Identification of H4K20me3- and H3K4me3-associated RNAs using CARIP-Seq expands the transcriptional and epigenetic networks of embryonic stem cells

Jiji T. Kurup and Benjamin L. Kidder

From the Department of Oncology and the Karmanos Cancer Institute, Wayne State University School of Medicine, Detroit, Michigan 48201

Regulation of gene expression is a dynamic process that is governed in part by 1) sequence-specific transcription factors and epigenetic regulators binding to regions of accessible chromatin; 2) post-transcriptional splicing events, where RNA binding proteins bind nascent RNA and facilitate splicing and translation (1–4); and 3) post-transcriptional events, such as modification of histone tails (e.g. lysine methylation) (5–7), which regulates chromatin structure (8) (e.g. packaging of DNA into chromatin) and gene expression by influencing the activity of epigenetic modifiers and the transcriptional state of the underlying DNA sequence. Combined, these processes are important for controlling gene expression networks that promote self-renewal or differentiation of embryonic stem (ES)2 cells. The advent of technologies to profile RNA–protein interactions has demonstrated that DNA-binding factors associate with RNAs to regulate transcriptional and post-transcriptional processes (9–11). Although the functional role of RNA in chromatin organization has been a focus of previous work (12, 13), recent studies have uncovered interactions between RNA and proteins, including long noncoding RNAs (lncRNAs), which influence chromatin structure by binding and regulating activity of chromatin modifying enzymes (14–18). Xist is an example of an lncRNA that facilitates chromatin organization by interacting with epigenetic repressors to inactivate one X chromosome in female mammalian cells (19). Recent work has also identified RNA binding by proteins associated with chromatin (18), suggesting that a broad set of RNAs may interact with chromatin constituents. Although these studies provide insight into RNA–protein interactions, it is unclear whether RNAs associate with post-translationally modified histones globally. Identifying interactions between RNAs and histone modifications may provide insight into mechanisms of regulation of gene expression in ES cells. Association of RNAs with histone modifications may regulate genome structure and expression patterns by mediating communication networks between the genome and chromatin-associated proteins. Also, interactions...
between RNAs and chromatin may demarcate euchromatin or heterochromatin domains to facilitate maintenance of self-renewal or differentiation, thus influencing cell fate decisions. For example, chromatin compaction and gene repression may be regulated by heterochromatin-associated transcripts, whereas euchromatin formation and gene activation may be regulated by euchromatin-associated transcripts. In addition, association of RNAs with multiple histone modifications may also regulate gene activity.

Post-translational modification of histones is thought to regulate chromatin structure (20) in part by marking active and inactive gene regulatory networks and by influencing the transcriptional state of the underlying DNA sequence. H4K20 methylation is associated with several cellular processes including transcriptional regulation (21), DNA repair (22, 23), DNA replication (24), chromosome condensation (25), and genome stability (23, 26). H4K20me1 is enriched at active genes (27, 28), whereas H4K20me3 is predominantly associated with heterochromatin regions. However, recent findings suggest that H4K20me3 co-localizes with H3K4me3 at a subset of regions (29, 30), suggesting that H4K20me3 marks transcriptionally dynamic regions in ES cells. In addition, H3K4me3 is enriched at the transcriptional start sites (TSSs) of active genes (28, 31–34), where it serves in part as a platform for RNA polymerase II (RNApolII) binding and gene activation (7, 35, 36).

Here, we identified RNAs associated with H4K20me3- and H3K4me3-marked chromatin on a global scale in ES cells by performing chromatin-associated RNA immunoprecipitation (CARIP) followed by next-generation sequencing (CARIP-Seq). Using this approach, we identified common and distinct sets of H4K20me3- and H3K4me3-associated RNAs involved in a diverse set of biological processes. We also found that H4K20me3 and H3K4me3 RNAs co-localize in promoter and gene body regions of active genes with histone modifications such as H3K36me3 and transcriptional machinery. Moreover, H4K20me3 and H3K4me3 are associated with distinct gene features including transcripts of greater length and exon number relative to unoccupied transcripts. H4K20me3 and H3K4me3-marked chromatin is also associated with processed RNAs (exon transcripts) relative to unspliced pre-mRNA transcripts. Altogether, our results support a model whereby H4K20me3- and H3K4me3-associated RNAs constitute a distinct subnetwork of the ES cell transcriptional repertoire.

**Results**

**Use of CARIP-Seq to identify H4K20me3- and H3K4me3-associated RNAs on a global scale in ES cells**

Native RNA immunoprecipitations (RNA) previously demonstrated that the histone methyltransferase, Suv420h2, which deposits H4K20me3, interacts with IncRNAs to facilitate recruitment to chromatin and condensation (37). However, it is unclear whether H4K20me3 directly interacts with RNAs. Moreover, whereas H4K20me3 has recently been shown to co-localize with activating histone modifications, such as H3K4me3, at a subset of regions (30), it is unknown whether H3K4me3 also interacts with RNAs. Therefore, to identify RNAs associated with H4K20me3 and H3K4me3, we surveyed H4K20me3- and H3K4me3–RNA interactions on a global scale in ES cells using a newly developed method, which we have termed CARIP-Seq (see “Experimental procedures”) (38), by modifying existing ChIP-Seq (39, 40) and RIP (41, 42) protocols. CARIP-Seq employs a formaldehyde cross-linking step to preserve RNA–protein interactions postlysis in a similar manner as RIP-ChIP (41) and RIP-Seq (18). However, although RIP-ChIP and fRIP-Seq utilize a lower concentration of formaldehyde (0.1%), CARIP-Seq utilizes a similar concentration of formaldehyde (1%) as ChIP-Seq. To avoid over-cross-linking, which may lead to lower RNA recovery, we reduced the cross-linking time typically used in ChIP-Seq experiments from 10 to 1 min. We hypothesized that a temporal reduction in cross-linking combined with a reduced number of sonication cycles (8–10) to lyse cells and shear chromatin, relative to ChIP-Seq (14–18 cycles), would preserve RNA–protein interactions at condensed heterochromatin regions, which are more refractory to fragmentation by sonication relative to open euchromatic regions such as promoters and enhancers. Following sonication, sheared chromatin is incubated with a ChIP-qualified antibody and paramagnetic beads and washed, and RNA is subsequently purified. H4K20me3- and H3K4me3-associated RNAs were immunoprecipitated using the newly developed protocol CARIP using ES cells, followed by construction of cDNAs libraries. CARIP-Seq libraries were then generated using a modified RNA-Seq protocol (see “Experimental procedures”) and subjected to next-generation sequencing. To identify RNAs that associate with H4K20me3 and H3K4me3 on a genome-wide scale in ES cells, we performed CARIP-Seq using 4 million ES cells (see “Experimental procedures”). Spatial Clustering for Identification of ChIP-Enriched Regions (SICER) software (43) was used to identify CARIP-enriched islands (FDR < 0.05; fold change > 1.5; see “Experimental procedures”) relative to input RNA. Using this approach, we identified 6,105 H4K20me3 and 5,863 H3K4me3 CARIP-Seq peaks in the mouse ES cell genome. To test for specific enrichment, we compared H4K20me3 and H3K4me3 CARIP-Seq data to IgG CARIP-Seq data. These results showed that H4K20me3 and H3K4me3 antibodies enriched for bulk RNA relative to IgG control (Fig. S1, A and B). We observed greater enrichment of RNA in H4K20me3 CARIP samples at H4K20me3 CARIP-defined peaks, whereas we observed increased enrichment of RNA in H3K4me3 CARIP samples at H3K4me3 CARIP-defined peaks (Fig. S1, A and B). A direct comparison of RNA-Seq data with H4K20me3 and H3K4me3 CARIP-Seq data demonstrated that H4K20me3 and H3K4me3 CARIP-Seq peaks overlap a subset of expressed transcripts in ES cells (Fig. 1A). The CARIP-Seq signal demonstrates that although the RNA is associated with H4K20me3 or H3K4me3, it does not necessarily indicate that it interacts with the locus shown. This is evident when comparing CARIP-Seq signals to H4K20me3 or H3K4me3 ChIP-Seq peaks (Fig. 1A). Moreover, scatter plots demonstrated that H4K20me3 and H3K4me3 CARIP-Seq samples exhibit distinct signals (Fig. 1B). A comparison with RNA-Seq data revealed that H4K20me3 and H3K4me3 CARIP-Seq samples are enriched at a subset of ESC-expressed genes (Fig. 1C). Annotation of H4K20me3 and H3K4me3 CARIP-Seq enriched islands using HOMER software (44) revealed that
H4K20me3- and H3K4me3-associated RNAs in ES cells

A

B

C

D

E

F

G

H

I

J

K
H4K20me3 and H3K4me3 CARIP peaks are predominantly enriched nearby intronic/gene body regions (61 and 67%) (±2 kb of TSS) and exonic regions (10 and 5%), relative to random genomic sequences of comparable size and frequency, which were used as controls. Moreover, although H4K20me3 and H3K4me3 CARIP peaks are also located in intergenic regions (18 and 19%), their enrichment in these regions is lower than random genomic sequences (Fig. 1D). A region to gene association analysis also showed that H4K20me3 and H3K4me3 CARIP-Seq peaks are largely enriched in gene body regions (downstream of TSS regions) (Fig. 1E). These findings indicate that H4K20me3 and H3K4me3 CARIP-Seq signals originate mainly from genic regions.

We also evaluated the relative enrichment of activating (H3K4me3 and H3K36me3) and repressive (H4K20me3 and H3K9me3) histone modifications and transcriptional machinery constituents at genes whose transcripts interact with H4K20me3 or H3K4me3, referred to hereafter as H4K20me3 and H3K4me3 CARIP-Seq peaks. H3K4me3 and RNAPII are enriched at promoter regions of actively transcribed genes in ES cells (45), H3K36me3 and RNAPII Ser2P (46) are enriched in gene body regions of actively transcribed genes, and H4K20me3 and H3K9me3 are enriched at inactive genes or nongenic regions (47). An evaluation of signals at H4K20me3 or H3K4me3 CARIP-Seq peaks showed that densities of features enriched in gene body regions, including H3K36me3 and RNAPII Ser2P, were higher relative to densities of histone modifications enriched at heterochromatin regions (H4K20me3 and H3K9me3) (Fig. 1F). Moreover, densities of features enriched at promoter regions, including H3K4me3 and RNAPII, exhibited lower densities relative to gene body features, but elevated levels relative to heterochromatin features (Fig. 1F). In contrast, an evaluation of densities at H4K20me3 ChIP-Seq peaks revealed decreased levels of H4K20me3 and H3K4me3 CARIP-Seq signals, H3K36me3, RNAPII, RNAPII Ser2P, and H3K4me3 ChIP-Seq levels relative to H4K20me3 and H3K9me3 ChIP-Seq levels (Fig. 1G, left panel). In addition, analysis of densities at H3K4me3 ChIP-Seq peaks revealed similar levels of H4K20me3 and H3K4me3 CARIP-Seq signals relative to H3K36me3 and RNAPII Ser2P but lower levels relative to H3K4me3 and RNAPII (Fig. 1G, right panel). Moreover, heat maps of H4K20me3 and H3K4me3 CARIP-Seq densities at TSSs to transcription end sites (TESs) also revealed greater enrichment of H3K4me3 CARIP signals relative to H4K20me3 CARIP signals in gene body regions of active genes (Fig. 1H). To evaluate the expression state of H4K20me3- and H3K4me3-associated RNAs in ES cells, we compared these transcripts with RNA-Seq transcriptome data from ES cells and embryoid body (EB) differentiated cells (48) using gene set enrichment analysis (49). Results from these analyses demonstrate that expression of H3K4me3-associated RNAs is enriched in ES cells relative to EB differentiated cells (Fig. 1I). These results also show that although expression of H4K20me3-associated RNAs is mainly enriched in ES cells, some are enriched in differentiated cells (Fig. 1I). To further investigate the distribution of expression of H4K20me3- and H3K4me3-associated transcripts, we evaluated their expression state relative to all genes and ES cell–expressed genes (RPKM > 1). Although these results confirm that most H4K20me3- and H3K4me3-associated transcripts are expressed in ES cells, these results also demonstrate that H4K20me3- and H3K4me3-associated transcripts exhibit a similar distribution of expression relative to all ES cell–expressed genes (Fig. 1I). In addition, these results demonstrate that CARIP-Seq captures a dynamic range of abundantly expressed transcripts and RNAs with a lower expression level.

We also evaluated the breadth of H4K20me3 and H3K4me3 CARIP-Seq peaks relative to other histone modifications, including H4K20me3 and H3K36me3 and elongating RNAPII Ser2P. These results reveal that the distribution of breadth is overall similar between H4K20me3 and H3K4me3 CARIP-Seq peaks and the H4K20me3 and H3K36me3 histone modifications and transcriptional machinery component, RNAPII Ser2P (Fig. 1K).

**Transcriptome analysis of H4K20me3- and H3K4me3-associated RNAs**

To determine whether H4K20me3- and H3K4me3-associated RNAs represent a distinct subset of the ES cell transcriptional repertoire that is expressed in pluripotent ES cells or lineage-specific cells, we compared the RNA-Seq transcriptomes of undifferentiated ES cells, epiblast stem cells (EpiSCs), trophoblast (TS) stem cells, EB differentiated ES cells, differentiated trophoblast cells, and mouse embryonic fibroblasts (MEFs) with H4K20me3- and H3K4me3-associated RNAs. ES cells, EpiSCs, and TS cells can be derived from blastocyst-stage embryos. ES cells are fully pluripotent stem cells, EpiSCs express OCT4 and retain pluripotent characteristics, and TS cells represent a multipotent stem cell. Hierarchical clustering analysis (HCA) followed by K-means clustering (KMC) to further refine major patterns of gene expression variability demonstrated that although overall H4K20me3 and

---

**Figure 1. CARIP-Seq reveals genome-wide association of RNAs with H4K20me3- and H3K4me3-marked chromatin in ES cells.** A, circular genome map showing distributions of H4K20me3 and H3K4me3 CARIP-Seq, and RNA-Seq densities (ESC RNA-Seq: black, inner histogram; H4K20me3 CARIP-Seq: gray, middle histogram), and H4K20me3 and H3K4me3 CARIP-Seq peaks determined using SICER software (H3K4me3 CARIP-Seq peaks: black outer ring; and H4K20me3 CARIP-Seq peaks: red ring). B and C, scatter plots of H4K20me3 and H3K4me3 CARIP-Seq densities (log2 RPKM 10−2) at H4K20me3 (B, left panel) and H3K4me3 (B, right panel) peaks and at all RefSeq genes (C). Log2 adjusted RPKM (10−2) densities are plotted. D, annotation of H4K20me3 and H3K4me3 CARIP-Seq regions (red) relative to random genomic regions (gray) using Homer (44) software. E, GREAT software (73) was used to perform region-gene association analysis showing distance of H3K4me3 (top panel) and H4K20me3 (bottom panel) CARIP-Seq peaks (regions) to TSSs (gene) in ES cells. F and G, box plots of H3K4me3 and H4K20me3 CARIP-Seq, H3K36me3, RNAPII Ser2P, H3K4me3, RNAPII, H4K20me3, and H3K9me3 ChiP-Seq densities at H4K20me3 (F, left) and H3K4me3 (F, right) CARIP-Seq peaks (left panel) and H4K20me3 (G, left panel) and H3K4me3 (G, right panel) ChiP-Seq peaks. H, heat map of H4K20me3 and H3K4me3 CARIP-Seq density in gene body regions (TSS to TES). I, gene set enrichment analysis of genes associated with H3K4me3 and H4K20me3 CARIP-Seq peaks, H3K4me3-only and H4K20me3-only CARIP-Seq peaks, and H4K20me3 and H3K4me3 overlapping CARIP-Seq peaks in undifferentiated ES cells relative to differentiated EBs. J, box plot of cell RNA-Seq expression data for mm9 mouse genes, ESC-expressed genes (RPKM > 1), H3K4me3-associated transcripts, and H4K20me3-associated transcripts. Genes were divided into quartiles based on their expression level in ES cells. K, distribution of H4K20me3 and H3K4me3 CARIP, H3K36me3, RNAPII-Ser2P, and H4K20me3 ChiP-Seq SICER-defined peaks in ES cells.
H4K20me3- and H3K4me3-associated RNAs in ES cells

H3K4me3 CARIP-Seq samples clustered away from undifferentiated or differentiated cells (Fig. 2A), KMC identified clusters of H4K20me3- and H3K4me3-associated transcripts whose expression in enriched pluripotent ES cells and other blastocyst-derived stem cells. Furthermore, annotation of H4K20me3- and H3K4me3-associated RNAs revealed clusters of genes bound by polycomb factors (SUZ12, JARID2, RNF2, EZH2, EED, and PHC1) and core
transcriptional circuitry (OCT4/POU5F1, NANOG, and SOX2) (Fig. 2A, right panels).

Next, we identified transcripts that are differentially associated with H4K20me3 or H3K4me3 (RPKM > 1, fold change > 1.5) and clustered these genes using HCA (Fig. 2B). DAVID (50) was used to functionally annotate transcripts associated with H4K20me3 and H3K4me3. Using this approach, we identified gene ontology (GO) terms enriched in H4K20me3-associated transcripts such as cell redox homeostasis, regulation of cell motility/migration, placental development, and epithelial cell differentiation (Fig. 2C, top panel) and GO terms enriched in H3K4me3-associated transcripts such as negative regulation of skeletal muscle differentiation, metabolic processes (e.g. macromolecular, ncRNA), and RNA processing (Fig. 2C, bottom panel).

GREAT gene ontology software was used to functionally annotate transcripts associated with H4K20me3 only and with H3K4me3 only. Results from these analyses reveal enrichment of multiple GO terms, including biological processes enriched in H3K4me3-associated transcripts such as chromosome organization, chromatin modification, histone modification, ncRNA processing, and DNA repair, whereas H4K20me3-associated transcripts were enriched with biological process GO terms such as RNA processing, noncoding RNA (ncRNA) metabolic processing, DNA repair, and histone modification. In addition, molecular function GO terms enriched in H3K4me3-associated transcripts include Ran GTPase binding, histone binding, and demethylase activity (Fig. 2D, top right panel), whereas H4K20me3-associated transcripts were enriched with GO terms such as histone acetyltransferase activity, repressing transcription factor binding (Fig. 2D, bottom right panel).

Relationship between H4K20me3- and H3K4me3-associated transcripts and local chromatin

Next, we investigated the overlap between H4K20me3 or H3K4me3 CARIP-Seq peaks and transcription factor (TF) or epigenetic regulator binding, or occupancy of histone modifications. To this end, we evaluated the percentage of overlap between H4K20me3 or H3K4me3 CARIP-Seq peaks and random genomic sequences of comparable size and frequency, which were used as controls, with a compendium of TF and epigenetic regulator (e.g. histone methyltransferase and histone demethylase) and histone modification ChIP-Seq peaks. Some of the features are positively correlated with promoter or gene body regions, respectively. Our results demonstrate that H4K20me3 and H3K4me3 CARIP-Seq peaks highly overlap binding sites of components of the transcriptional machinery such as RNAPII, initiated RNAPII (Ser5P), elongating RNAPII (Ser2P), E2f1, H3K36me3, histone modifications enriched in promoter (H3K4me3, H3K4me2) and enhancer (H3K27ac) regions, histone-modifying enzymes including KDM5B, and multiple transcription factors including ESRRB, KLF4, MYC, STAT3, and OCT4 (Fig. 2A). However, we observed a greater overlap of ZFX, NANOG, N-MYC, MYC, and KLF4 binding sites with H3K4me3 CARIP-Seq peaks relative to H4K20me3 CARIP-Seq peaks (Fig. 2A). These results further suggest that H3K4me3-associated transcripts are aligned with transcriptionally active features in ES cells.

Venn diagrams further show that H4K20me3 and H3K4me3 CARIP-Seq peaks overlap highly with RNAPII-Ser5P, RNAPII-Ser2P, H3K36me3, H3K4me3, H3K4me2, and H3K27ac ChIP-Seq peaks (Fig. 3, B and C). We also found that 44% of H3K4me3 CARIP-Seq peaks overlap with H4K20me3 CARIP-Seq peaks, and 46% of H4K20me3 CARIP-Seq peaks overlap with H3K4me3 CARIP-Seq peaks (Fig. 3B). These results suggest that H4K20me3 and H3K4me3 associate with common and distinct sets of transcripts. This finding is consistent with a previous report demonstrating that H4K20me3 and H3K4me3 co-localize at a subset of regions in ES cells (30). Heat maps also show enrichment of transcriptional and epigenetic regulator binding (RNAPII, RNAPII-Ser5P, RNAPIII-Ser2P, and KDM5B), histone modification occupancy (H3K36me3, H3K4me3, and H3K4me2), and TF binding (MYC, E2f1, OCT4, ZFX, KLF4, SOX2, NANOG, ESRRB, and STAT3) at H4K20me3 (Fig. 3D) and H3K4me3 CARIP-Seq peaks (Fig. 3E). In contrast, the density of repressive histone modifications such as H3K27me3, H4K20me3, and H3K9me3 was lower at these regions (Fig. 3, D and E). Together, these results show that chromatin constituents associated with active transcription are enriched at genic regions where H4K20me3- or H3K4me3-associated RNAs are transcribed.

We then evaluated whether H4K20me3 or H3K4me3 exhibits sequence composition binding preference in transcripts. To this end, we searched for motifs enriched in transcript sequences that associate with H4K20me3 or H3K4me3 using MEME-ChIP (51). Using this approach, we discovered several statistically significant motifs in a transcript-wide search, including motifs resembling STAT3 and MYC DNA-binding motifs in H4K20me3-associated transcripts, and sequences resembling NANOG DNA-binding motifs in H3K4me3-associated transcripts (Fig. 3, F and G) enriched in transcript sequences that associate with H4K20me3 and H3K4me3.

The enrichment of H3K4me3, H3K36me3, and RNAPIII-Ser2P in promoter and regions spanning gene bodies at genes where H4K20me3- and H3K4me3-associated RNAs are transcribed is exemplified at the ES cell enriched gene, Zfp42 (Fig. 4A). In this case, H3K4me3 is enriched nearby the CARIP-Seq peaks, suggesting that the H3K4me3-Zfp42-associated tran-
H4K20me3- and H3K4me3-associated RNAs in ES cells

Figure 3. Association of active histone modifications and transcriptional machinery at genes where H4K20me3- and H3K4me3-associated RNAs are transcribed. A, heat map showing the percentage of overlap of H4K20me3 and H3K4me3 CARIP-Seq peaks and random genomic sequences with post-translational modifications of histones, RNA transcriptional machinery (e.g. RNApol, RNAPII, RNAPII-Ser5P; elongating RNAPII, RNAPII-Ser2P), transcription factors, and epigenetic regulators. B and C, Venn diagrams showing overlap between H3K4me3 (B) or H4K20me3 (C) CARIP-Seq peaks and histone modifications and transcriptional machinery components. D and E, heat maps of H4K20me3 (D) and H3K4me3 (E) CARIP-Seq and ChIP-Seq densities around H4K20me3 or H3K4me3 CARIP-Seq peaks. F and G, enriched DNA-binding motifs in H4K20me3 (F) or H3K4me3 (G) CARIP-Seq peak regions identified using MEME-ChIP software.

script may interact directly with the locus shown. However, H4K20me3 is enriched downstream of the CARIP-Seq peaks, suggesting that the H4K20me3-Zfp42-associated transcript does not interact with the locus shown. At another locus, H4K20me3 and H3K4me3 ChIP-Seq signals overlap with H4K20me3 and H3K4me3 CARIP-Seq signals in promoter and gene body regions (Fig. 4B). In addition, we observed H4K20me3-only (Fig. 4C) and H3K4me3-only associated transcripts (Fig. 4D). These results suggest that H4K20me3 and H3K4me3 CARIP-Seq signals are enriched at various genic
features including promoters, exons, introns, and gene body regions.

**H4K20me3 and H3K4me3 associate with various features of transcripts**

An abundance of recent studies have suggested that chromatin and co-transcriptional splicing are interconnected (46, 52–58). To investigate whether H4K20me3 or H3K4me3 has a binding preference for certain structural features of transcripts, we computed the density of H4K20me3 and H3K4me3 CARIP-Seq signals in exon, intron, and gene body regions. Specifically, we compared the density of H4K20me3 and H3K4me3 CARIP-Seq signals at exon versus intron regions (Fig. 5A), exon versus gene body regions (Fig. 5B), and intron versus gene body regions (Fig. 5C). These comparisons revealed that both H4K20me3 and H3K4me3 CARIP-Seq signals are greater in exons relative to introns, higher in exons relative to gene body regions, and higher in gene body regions relative to introns.

We next investigated properties of RNA that determine H4K20me3 and H3K4me3 binding. Previous work demonstrated that histone-modifying enzymes, such as EZH2 and HDAC1, exhibit greater affinity for longer transcripts (18, 59, 60), whereas other chromatin constituents such as DNMT1, CHD4, and CTCF bind to shorter genes (18). To evaluate the preference for length of transcript binding by H4K20me3, we compared the length of gene isoforms with or without H4K20me3 binding. We found that H4K20me3 and H3K4me3 exhibit a binding preference for longer genes (Fig. 5D). We further investigated a correlation between expression level and transcript length for all genes in the mouse genome, genes expressed in ES cells, and H4K20me3- and H3K4me3-associated transcripts. The results from these analyses reveal that ESC-enriched genes with a longer length were expressed at a slightly lower level than genes with a shorter length (Fig. S1C). A similar trend was observed for H4K20me3- and H3K4me3-associated RNAs in ES cells.
H4K20me3- and H3K4me3-associated RNAs in ES cells

F

Length vs Density

E

H4K20me3 CARIP

P<2.2e-16

H3K4me3 CARIP

P<2.2e-16

H4K20me3 CARIP

H3K4me3 CARIP

H4K20me3 CARIP

H3K4me3 CARIP

R2 = 0.23

All genes

H4K20me3 CARIP

Intron (log2 rpkms)

H3K4me3 CARIP

Intron (log2 rpkms)

R2 = 0.72

All genes

H4K20me3 CARIP

Gene body (log2 rpkms)

H3K4me3 CARIP

Gene body (log2 rpkms)

R2 = 0.60

All genes

H4K20me3 CARIP

Intron (log2 rpkms)

H3K4me3 CARIP

Intron (log2 rpkms)

R2 = 0.59

All genes

H4K20me3 CARIP

Gene body (log2 rpkms)

H3K4me3 CARIP

Gene body (log2 rpkms)
associated transcripts, where, in general, H4K20me3- and H3K4me3-associated transcripts with a longer length were expressed at a slightly lower level than transcripts with a shorter length (Fig. S1C). Because longer genes usually have a greater number of exons, we next asked whether the RNA length preference for H4K20me3 and H3K4me3 associations is due to a greater number of exons in the occupied transcripts. These results demonstrated that H4K20me3 and H3K4me3 preferentially associate with transcripts with a greater number of exons relative to transcripts that do not associate with H4K20me3 or H3K4me3 (Fig. 5E).

To distinguish the roles for exon number and transcript length in H4K20me3- and H3K4me3-RNA binding preference, we performed a regression for transcript length and H4K20me3 or H3K4me3 CARIP-Seq densities for genes with varying numbers of exons. For genes with a greater number of exons, we observed a positive correlation between H4K20me3- and H3K4me3-RNA-binding signals and transcript length (Fig. 5F). In contrast, for genes with few exons, transcript length and H4K20me3 and H3K4me3-RNA-binding signals were negatively correlated (Fig. 5F). Overall, these results suggest that binding of H4K20me3 and H3K4me3 to RNA is influenced by the structural properties of RNA such as exon number and transcript length. Although our results are in agreement with previous results, which demonstrate that the PRC2 subunit EZH2 exhibits greater binding to longer transcripts, H4K20me3 and H3K4me3 also bind to shorter RNAs with fewer numbers of exons.

**H4K20me3 and H3K4me3 associate with noncoding RNAs**

Previous work has shown that lncRNAs interact with the H4K20me3 methyltransferase, Suv420h2, where lncRNAs recruit Suv420h2 to chromatin (37). In addition, the H3K4 methyltransferase LML1 has been shown to interact with ncRNAs (61), and WDR5, a component of the complex that mediates H3K4me3, has been shown to interact with the lncRNA HOTTIP (62). Although our results described above demonstrate that H4K20me3 and H3K4me3 interact with various transcripts in ES cells, to investigate whether H4K20me3 and H3K4me3 binding preference for certain structural features of transcripts extends to ncRNAs, we computed the density of H4K20me3 and H3K4me3 CARIP-Seq signals in exon, intron, and gene body regions as described above. To this end, we evaluated the densities of H4K20me3 and H3K4me3 CARIP-Seq signals at exon versus introns (Fig. 6A), exon versus gene body regions (Fig. 6B), and intron versus gene body regions (Fig. 6C) of ncRNAs. In a similar manner as described above for all genes, H4K20me3 and H3K4me3 CARIP-Seq signals are greater in exons relative to introns, similar between exons and gene body regions, and similar between introns and gene body regions of ncRNAs.

In addition, we investigated properties of ncRNAs that determine H4K20me3 and H3K4me3 binding. We evaluated the preference for length of transcript binding by H4K20me3 or H3K4me3 by comparing the length of ncRNAs with or without an H4K20me3 or H3K4me3 association. Using this approach, we discovered that H4K20me3 and H3K4me3 have a preference for binding ncRNAs of longer length (Fig. 6D). We also investigated whether ncRNA length preferences for H4K20me3 or H3K4me3 associations is related to a greater number of exons in the bound ncRNAs. Our results show that H4K20me3 and H3K4me3 preferentially associated with ncRNAs with a greater number of exons relative to unoccupied ncRNAs (Fig. 6E).

We also evaluated the roles for exon number and transcript length in determining H4K20me3- or H3K4me3-ncRNA binding preferences. To this end, we performed a regression for transcript length and H4K20me3 and H3K4me3 CARIP-Seq densities for ncRNAs with varying numbers of exons. In contrast to all H4K20me3–RNA associations, for genes with a low or high number of exons, we observed a negative correlation between H4K20me3–ncRNA-binding signals and transcript length (Fig. 6F, left panel). In addition, for H3K4me3–RNA associations, for genes with a low number of exons, we observed a positive correlation between H4K20me3–ncRNA-binding signals and transcript length (Fig. 6F, right panel). Enrichment of H4K20me3 and H3K4me3 CARIP-Seq signals in exon and intron regions of noncoding genes is exemplified at the lncRNA Xist (Fig. 6G). Altogether, these results further suggest that associations between H4K20me3 or H3K4me3 and RNA are affected by structural properties of RNAs including exon number and transcript length.

**Discussion**

Here, we have utilized a novel protocol, CARIP-Seq, to identify RNAs associated with the histone modifications, H4K20me3 and H3K4me3, on a global scale in ES cells. CARIP-Seq is a useful method to interrogate genome-wide association of RNAs with chromatin constituents. Using this approach, we identified a diverse set of mRNAs and ncRNAs that associate with H4K20me3 and H3K4me3. These results describing thousands of RNA co-factors for H4K20me3 and H3K4me3 may represent a feature of heterochromatin and euchromatin regulation in ES cells.

**Figure 5. H4K20me3 and H3K4me3 associate with specific transcript features.** H4K20me3- and H3K4me3-marked chromatin enriches for exon isoforms relative to unspliced pre-mRNA isoforms. A–C, the scatter plots shows the gene’s RPM contribution to exon versus unspliced pre-mRNA intron isoforms (A), exon versus unspliced pre-mRNA isoforms (gene body; TSS to TES) (B), and unspliced intron isoforms versus full-length unspliced pre-mRNA isoforms (gene body; TSS to TES) (C). D, cumulative distribution function for the transcript length of H4K20me3- or H3K4me3-occupied or unoccupied RNAs. Note that H4K20me3- and H3K4me3-marked chromatin is associated with longer RNAs. p value < 2.2E-16 (Kolmogorov–Smirnov test). E, cumulative distribution function for the number of exons in H4K20me3- or H3K4me3-occupied or unoccupied RNAs. Note that H4K20me3- or H3K4me3-marked chromatin is associated with RNAs with a greater number of exons. p value < 2.2E-16 (Kolmogorov–Smirnov test). F, correlation between transcript length of H4K20me3- or H3K4me3-associated RNAs and density. We regressed density against transcript length for genes with two, four, six, or eight exons. These results demonstrate that the density of H4K20me3- and H3K4me3-associated RNAs correlates with transcript length and exon number.
**H4K20me3- and H3K4me3-associated RNAs in ES cells**

Figure 6. H4K20me3 and H3K4me3 associate with specific noncoding RNA features. H4K20me3- and H3K4me3-marked chromatin enriches for exon isoforms relative to unspliced ncRNA isoforms. A–C, the scatter plots show the gene’s RPKM contribution to exon versus unspliced ncRNA intron isoforms (A), exon versus unspliced ncRNA isoforms (gene body; TSS to TES) (B), and unspliced intron isoforms versus full-length unspliced ncRNA isoforms (gene body; TSS to TES) (C). D, cumulative distribution function for the transcript length of H4K20me3- or H3K4me3-occupied or unoccupied ncRNAs. Note that H4K20me3- and H3K4me3-marked chromatin is associated with longer ncRNAs. *p* value < 2.2E-16 (Kolmogorov–Smirnov test). E, cumulative distribution function for the number of exons in H4K20me3- or H3K4me3-occupied or unoccupied ncRNAs. Note that H4K20me3- and H3K4me3-marked chromatin is associated with ncRNAs with a greater number of exons. *p* value < 2.2E-16 (Kolmogorov–Smirnov test). F, correlation between transcript length of H4K20me3-associated ncRNAs and density. We regressed density against transcript length for genes with two, four, six, or eight exons. These results demonstrate that the density of H4K20me3- and H3K4me3-associated RNAs correlates with transcript length and exon number. G, University of California Santa Cruz browser view of H4K20me3 and H3K4me3 CARIP-Seq signals and RNAPII-Ser2P, H3K36me3, and H3K4me3 ChIP-Seq data at a ncRNA locus.

**Figure 6.** H4K20me3 and H3K4me3 associate with specific noncoding RNA features. H4K20me3- and H3K4me3-marked chromatin enriches for exon isoforms relative to unspliced ncRNA isoforms. A–C, the scatter plots show the gene’s RPKM contribution to exon versus unspliced ncRNA intron isoforms (A), exon versus unspliced ncRNA isoforms (gene body; TSS to TES) (B), and unspliced intron isoforms versus full-length unspliced ncRNA isoforms (gene body; TSS to TES) (C). D, cumulative distribution function for the transcript length of H4K20me3- or H3K4me3-occupied or unoccupied ncRNAs. Note that H4K20me3- and H3K4me3-marked chromatin is associated with longer ncRNAs. *p* value < 2.2E-16 (Kolmogorov–Smirnov test). E, cumulative distribution function for the number of exons in H4K20me3- or H3K4me3-occupied or unoccupied ncRNAs. Note that H4K20me3- and H3K4me3-marked chromatin is associated with ncRNAs with a greater number of exons. *p* value < 2.2E-16 (Kolmogorov–Smirnov test). F, correlation between transcript length of H4K20me3-associated ncRNAs and density. We regressed density against transcript length for genes with two, four, six, or eight exons. These results demonstrate that the density of H4K20me3- and H3K4me3-associated RNAs correlates with transcript length and exon number. G, University of California Santa Cruz browser view of H4K20me3 and H3K4me3 CARIP-Seq signals and RNAPII-Ser2P, H3K36me3, and H3K4me3 ChIP-Seq data at a ncRNA locus.
Our results also demonstrate that H4K20me3- and H3K4me3-associated transcripts originate from genes that colocalize with features of active chromatin in proliferator and gene body regions, such as the histone modification, H3K4me3, and transcriptional machinery component RNA Poly II in promoter regions, and elongating RNA Poly II (RNAPII Ser2P) and H3K36me3 in gene body regions of actively transcribed genes. We also show that H4K20me3 and H3K4me3 interact with specific features of RNAs, such as exons and introns. In addition, we demonstrate that H4K20me3 and H3K4me3 exhibits a preference for binding to RNA transcripts with a greater number of exons. Previous work revealed that histone-modifying enzymes such as the PRC2 component EZH2 and the histone demethylase HDAC1 bind longer transcripts (18, 59, 60), whereas other epigenetic regulators such as the DNA methyltransferase DNMT1, the chromatin remodeler CHD4, and the insulator CTCF bind shorter transcripts (18). Our results are in alignment with the binding preference length of the aforementioned histone modifying enzymes, where we found that H4K20me3 and H3K4me3 exhibit a binding preference for longer transcripts. Combined, these findings suggest that the analyzed histone-modifying enzymes and the H4K20me3 and H3K4me3 histone modification have a binding preference for longer RNA transcripts.

Although the vast majority of previously described RNA and chromatin-associated protein interactions involve ncRNAs, strikingly, we identified more mRNAs associated with H4K20me3 and H3K4me3 and with greater enrichment relative to ncRNAs, implicating a potential novel role for mRNAs as regulators of the genome. We also discovered RNA binding preferences for H4K20me3 and H3K4me3 using a genome-wide transcript analysis and several motifs enriched in H4K20me3- or H3K4me3-associated transcripts. For example, we identified motifs resembling STAT3 and MYC DNA-binding motifs in H4K20me3-associated transcripts and sequences resembling NANOG DNA-binding motifs in H3K4me3-associated transcripts.

Although H4K20me3-marked chromatin is thought to be a repressive histone modification, where it is enriched at pericentric heterochromatin and is involved in repressing transcription of repetitive DNA elements (23, 63, 64), our data overall suggest that H4K20me3 interacts with many transcripts from loci that are actively transcribed in ES cells. Although H4K20me3 is mainly associated with heterochromatin regions, a recent study demonstrated that a subset of H4K20me3-marked regions contain H3K4me3 (30). These previous findings demonstrating that H4K20me3 marks transcriptionally dynamic regions in ES cells is in alignment with results from this study, which show that expression of H4K20me3-associated RNAs is predominantly enriched in undifferentiated ES cells relative to differentiated cells (Fig. 1). H4K20me3- and H3K4me3-associated RNAs may serve as regulators of genome structure and gene expression by facilitating communication between the genome and chromatin-associated proteins. Interactions between H4K20me3 or H3K4me3 and pluripotency or lineage-specific transcripts may also facilitate distinct heterochromatin or euchromatin boundaries during self-renewal and differentiation, and interactions between H4K20me3 or H3K4me3 and ESC-enriched transcripts may also dictate ES cell fate decisions by supporting a positive feedback loop by communicating external LIF signals and an intrinsic OCT4-centric transcriptional network to the H4K20me3-marked genome. It is also possible that the RNA–chromatin interactions described in this study occur transiently before the RNAs are translated or that some RNAs associate with chromatin on a more permanent basis and avoid translation, which has previously been suggested (18). However, additional studies are needed to determine the role for chromatin-associated RNAs in regulating chromatin. Moreover, although it is possible that some histone modification–RNA interactions captured using CARIP-Seq may represent co-localization of nascent RNAs with chromatin containing specific histone modifications, because we observed enrichment of CARIP-Seq signals throughout the gene bodies of many genes (Fig. 1H) and not increased enrichment near the 5’ region of genes (TSS), these results argue against indirect pulldown of nascent RNAs as the primary explanation for these interactions. It is also unlikely that CARIP-Seq signals represent association of H4K20me3 RNA with non-chromatin–bound histones, because the sonication protocol used for CARIP-Seq would likely result in oversonication of nonchromatin-bound RNAs, because of increased sensitivity of naked DNA or RNA to shearing. Our results show that RNA fragments used for CARIP-Seq are 500–700 bp (Fig. S1D), a size that is unlikely to be achieved following sonication of nonchromatin-bound RNAs, thus arguing against the possibility of CARIP-Seq signals representing association of H4K20me3 RNA with non-chromatin–bound histones.

In addition, H4K20me3-mediated heterochromatin formation and gene silencing may occur in the presence of H4K20me3-associated transcripts. Along this line, previous work showed that a lncRNA-mediated mechanism guides the H4K20me3-methyltransferase, Suvi200h2, to specific loci to facilitate chromatin compaction (37). Although these results show that interaction between a lncRNA and Suvi200h2 triggers H4K20me3 deposition, it is unclear from these results whether H4K20me3 interacts with RNAs to regulate heterochromatin formation. Additional studies will be required to investigate the role for H4K20me3-associated RNAs in these processes.

Conclusions

In summary, we describe the application of a novel method, CARIP-Seq, to identify H4K20me3 and H3K4me3 associated transcripts on a genome-wide scale in ES cells. These results provide novel insight into the epigenetic and transcriptional repertoires of ES cells.

Experimental procedures

ES cell culture

ES cells were cultured as previously described with minor modifications (45, 46, 48). Briefly, R1 ES cells were cultured on irradiated MEFs in DMEM, 15% FBS medium containing Leukemia inhibitory factor (LIF) (ESGRO) at 37 °C with 5% CO2. For CARIP-Seq experiments, ES cells were cultured on gelatin-coated dishes in ES cell medium containing 1.5 μM CHIR99021 (GSK3 inhibitor) for several passages to remove feeder cells. ES cells were passed by washing with PBS (Gibco), and dissociating
with 0.25% trypsin (Gibco) using serological pipettes (Santa Cruz Biotechnology; sc-200279 and sc-200281).

**CARIP-Seq analysis**

CARIP-Seq experiments were performed by modifying ChIP-Seq (39, 40, 45, 48) and RNA-Seq experiments. ES cells were harvested by washing with room temperature PBS and dissociating using 0.25% trypsin. Following centrifugation, ES cells were resuspended in prewarmed medium (15% FBS (Millipore) and high-glucose DMEM with glutamine (Gibco)) at a concentration of 2 million cells/ml. To cross-link RNA and chromatin, formaldehyde (Sigma) was added to a final concentration of 1%, and the cells were incubated for 1 min at 37 °C in a water bath. Cross-linking was quenched by addition of chilled glycine (Sigma) to a final concentration of 125 mM, and rotated at room temperature for 5 min. The cells were pelleted by centrifuging at 1500 × g for 5 min, the medium/cross-linking solution was aspirated, and the pellet was washed with ice-cold PBS.

The cells were washed a second time with PBS. Fixed cell pellets were flash frozen using liquid nitrogen and stored at −80 °C.

**Sonication**

The cell pellet was thawed on ice, and subsequently resuspended in 1.25 ml of sonication buffer (10 mM Tris-HCl, 1 mM EDTA (TE), pH 8.0 (Quality Biological), 0.1% SDS (Quality Biological), 1 mM phenylmethylsulfonyl fluoride (Sigma), and 1× proteinase inhibitor mixture (Sigma)) made fresh. The cells were sonicated on ice using a Qsonica Q125 sonifier (eight cycles of 30 s on and 30 s off). Following sonication, to solubilize chromatin, 28 μl of 10% SDS, 28 μl of sodium deoxycholate (Sigma), 0.28 ml of 10% Triton X-100(Sigma), and 100 units/ml RNaseOUT was added. To preclear the chromatin, samples were sonicated for 10 min at 10,000 × g at 4 °C. To check the size of chromatin-associated RNA, DNA was removed using DNase (Ambion), and CARIP RNA was run on a 2% agarose gel (Fig. S1D).

The size of RNA fragments used for CARIP-Seq is ~500–700 bp (Fig. S1D).

The monoclonal H3K4me3 antibody (Millipore, 17-614) and the H4K20me3 antibody (Millipore, 07-463) were obtained from EMD Millipore. Sonicated cell extracts equivalent to 5 million cells were used for CARIP assays. To bind antibody to paramagnetic beads, 40 μl of protein G paramagnetic Dynabeads (Invitrogen) were first washed with PBS and incubated in PBS with 4 μg of H4K20me3 antibody or 4 μg of H3K4me3 antibody at room temperature for 40 min while rotating. The beads were washed twice with PBS, and sonicated chromatin extract was added to the beads and incubated overnight at 4 °C while rotating to immunoprecipitate H4K20me3- and H3K4me3-associated factors and RNA. The next day, the beads were washed twice with radioimmune precipitation assay buffer (TE, pH 8.0, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS), twice with radioimmune precipitation assay buffer with 0.3 mM NaCl (Sigma), twice with LiCl buffer (0.5% sodium deoxycholate, 0.5% NP40 (Sigma), 0.25 mM LiCl (Sigma), once with TE and 0.2% Triton X-100 (Sigma), and once with TE. The samples were then resuspended in TE buffer. To de-cross-link, the samples were incubated for 4 h at 65 °C in TE buffer (pH 8.0) with 1% SDS and 5 μl of proteinase K (20 mg/ml; Roche). RNA was first using QIAzol (Qiagen) following the manufacturer’s instructions and eluted in nucleic-free water.

Double-stranded cDNA was generated using a Super-Script double-stranded cDNA synthesis kit (Invitrogen) following the manufacturer’s instructions. cDNA was fragmented using a water bath sonifier (Diagenode Bioruptor Pico) and subjected to library preparation as described previously (29, 45–48).

Briefly, CARIP-enriched cDNA was end-repaired using the End-It DNA end-repair kit (Epigenics), followed by addition of a single A nucleotide and ligation of custom Illumina adapters, as previously described (29, 40, 48). PCR was performed using Phusion High Fidelity PCR master mix. CARIP libraries were sequenced on Illumina HiSeq platforms according to the manufacturer’s protocol. Sequence reads were mapped to the mouse genome (mm9) using bowtie2 (65) with default settings.

CARIP-Seq read enriched regions (peaks) were identified using SICER software (43) with a window size setting of 200 bp, a gap setting of 400 bp, and a FDR setting of 0.001. The SICER compare function was used to compare multiple samples (FDR < 0.001, fold change > 1.5). The read per base per million reads measure was used to quantify the density at genomic regions from ChIP-Seq data sets. We have also applied the Kolmogorov–Smirnov test to obtain p value statistics and compare densities at genomic regions. The RPMK measure (66) was also used to quantify RNA levels of a gene from CARIP-Seq data.

**RNA-Seq analysis**

Poly(A) mRNA was purified using a Dynabeads mRNA purification kit. Double-stranded cDNA was generated using a Super-Script double-stranded cDNA synthesis kit (Invitrogen). cDNA was subjected to library preparation as described above. RNA-Seq libraries were sequenced on an Illumina HiSeq platform according to the manufacturer’s protocol. The RPMK measure (66) was used to quantify the mRNA expression level of a gene from RNA-Seq data. Differentially expressed genes were identified using edgeR (FDR < 0.001; fold change > 1.5; RPMK > 3) (67).

**ChIP-Seq analysis**

ChIP-Seq experiments were performed as previously described with minor modifications (29, 39, 40, 45, 47, 48). The H3K36me3 antibody (ab9050) was obtained from Millipore. Briefly, 1566 mouse ES cells (R1) were harvested and chemically cross-linked with 1% formaldehyde (Sigma) for 5–10 min at 37 °C and subsequently sonicated. Sonicated cell extracts equivalent to 5e6 cells were used for ChIP assays. ChIP-enriched DNA was end-repaired using the End-It DNA end-repair kit (Epigenics), followed by addition of a single A nucleotide, and ligation of custom Illumina adapters. PCR was performed using Phusion high-fidelity PCR master mix. ChIP libraries were sequenced on Illumina HiSeq platforms according to the manufacturer’s protocol. Sequence reads were mapped to the mouse genome (mm9) using bowtie2 (65) with default settings. ChIP-Seq read enriched regions (peaks) were identified relative to control input DNA using SICER software (43) with a window size setting of 200 bp, a gap setting of 400 bp, and a FDR setting of 0.001. The read per base per million reads
measure was used to quantify the density at genomic regions from ChIP-Seq data sets. Additional ChIP-Seq data sets were also analyzed in this manuscript, including RNApII and RNApII Ser2P (GSE94739), H4K20me3 and H3K4me3 (GSE94086), and H3K4me3 (GSE53087) ChIP-Seq data.

**EpiSC culture**

E3 EpiSCs were culture as previously described with minor modifications (68). Briefly, EpiSCs were cultured in EpiSC medium containing knockout DMEM supplemented with 20% knockout serum replacement, L-glutamine, penicillin/streptomycin, and 1 mM sodium pyruvate (Gibco), 25 ng/ml recombinant human FGF4 (1 M Cytoskeleton), and 1 g/ml heparin (1 M Sigma). EpiSCs were passed by washing with PBS and dissociating with 0.05% EDTA. Following passage, EpiSCs were transiently cultured with 10 M ROCK inhibitor (Y27632).

**TS cell culture**

TS cells were cultured as previously described with minor modifications (70–72). Briefly, TSC-BK12 TS cells (70) were cultured on irradiated MEFs in RPMI 1640 medium supplemented with 20% FBS (Gibco), 1 mM sodium pyruvate (Gibco), 100 M M-mercaptoethanol (Chemicon), 2 mM L-glutamine (Gibco), 25 ng/ml recombinant human FGF4 (1 X R&D Systems), and 1 g/ml heparin (1 X Sigma). For RNA-Seq experiments, TS cells were transitioned to dishes containing TS cell medium, 70% iMEF-conditioned medium, and 1.5 X FGF4/heparin and cultured at 37 °C with 5% CO2. TS cells were passed at least three times to remove iMEFs by washing with PBS and dissociating with trypsin using serological pipettes (sc-200279 and sc-200281). For differentiation experiments, TS cells were cultured in feeder-free dishes containing TS cell medium without FGF4 and heparin and fed every 2–3 days for 14 days as previously described (70).

**Author contributions**—J. T. K. and B. L. K. conceptualization; B. L. K. formal analysis; B. L. K. investigation; B. L. K. methodology; B. L. K. writing—original draft; B. L. K. project administration; B. L. K. writing—review and editing.

**Acknowledgment**—We thank the Wayne State University High Performance Computing Grid for computational resources.

**References**

1. Hogan, D. J., Riordan, D. P., Gerber, A. P., Herschlag, D., and Brown, P. O. (2008) Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *Plos Biol.* 6, e255 [CrossRef Medline]

2. Keene, J. D. (2001) Ribonucleoprotein infrastructure regulating the flow of genetic information between the genome and the proteome. *Proc. Natl. Acad. Sci. U.S.A.* 98, 7018–7024 [CrossRef Medline]

3. Keene, J. D., and Tenenbaum, S. A. (2002) Eukaryotic mRNPs may represent posttranscriptional operons. *Mol. Cell* 9, 1161–1167 [CrossRef Medline]

4. Moore, M. J. (2005) From birth to death: the complex lives of eukaryotic mRNAs. *Science* 309, 1514–1518 [CrossRef Medline]

5. Zhang, Y., and Reinberg, D. (2001) Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.* 15, 2343–2360 [CrossRef Medline]

6. Lachner, M., and Jenuwein, T. (2002) The many faces of histone lysine methylation. *Curr. Opin. Cell Biol.* 14, 286–298 [CrossRef Medline]

7. Sims, R. J., 3rd, Nishioika, K., and Reinberg, D. (2003) Histone lysine methylation: a signature for chromatin function. *Trends Genet.* 19, 629–639 [CrossRef Medline]

8. Kouzarides, T. (2007) Chromatin modifications and their function. *Cell* 128, 693–705 [CrossRef Medline]

9. Baltz, A. G., Munschauer, M., Schwahnhaeuser, B., Vasile, A., Murakawa, Y., Schuerer, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., Wyler, E., Bonnaeau, R., Selbach, M., Dieterich, C., and Landthaler, M. (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* 46, 674–690 [CrossRef Medline]

10. Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., Davey, N. E., Humphreys, D. T., Preiss, T., Steinmetz, L. M., Krijgsfeld, J., and Hentze, M. W. (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 149, 1393–1406 [CrossRef Medline]

11. Di Ruscio, A., Ebralidze, A. K., Benoukraf, T., Amabile, G., Goff, L. A., Terragni, I., Figueroa, M. E., De Figueiredo Pontes, L. L., Alberich-Jorda, M., Zhang, P., Wu, M., D’Alò, F., Melnick, A., Leone, G., Ebralidze, K. K., et al. (2013) DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* 503, 371–376 [CrossRef Medline]

12. Bernstein, E., and Allis, C. D. (2005) RNA meets chromatin. *Genes Dev.* 19, 1635–1655 [CrossRef Medline]

13. Johnson, W. L., and Straight, A. F. (2017) RNA-mediated regulation of heterochromatin. *Curr. Opin. Cell Biol.* 46, 102–109 [CrossRef Medline]

14. Tsai, M. C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lao, F., Shi, Y., Segal, E., and Chang, H. Y. (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329, 689–693 [CrossRef Medline]

15. Kelley, R. L., and Kuroda, M. I. (2000) Noncoding RNA genes in dosage compensation and imprinting. *Cell* 103, 9–12 [CrossRef Medline]

16. Mercer, T. R., and Mattick, J. S. (2013) Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.* 20, 300–307 [CrossRef Medline]

17. Rinn, J. L., and Chang, H. Y. (2012) Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* 81, 145–166 [CrossRef Medline]

18. G. Hendrickson, D., Kelley, D. R., Tenen, D., Bernstein, B., and Rinn, J. L. (2016) Widespread RNA binding by chromatin-associated proteins. *Genome Biol.* 17, 28 [CrossRef Medline]

19. Chu, C., Zhang, Q. C., da Rocha, S. T., Flynn, R. A., Bharadwaj, M., Calabrese, J. M., Magnuson, T., Heard, E., and Chang, H. Y. (2015) Systematic discovery of Xist RNA binding proteins. *Cell* 161, 404–416 [CrossRef Medline]

20. Jenuwein, T., and Allis, C. D. (2001) Translating the histone code. *Science* 293, 1074–1080 [CrossRef Medline]

21. Karachentsev, D., Sarma, K., Reinberg, D., and Steward, R. (2005) PR-Set7-dependent methylation of histone H4 Lys 20 functions in repression of gene expression and is essential for mitosis. *Genes Dev.* 19, 431–435 [CrossRef Medline]

22. Botuyan, M. V., Lee, J., Ward, I. M., Kim, J. E., Thompson, J. R., Chen, J., and Mer, G. (2006) Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Cbx2 in DNA repair. *Cell* 127, 1361–1373 [CrossRef Medline]

23. Schotta, G., Sengupta, R., Kubicek, S., Malin, S., Kauer, M., Callen, E., Celeste, A., Pagani, M., Opravil, S., De La Rosa-Velazquez, I. A., Espejo, A., Bedford, M. T., Nussenzwieg, A., Busslinger, M., and Jenuwein, T. (2008) A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. *Genes Dev.* 22, 2048–2061 [CrossRef Medline]

24. Vermeulen, M., Eberl, H. C., Matarrese, F., Marks, H., Denissov, S., Butter, F., Lee, K. K., Olsen, J. V., Hyman, A. A., Stunnenberg, H. G., and Mann, M. (2010) Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* 142, 967–980 [CrossRef Medline]

25. Beck, D. B., Oda, H., Shen, S. S., and Reinberg, D. (2012) PR-Set7 and H4K20me1 at the crossroads of genome integrity, cell cycle, chromosome
condensation, and transcription. Genes Dev. 26, 325–337

26. Oda, H., Okamoto, I., Murphy, N., Chu, J., Price, S. M., Shen, M. M., Torres-Padilla, M. E., Heard, E., and Reinberg, D. (2009) Monomethylation of histone H4-lysine 20 is involved in chromosome structure and stability and is essential for mouse development. Mol. Cell. Biol. 29, 2278–2295

27. Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T. K., Koche, R. P., Lee, W., Mendenhall, E., O’Donovan, A., Presser, A., Russ, C., et al. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553–560

28. Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome. Cell 129, 823–837

29. Kidder, B. L., He, R., Wangsa, D., Padilla-Nash, H. M., Bernardo, M. M., Sheng, S., Ried, T., and Zhao, K. (2017) SMYD5 controls heterochromatin and chromosome integrity during embryonic stem cell differentiation. Cancer Res. 77, 6729–6745

30. Xu, J., and Kidder, B. L. (2018) H4K20me3 co-localizes with activating histone modifications at transcriptionally dynamic regions in embryonic stem cells. BMC Genomics 19, 514

31. Ringrose, L., and Paro, R. (2004) Epigenetic regulation of cellular memory and transcriptional landscapes in embryonic stem cells. Methods Mol. Biol. 33, 413–443

32. Schübeler, D., MacAlpine, D. M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gotschling, D. E., O’Neill, L. P., Turner, B. M., Derrow, J., Bell, S. P., and Groudine, M. (2004) The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes Dev. 18, 1263–1271

33. Pokholok, D. K., Harbison, C. T., Levine, S., Cole, M., Hannett, N. M., Lee, T. I., Bell, G. W., Walker, K., Rolfe, P. A., Herboldsberg, E., Zeitlinger, J., Lewitter, F., Gifford, D. K., and Young, R. A. (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 122, 517–527

34. Kim, T. H., Barrera, L. O., Zheng, M., Qu, C., Singer, M. A., Richmond, T. A., Wu, Y., Green, R. D., and Ren, B. (2005) A high-resolution map of active promoters in the human genome. Nature 436, 876–880

35. Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., and Misteli, T. (2002) Active genes are tri-methylated at K4 of histone H3. Nature 419, 407–411

36. Schneider, R., Bannister, A. J., Myers, F. A., Thorne, A. W., Crane-Robinett, M. A., Brocks, D., Dambacher, S., Schotta, G., and Saint-André, V., Batsché, E., Rachez, C., and Muchardt, C. (2011) Histone H3 lysine 9 trimethylation and HP1γ favor inclusion of alternative exons. Nat. Struct. Mol. Biol. 18, 337–344

37. Zhou, H. L., Luo, G., Wise, J. A., and Lou, H. (2014) Regulation of alternative splicing by local histone modifications: potential roles for RNA-guided mechanisms. Nucleic Acids Res. 42, 701–713

38. Davidovich, C., Wang, X., Cifuentes-Rojas, C., Goodrich, K. J., Godding, A. R., Lee, J. T., and Cech, T. R. (2015) Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. Mol. Cell 57, 552–558

39. Dinger, M. E., Amaral, P. P., Mercer, T. R., Pang, K. C., Bruce, S. J., Gardiner, B. B., Askarian-Amiri, M. E., Ru, K., Solda, G., Simons, C., Sunkin, S. M., Crowe, M. L., Grimmond, S. M., Perkins, A. C., and Mattick, J. S.
Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* **18**, 1433–1445 [CrossRef Medline](https://doi.org/10.1101/gr.085391.108)

62. Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B. R., Protacio, A., Flynn, R. A., Gupta, R. A., Wysocka, J., Lei, M., Dekker, J., Helms, J. A., and Chang, H. Y. (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 [CrossRef Medline](https://doi.org/10.1038/nature09968)

63. Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004) A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* **18**, 1251–1262 [CrossRef Medline](https://doi.org/10.1101/gad.2167604)

64. Fodor, B. D., Shukeir, N., Reuter, G., and Jenuwein, T. (2010) Mammalian Su(var) genes in chromatin control. *Annu. Rev. Cell Dev. Biol.* **26**, 471–501 [CrossRef Medline](https://doi.org/10.1146/annurev.cellbio.042408.134127)

65. Langmead, B., and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 [CrossRef Medline](https://doi.org/10.1038/nmeth.1919)

66. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 [CrossRef Medline](https://doi.org/10.1038/nmeth.1208)

67. Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 [CrossRef Medline](https://doi.org/10.1093/bioinformatics/btp610)

68. Greber, B., Wu, G., Bernemann, C., Joo, J. Y., Han, D. W., Ko, K., Tapia, N., Sabour, D., Sterneckert, J., Tesar, P., and Schöler, H. R. (2010) Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells. *Cell Stem Cell* **6**, 215–226 [CrossRef Medline](https://doi.org/10.1016/j.stem.2010.02.009)

69. Rao, J., and Greber, B. (2014) Conversion of epiblast stem cells to embryonic stem cells using growth factors and small molecule inhibitors. *Methods Mol. Biol.* **1150**, 215–226 [CrossRef Medline](https://doi.org/10.1007/978-1-4939-0119-6_13)

70. Kidder, B. L., and Palmer, S. (2010) Examination of transcriptional networks reveals an important role for TCFAP2C, SMARCA4, and EOMES in trophoblast stem cell maintenance. *Genome Res.* **20**, 458–472 [CrossRef Medline](https://doi.org/10.1101/gr.109508.109)

71. Kidder, B. L., and Palmer, S. (2012) HDAC1 regulates pluripotency and lineage specific transcriptional networks in embryonic and trophoblast stem cells. *Nucleic Acids Res.* **40**, 2925–2939 [CrossRef Medline](https://doi.org/10.1093/nar/gks074)

72. Xu, J., and Kidder, B. L. (2018) KDM5B decommissions the H3K4 methylation landscape of self-renewal genes during trophoblast stem cell differentiation. *Biol. Open* **7**

73. McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M., and Bejerano, G. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 [CrossRef Medline](https://doi.org/10.1038/nbt.1679)