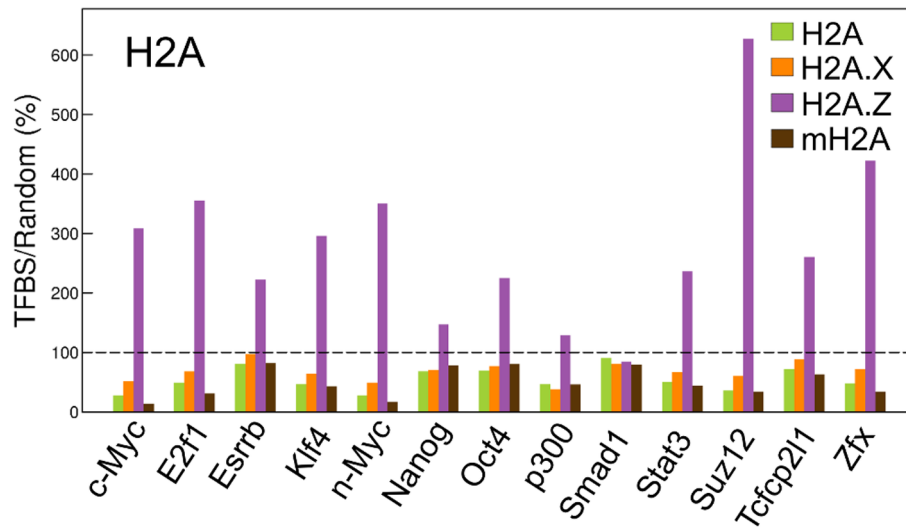
incubated at room temperature for 10 min to stop fixation. Cells were transferred to 15 mL tubes and washed twice with cold phosphate buffered saline (PBS). All buffers used in the following procedures were supplemented with protease inhibitor solution (final concentration 1 \times ; complete protease inhibitor cocktail tablet; Roche) before use. Cells were suspended in lithium borate (LB) buffer (10 mM Tris-HCl [pH 8.0], 100 mM NaCl, 1 mM EDTA [pH 8.0], 0.5 mM EGTA [pH 8.0], 0.1% Na-deoxycholate, 0.5% N-lauroylsarcosine), and sonicated (30 s pulse +1 min break) seven times on ice using a SONIFER 250 (Branson) to obtain 150–300-bp chromatin fragments. Triton X-100 was added into the tubes to a final concentration of 1%, and then the tubes were centrifuged (15,000 rpm, 4°C, 10 min) to obtain the supernatant used for immunoprecipitation. For the input sequencing, DNA was purified—without being immunoprecipitated—from cells expressing Flag-mH2A. To bind the antibodies to beads, 50 μ L of Dynabeads Protein G (Invitrogen) were washed twice with PBS, 5 μ g of the antibody in blocking solution (0.5% bovine serum albumin [BSA]/PBS) was added and gently flipped overnight at 4°C. Anti-FLAG M2 (Sigma, cat# F1804), anti-H3K4me3 (Millipore, cat#07-473), anti-H3K27me3 (Millipore, cat#07-449), and anti-H3K36me3 (Abcam, cat# ab9050) antibodies were used for immunoprecipitation. After three PBS washes, the antibody-bound beads were added to the supernatant containing the chromatin fragments, and gently flipped overnight at 4°C. The beads were washed once with a low salt buffer (20 mM Tris-HCl [pH 8.0], 150 mM NaCl, 2 mM EDTA [pH 8.0], 0.1% SDS, 1% Triton X-100), twice with a high salt buffer (20 mM Tris-HCl [pH 8.0], 400 mM NaCl, 2 mM EDTA [pH 8.0], 0.1% SDS, 1% Triton x-100), five times with RIPA

buffer (50 mM HEPES-KOH [pH 7.5], 500 mM LiCl, 1 mM EDTA [pH 8.0], 1% NP-40, 0.7% Na-Deoxycholate), and once with TE (pH 8.0) supplemented with 50 mM NaCl. To elute the chromatin, the beads were suspended in elution buffer (50 mM Tris-HCl [pH 8.0], 10 mM EDTA [pH 8.0], 1% SDS), and incubated at 65°C for 15 min. After centrifugation (15,000 rpm, room temperature, 15 min), the supernatant was collected and incubated overnight at 65°C to de-crosslink the chromatin. At this time, the samples for input sequencing were added to the elution buffer and then subjected to the same treatments as those for the immunoprecipitated samples in the following procedures. After de-crosslinking, RNase A (25 mg, Roche) was added to all samples, which were incubated at 37°C for 2 h, followed by the addition of proteinase K (20 mg/mL; Takara Bio) and incubation at 55°C for 2 h. Finally, DNA was isolated by the phenol-chloroform method followed by ethanol precipitation; isolated DNA was dissolved in Milli-Q water. Adapters were ligated to both ends of the DNA following the Illumina protocol for “Preparing Samples for ChIP Sequencing of DNA.” The ligated DNA was subjected to electrophoresis, and 250–400-bp DNA fragments were excised from the gel and purified. The purified DNA was subjected to PCR amplification following the Illumina protocol and then sequenced using the Illumina Genome Analyzer (Illumina).

Sequencing and read mapping

DNA fragments obtained from ChIP were applied to the deep sequencer following the manufacturer’s protocol. Sequences (36 bp) obtained by high-throughput sequencing were aligned against the mouse reference genome (mm9) using Bowtie. The

A



B

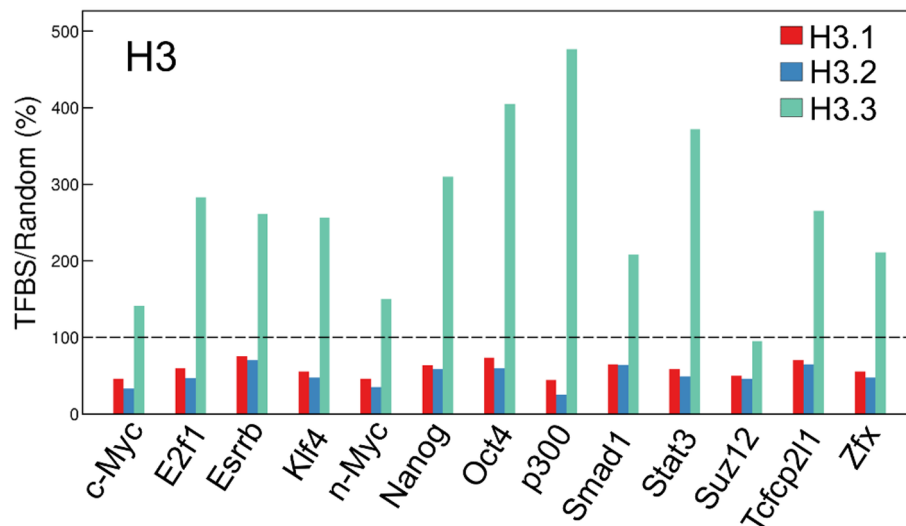


Figure 3. Enrichment of histone variants in enhancer regions. Bars represent the ratios (%) of enrichment of the histone H2A (A) and H3 (B) variants ± 200 bp from the center of the transcription factor binding site (TFBS) as defined using Chen's TF ChIPseq data [35] to the average enrichment in 5,000 400-bp regions randomly selected from the whole genome, excluding TFBSs. The dotted line indicates 100%, representing the average value of the enrichment in randomly selected regions.
doi:10.1371/journal.pone.0092689.g003

total number of uniquely mapped reads in each sample is shown in Table S2.

Determination of the enrichment of each histone variant at a single-nucleotide resolution

Uniquely mapped 36-bp sequences in each ChIP sample were extended to 150–300 bp, which is the length appropriate for sequencing without adapters. At each mapped nucleotide position, the number of reads mapped there $\times 1,000,000$ /total number of reads was calculated. For input samples, the calculated value at each position was added to the minimum calculated value, except for zero, to avoid a value of zero at any position. The calculated

values of ChIP samples were divided by those of input samples to determine the enrichment at each nucleotide position in the genome.

Preferential combinations of H2A and H3 variants

The whole genome was compartmentalized into 150-bp bins (Figure S4). The ratio of the number of reads in a bin to the averaged read number of 3,333 bins (total length, ~ 500 kb) around the corresponding bin was calculated for each variant, and the variant with the highest ratio value was selected. The read number of the selected variant was analyzed statistically using a Poisson distribution with the value of λ , which was determined by

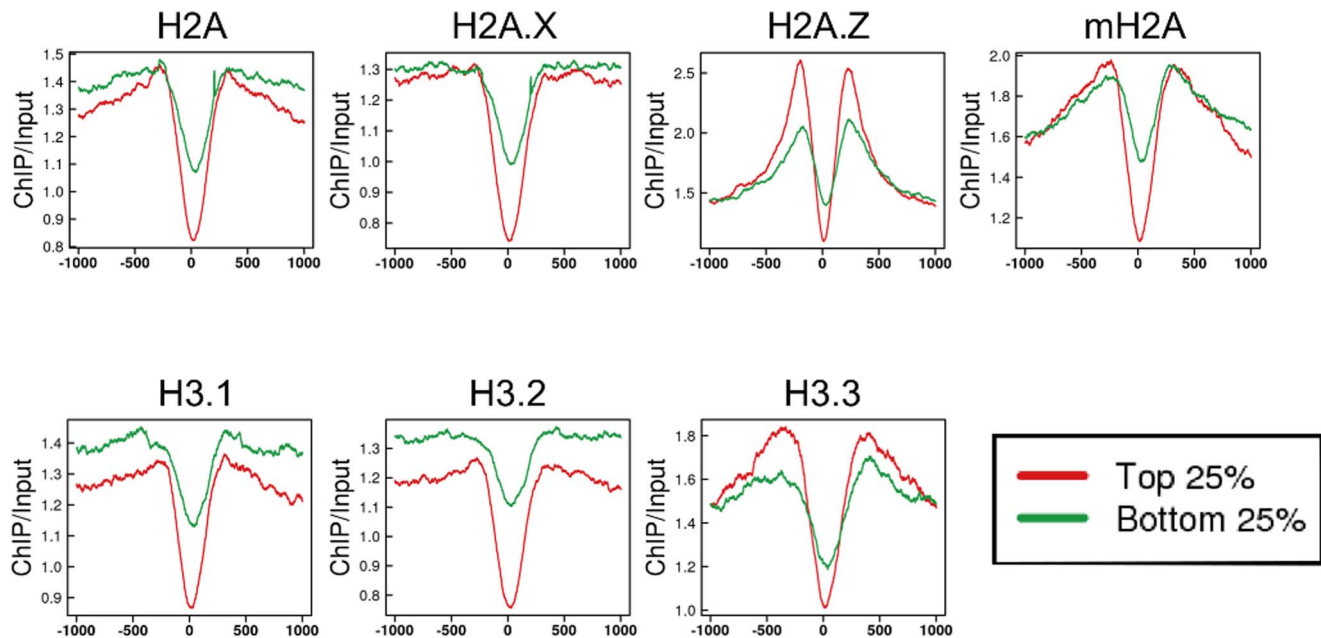


Figure 4. Enrichment of histone variants in insulator regions. Profile of the average enrichment of H2A (upper) and H3 (lower) variants in insulator regions defined as CTCF binding sites in genomic regions except for gene body obtained from Chen's ChIPseq data [35]. The 0 position is the center of the CTCF binding site. CTCF binding sites were ordered based on the amount of CTCF bound; the average enrichment of histone variants in the top (red) and bottom (green) 25% is shown.
doi:10.1371/journal.pone.0092689.g004

multiplying the averaged read number of the selected variant by the second highest ratio in the corresponding bin. If this was significant ($FDR < 0.05$), the bin was judged to be enriched for that variant. The nucleosome around a bin enriched for specific H2A and H3 variants was considered to be preferentially composed of these two variants.

Determination of major isoform among transcripts from a gene

To determine major isoform among transcripts from a gene, we determined the transcript with the highest RPKM value in transcripts from a gene registered in Ensemble (GRC38) as a major isoform of the gene.

Gene classification according to CpG density in the promoter

Genes registered in Ensemble (GRC38) were classified into HCP, ICP, and LCP groups as described by Mikkelsen *et al.* [34].

CpG islands

The CpG island annotation file was downloaded from the UCSC Genome Bioinformatics site (<http://genome.ucsc.edu/>), and used for the determination of CpG islands.

Definition of TFBS

ChIPseq of transcription factor (TF) and CTCF data were obtained from Chen *et al.* [35]. TF and CTCF binding sites were identified using Cisgenome peak caller with cutoff ($FDR < 0.01$). The numbers of binding sites thus identified were 2506, 22533, 55527, 17057, 8422, 6851, 3157, 15, 771, 1520, 3149, 28819, and 20975 for c-MYC, E2F1, ESRRB, KLF4, n-MYC, NANOG, OCT4, P300, SMAD1, STAT3, SUZ12, TCF21, and ZFX, respectively. For CTCF, the numbers of binding sites in genomic regions except for gene body were 16612.

Repeat element analysis

Repeat element sequences were obtained from Repbase (<http://www.girinst.org/repbase/>). The number of the reads from ChIPseq data mapped on each repeat element was determined and corrected for the length of the repeat element and the number of reads for the input sample.

Data access

Gene expression and sequencing data have been deposited in the DDBJ Sequence Read Archive (DRA) (<http://trace.ddbj.nig.ac.jp/dra/index.html>) under accession number DRP001103.

Results

Distribution of histone variants in genes

Because no ChIP antibody that reacts specifically with H3.1 or H3.2 is available, we prepared mouse ES cells (ZHBTc4) that expressed Flag-tagged H3.1, H3.2 and H3.3. We also prepared ES cells expressing Flag-tagged H2A variants to conduct ChIP with identical specificity and efficiency for H2A and H3 variants. Thus, we established seven cell lines that constitutively expressed Flag-tagged H2A, H2A.X, H2A.Z, mH2A, H3.1, H3.2 or H3.3.

We first investigated the distribution of H2A and H3 variants both in genes and in their vicinities. Genes were classified as active or inactive based on RNA sequencing data (Figure S5). We found that all histone variants examined, except for H2A.Z, were markedly depleted at the TSS of active genes (Figure 1). In inactive genes, depletion at the TSS was also observed, except for H2A.Z and H3.3, but the degree of depletion was less than in active genes. In the H2A.Z and H3.3 ChIP data, peaks of abundant H2A.Z and H3.3 deposition were observed in the vicinity of the TSS. H3.3 was enriched in gene bodies as well as in the upstream regions of active genes compared to inactive genes. In contrast, mH2A enrichment was lower in the bodies of active genes than inactive

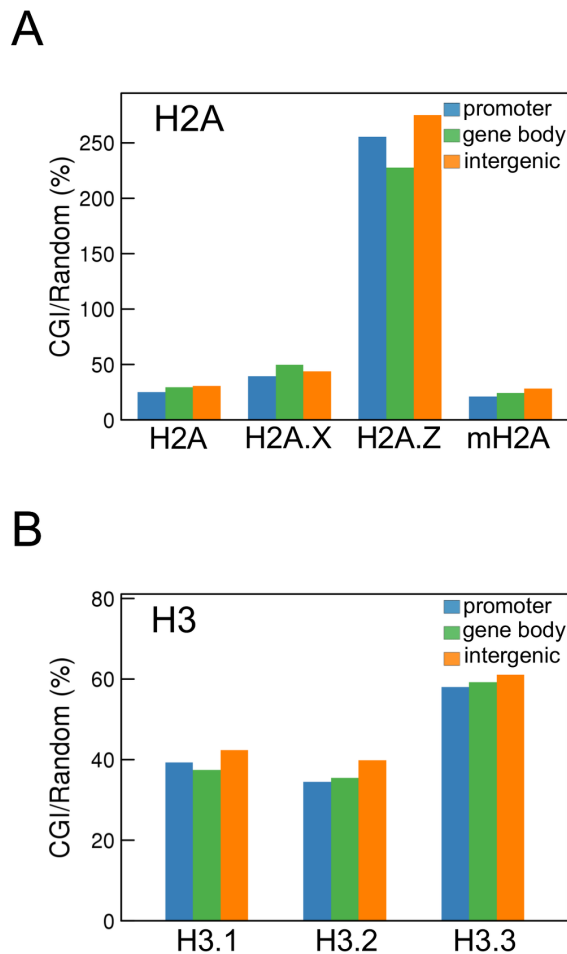


Figure 5. Enrichment of histone variants in CpG islands (CGIs). The enrichment of histone H2A (A) and H3 (B) variants in CGIs defined in the UCSC Genome Bioinformatics Site was determined in promoters (blue), gene bodies (green), and intergenic regions (orange). The ratios (%) of the enrichment in CGIs to the averaged enrichment in 5,000 randomly selected 656-bp regions, which is the average length of all CGIs analyzed, from the whole genome excluding CGIs were calculated. doi:10.1371/journal.pone.0092689.g005

genes, suggesting that chromatin in the bodies of active genes is depleted in mH2A. H2A.Z enrichment was also less in the bodies of active genes, while it was higher in the vicinity of the TSS.

To further assess the correlation between the expression level and the enrichment of histone variants in genes, genes were classified into six groups according to their expression levels (Figure S6). In these six groups, the enrichment of H3K4me3 and H3K27me3 at promoters, which are known to be positively and negatively correlated with gene expression level, respectively, and H3K36me3 enrichment in gene bodies, which is correlated with gene expression level [16], were altered in the expected manner (Figure S7, S8). In terms of histone variants, H3.3 and mH2A enrichment were positively and negatively correlated with gene expression, respectively, in both promoters and gene bodies (Figure S7, S8). A negative correlation was observed between gene expression level and H2A.Z enrichment in gene bodies (Figure S8). Interestingly, although H2A.Z enrichment at promoters was positively correlated with gene expression at a lower expression level (group 1–3), it was negatively correlated at a higher level (group 4–6) (Figure S7). These results suggest that H2A.Z, mH2A,

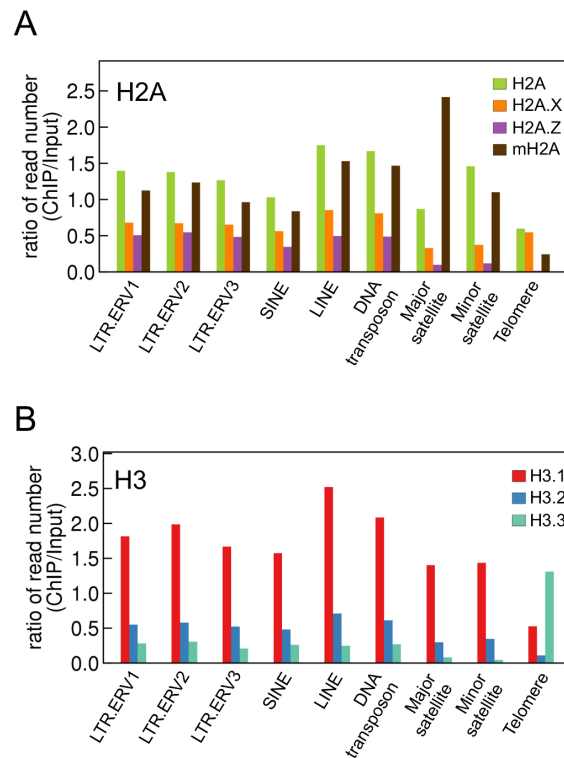


Figure 6. Enrichment of histone variants in repeat elements. Bars represent the ratios of the number of ChIP reads to that of input reads in repeat elements. doi:10.1371/journal.pone.0092689.g006

and H3.3 deposition in chromatin plays an important role in the regulation of gene expression.

Chromatin structure and the regulation of transcription are known to differ markedly between genes with promoters of high and low CpG densities [36–39]. To investigate the association of CpG density at promoters with the enrichment of histone variants in genes, we classified genes according to their promoter CpG density and expression level (Table S1). First, genes were classified into those with high, intermediate and low CpG density promoters (*i.e.*, HCP, ICP and LCP, respectively) according to Mikkelsen's criteria [34], and then as active and inactive using the criteria described above (Figure S5). Although in the gene body the enrichment of no variant differed markedly between genes with HCP and LCP promoters, enrichment of H2A.Z and H3.3 in the promoter regions differed markedly between these two types of genes (Figure S9, S10). The enrichment of H2A.Z was higher in genes with HCP promoters regardless of gene activity. In H3.3, enrichment was similar between active genes with LCP promoters and inactive genes with HCP promoters, and was higher in active genes with HCP promoters than the others. The enrichment of variants in the promoters was thus more dependent on the promoter type, HCP or LCP, than the gene expression level (Figure S9). The only exceptions were mH2A and H3.3 in the promoter region, enrichment of which seemed to be associated with gene expression (Figure S10).

A detailed analysis indicated that all H2A and H3 variants—with the exception of H2A.Z—were depleted from the nucleosome at the TSS of genes with HCP promoters, regardless of their expression levels (Figure 2). H2A.Z was also depleted from the nucleosome at the TSS in active, but not inactive, genes with HCP promoters. All of variants, except for H2A.Z, their levels were

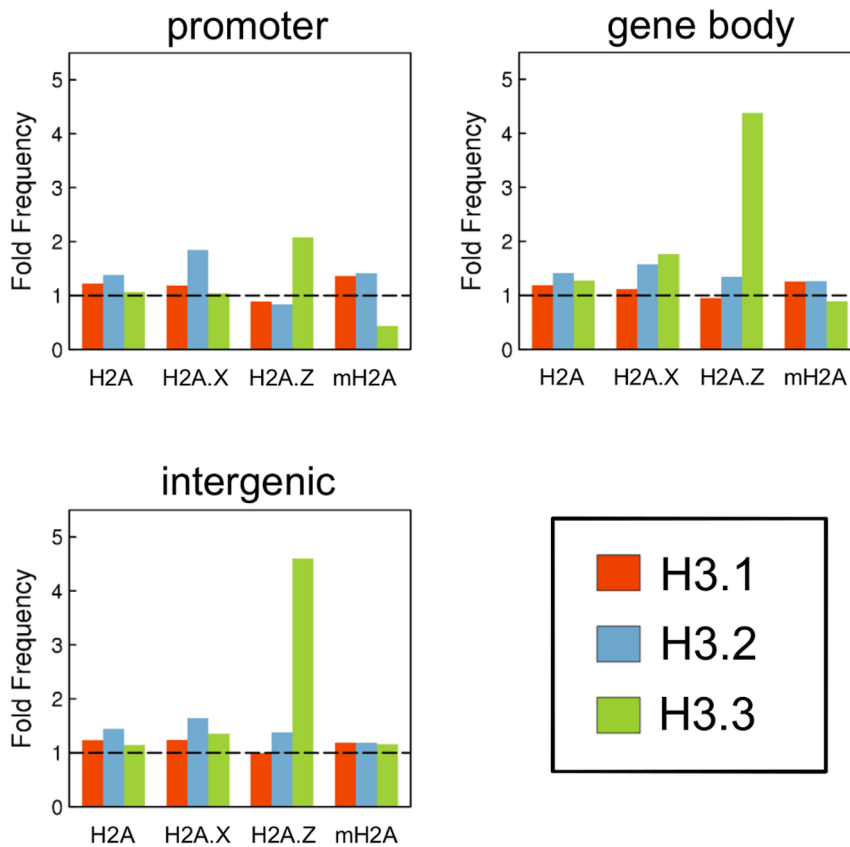


Figure 7. Preferred combination of H2A and H3 variants in various genome regions. The genome was divided into 150 nucleotide bins, and the bins in which a particular variant was significantly enriched were determined as described in the Materials and Methods (see Figure S4). The frequency of bins in which two H2A and H3 variants were simultaneously enriched was determined for all combinations of H2A and H3 variants. Bars represent the ratios of the frequencies determined as described above to those of the expected values calculated from the number of bins in which each variant was enriched. Therefore, values >1.0 indicate a preference for that particular H2A and H3 variant combination. doi:10.1371/journal.pone.0092689.g007

always higher in genes with LCP promoters than those with HCP ones, regardless of their gene expression levels. On both sides of the histone-depleted region at the TSS, H2A.Z was highly enriched in genes with HCP promoters, but not those with LCP promoters, regardless of their expression levels. Although H3.3 was also highly enriched in the vicinity of the TSS in genes with HCP promoters, it was present in only active genes. These results suggest that CpG density at the promoter, but not gene activity, is associated with the enrichment of the majority of H2A and H3 variants at the region around the TSS.

Composition of histone variants in regulatory elements

Many regulatory elements have specific functions in the genome. We thus investigated the enrichment of H2A and H3 variants in these elements to evaluate the involvement of histone variants in their functions. Both H2A.Z and H3.3 were enriched in all of enhancers defined as transcription factor binding sites (TFBSs) [40], except for *Smad1*, suggesting that the nucleosomes at enhancer elements comprise mainly H2A.Z and H3.3 (Figure 3).

Next, we examined histone variant enrichment in insulators. Insulators bind to CCCTC-binding factor (CTCF), which contributes to higher-order chromatin architecture for long-range interactions [41,42]. These interactions regulate enhancer-promoter binding, and therefore are involved in the regulation of gene expression [43–45]. CTCF also contributes to pol II stalling and splicing in bodies of genes [46]. Therefore, we defined insulators as

the sites to which CTCF was bound in genomic regions except for gene body (see Materials and Methods) and compared the enrichment of histone variants at sites with the highest and lowest quartile of bound CTCF. All variants were absent from the CTCF binding sites in both the highest and lowest quartiles, although the extent of the removal was greater in the highest quartile (Figure 4), suggesting that nucleosomes are almost completely depleted from insulator elements when they bind CTCF. In addition, on both sides of the binding site, abundant H2A.Z deposition, as evidenced by the peak height, was observed. The peaks for H3.3 and mH2A were slightly wider at these regions. The enrichment of H3.1 and H3.2 was low in a wide region around the CTCF binding sites.

Many genes have CpG islands (CGIs) in their upstream regulatory regions; moreover, a number of CGIs are present in gene bodies and intergenic regions [47]. We found that H2A.Z was more enriched in CGIs compared with other H2A variants (Figure 5A). Among the H3 variants, H3.3 was relatively highly enriched in CGIs (Figure 5B). This enrichment of H2A.Z and H3.3 at CGIs was true for all genome regions (*i.e.*, promoters, gene bodies, and intergenic regions), suggesting that their enrichment at CGIs is not regulated by a particular epigenetic environment but is dependent only on the presence of a CG-rich DNA sequence.

Composition of histone variants in repeat elements

Repeat elements are generally enriched in H3K9me3 and H4K20me3, and are constitutively heterochromatic [34]. We

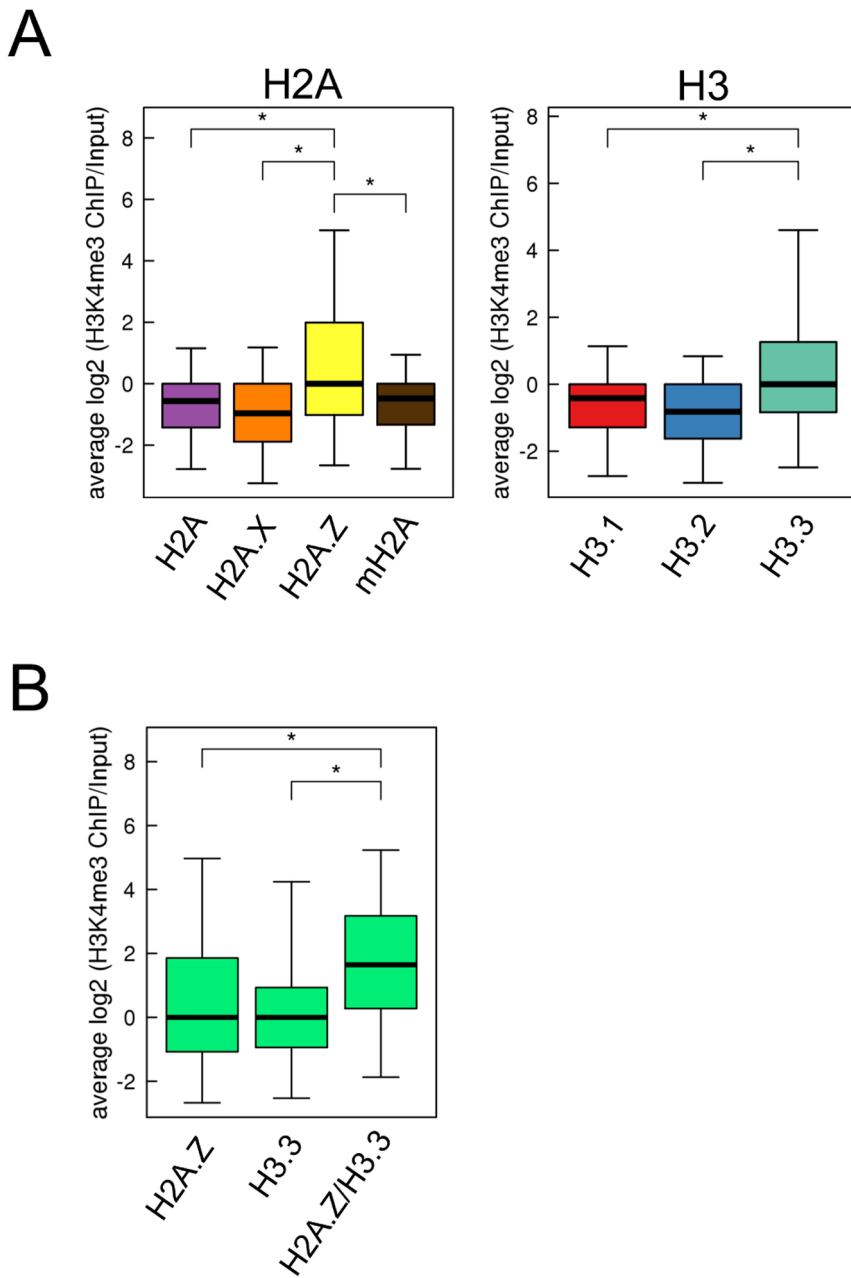


Figure 8. H3K4me3 level in H2A.Z- and H3.3-enriched regions. The genome was divided into 150 nucleotide bins, and the bins in which a particular variant was significantly enriched was determined as described in Materials and Methods (see Figure S4). (A) Boxplot of H3K4me3 levels in H2A- and H3-variant-enriched bins. (B) Boxplot of H3K4me3 levels in H2A.Z- or H3.3-, and both H2A.Z and H3.3 (H2A.Z/H3.3)-enriched bins. The bottom and top of the box are the 25th and 75th percentile, respectively, and the upper and lower whiskers represent 2.5th and 97.5th percentile, respectively. *P*-values were calculated using Mann-Whitney U-tests; (*) *P*<0.001. doi:10.1371/journal.pone.0092689.g008

found that of the H2A and H3 variants, H2A, mH2A, and H3.1 were relatively enriched in most repeat elements (Figure 6). The only exception was the telomeres, in which H3.3 was enriched, but where no particular H2A variant was enriched. Remarkably, H2A.Z enrichment in this region was extremely low.

Preferential combination of H2A and H3 variants in various genomic regions

To investigate the existence of a preferential combination of H2A and H3 variants comprising nucleosomes in various genome

regions, we compartmentalized the genome into 150-bp bins, and identified those in which the particular histone H2A or H3 variant was enriched (Figure S4). Then, we determined the frequencies of bins in which a pair of H2A and H3 variants was enriched simultaneously. As shown in Figure 7, the frequency of bins in which H2A.Z and H3.3 enrichment overlapped was high in all genome regions. These results suggest that nucleosomes tend to comprise a combination of H2A.Z and H3.3 variants, and that this property is independent of the genome region.

Both H2A.Z and H3.3 are associated with H3K4me3 [7,48]; thus, we investigated whether the combination of H2A.Z and

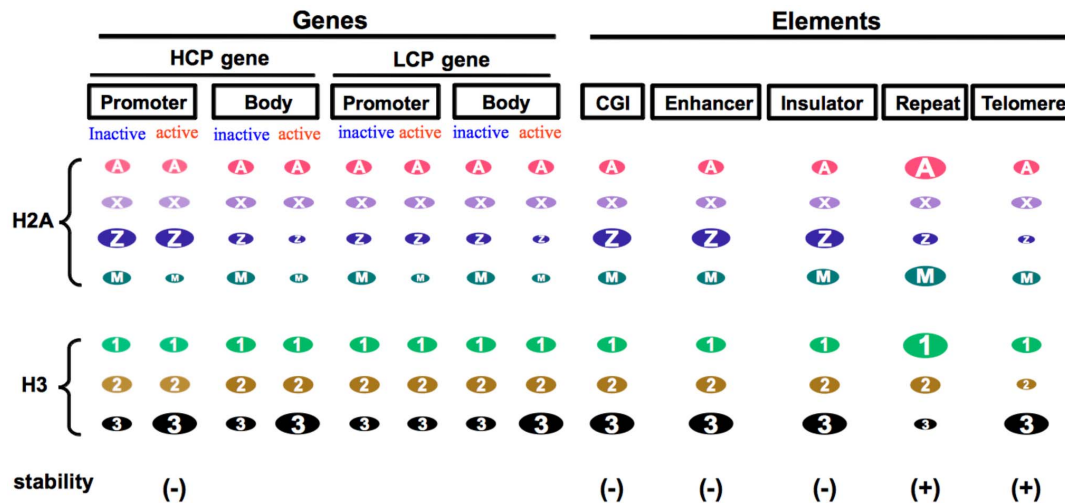


Figure 9. Schematic view of histone variant enrichment in various genome regions. Relative enrichment of each histone variant in various genome regions is indicated by the ellipse magnitude. Stability (+) and (-) indicate stable and unstable chromatin, respectively. Chromatin was unstable only in regions enriched for both H2A.Z and H3.3. doi:10.1371/journal.pone.0092689.g009

H3.3 influences the level of H3K4me3. Consistent with previous reports, the H3K4me3 level was highest in the H2A.Z- and H3.3-enriched bins (Figure 8A). The H3K4me3 level was higher in bins in which both H2A.Z and H3.3 were enriched simultaneously than those in which only H2A.Z or H3.3 was enriched (Figure 8B). These results suggest that nucleosomes composed of H2A.Z and H3.3 are preferentially subjected to the H3K4me3 modification.

Discussion

Several reports of the genome-wide distribution of various histone variants have been published [3,4,7,22,26,28,31,49]. However, most examined a single histone variant. In the present study, we investigated the majority of H2A and H3 variants expressed in mouse ES cells by genome-wide analysis using ChIP sequencing, which revealed enrichment of H2A and H3 variants in various genome regions (Figure 1–6). Furthermore, analysis of the combinations of H2A and H3 variants suggests that H2A.Z and H3.3 preferentially associate with each other to generate nucleosomes throughout the genome (Figure 7). Such data could only be obtained by simultaneous analysis of several variants. During this study, a report of the genome-wide distribution of most H2A variants (*i.e.*, H2A, H2A.Z, mH2A, and H2A.Bbd) was published [50]; however, H2A.X was not included. Of all H2A variants, H2A.X is evolutionarily closest to H2A. H2A.X and H2A are not distinguished as separate proteins in phylogenetically ancient species (*e.g.*, insects and fungi). Therefore, H2A and H2A.X must be investigated simultaneously to understand their roles in vertebrate genomic function. In addition, the amino acid sequence of H3.1 is similar to that of H3.2, and is present only in mammals. Because no antibody that can distinguish these proteins is available, the difference in their genome distribution and function has not been determined. Here we report that the genome distributions of H2A and H2A.X, and H3.1 and H3.2 differ, suggesting that each of these variants has a specific function that is different from their counterparts.

We found that H3.1 but not H3.2 was enriched in repeat elements other than telomeres (Figure 6B). These elements are enriched in H3K9me3 [34,51], and so constitutive heterochromatin is generated [52,53]. In addition, H3K9me2 is enriched at

centromere minor repeats [54,55]. H3.1 is preferentially modified with H3K9me2 and H3K9me3 as compared to other H3 variants [24,25]. Therefore, repeat elements are enriched in H3.1 with H3K9me2 and H3K9me3, which contributes to heterochromatin formation. Furthermore, enrichment of H2A and H2A.X was different in repeat elements other than telomeres (Figure 6A). These data demonstrate the difference between H2A and H2A.X as well as H3.1 and H3.2, and suggest that nucleosomes composed of H2A and H3.1 contribute to the formation of constitutive heterochromatin at repeat elements.

Histone variants play important roles in the regulation of transcription. Previous reports showed that H2A.Z and H3.3 enrichment in promoters and gene bodies was well correlated with the expression levels of the genes [3,7,16], and that mH2A enrichment in gene bodies was inversely correlated with gene expression [6]. In the present study, the enrichment of these variants in promoters and gene bodies was correlated with gene expression levels, as reported previously (Figure 1 and Figure S7, S8). H3.2 enrichment is negatively correlated with gene expression in *Arabidopsis thaliana* [28]. However, this finding was not replicated here (Figure S7, S8). This is consistent with the study by Goldberg *et al.* (2010), who reported that H3.2 enrichment was not correlated with gene expression in mouse ES cells. Therefore, the role of H3.2 in gene expression seems to differ between mice and plants.

Previous studies have shown that nucleosomes are depleted at the TSSs of actively expressed genes [56–58]. However, one recent report showed that TSSs are sensitive to MNase and that the sensitivity is more dependent on CpG density in promoters than on gene expression [59,60], which suggests that nucleosomes are unstable or depleted at TSSs of HCP genes. Supporting this hypothesis is our finding that almost all histone variants were depleted at TSSs of HCP, but not LCP, genes regardless of their expression levels (Figure 2). The discrepancy between previous, and recent studies—including ours—may be due to the difference in the occupancies of genes with HCP and LCP promoters between the active and inactive gene groups. Genes with HCP and LCP promoters comprise major proportions of active and inactive genes, respectively (Table S1). Therefore, the results of previous analyses of active and inactive genes likely reflect genes with HCP

and LCP promoters, which represent active and inactive genes, respectively.

Nucleosomes containing H2A.Z and H3.3 are preferentially subjected to the H3K4me3 modification. Previous studies showed that regions of H2A.Z enrichment overlapped substantially with those of H3K4me3 in promoter and enhancer regions [48]. The distribution of H3.3-enriched regions is similar to that of H3K4me3 [7]. Consistent with these results, we showed that H3K4me3 enrichment in nucleosomes containing H2A.Z and H3.3 was higher than in those containing other H2A and H3 variants (Figure 8A). Furthermore, we suggest that the nucleosomes composed of H2A.Z and H3.3 were more highly enriched in H3K4me3 than those containing only H2A.Z or H3.3 (Figure 8B). Thus, the combination of H3.3 and H2A.Z is important in terms of constituting the structure of nucleosomes that are preferentially subjected to H3K4me3, and are involved in the formation of transcriptionally active chromatin. Therefore, information from only one of each H2A and H3 variant would not be enough to understand the structure of nucleosomes and chromatin, and analyzing the combination of H2A and H3 variants composing the nucleosomes is important.

H2A.Z and H3.3 were enriched in regulatory elements (*i.e.*, enhancers, insulators, and CGIs; Figure 3, 4, 5). Various chromatin proteins bind to these regions, including the transcription factors, CTCF and Cfp1, which recruit histone modification enzymes [61,62]. Therefore, H2A.Z and H3.3 might be recruited to these regions to form nucleosomes that are readily accessible to chromatin proteins. In insulator regions, mH2A peaks were slightly wider. Insulators form chromatin loop, and contribute to higher-order chromatin [41,42]. Therefore, mH2A, which contributes to formation of chromatin loop, may be abundant both up- and downstream of insulator regions.

Previous studies suggested that nucleosomes containing both H2A.Z and H3.3 compose unstable chromatin. Jin and Felsenfeld [32] reported that nucleosomes containing both of these variants were more highly unstable than those containing other combinations of the H2A and H3 variants under high-salt conditions *in vitro*. Genome-wide analysis of the distribution of H2A.Z and H3.3 revealed that both of these variants are enriched around the TSSs of active genes, enhancers and insulators, which suggests that in these regions, nucleosomes are so unstable that the variants are easily removed and displaced by DNA-binding proteins (*e.g.*, transcription factors and CTCF) [26]. In the present study, we found that both H2A.Z and H3.3 are abundant in genome regions known to have unstable chromatin (*i.e.*, the promoters of active HCP genes, insulators, enhancers and CGIs; Figure 9). Ramirez-Carrozzi [63] reported that the expression of LCP genes depended on the SWI/SNF chromatin remodeling factors, whereas that of HCP genes was independent of SWI/SNF, suggesting that the chromatin of HCP genes is so unstable that transcription factors can easily access DNA without remodeling the chromatin structure. It has also been shown that the turnover rate of H3.3 was high around the TSS of active genes [64]. Insulators, enhancers and CGIs are DNase I hypersensitive, which suggests that the chromatin of these elements is loosened and unstable [65–68]. In contrast, it is interesting that in regions in which only H3.3 is enriched, chromatin is not unstable but rather stable (Figure 9). In the bodies of transcribed genes in which only H3.3 is enriched, chromatin is so stable that histone exchange is suppressed, which prevents the cryptic initiation of transcription of the gene bodies [69–72], although a report showed that the turnover of chromatin proteins was rapid in the bodies of active genes [73]. In telomeres, in which stable constitutive heterochromatin is formed [74], only H3.3 is enriched. Therefore, the histone variant composition is an

important determinant of chromatin stability, which is associated with various genomic functions.

In summary, we clarified the histone variant composition of various genome regions by genome-wide analysis of the majority of H2A and H3 variants expressed in mouse ES cells. We found that in HCP, but not LCP, genes, regardless of their expression levels, most of the variants were depleted at the TSS but that H2A.Z and H3.3 were enriched in the vicinity of the TSS. H2A.Z and H3.3 were also enriched in enhancers, CpG islands and insulators. We found that H2A and H3.1 were enriched in repeat elements, which is the first knowledge of the differential distribution between H2A and H2A.X, and between H3.1 and H3.2 on specific genome regions. The analysis of the combinations of H2A and H3 variants that comprise nucleosomes revealed that the frequency of the H2A.Z and H3.3 combination was high throughout the genome when compared to other combinations, and that nucleosomes containing this combination were preferentially trimethylated at the lysine 4 residue. None of these results could have been obtained without simultaneous analysis of the majority of histone variants. Taken together, these data suggest that specific combinations of histone variants influence the regulation of various genomic functions (*e.g.*, gene expression and heterochromatin formation) and histone modifications.

Supporting Information

Figure S1 The genome distribution pattern of Flag-H2A.X among various cell lines. The enrichment of Flag-H2A.X in four genome regions (*i.e.*, *Nanog* promoter, *Cdx* 3'UTR, *Pml1* promoter and *Sry* promoter) was examined in four cell lines expressing Flag-H2A.X by ChIP-qPCR.

(TIF)

Figure S2 Morphology of ES cells expressing Flag-tagged histone variants. ZHBTc4 cells expressing Flag-tagged H2A, H2A.X, H2A.Z, macroH2A (mH2A), H3.1, H3.2 or H3.3, and wild-type (WT) cells, were observed before (0 h) and after (72 h) induction of differentiation.

(TIF)

Figure S3 Clustering analysis of gene expression profiles in ES cells expressing Flag-histone variants and wild-type cells before and after the induction of differentiation. cDNA microarray data sets of wild-type ZHBTc4 cells before (WT_0 h) and 24, 48, 72, and 96 h after induction of differentiation (WT_24 h, WT_48 h, WT_72 h, and WT_96 h, respectively) and cells expressing H2A, H2A.X, H2A.Z, macroH2A (mH2A), H3.1, H3.2 or H3.3 without differentiation induction were subjected to hierarchical clustering. Results are shown as a tree view.

(TIF)

Figure S4 Identification of the regions in which a particular histone H2A or H3 variant is enriched. As an example, the identification of the bin in which H2A.Z is enriched is shown.

(TIF)

Figure S5 Histogram of the expression levels of protein-coding genes in mouse ES cells. The expression levels of coding genes were obtained from the RNA sequence data. The histogram of gene expression levels was smoothed by kernel density estimation, which is depicted by bimodal peaks. The genes in the first and second peaks are defined as inactive and active, respectively.

(TIF)

Figure S6 Classification of genes according to expression level. The expression level of coding genes was obtained from the RNA sequence data. The histogram of gene expression levels was smoothed by kernel density estimation. The genes were classified into six groups according to their expression levels. Groups 1, 2, 3, 4, 5 and 6 contained 6728, 447, 1161, 2526, 2760 and 1009 genes, respectively.
(TIF)

Figure S7 Correlation between the expression level and enrichment of histone variants in gene promoters. Genes were classified into six groups according to their expression levels as shown in Figure S6, and the enrichment of histone variants and histone modifications in the promoters (TSS±2 kb) of genes in each group are shown as box plots. The bottom and top of the box are the 25th and 75th percentile, respectively, and the upper and lower whiskers represent 2.5th and 97.5th percentile, respectively.
(TIF)

Figure S8 Correlation between the expression level and enrichment of histone variants in gene bodies. Genes were classified into six groups according to their expression levels as shown in Figure S6, and the enrichment of histone variants and histone modifications in the gene body (TSS+2 kb~TES) in each group are shown as box plots.
(TIF)

Figure S9 Enrichment of histone variants in the promoters of HCP and LCP genes. Genes were classified into HCP & active, HCP & inactive, LCP & active, and LCP & inactive on the basis of CpG density at the promoter and gene

expression levels. Enrichment of histone variants at promoters (TSS±2 kb) in each group is shown as a box plot.
(TIF)

Figure S10 Enrichment of histone variants in the bodies of HCP and LCP genes. Genes were classified into HCP & active, HCP & inactive, LCP & active, and LCP & inactive on the basis of CpG density at the promoter and gene expression levels. Enrichment of histone variants at gene bodies (TSS+2 kb~TES) in each group is shown as a box plot.
(TIF)

Table S1 Classification of the genes by their expression levels and the CpG density at promoter.
(TIF)

Table S2 The number of reads uniquely mapped to the genome for all ChIPseq and Inputseq data sets.
(TIF)

Acknowledgments

We thank H. Niwa for providing ZHBTc4 ES cells. Computation time was provided by the Supercomputer System, Human Genome Center, the Institute of Medical Science, the University of Tokyo.

Author Contributions

Conceived and designed the experiments: MY TA FA. Performed the experiments: MY TA. Analyzed the data: MY TA VF MGS KV KA HA FA. Contributed reagents/materials/analysis tools: NM TI YS KA HA. Wrote the paper: MY FA.

References

- Margueron R, Reinberg D (2010) Chromatin structure and the inheritance of epigenetic information. *Nat Rev Genet* 11: 285–296. doi:10.1038/nrg2752.
- Zhou VW, Goren A, Bernstein BE (2010) Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* 12: 7–18. doi:10.1038/nrg2905.
- Mito Y, Henikoff JG, Henikoff S (2005) Genome-scale profiling of histone H3.3 replacement patterns. *Nat Genet* 37: 1090–1097. doi:10.1038/ng1637.
- Creyghton MP, Markoulaki S, Levine SS, Hanna J, Lodato MA, et al. (2008) H2AZ Is Enriched at Polycomb Complex Target Genes in ES Cells and Is Necessary for Lineage Commitment. *Cell* 135: 649–661. doi:10.1016/j.cell.2008.09.056.
- Buschbeck M, Urbesalgo I, Wibowo I, Rué P, Martin D, et al. (2009) The histone variant macroH2A is an epigenetic regulator of key developmental genes. *Nat Struct Mol Biol* 16: 1074–1079. doi:10.1038/nsmb.1665.
- Gamble MJ, Frizzell KM, Yang C, Krishnakumar R, Kraus WL (2010) The histone variant macroH2A1 marks repressed autosomal chromatin, but protects a subset of its target genes from silencing. *Genes & Development* 24: 21–32. doi:10.1101/gad.1876110.
- Goldberg AD, Banaszynski LA, Noh K-M, Lewis PW, Elsaesser SJ, et al. (2010) Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions. *Cell* 140: 678–691. doi:10.1016/j.cell.2010.01.003.
- Wollmann H, Holec S, Alden K, Clarke ND, Jacques P-É, et al. (2012) Dynamic deposition of histone variant H3.3 accompanies developmental remodeling of the Arabidopsis transcriptome. *PLoS Genetics* 8: e1002658. doi:10.1371/journal.pgen.1002658.
- Talbert PB, Henikoff S (2010) Histone variants — ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol* 11: 264–275. doi:10.1038/nrm2861.
- West MH, Bonner WM (1980) Histone 2A, a heteromorphous family of eight protein species. *Biochemistry* 19: 3238–3245.
- Sauvageau M, Sauvageau G (2010) Polycomb Group Proteins: Multi-Faceted Regulators of Somatic Stem Cells and Cancer. *Cell Stem Cell* 7: 299–313. doi:10.1016/j.stem.2010.08.002.
- Rogakou EP, Pilch DR, Orr AH, Ivanova VS, Bonner WM (1998) DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* 273: 5858–5868.
- Kinner A, Wu W, Staudt C, Iliakis G (2008) Gamma-H2AX in recognition and signaling of DNA double-strand breaks in the context of chromatin. *Nucleic Acids Research* 36: 5678–5694. doi:10.1093/nar/gkn550.
- Shechter D, Chitta RK, Xiao A, Shabanowitz J, Hunt DF, et al. (2009) A distinct H2A. X isoform is enriched in *Xenopus laevis* eggs and early embryos and is phosphorylated in the absence of a checkpoint. *Proceedings of the National Academy of Sciences* 106: 749–754.
- Nashun B, Yukawa M, Liu H, Akiyama T, Aoki F (2010) Changes in the nuclear deposition of histone H2A variants during pre-implantation development in mice. *Development* 137: 3785–3794. doi:10.1242/dev.051805.
- Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, et al. (2007) High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* 129: 823–837. doi:10.1016/j.cell.2007.05.009.
- Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S (2008) Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* 456: 125–129. doi:10.1038/nature07324.
- Conerly ML, Teves SS, Diolaiti D, Ulrich M, Eisenman RN, et al. (2010) Changes in H2A.Z occupancy and DNA methylation during B-cell lymphomagenesis. *Genome Research* 20: 1383–1390. doi:10.1101/gr.106542.110.
- Chakravarthy S, Gundimella SKY, Caron C, Perche PY, Pehrson JR, et al. (2005) Structural Characterization of the Histone Variant macroH2A. *Molecular and Cellular Biology* 25: 7616–7624. doi:10.1128/MCB.25.17.7616-7624.2005.
- Sarma K, Reinberg D (2005) Histone variants meet their match. *Nat Rev Mol Cell Biol* 6: 139–149. doi:10.1038/nrm1567.
- Hernández-Muñoz I, Lund AH, Van Der Stoep P, Boutsma E, Muijers I, et al. (2005) Stable X chromosome inactivation involves the PRC1 Polycomb complex and requires histone MACROH2A1 and the CULLIN3/SPOP ubiquitin E3 ligase. *Proceedings of the National Academy of Sciences* 102: 7635–7640.
- Changolkar LN, Singh G, Cui K, Berletch JB, Zhao K, et al. (2010) Genome-Wide Distribution of MacroH2A1 Histone Variants in Mouse Liver Chromatin. *Molecular and Cellular Biology* 30: 5473–5483. doi:10.1128/MCB.00518-10.
- Szenker E, Ray-Gallet D, Almouzni G (2011) The double face of the histone variant H3.3. *Nature Publishing Group* 21: 421–434. doi:10.1038/cr.2011.14.
- Hake SB, Garcia BA, Duncan EM, Kauer M, Dellaire G, et al. (2006) Expression patterns and post-translational modifications associated with mammalian histone H3 variants. *J Biol Chem* 281: 559–568. doi:10.1074/jbc.M509266200.
- Loyola A, Bonaldi T, Roche D, Imhof A, Almouzni G (2006) PTMs on H3 Variants before Chromatin Assembly Potentiate Their Final Epigenetic State. *Molecular Cell* 24: 309–316. doi:10.1016/j.molcel.2006.08.019.
- Jin C, Zang C, Wei G, Cui K, Peng W, et al. (2009) H3.3/H2A.Z double variant-containing nucleosomes mark “nucleosome-free regions” of active promoters and other regulatory regions. *Nat Genet* 41: 941–945. doi:10.1038/ng.409.
- Hake SB, Allis CD (2006) Histone H3 variants and their potential role in indexing mammalian genomes: the “H3 barcode hypothesis”. *Proceedings of the National Academy of Sciences* 103: 6428–6435. doi:10.1073/pnas.0600803103.

28. Stroud H, Otero S, Desvoyes B, Ramirez-Parra E, Jacobsen SE, et al. (2012) Genome-wide analysis of histone H3.1 and H3.3 variants in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* 109: 5370–5375. doi:10.1073/pnas.1203145109.
29. Wollmann H, Holec S, Alden K, Clarke ND, Jacques P-É, et al. (2012) Dynamic Deposition of Histone Variant H3.3 Accompanies Developmental Remodeling of the *Arabidopsis* Transcriptome. *PLoS Genetics* 8: e1002658. doi:10.1371/journal.pgen.1002658.t001.
30. Hardy S, Jacques P-É, Gévy N, Forest A, Fortin M-È, et al. (2009) The Euchromatic and Heterochromatic Landscapes Are Shaped by Antagonizing Effects of Transcription on H2A.Z Deposition. *PLoS Genetics* 5: e1000687. doi:10.1371/journal.pgen.1000687.g005.
31. Seo J, Kim SC, Lee HS, Kim JK, Shon HJ, et al. (2012) Genome-wide profiles of H2AX and -H2AX differentially endogenous and exogenous DNA damage hotspots in human cells. *Nucleic Acids Research* 40: 5965–5974. doi:10.1093/nar/gks287.
32. Jin C, Felsenfeld G (2007) Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes & Development* 21: 1519–1529. doi:10.1101/gad.1547707.
33. Niwa H, Miyazaki J, Smith AG (2000) Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* 24: 372–376. doi:10.1038/74199.
34. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, et al. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448: 553–560. doi:10.1038/nature06008.
35. Chen X, Xu H, Yuan P, Fang F, Huss M, et al. (2008) Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. *Cell* 133: 1106–1117. doi:10.1016/j.cell.2008.04.043.
36. Juven-Gershon T, Hsu J-Y, Theisen JW, Kadonaga JT (2008) The RNA polymerase II core promoter — the gateway to transcription. *Current Opinion in Cell Biology* 20: 253–259. doi:10.1016/j.ccb.2008.03.003.
37. Collas P (2009) Epigenetic states in stem cells. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1790: 900–905. doi:10.1016/j.bbagen.2008.10.006.
38. Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. *Genes & Development* 25: 1010–1022. doi:10.1101/gad.2037511.
39. Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* 13: 484–492. Available: <http://www.nature.com/doifinder/10.1038/nrg3230>.
40. Hallikas O, Palin K, Sinjushina N, Rautiainen R, Partanen J, et al. (2006) Genome-wide Prediction of Mammalian Enhancers Based on Analysis of Transcription-Factor Binding Affinity. *Cell* 124: 47–59. doi:10.1016/j.cell.2005.10.042.
41. Lee BK, Iyer VR (2012) Genome-wide Studies of CCCTC-binding Factor (CTCF) and Cohesin Provide Insight into Chromatin Structure and Regulation. *Journal of Biological Chemistry* 287: 30906–30913. doi:10.1074/jbc.R111.324962.
42. Phillips-Cremins JE, Corces VG (2013) Chromatin Insulators: Linking Genome Organization to Cellular Function. *Molecular Cell* 50: 461–474. doi:10.1016/j.molcel.2013.04.018.
43. Handoko L, Xu H, Li G, Ngan CY, Chew E, et al. (2011) CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet* 43: 630–638. doi:10.1038/ng.857.
44. Liu Z, Scannell DR, Eisen MB, Tjian R (2011) Control of Embryonic Stem Cell Lineage Commitment by Core Promoter Factor, TAF3. *Cell* 146: 720–731. doi:10.1016/j.cell.2011.08.005.
45. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, et al. (2012) A map of the cis-regulatory sequences in the mouse genome. *Nature* 488: 116–120. doi:10.1038/nature11243.
46. Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, et al. (2011) CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479: 74–79. doi:10.1038/nature10442.
47. Illingworth R, Kerr A, DeSousa D, Jorgensen H, Ellis P, et al. (2008) A Novel CpG Island Set Identifies Tissue-Specific Methylation at Developmental Gene Loci. *Plos Biol* 6: e22. doi:10.1371/journal.pbio.0060022.st004.
48. Hu G, Cui K, Northrup D, Liu C, Wang C, et al. (2013) H2A.Z Facilitates Access of Active and Repressive Complexes to Chromatin in Embryonic Stem Cell Self-Renewal and Differentiation. *Cell Stem Cell* 12: 180–192. doi:10.1016/j.stem.2012.11.003.
49. Guillemette B, Bataille AR, Gévy N, Adam M, Blanchette M, et al. (2005) Variant Histone H2A.Z Is Globally Localized to the Promoters of Inactive Yeast Genes and Regulates Nucleosome Positioning. *Plos Biol* 3: e384. doi:10.1371/journal.pbio.0030384.st008.
50. Tolstorukov MY, Goldman JA, Gilbert C, Ogryzko V, Kingston RE, et al. (2012) Histone variant H2A.Bbd is associated with active transcription and mRNA processing in human cells. *Molecular Cell* 47: 596–607. doi:10.1016/j.molcel.2012.06.011.
51. Lehnertz B, Ueda Y, Derijck AAHA, Braunschweig U, Perez-Burgos L, et al. (2003) Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Current Biology* 13: 1192–1200.
52. Dimitri P (1997) Constitutive heterochromatin and transposable elements in *Drosophila melanogaster*. *Genetica* 100: 85–93.
53. Almouzni G, Probst AV (2011) Heterochromatin maintenance and establishment: Lessons from the mouse pericentromere. *nucleus* 2: 332–338. doi:10.4161/nucl.2.5.17707.
54. Peters AHFM, Kubicek S, Mechtler K, O'Sullivan RJ, Derijck AAHA, et al. (2003) Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Molecular Cell* 12: 1577–1589.
55. Santos F, Peters AH, Otte AP, Reik W, Dean W (2005) Dynamic chromatin modifications characterise the first cell cycle in mouse embryos. *Developmental Biology* 280: 225–236. doi:10.1016/j.ydbio.2005.01.025.
56. Lee C-K, Shibata Y, Rao B, Strahl BD, Lieb JD (2004) Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat Genet* 36: 900–905. doi:10.1038/ng1400.
57. Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, et al. (2005) Genome-wide Map of Nucleosome Acetylation and Methylation in Yeast. *Cell* 122: 517–527. doi:10.1016/j.cell.2005.06.026.
58. Tirosh I, Barkai N (2008) Two strategies for gene regulation by promoter nucleosomes. *Genome Research* 18: 1084–1091. doi:10.1101/gr.076059.108.
59. Fenouil R, Cauchy P, Koch F, Descostes N, Cabeza JZ, et al. (2012) CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Research* 22: 2399–2408. doi:10.1101/gr.138776.112.
60. Teif VB, Vainshtein Y, Caudron-Herger M, Mallm J-P, Marth C, et al. (2012) Genome-wide nucleosome positioning during embryonic stem cell development. *Nat Struct Mol Biol* 19: 1185–1192. doi:10.1038/nsmb.2419.
61. Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, et al. (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* 464: 1082–1086. Available: <http://www.nature.com/doifinder/10.1038/nature08924>.
62. Clouaire T, Webb S, Skene P, Illingworth R, Kerr A, et al. (2012) Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes & Development* 26: 1714–1728. doi:10.1101/gad.194209.112.
63. Ramirez-Carrozzi VR, Braas D, Bhatt DM, Cheng CS, Hong C, et al. (2009) A Unifying Model for the Selective Regulation of Inducible Transcription by CpG Islands and Nucleosome Remodeling. *Cell* 138: 114–128. doi:10.1016/j.cell.2009.04.020.
64. Huang C, Zhang Z, Xu M, Li Y, Li Z, et al. (2013) H3.3-H4 Tetramer Splitting Events Feature Cell-Type Specific Enhancers. *PLoS Genetics* 9: e1003558. doi:10.1371/journal.pgen.1003558.s005.
65. West AG, Gaszner M, Felsenfeld G (2002) Insulators: many functions, many mechanisms. *Genes & Development* 16: 271–288. doi:10.1101/gad.954702.
66. Crawford GE, Holt IE, Mullikin JC, Tai D, Blakesley R, et al. (2004) Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. *Proceedings of the National Academy of Sciences* 101: 992–997. doi:10.1073/pnas.0307540100.
67. Sabo PJ, Humbert R, Hawrylycz M, Wallace JC, Dorschner MO, et al. (2004) Genome-wide identification of DNaseI hypersensitive sites using active chromatin sequence libraries. *Proceedings of the National Academy of Sciences* 101: 4537–4542. doi:10.1073/pnas.0400678101.
68. Wang Y-M, Zhou P, Wang L-Y, Li Z-H, Zhang Y-N, et al. (2012) Correlation Between DNase I Hypersensitive Site Distribution and Gene Expression in HeLa S3 Cells. *PLoS ONE* 7: e42414. doi:10.1371/journal.pone.0042414.s011.
69. Smolle M, Venkatesh S, Gogol MM, Li H, Zhang Y, et al. (2012) Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nat Struct Mol Biol* 19: 884–892. doi:10.1038/nsmb.2312.
70. Venkatesh S, Smolle M, Li H, Gogol MM, Saint M, et al. (2012) Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature* 489: 452–455. doi:10.1038/nature11326.
71. Dion MF, Kaplan T, Kim M, Buratowski S, Friedman N, et al. (2007) Dynamics of Replication-Independent Histone Turnover in Budding Yeast. *Science* 315: 1405–1408. doi:10.1126/science.1134053.
72. Jamai A, Imoberdorf RM, Strubin M (2007) Continuous Histone H2B and Transcription-Dependent Histone H3 Exchange in Yeast Cells outside of Replication. *Molecular Cell* 25: 345–355. doi:10.1016/j.molcel.2007.01.019.
73. Deal RB, Henikoff JG, Henikoff S (2010) Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* 328: 1161–1164. doi:10.1126/science.1186777.
74. Dimitri P, Caizzi R, Giordano E, Carmela Accardo M, Lattanzi G, et al. (2009) Constitutive heterochromatin: a surprising variety of expressed sequences. *Chromosoma* 118: 419–435. doi:10.1007/s00412-009-0211-y.