



# HHS Public Access

Author manuscript

*Nat Struct Mol Biol.* Author manuscript; available in PMC 2011 June 01.

Published in final edited form as:

*Nat Struct Mol Biol.* 2010 December ; 17(12): 1495–1499. doi:10.1038/nsmb.1924.

## Reciprocal intronic and exonic histone modification regions in humans

Jason T. Huff<sup>1,3</sup>, Alex M. Plocik<sup>2,3</sup>, Christine Guthrie<sup>2</sup>, and Keith R. Yamamoto<sup>1</sup>

<sup>1</sup>Department of Cellular and Molecular Pharmacology, University of California, San Francisco, 600 16<sup>th</sup> Street, San Francisco, California 94143, USA

<sup>2</sup>Department of Biochemistry and Biophysics, University of California, San Francisco, 600 16<sup>th</sup> Street, San Francisco, California 94143, USA

### Abstract

While much attention has been focused on chromatin at promoters and exons, human genes are mostly composed of intronic sequences. Analyzing published surveys of nucleosomes and 41 chromatin marks in humans, we identified histone modifications specifically associated with 5' intronic sequences, distinguishable from promoter marks and bulk nucleosomes. These intronic marks were spatially reciprocal to H3K36me<sub>3</sub>, typically transitioning near internal exons. Several marks transitioned near *bona fide* exons, but not near nucleosomes at exon-like sequences. Thus, we interrogated splicing for a role in histone marking. Despite dramatic changes in regulated alternative splicing, histone marks were stable. Notably, these findings are consistent with a role for exon definition in influencing histone marks. In summary, we demonstrate that the location of many intragenic marks in humans can be distilled into a simple organizing principle: association with 5' intronic or 3' exonic regions.

### Keywords

DOT1; BRE1; RNF20; SET2; CD45

### Introduction

Nucleosomes, the basic units of eukaryotic chromatin structure, are marked by numerous post-translational modifications and compositional variants that influence DNA and RNA metabolism<sup>1-3</sup>. For example, transcription by RNA polymerase II is associated with particular modifications and variants at promoters, including histone H3 lysine 4 methylations, numerous acetylations, H2A.Z, and H3.3 (refs. 4,5). Of these, H3K4me<sub>3</sub> acts as

---

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

Corresponding authors: Christine Guthrie, 600 16th St, GH-N374, San Francisco, CA 94143-2200, +1 (415) 476-4705 tel, +1 (415) 502-4315 fax, christineguthrie@gmail.com, Keith R. Yamamoto, 600 16th St, GH-S574, San Francisco, CA 94143-2280, +1 (415) 476-3128 tel, +1 (415) 476-6129 fax, yamamoto@cmp.ucsf.edu.

<sup>3</sup>These authors contributed equally to this work.

**Author Contributions:** J.T.H. and A.M.P. designed and performed the analyses and experiments. J.T.H., A.M.P., C.G., and K.R.Y. wrote the manuscript.

a binding determinant for the TAF1 subunit of TFIID, mechanistically linking this modification to transcription initiation<sup>6</sup>.

By analogy to chromatin marks at promoters, marks positioned within genes physically coincide with various co-transcriptional processes and may influence or be influenced by RNA processing<sup>7,8</sup>. For example, depletion of human BRE1A/RNF20, which is required to catalyze H2B lysine 120 ubiquitylation (H2Bub), results in decreased chromatin association of the transcription elongation and splicing factor SKIP/SNW1 (ref. 9). Similarly, H3K36me3 is catalyzed by a methyltransferase associated with elongating, hyperphosphorylated RNA polymerase II<sup>10</sup> and depletion of the human enzyme, SETD2/HYPB/KMT3A, leads to a defect in mRNA export and changes in alternative splicing<sup>11,12</sup>.

Most provocatively, bioinformatic analyses of data from diverse animals indicate that certain histone modifications, particularly H3K36me3, are not only elevated within genes, but specifically at exons<sup>13-19</sup>. It has been suggested, though, that H3K36me3 levels simply reflect higher nucleosome occupancy at exons in Metazoa<sup>17,19</sup>. In any case, how the bulk of chromatin within genes compares to the one or few nucleosomes at each exon has not been fully elucidated.

To explore more broadly the relationships between chromatin and gene architecture, we began by analyzing published human epigenomic data. We detected regions located after active transcription start sites and extending through 5' introns of variable length and the first few internal exons, enriched in H2Bub, H3K23ac, and eight different methylations including H3K79me2/3. Regions in the 3' portions of active genes displayed H3K36me3 peaking at internal exons with elevated levels extending into each downstream intron. Intriguingly, regulated alternative splicing failed to change these histone modification profiles.

## Results

### Chromatin marks in 5' introns

To discover new epigenomic features related to gene expression, we analyzed published sequencing surveys of human mononucleosomes prepared by native micrococcal nuclease digestion (MNase-seq), mononucleosome ChIP (ChIP-seq) for 20 histone methylations, 18 histone acetylations, and the histone variants H2A.Z and H3.3, as well as cross-linked sonicated ChIP-seq of H2B lysine 120 ubiquitylation<sup>20-24</sup>. We used principal component analysis (PCA) to reduce the complexity of this large data set, producing a smaller number of dimensions, or principal components, that retain much of the variance in the original data set (see Methods). Thus, we aimed to identify subsets of the 41 chromatin marks that occurred at similar regions of the genome.

To identify relationships specific to genes, we applied PCA to data from 10 kb upstream to 10 kb downstream of transcription start sites (TSSs) (Fig. 1 and Supplementary Fig. 1). The first principal component corresponded to the 17 marks previously characterized as a "backbone" at active promoters<sup>23</sup> (H2A.Z, H2BK5ac, H2BK12ac, H2BK20ac, H2BK120ac, H3K4ac, H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K9me1, H3K18ac, H3K27ac,

H3K36ac, H4K5ac, H4K8ac, and H4K91ac). We also identified four more acetylations at active promoters (H2AK9ac, H3K23ac, H4K12ac, and H4K16ac). The second principal component included nucleosomes and marks that were generally depleted near active promoters, but had elevated occupancy within genes. The third principal component included high levels of H2BK5me1, H2Bub, H3K4me1, H3K4me2, H3K9me1, H3K23ac, H3K79me1, H3K79me2, H3K79me3, and H4K20me1.

Intriguingly, we noticed that several marks, particularly those enriched in the third principal component, were enriched in the 5' regions of active genes. This was reminiscent of their spatial distribution in yeast<sup>10</sup>, but in humans these marks extended downstream into 5' introns of varying length, decreasing in the vicinity of exons (Fig. 2a). To assess if this were a genome-wide phenomenon, we aligned thousands of genes by TSSs and sorted them by distances from the TSSs to the first 3' splice sites, plotting each gene as a row with sequencing read density as a heatmap. Thus, this perspective simultaneously displayed three different aspects of gene architecture: TSS positions, first introns, and first internal exons. In this view, principal component three, and to a lesser extent component two, displayed a 5' bias that decreased near internal exons, with enrichment more apparent at genes with higher expression (Supplementary Fig. 1).

Analyzing individual chromatin marks in this way revealed specific relationships with features of gene architecture (Fig. 2b and Supplementary Fig. 2). As expected, promoter marks high in principal component one, such as H3K4me3, surrounded TSSs. Nucleosomes, high in component two, appeared slightly higher near TSSs and internal exons than within neighboring introns<sup>13,16-19</sup>. H3K79me2, representative of the 10 marks high in component three, showed enrichment that extended downstream of TSSs into first introns until reaching internal exons. In contrast, H3K36me3 exhibited a spatially reciprocal relationship with H3K79me2, clearly peaking at internal exons. As a control, H3K27me1 was present within active genes<sup>20</sup>, but exhibited little if any change with respect to exons (Supplementary Fig. 2).

### Histone modifications separable from nucleosome occupancy

Interpreting the distribution of histone modifications from MNase- and ChIP-seq data is complicated by potential technical artifacts<sup>25-27</sup> and nucleosome occupancy biases<sup>13,16-19,27</sup>. Therefore, we independently assessed intronic and exonic marks by quantitative PCR (ChIP-qPCR) at the *STIM1* gene, as it has internal exons separated by large 5' introns (Supplementary Fig. 3). Validating the ChIP-seq profiles, we observed enrichment for H3K79me2 in the first three long introns and H3K36me3 near exons. H3 displayed some heterogeneity along *STIM1*, which may reflect differential nucleosome occupancy or positioning, especially at exons<sup>13,16-19</sup>. However, H3K36me3 near the exons of *STIM1* remained enriched after normalization to H3, although reduced in dynamic range.

Across the genome, normalizing ChIP-seq data by MNase-seq data largely removed exon-centered enrichment for most marks (e.g. H3K79me2), but H3K36me3 remained enriched after normalization, peaking near exons and extending on average a few kilobases downstream (Supplementary Fig. 4). Also, performing PCA and heatmap analyses after

normalizing each ChIP-seq dataset by MNase-seq (Supplementary Fig. 4) supported similar conclusions as the analyses without normalization above (Figs. 1 and 2).

### Histone modifications transitioning near *bona fide* exons

We considered the possibility that nucleosome occupancy at internal exons could function as the signal delineating intronic and exonic regions. *Bona fide* internal exons are annotations of spliced RNAs, minimally including 3' and 5' splice sites with additional sequence constraints to encode proteins and include splicing enhancer and silencer sites. Spies *et al.* identified intronic exon-like sequence composition regions (ECRs) that are not spliced, but which have high nucleosome occupancy like that at *bona fide* exons<sup>18</sup>.

We compared ECRs to *bona fide* exons to determine if the general sequence composition sufficient to direct high nucleosome occupancy<sup>18</sup> were also sufficient to pattern histone modifications. To control for biases in number and position of ECRs relative to all exons, we compared each intronic ECR (12,687 total) to its nearest *bona fide* exon. Similar to genome-wide averages of H3K36me3 at exons (Supplementary Fig. 4), H3K36me3 peaked at ECR-matched exons (Supplementary Fig. 5). However, H3K36me3 did not peak at ECR positions. On average, marks in the third principal component were higher upstream of exons than downstream and a similar pattern was found near ECRs, although the marks variably displayed decreased spatial response to ECRs compared to exons (e.g. H3K4me1, H3K4me2, H3K9me1, and H3K23ac). As a control for comparison, H3K27me1 showed little difference up- and downstream of exons or ECRs (Supplementary Fig. 5).

### Histone modifications stable despite changes in splicing

Finding that nucleosome occupancy could not fully account for the spatial dynamics of histone modifications that change near exons, we considered a model in which RNA splicing events are mechanistically linked to co-transcriptional histone modification, thus coupling exon positions to chromatin states. Splicing is a complex, multistep process that involves recognition of splice sites and progression of spliceosome assembly, ultimately resulting in intron excision<sup>28</sup>. Both splice site recognition and spliceosome assembly can be regulated to produce alternative RNA splicing outcomes<sup>29</sup>. To test if RNA splicing itself could direct histone modification profiles, we examined two cases of alternative splicing in which the frequency of exon inclusion was experimentally manipulable without altering the DNA/RNA sequence.

In our first example, YPEL5 exon 2 inclusion increases when cells are treated with caffeine<sup>30</sup>. To assess both YPEL5 expression and alternative splicing, we measured total mRNA by qPCR and the ratio of mRNA isoforms by RT-PCR in various concentrations of caffeine. After 7 hours exposure, mRNA levels were similar between 6 mM and 18 mM caffeine (Supplementary Fig. 6), while exon 2 inclusion increased from 1% inclusion to 66% inclusion (Fig. 3a and Supplementary Fig. 6). We verified that our ChIP-qPCR at YPEL5 spatially recapitulated the profiles of H3K79me2, H3K36me3, and nucleosomes in the ChIP-seq and MNase-seq data (Supplementary Fig. 6). Notably, H3K79me2 and H3K36me3 profiles were similar at the low and high caffeine concentrations (Fig. 3b). Therefore, the

dramatic change in splicing of YPEL5 RNA was not sufficient to direct changes in these histone modifications.

As with many alternative splicing events, little is known about the mechanism of YPEL5 exon 2 inclusion. Moreover, we were constrained to short treatment periods in the experiments, as high levels of caffeine are toxic. Hence, we sought a second example of regulated alternative splicing, in which changes in alternative splicing could be maintained over many cell divisions. In lymphocytes, the RNA-binding protein hnRNPLL binds to CD45 pre-mRNA to promote cell-type specific exon-skipping of multiple CD45 exons<sup>31-33</sup>. Although B cells typically express long isoforms of CD45, the B cell lymphoma lines BJAB and BL41 express different levels of hnRNPLL, and consequently, maintain different CD45 isoforms<sup>31</sup>.

We obtained a set of matched BJAB and BL41 cell lines in which hnRNPLL levels are stably decreased by short hairpin RNA interference or increased by hnRNPLL over-expression, respectively<sup>31</sup>. BJAB and BL41 cells incidentally expressed 100-fold different levels of CD45 mRNA, so we compared matched lines to control for variable CD45 expression (Fig. 4a). RT-PCR analysis to assay variable inclusion of exons 4-6 confirmed that depletion of hnRNPLL in BJAB cells increased inclusion, whereas overexpression of hnRNPLL in BL41 cells decreased inclusion (Fig. 4b). H3K36me3, H3K79me2, and nucleosome levels at CD45 measured by ChIP-qPCR were generally similar to published ChIP-seq and MNase-seq data in T cells (Supplementary Fig. 7). If RNA splicing were sufficient to change histone modifications, increased exon inclusion at exons 4-6 should decrease H3K79me2 and/or increase H3K36me3 levels. However, we observed no significant differences in histone modifications between matched pairs of BJAB or BL41 cells expressing different CD45 isoforms (Fig. 4c,d).

## Discussion

We used PCA to help identify groups of histone marks (Fig. 1) that displayed distinct profiles within active genes (Supplementary Fig. 1). Among these, we identified a group of 10 marks biased for 5' introns and roughly reciprocal with H3K36me3 near exons (Fig. 2 and Supplementary Fig. 2). Sequence-related nucleosome occupancy alone was not sufficient to explain transitions between intronic and exonic histone modification regions (Supplementary Fig. 5). The genome-wide phenomenon of intronic and exonic marks suggests that a general mechanism associated with splicing influences histone marking, as speculated previously<sup>15</sup>. Thus, we tested if levels of RNA splicing influenced the profiles of these histone modifications. Neither inducible changes in exon inclusion in YPEL5 after several hours (Fig. 3), nor changes in inclusion of multiple exons in CD45 over numerous cell divisions (Fig. 4) resulted in significant differences in H3K79me2 or H3K36me3. This contrasts starkly with transcriptional induction, in which H3K79me2 and H3K36me3 levels can increase within one hour<sup>34</sup>.

Our results rule out a simple correlation between the extent of RNA splicing and histone modification levels. Intriguingly, regulated skipping of CD45 exon 4 is achieved by stalling spliceosome assembly at an early step following exon definition<sup>35</sup>; therefore, spliceosome

assembly may be sufficient to pattern histone modifications at *CD45* exon 4, regardless of splicing outcomes. As it is now recognized that promoter marks are relatively stable<sup>36,37</sup> and may reflect transcription initiation<sup>38</sup> rather than productive transcription, we suggest by analogy that intronic and exonic marks are relatively stable and may reflect an aspect of the splicing process, such as exon definition, rather than productive splicing *per se*.

If a feature related to splicing were required to influence histone modifications, we would expect a decrease in 5' intron marks and an increase in H3K36me3 specifically at *bona fide* exons, but not at ECRs, which have similar sequence composition, but are not spliced. Indeed, ECRs do not direct increased levels of H3K36me3 as at exons or decreased levels of certain 5' intronic marks (i.e. H3K4me1, H3K4me2, H3K9me1, H3K23ac) (Supplementary Fig. 5). In contrast, some 5' intronic marks (e.g. H3 lysine 79 methylations) have nearly equivalent decreases near exons and ECRs, suggesting that the local sequence composition of exons (and ECRs) may be sufficient to influence these histone modification profiles. Thus, *cis*-acting sequences, functioning as DNA or RNA, could signal to change histone modifications. For example, SR protein binding sites at exons could affect chromatin marking through "reverse" coupling<sup>8</sup>. Additionally, elevated nucleosome occupancy at exons<sup>13,16,19</sup> may play a role, independently or in conjunction with *cis*-acting sequences. Extensive mutagenesis will be essential to distinguish between these possibilities and identify necessary sequences.

How distinct histone modification patterns are determined remains largely unknown. In agreement with Spies *et al.*<sup>18</sup>, we find that histone modification patterns do not merely reflect nucleosome enrichment at exons, indicating the need for additional steps of regulation beyond nucleosome occupancy such as chromatin modifying enzyme action. Spatial similarities among the intronic marks could reflect cross-regulation between histone modifications<sup>39</sup>. For example, the extensive overlap between H2Bub and H3K79me2/3 (Supplementary Fig. 2) is consistent with direct stimulation by H2Bub of H3 lysine 79 methyltransferase activity by DOT1L and a DOT1L-containing protein complex<sup>40,41</sup>. Also, H2Bub can directly stimulate human SET1-mediated methylation of H3 lysine 4 *in vitro*, so H2Bub within 5' introns may help to increase H3K4me1 and H3K4me2 (refs. 42,43).

We propose that promoter, intronic, and exonic intragenic histone modification regions constitute three distinct chromatin zones at active human genes<sup>4,23</sup>, analogous to those found in yeast<sup>10,44</sup>: i) the "backbone" or promoter zone contains H3K4me3, ii) a 5' zone includes H3K4me2 (and 9 other marks in humans, including H3K79me2), and iii) a 3' zone contains H3K36me3 (Fig. 2a). However, most yeast genes are short and intronless, in contrast to human genes, which frequently contain large first introns. Also, we found occasionally that these different zones may interdigitate, *e.g.* there was switching back and forth between H3K79me2 and H3K36me3 within *STIM1* (Supplementary Fig. 3). This implies that these zones are not strictly 5' and 3' with respect to genic positioning, but rather that introns and exons act to organize intragenic histone modifications.

Intronic and exonic histone modifications can, in principle, affect many aspects of gene expression<sup>7</sup>. For example, H3K36me3-modified nucleosomes could directly influence splicing<sup>12</sup>. Furthermore, H3K36me3 levels on average are slightly lower at alternatively

spliced exons compared to nearby constitutive exons<sup>14,15,17</sup>. An open question for future studies remains: how are these marks established and maintained at particular introns and exons within the same transcribed unit? A working knowledge of such localization mechanisms will be paramount to understanding temporal and spatial regulation of chromatin marks in humans.

## Methods

### Epigenomic analyses

Analyses were performed and plots were generated with custom R scripts (<http://www.r-project.org>). Human gene models were obtained from the UCSC genome browser RefSeq (hg18) annotations (<http://genome.ucsc.edu/>). We offset sequencing read end-positions by 73 bases (half the width of the canonical nucleosome) to approximate the position of nucleosomal centers. MNase- and ChIP-seq reads were aggregated in non-overlapping 1-kb windows from 10 kb upstream to 10 kb downstream of TSSs and data across these genomic segments were subjected to PCA. PCA was performed by singular value decomposition with the data centered and scaled using the R function `prcomp()`. Figure 1 was rendered in PyMOL (<http://www.pymol.org/>). For all plots only one gene model was shown for genes with multiple models. For Figure 2a and Supplementary Figures 6 and 7, a Gaussian kernel density estimator was used with both the kernel width and bin size set to 146 bp (size of one canonical nucleosome). The kernel density scale was adjusted so that uniform density would be equal to 1 at all positions. We did not precisely map the positions of individual nucleosomes, because our analysis was generally on the scale of tens to hundreds of nucleosomes. However, we verified that the plots were robust to changes in nucleosome phasing and average width in the kernel density estimator.

### Cell culture and regulated alternative splicing

SW620 human colon cancer cells (ATCC CCL-227) were grown in DMEM supplemented with 10% FBS (v/v), Pen-Strep, and Normocin (Amaya). They were grown to near confluence for the experiments in Supplementary Figure 3 and were seeded at a density of  $1 \times 10^5$  cells per  $\text{cm}^2$  24 hours prior to treatment with media supplemented with the indicated concentration of caffeine for 7 hours in Figure 3. BJAB and BL41 B cell lines<sup>31</sup> were grown in RPMI-1640 supplemented with 10% FBS (v/v), glutamine, Pen-Strep, and Normocin (and  $5 \mu\text{g ml}^{-1}$  puromycin for the BJAB lines) to a density of about  $1 \times 10^6$  cells  $\text{ml}^{-1}$ . The earliest passages available for the B cell experiments were used, because the effect on CD45 alternative splicing decreased over continual passaging.

### Cross-linked mononucleosome ChIP

ChIP was performed as described<sup>34</sup> with the following modifications. Nuclei from  $1 \times 10^7$  crosslinked cells were digested with 100 units of MNase (Worthington) in 500  $\mu\text{l}$  of Buffer N supplemented with 3 mM  $\text{CaCl}_2$  for 10 min at  $37^\circ\text{C}$ . We previously determined the concentration of MNase to be optimal for digestion to mostly mononucleosomes by agarose gel electrophoresis of the cross-link reversed input chromatin. Chromatin extract equivalent to  $2 \times 10^6$  cells in 500  $\mu\text{l}$  of RIPA was incubated with 4  $\mu\text{g}$  of one of the following antibodies for 2 hours at  $4^\circ\text{C}$ : H3K79me2, H3K36me3, or H3 (Abcam ab3594, ab9050, or

ab1791). Instead of Protein A/G agarose beads, 50  $\mu$ l of Protein G Dynabeads (Invitrogen) were added to each tube and incubated overnight at 4°C. After the chromatin elution step, DNA was purified with QIAquick PCR columns (QIAGEN) and quantified by qPCR as described below.

To analyze ChIP-qPCR data, we used MNase-digested input material as a quantitative standard, thereby measuring the efficiency of immunoprecipitation, and internally controlling for digestion biases. In addition, we normalized H3K79me2 and H3K36me3 ChIP to an H3 ChIP to control for differences in nucleosome occupancy (Supplementary Fig. 3). It should be noted that ChIP-qPCR is subject to particular caveats, most notably the dependence on fixed primer positions. As such, MNase digestion of DNA around strongly positioned nucleosomes can negatively affect the amplification efficiency of partially overlapping amplicons. Similarly, histone modifications or variants may differentially influence the extent of MNase digestion and, consequently, bias amplification efficiency. We attempted to minimize these effects by tiling genes with primer sets designed to amplify small segments of DNA (62 bp median amplicon size, Supplementary Table 1).

### Quantitative PCR

PCR primers (Supplementary Table 1) were designed with BatchPrimer3 v1.0 (<http://wheat.pw.usda.gov/demos/BatchPrimer3/>) from human repeat-masked sequence obtained from the UCSC genome browser (hg18). Serial dilutions of DNA ranging from 50% to 0.4% of the input chromatin were used to obtain calibration curves, measure primer efficiencies, and ensure that quantification was in a linear dynamic range. Primer sets yielding multiple amplification products or calibration curves with R-squared values of < 0.96 were excluded. For each qPCR sample, 0.5  $\mu$ l of 50  $\mu$ l eluted ChIP DNA was amplified in 25  $\mu$ l volume reactions containing 250  $\mu$ M dNTPs, 1  $\times$  (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> buffer (Fermentas), 0.5  $\mu$ M primer, 1.5 mM MgCl<sub>2</sub>, 1.25 units Dynazyme II (Finnzymes), and Sybr Green I fluorescent dye (Sigma). Fluorescence was measured on a BioRad CFX-96 machine using standard cycling conditions (3 min at 95°C, 40 cycles of 15 s at 95°C, 30 s at 55°C, and 15 s at 72°C). Each ChIP value was calculated as the median of 2-4 technical qPCR replicates. For each biological replicate consisting of qPCR replicates, median ChIP values for H3K79me2 and H3K36me3 were then divided by median H3 ChIP values.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

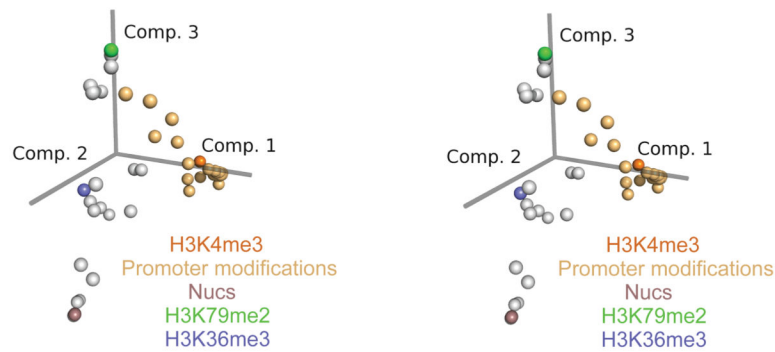
We thank the following for kindly providing reagents and data: Shalini Oberdoerffer and Anjana Rao (Harvard Medical School) for B cell lines, Noah Spies and Chris Burge (MIT) for ECR locations, and Efrat Shema and Moshe Oren (Weizmann Institute of Science) for H2Bub ChIP-seq data. We also thank Hiten Madhani and Joan Steitz for critical reading of the manuscript and the Guthrie, Yamamoto, and Panning groups for helpful discussions. J.T.H. and A.M.P. were supported by individual ARCS Foundation Scholarships. Research support was provided by NIH grants GM21119 to C.G. and CA020535 to K.R.Y. C.G. is an American Cancer Society Research Professor of Molecular Genetics and K.R.Y. is a consultant with Merck & Co., Inc.



## References

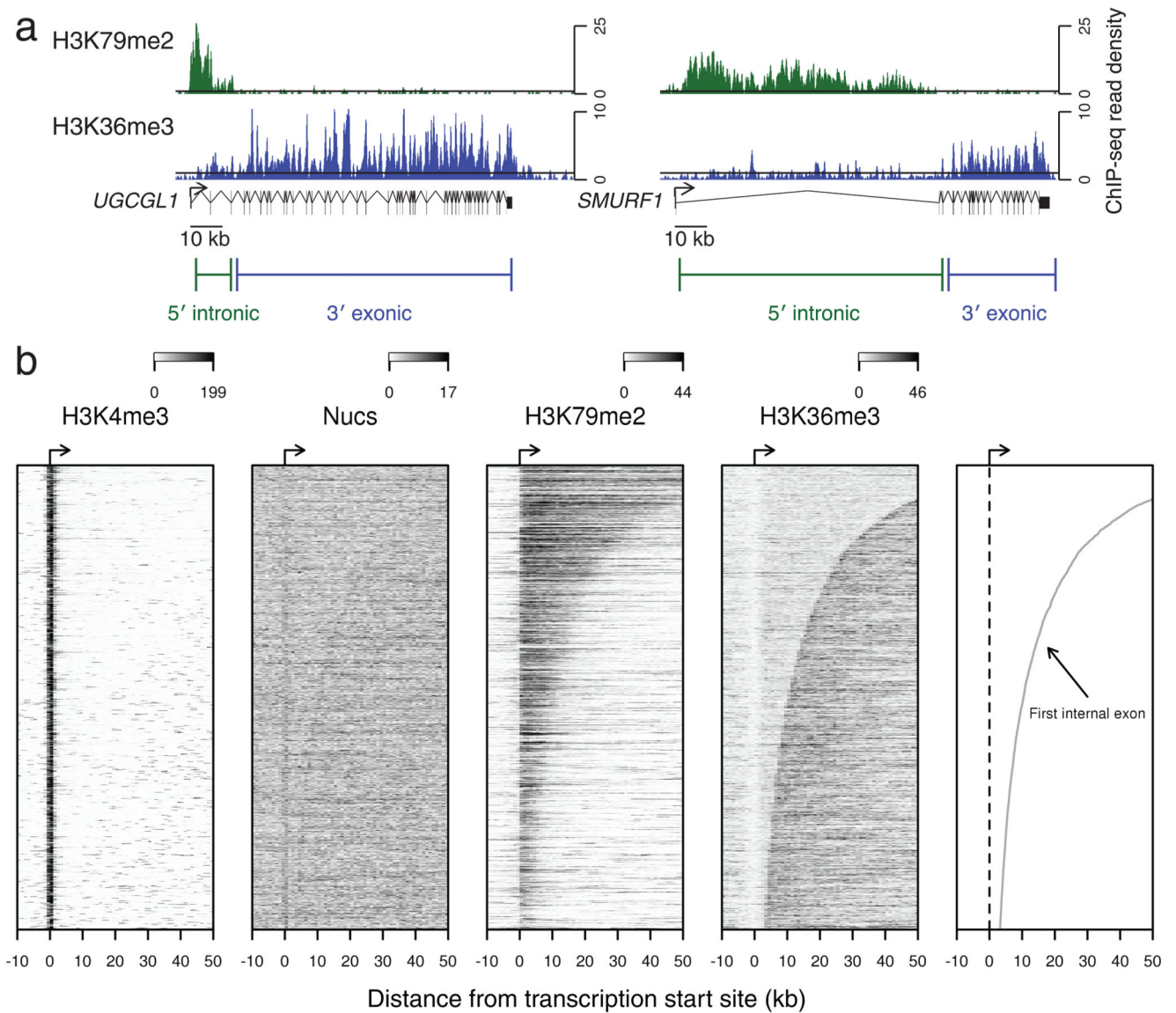
1. Berger SL. The complex language of chromatin regulation during transcription. *Nature*. 2007; 447:407–12. [PubMed: 17522673]
2. Kouzarides T. Chromatin modifications and their function. *Cell*. 2007; 128:693–705. [PubMed: 17320507]
3. Loyola A, Almouzni G. Marking histone H3 variants: how, when and why? *Trends Biochem Sci*. 2007; 32:425–33. [PubMed: 17764953]
4. Lee BM, Mahadevan LC. Stability of histone modifications across mammalian genomes: Implications for ‘epigenetic’ marking. *J Cell Biochem*. 2009; 108:22–34. [PubMed: 19623574]
5. Talbert PB, Henikoff S. Histone variants--ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol*. 2010; 11:264–275. [PubMed: 20197778]
6. Vermeulen M, et al. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell*. 2007; 131:58–69. [PubMed: 17884155]
7. Perales R, Bentley D. “Cotranscriptionality”: the transcription elongation complex as a nexus for nuclear transactions. *Mol Cell*. 2009; 36:178–191. [PubMed: 19854129]
8. Zhong X, Wang P, Han J, Rosenfeld MG, Fu X. SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Mol Cell*. 2009; 35:1–10. [PubMed: 19595711]
9. Brès V, Yoshida T, Pickle L, Jones KA. SKIP interacts with c-Myc and Menin to promote HIV-1 Tat transactivation. *Mol Cell*. 2009; 36:75–87. [PubMed: 19818711]
10. Li B, Carey M, Workman JL. The role of chromatin during transcription. *Cell*. 2007; 128:707–19. [PubMed: 17320508]
11. Yoh SM, Lucas JS, Jones KA. The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev*. 2008; 22:3422–34. [PubMed: 19141475]
12. Luco RF, et al. Regulation of alternative splicing by histone modifications. *Science*. 2010; 327:996–1000. [PubMed: 20133523]
13. Andersson R, Enroth S, Rada-Iglesias A, Wadelius C, Komorowski J. Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res*. 2009; 19:1732–1741. [PubMed: 19687145]
14. Hon G, Wang W, Ren B. Discovery and annotation of functional chromatin signatures in the human genome. *PLoS Comput Biol*. 2009; 5:e1000566. [PubMed: 19918365]
15. Kolasinska-Zwierz P, et al. Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet*. 2009; 41:376–381. [PubMed: 19182803]
16. Nahkuri S, Taft RJ, Mattick JS. Nucleosomes are preferentially positioned at exons in somatic and sperm cells. *Cell Cycle*. 2009; 8:3420–3424. [PubMed: 19823040]
17. Schwartz S, Meshorer E, Ast G. Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol*. 2009; 16:990–995. [PubMed: 19684600]
18. Spies N, Nielsen CB, Padgett RA, Burge CB. Biased chromatin signatures around polyadenylation sites and exons. *Mol Cell*. 2009; 36:245–254. [PubMed: 19854133]
19. Tilgner H, et al. Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol*. 2009; 16:996–1001. [PubMed: 19684599]
20. Barski A, et al. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007; 129:823–837. [PubMed: 17512414]
21. Schones DE, et al. Dynamic regulation of nucleosome positioning in the human genome. *Cell*. 2008; 132:887–898. [PubMed: 18329373]
22. Shema E, et al. The histone H2B-specific ubiquitin ligase RNF20/hBRE1 acts as a putative tumor suppressor through selective regulation of gene expression. *Genes & Development*. 2008; 22:2664–2676. [PubMed: 18832071]
23. Wang Z, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet*. 2008; 40:897–903. [PubMed: 18552846]

24. Jin C, et al. H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat Genet.* 2009; 41:941–945. [PubMed: 19633671]
25. McGhee JD, Felsenfeld G. Another potential artifact in the study of nucleosome phasing by chromatin digestion with micrococcal nuclease. *Cell.* 1983; 32:1205–1215. [PubMed: 6301684]
26. Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.* 2008; 36:e105. [PubMed: 18660515]
27. Jin C, Felsenfeld G. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes & Development.* 2007; 21:1519–1529. [PubMed: 17575053]
28. Wahl MC, Will CL, Lührmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell.* 2009; 136:701–718. [PubMed: 19239890]
29. Chen M, Manley JL. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol.* 2009; 10:741–754. [PubMed: 19773805]
30. Shi J, Hu Z, Pabon K, Scotto KW. Caffeine regulates alternative splicing in a subset of cancer-associated genes: a role for SC35. *Mol Cell Biol.* 2008; 28:883–895. [PubMed: 18025108]
31. Oberdoerffer S, et al. Regulation of CD45 alternative splicing by heterogeneous ribonucleoprotein, hnRNPLL. *Science.* 2008; 321:686–691. [PubMed: 18669861]
32. Topp JD, Jackson J, Melton AA, Lynch KW. A cell-based screen for splicing regulators identifies hnRNP LL as a distinct signal-induced repressor of CD45 variable exon 4. *RNA.* 2008; 14:2038–2049. [PubMed: 18719244]
33. Wu Z, et al. Memory T cell RNA rearrangement programmed by heterogeneous nuclear ribonucleoprotein hnRNPLL. *Immunity.* 2008; 29:863–875. [PubMed: 19100700]
34. Edmunds JW, Mahadevan LC, Clayton AL. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J.* 2008; 27:406–20. [PubMed: 18157086]
35. House AE, Lynch KW. An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition. *Nat Struct Mol Biol.* 2006; 13:937–944. [PubMed: 16998487]
36. Raisner RM, et al. Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell.* 2005; 123:233–248. [PubMed: 16239142]
37. Kouskouti A, Talianidis I. Histone modifications defining active genes persist after transcriptional and mitotic inactivation. *EMBO J.* 2005; 24:347–357. [PubMed: 15616580]
38. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell.* 2007; 130:77–88. [PubMed: 17632057]
39. Latham JA, Dent SYR. Cross-regulation of histone modifications. *Nat Struct Mol Biol.* 2007; 14:1017–1024. [PubMed: 17984964]
40. McGinty RK, Kim J, Chatterjee C, Roeder RG, Muir TW. Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature.* 2008; 453:812–816. [PubMed: 18449190]
41. Mohan M, et al. Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). *Genes Dev.* 2010; 24:574–589. [PubMed: 20203130]
42. Kim J, Hake SB, Roeder RG. The human homolog of yeast BRE1 functions as a transcriptional coactivator through direct activator interactions. *Mol Cell.* 2005; 20:759–770. [PubMed: 16337599]
43. Kim J, et al. RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. *Cell.* 2009; 137:459–471. [PubMed: 19410543]
44. Kim T, Buratowski S. Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. *Cell.* 2009; 137:259–272. [PubMed: 19379692]



**Figure 1.**

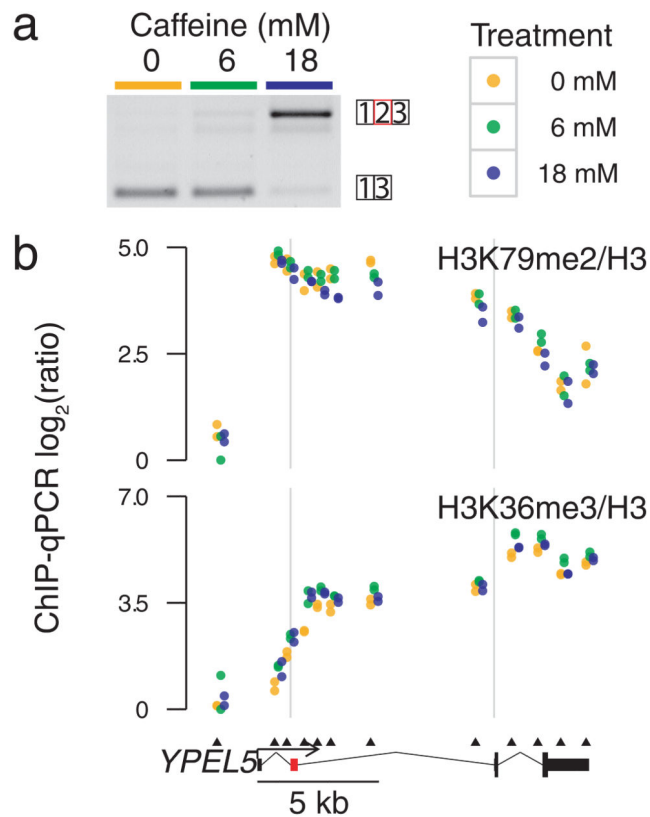
Groups of histone marks revealed by PCA. PCA of all ChIP-seq and MNase-seq data 10 kb up- and downstream of RefSeq TSSs is presented in a 3D stereoscopic image. Gray points represent the PCA loadings for each of the 41 histone marks and nucleosome occupancy (Nucs). Principal components (Comp.) 1, 2, and 3 are projected to the lower right, lower left, and top, respectively, with the orientation of the axes shown as gray bars radiating from the origin. Individual points representative of each component are colored: H3K4me3 (orange) for Comp. 1, Nucs (purple) for Comp. 2, and H3K79me2 (green) for Comp. 3. For reference, marks enriched at promoters are colored light orange and H3K36me3 is colored blue. Supplementary Figure 1 shows the fully labeled and separated PCA loadings and the transformed PCA scores plotted on the genome with respect to transcription start sites.



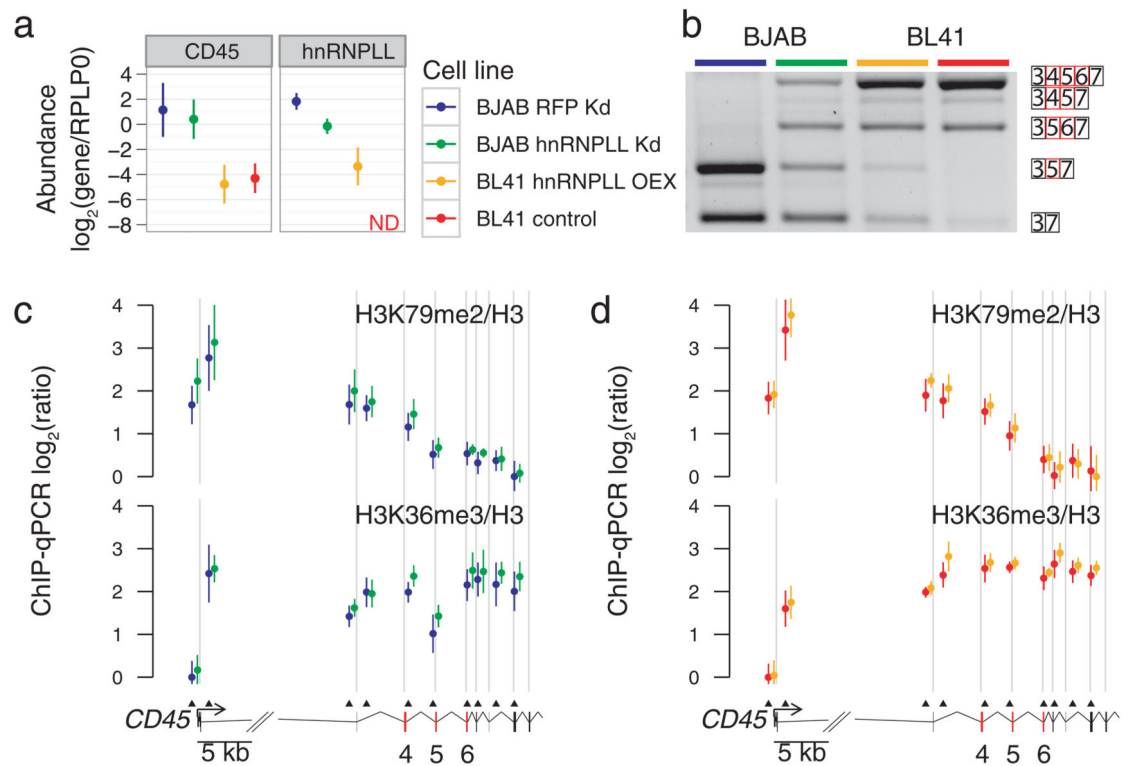
**Figure 2.**

Intragenic histone modification regions reflect gene architecture. (a) H3K79me2, and H3K36me3 ChIP-seq reads plotted on two genes (*UGCGL1* and *SMURF1*) of similar length, but different gene architectures. A black line in each density plot shows where a uniform distribution of sequence reads would be. The proposed 5' intronic, and 3' exonic histone modification zones are diagrammed below each gene. (b) Heatmaps show the density for H3K4me3, nucleosomes (Nucs), H3K79me2, and H3K36me3 ChIP- and MNase-seq, plotted with respect to transcription start sites (bent arrows) with genes sorted by distances to the beginning of the first internal exons (i.e. first 3' splice sites). For clarity only the 4286 annotated RefSeq genes that are in the top 50% of expression and with the 50% longest distances to the first internal exon are shown. Full plots are shown in Supplementary Figure 2. Each gene is displayed as a row and columns are 1-kb bins from 10 kb upstream to 50 kb downstream of transcription start sites (bottom). The value of each bin is shaded by the

number of ChIP- or MNase-seq reads per kb, scaled to cover data between 0 reads (white) and the 99th percentile of each sequencing experiment (black), indicated at the top right of each heatmap. At far right the aligned transcription start sites (broken line) and the positions of first internal exons (gray line) are shown as a visual reference.



**Figure 3.** Histone modification profiles at the alternatively spliced exon of *YPEL5* are similar between caffeine-treated SW620 cells expressing different *YPEL5* mRNA isoforms. **(a)** Increasing caffeine concentrations result in higher exon inclusion of *YPEL5* exon 2. RT-PCR of SW620 cells treated for 7 hours with 0, 6, or 18 mM caffeine. Alternatively (red) and constitutively (black) spliced exons are numbered and diagrammed to the right of the corresponding RT-PCR product. **(b)** Normalized H3K79me2 and H3K36me3 profiles in SW620 cells treated with 0 (orange), 6 (green), and 18 (blue) mM caffeine for 7 hours. Points indicate  $\log_2(\text{ratios})$  of H3K79me2/H3 (top) and H3K36me3/H3 (bottom) ChIP-qPCR measurements, relative to the lowest value, for each of two SW620 biological replicates. Vertical gray lines provide a visual reference for the 3' splice site of each internal exon. Arrowheads indicate the positions of qPCR amplicons along the gene (bottom). The alternatively spliced cassette exon 2 is indicated in red. The gene schematic and markers are drawn to scale according to their genomic coordinates except for the second (green) and third (blue) data points at each amplicon, which are offset to the right to facilitate comparison between conditions.

**Figure 4.**

Histone modification profiles at the alternatively spliced exons of *CD45* are similar between B cell lines stably expressing different *CD45* mRNA isoforms. In all panels, colored points and lines refer to matched BJAB B cell lines expressing a short hairpin targeting knockdown of either RFP as a control (blue) or hnRNPLL (green) or to matched BL41 B cell lines overexpressing hnRNPLL (orange) or mock-treated as a control (red). **(a)** BJAB and BL41 B cell lines differ in the relative abundance of *CD45* and hnRNPLL mRNA. Points indicate mean  $\log_2(\text{ratios}) \pm \text{s.d.}$  of *CD45*/RPLP0 and hnRNPLL/RPLP0 qRT-PCR measurements from (n=4) biological replicates (not detectable; ND). **(b)** Confirmation that passaged BJAB and BL41 B cell lines express different *CD45* isoforms. Alternatively (red) and constitutively (black) spliced exons are diagrammed to the right of the corresponding RT-PCR product. **(c and d)** Normalized H3K79me2 and H3K36me3 profiles in matched B cell lines. Points indicate the mean  $\log_2(\text{ratios}) \pm \text{s.d.}$  of H3K79me2/H3 (top) and H3K36me3/H3 (bottom) ChIP-qPCR measurements, relative to the lowest value, from **(c)** BJAB (n=5) and **(d)** BL41 (n=6) biological replicates. Positions of the 3' splice sites and qPCR amplicons along the gene are displayed as vertical grey lines and arrowheads, respectively. All positions are drawn to scale, except for a gap in the second intron and the second point at each amplicon, which is offset to the right to facilitate comparison.