
A multiprotein occupancy map of the mRNP on the 3' end of histone mRNAs

LIONEL BROOKS III,¹ SHAWN M. LYONS,² J. MATTHEW MAHONEY,¹ JOSHUA D. WELCH,³ ZHONGLE LIU,¹ WILLIAM F. MARZLUFF,² and MICHAEL L. WHITFIELD¹

¹Department of Genetics, Dartmouth Geisel School of Medicine, Hanover, New Hampshire 03755, USA

²Integrative Program for Biological and Genome Sciences, University of North Carolina, Chapel Hill, North Carolina 27599, USA

³Department of Computer Science, University of North Carolina, Chapel Hill, North Carolina 27599, USA

ABSTRACT

The animal replication-dependent (RD) histone mRNAs are coordinately regulated with chromosome replication. The RD-histone mRNAs are the only known cellular mRNAs that are not polyadenylated. Instead, the mature transcripts end in a conserved stem-loop (SL) structure. This SL structure interacts with the stem-loop binding protein (SLBP), which is involved in all aspects of RD-histone mRNA metabolism. We used several genomic methods, including high-throughput sequencing of cross-linked immunoprecipitate (HITS-CLIP) to analyze the RNA-binding landscape of SLBP. SLBP was not bound to any RNAs other than histone mRNAs. We performed bioinformatic analyses of the HITS-CLIP data that included (i) clustering genes by sequencing read coverage using CVCA, (ii) mapping the bound RNA fragment termini, and (iii) mapping cross-linking induced mutation sites (CIMS) using CLIP-PyL software. These analyses allowed us to identify specific sites of molecular contact between SLBP and its RD-histone mRNA ligands. We performed *in vitro* crosslinking assays to refine the CIMS mapping and found that uracils one and three in the loop of the histone mRNA SL preferentially crosslink to SLBP, whereas uracil two in the loop preferentially crosslinks to a separate component, likely the 3'hExo. We also performed a secondary analysis of an iCLIP data set to map UPF1 occupancy across the RD-histone mRNAs and found that UPF1 is bound adjacent to the SLBP-binding site. Multiple proteins likely bind the 3' end of RD-histone mRNAs together with SLBP.

Keywords: RBPome; SLBP; histone mRNA; RNP; HITS-CLIP; CIMS; CLIP-seq; RIP-chip; RIP-seq; 3'hExo; UPF1; H2A.X; H2AFX

INTRODUCTION

In mammals, coordinate regulation of histone protein abundance and DNA replication is achieved primarily by post-transcriptional mechanisms that regulate the level of replication-dependent (RD) histone mRNAs (Harris et al. 1991). The metazoan RD-histone mRNAs are the only known eukaryotic cellular mRNAs that are not polyadenylated, ending instead in a conserved stem-loop (SL) (Marzluff et al. 2008). Their accumulation and degradation are coordinately regulated with DNA replication, with a 30–40-fold increase of mature RD-histone mRNA levels at the onset of S phase (Harris et al. 1991; Whitfield et al. 2000). In mammalian cultured cells, this increase in abundance is largely facilitated by an increase in 3'-end pre-mRNA-processing efficiency, which is a result of increased synthesis of stem-loop binding protein (SLBP) (Whitfield et al. 2000). SLBP binds to the SL at the 3' end of RD-histone mRNAs (Wang et al. 1996) and participates in all aspects of RD-histone mRNA metabolism

(Marzluff et al. 2008). After S phase or when DNA replication is inhibited, the RD-histone mRNA and SLBP are rapidly degraded (Harris et al. 1991; Zheng et al. 2003). The SL is also the *cis*-acting element that controls regulation of RD-histone mRNA half-life (Pandey and Marzluff 1987).

All RD-histone mRNAs are transcribed from approximately 80 distinct gene loci in humans, end in the conserved SL, and most of the genes are found in two large chromosomal clusters (Marzluff et al. 2002). The SL:SLBP complex facilitates stable interaction with several nuclear and cytoplasmic ribonucleoprotein (RNP) complexes including the U7 snRNP for processing (Dominski et al. 1999), TAP nuclear export complexes (Erkmann et al. 2005), 3'hExo (Yang et al. 2006), translation initiation factors (Sánchez and Marzluff 2002; Gorgoni et al. 2005; Cakmakci et al. 2008; von Moeller et al. 2013), and UPF1-associated complexes required for histone mRNA degradation (Kaygun and Marzluff 2005).

© 2015 Brooks et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Corresponding author: Michael.whitfield@dartmouth.edu

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.053389.115>.

It is not known whether SLBP has additional targets other than the RD-histone mRNAs. An analysis of SLBP's binding potential with an expansive ribonucleotide library has shown that SLBP demonstrates a wider in vitro RNA-binding repertoire than previously thought (Martin et al. 2012). Here we define the SLBP RNA interactome in asynchronously proliferating S3 HeLa cells. We used "high-throughput sequencing

of cross-linked immunoprecipitate" (HITS-CLIP) (Ule et al. 2005), which captures RNA:protein interactions in vivo and reveals the specific regions of the mRNAs that are bound (Fig. 1A). We developed a clustering method called coverage vector correlation analysis (CVCA) to sort genes into groups based on their read coverage distributions, which provides a comprehensive view of the RNA-binding protein

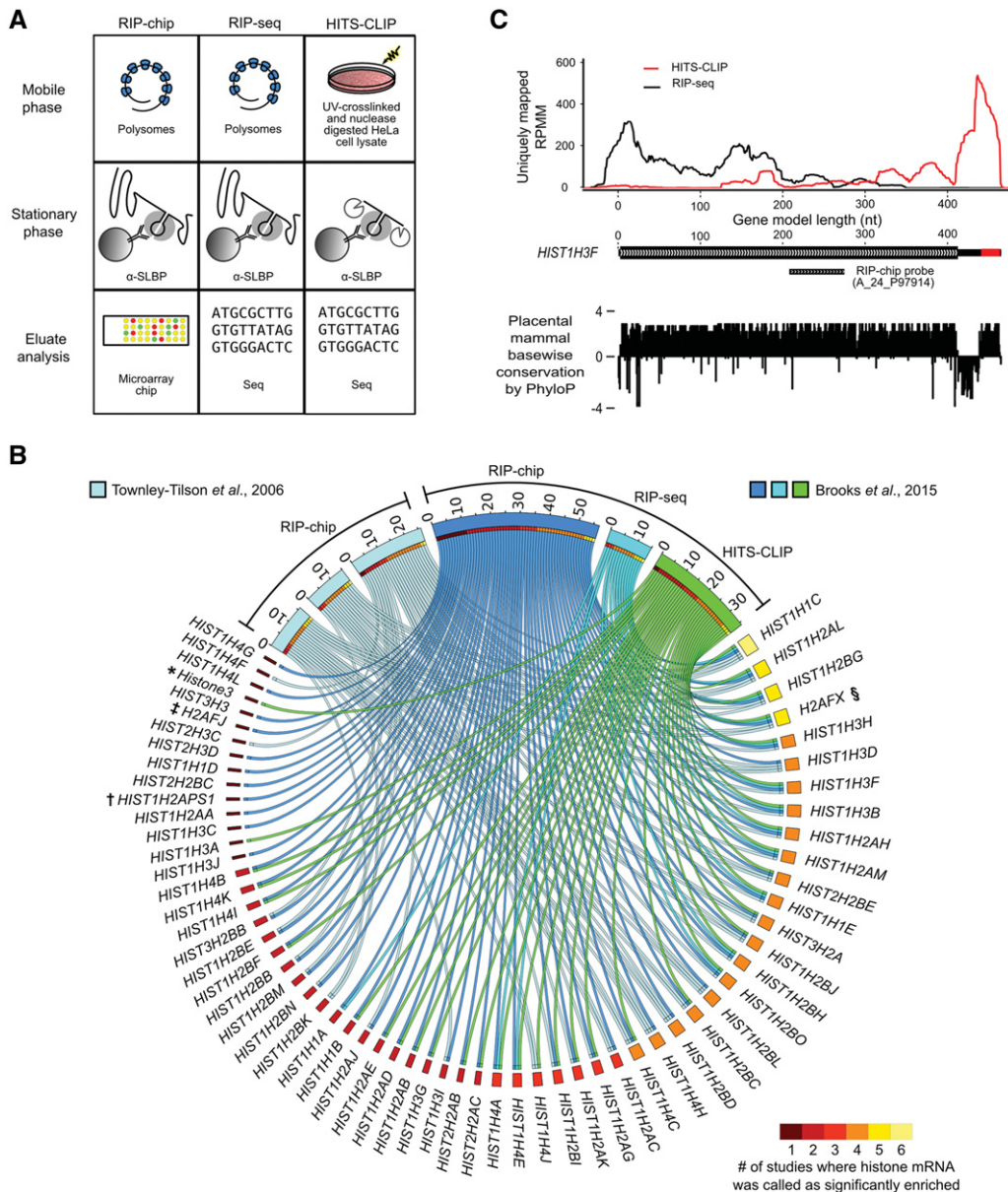


FIGURE 1. Genome-wide analysis of SLBP RNA ligands. (A) Affinity chromatography strategies for the three different genomic techniques that we utilized are shown. (B) Circos plot (Krzywinski et al. 2009) showing histone genes that have been identified as significantly enriched using RIP-chip, RIP-seq, and HITS-CLIP. *Histone3*, denoted by (*), indicates the accession number for this annotation in the Known Gene database (Hsu et al. 2006) is uc021yox.1. This is an Rfam entry that maps to the *HIST1H2BK* gene, which is misannotated. The *HIST1H2APS1* gene (†) is a pseudogene, which is not expressed, and the appearance of this gene in the significantly enriched probe set is certainly a microarray cross-hybridization artifact. The *H2AFJ* gene (‡) does not have a stem-loop, which indicates that this is also likely a microarray cross-hybridization artifact. The *H2AFX* gene (§) is a notable mRNA target of SLBP because it is bimorphic. (C) Read coverage distributions for the RIP-seq and HITS-CLIP experiments and the microarray probe location for the *HIST1H3F* gene. The stem-loop region is marked by a red block in the gene model schematic and is highly conserved across placental mammals (Pollard et al. 2010). HITS-CLIP is the only technique that provides direct evidence of the SLBP site. Coverage distribution is expressed as reads per million mapped (RPM).

landscape across all RD-histone mRNAs. We also developed a method to define the boundaries of nuclease-resistant RNA fragments by inferring the sites of micrococcal nuclease cleavage events from the HITS-CLIP data; the mapping algorithm is available as part of CLIP-PyL, an open-source software package that we developed. We analyzed existing iCLIP data, and mapped the binding sites for UPF1, a protein required for histone mRNA degradation, to the region between the SL and the stop codon (Zünd et al. 2013). Recent work suggests that crosslink sites can be identified by mapping small deletions, crosslink induced mutation sites (CIMS), in HITS-CLIP read data (Zhang and Darnell 2011). Mapping CIMS identified the uracil-rich loop region as the primary locus of crosslinking in our SLBP HITS-CLIP experiments.

There is a bimorphic histone gene *H2AFX*, which encodes the H2A.X protein variant. The *H2AFX* mRNA is alternatively processed to produce either a short (0.55 kb) RD-mRNA ending in the SL in S phase, or a long (1.6 kb) polyadenylated mRNA which is used to produce H2A.X protein outside of S phase (Mannironi et al. 1989). In addition to crosslinking of SLBP to the SL, we also find that SLBP contacts the *H2AFX* mRNA at both the SL and a conserved sequence 5' of the polyadenylation signal. Since SLBP and 3'hExo form a complex on the 3' end of histone mRNAs (Tan et al. 2013), SLBP and 3'hExo may both contribute to the accumulation of CIMS at the RD-histone mRNA SL. We propose that at least three distinct RNA-binding proteins can interact with the 3' end—SLBP and 3'hExo with the SL, and UPF1 at a site in the 3' UTR just upstream of the SL. We identify distinct nucleotides that determine cross-linking efficiency of SLBP and of 3'hExo to the SL fragment in vitro, consistent with crystallographic evidence (Tan et al. 2013) that a stable ternary complex containing SLBP and the 3'hExo forms on the SL.

RESULTS

Histone mRNA-sequence similarity affects both RIP-seq and RIP-chip output

Our initial RIP-chip analysis (Townley-Tilson et al. 2006) was inherently limited due to only a defined set of gene sequences being present on the microarray. The number of histone mRNA probes was limited and did not represent the entire set of histone mRNAs. We performed an expanded RIP-chip analysis with more probes that identified 72 histone mRNAs as potentially bound by SLBP (see Materials and Methods; Fig. 1B; Supplemental Fig. S1; Supplemental Table 2), although some fraction of these result from cross-hybridization due to extensive sequence similarity among different histone mRNAs.

A subset of samples analyzed by RIP-chip was also analyzed using RIP-seq. We performed a RIP-seq analysis (Fig. 1A) that showed mRNAs from a subset of histone genes are

bound to SLBP and expressed at high levels in HeLa cells (Fig. 1B; Supplemental Fig. S1). We identified enriched transcripts using edgeR (Robinson et al. 2010). Only a subset of RD-histone mRNAs was identified as significantly enriched (Fig. 1B; Supplemental Fig. S1; Supplemental Table 3). These correspond to the histone mRNAs that are most highly expressed in HeLa cells (Table 1; Yang et al. 2011; Djebali et al. 2012). In neither of these experiments did we observe significant amounts of any RNAs other than the RD-histone mRNAs.

The SLBP RIP-seq analysis does not yield reads that extend to the 3' ends of histone mRNAs (Fig. 1C). This is the result of failure to initiate reverse transcription at the histone mRNA 3' end during the library preparation procedure, as the highly stable SL structure renders the 3' end refractory to primer hybridization and initiation of reverse transcription. Truncation of histone 3' ends is also found in other high-throughput sequencing (HITS) data sets containing nonpolyadenylated histone mRNAs. We processed and aligned a public nonpolyadenylated mRNA-seq data set (Yang et al. 2011) and found that nonuniform distribution of read coverage across histone genes is also evident in those data (Supplemental Fig. S2).

Another challenge in analyzing histone mRNA expression is that three highly expressed genes in the HIST2 cluster are duplicated, as are two *HISTH4* genes in the Hist1 cluster. The *HIST2H4A–HIST2H4B*, *HIST2H3A–HIST2H3C*, and *HIST2H2AA3–HIST2H2AA4* genes are present in an inverted repeat containing two copies of all three genes (Braastad et al. 2004), and the *HIST1H4J–HIST1H4K* genes in the HIST1 cluster are repeated and expressed (Holmes et al. 2005). The HIST2 cluster repeat was listed as not expressed in some RNA-seq experiments (Yang et al. 2011), although they are clearly expressed at high levels as determined by S1 nuclease mapping analyses (Sullivan et al. 2009) and in the ENCODE data (Table 1; Djebali et al. 2012). Reads that map to these duplicated gene pairs cannot be uniquely mapped to a single gene. We built a coverage plot for the mRNA encoded by these pairs of duplicate loci. To assess the coverage for each duplicate pair, we selected only the reads that mapped solely to the two genes in the duplicate pair with 100% identity. This allowed us to build a coverage plot for each mRNA from the duplicate aligned reads (Supplemental Fig. S3), confirming that these genes are expressed at high levels in HeLa cells.

Isolation of nuclease-resistant SLBP RNP yields two discrete molecular weight populations

We used HITS-CLIP to capture and define the SLBP-binding site in vivo. HITS-CLIP (Licatalosi et al. 2008) involves amplifying the nuclease-resistant RNA fragments for high-throughput sequencing (Fig. 1A) using a ligated RNA adapter sequence as the site of annealing for the reverse transcriptase primer. We used micrococcal nuclease instead of RNase A because it is inactivated by EGTA, allowing control over

TABLE 1. Summary characteristics of RD-histone mRNAs

Gene	ENCODE HeLa poly(A) ⁻ (FPKM)	SLBP HITS-CLIP CIMS	UPF1 iCLIP untreated (RPMM)	UPF1 iCLIP + puromycin (RPMM)	CVCA CLASS
HIST1H2AA	0	–	0	0	B
HIST1H2AB	224	Y	2	6	A
HIST1H2AC	1632	Y	10	35	A
HIST1H2AD	64	Y	0	0	A
HIST1H2AE	1049	Y	19	58	A
HIST1H2AG	285	Y	11	58	A
HIST1H2AH	45	Y	24	44	A
HIST1H2AI	ND ^a	Y	3	5	AB
HIST1H2AJ	360	Y	8	27	A
HIST1H2AK	126	Y	3	4	A
HIST1H2AL	72	Y	11	96	A
HIST1H2AM	264	Y	20	57	AB
HIST2H2AB	162	Y	3	16	B
HIST2H2AC	590	Y	15	75	B
HIST3H2A	2	–	0	3	A
HIST1H2BA	0	–	0	0	B
HIST1H2BB	585	–	10	20	B
HIST1H2BC	1723	Y	16	58	AB
HIST1H2BD	831	Y	13	110	C
HIST1H2BE	<1	–	0	1	B
HIST1H2BF	133	Y	27	56	AB
HIST1H2BG	828	Y	19	53	A
HIST1H2BH	52	Y	2	29	B
HIST1H2BI	29	–	8	3	A
HIST1H2BJ	138	Y	12	47	A
HIST1H2BK	22	Y	26	154	C
HIST1H2BL	199	Y	0	5	AB
HIST1H2BM	68	–	8	11	AB
HIST1H2BN	642	–	0	11	AB
HIST1H2BO	403	Y	1	10	AB
HIST2H2BE	10	Y	20	38	C
HIST2H2BF	49	–	1	7	C
HIST3H2BB	<1	–	0	6	B
HIST1H3A	2	Y	4	16	A
HIST1H3B	1954	Y	11	69	B
HIST1H3C	2244	–	9	64	B
HIST1H3D	1835	Y	12	67	B
HIST1H3E	190	–	0	0	C
HIST1H3F	29	Y	15	54	AB
HIST1H3G	3	Y	6	13	A
HIST1H3H	572	Y	6	24	A
HIST1H3I	<1	–	4	28	BC
HIST1H3J	11	–	0	3	BC
HIST2H3D	84	–	0	5	A
HIST3H3	<1	–	0	0	BC
HIST1H4A	23	–	1	5	B
HIST1H4B	1095	Y	9	58	B
HIST1H4C	1043	Y	25	44	AB
HIST1H4D	654	–	6	9	B
HIST1H4E	1282	Y	16	66	B
HIST1H4F	<1	–	0	0	B
HIST1H4G	0	–	0	1	BC
HIST1H4H	414	Y	9	34	B
HIST1H4I	<1	Y	4	4	B
HIST1H4L	<1	–	0	1	B
HIST4H4	343	Y	3	17	B
HIST1H1A	1	–	15	31	A
HIST1H1B	34	Y	4	40	B
HIST1H1C	333	Y	12	75	B

Continued

TABLE 1. Continued

Gene	ENCODE HeLa poly(A) ⁻ (FPKM)	SLBP HITS-CLIP CIMS	UPF1 iCLIP untreated (RPMM)	UPF1 iCLIP + puromycin (RPMM)	CVCA CLASS
HIST1H1D	3	–	0	0	C
HIST1H1E	300	Y	13	67	B
HIST1H1T	<1	–	0	0	B
H2AFX	34	Y	146	90	C
HIST1H4K ^b	762	–	4	31	B
HIST1H4J ^b	259	Y	3	16	B
HIST2H4A/HIST2H4B ^c	904	ND	ND	ND	B
HIST2H3A/HIST2H3C ^c	696	ND	ND	ND	AB
HIST2H2AA3/HIST2H2AA4 ^c	876	ND	ND	ND	AB

(ND) Not determined, i.e., only uniquely aligning reads analyzed or were not measured on the DNA microarray.

^aExpressed at ~300 RPKM, misannotated in UCSC gene track which fused it with the HIST1H3H.

^bHIST1H4K/HIST1H4J are duplicate genes with minor mismatches in the 3' UTR.

^cHistone genes with duplicate copies present in the genome. Not analyzed for CIMS or UPF1 binding since only uniquely mapping reads were considered.

the activity of the nuclease (Yeo et al. 2009). The nuclease-digested SLBP RNPs were immunoprecipitated, treated with phosphatase, and the RNA 5' end labeled with [γ -³²P]-ATP using polynucleotide kinase. The radiolabeled RNPs were separated by SDS-PAGE and subsequently blotted to nitrocellulose for autoradiographic exposure (Fig. 2A). We constructed two size standards comprised of radiolabeled 30-bp SL probe (Williams and Marzluff 1995) crosslinked to *in vitro* translated SLBP in reticulocyte lysate or to HeLa S3 cell lysate (Fig. 2A, lanes 1,2). We observed a prominent complex at 45 kDa in the size standards (Fig. 2A, lanes 1,2). This is similar to the previously reported size of UV-crosslinked SL RNA:SLBP complex (Hanson et al. 1996). A similar complex was observed in our HITS-CLIP samples. In addition, a high-molecular weight complex that migrated between 80 and 260 kDa was present in the CLIP samples (Fig. 2A). The two complexes were present in similar amounts based on intensity of the radiolabeled CLIP complexes.

We prepared six HITS libraries from three CLIP experiments by excising the two distinct complexes at 30–80 kDa and 80–260 kDa from the nitrocellulose blotting membrane. The 30- to 80-kDa fraction will be referred to as “Low MW” (LMW) and the 80- to 260-kDa fraction as “High MW” (HMW). Two different MNase concentrations were used: 0.03 GU/ μ L or 3.0 GU/ μ L MNase. Despite the 100-fold difference in nuclease amounts, the qualitative pattern observed on the gel was the same (Fig. 2, lanes 3,4). As a control, we also performed one mock CLIP using α -GFP and analyzed the same two MW ranges. RNA isolated from the LMW and HMW regions were used to construct six HITS-CLIP-sequencing libraries. Each library was sequenced on an Illumina GA IIx instrument and yielded more than 21 million reads (Table 2). Low-quality base calls were removed with Sol-exaQA (Cox et al. 2010) and reads with identity to the sequencing adapters were trimmed using the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/).

The two nuclease-resistant fractions are not a result of differences in RNA fragment length

We hypothesized that the differential electrophoretic mobilities of the HMW and LMW α -SLBP HITS-CLIP fractions might be due to differences in the length of the crosslinked RNA fragments. Most of the fragments were larger than the 36-bp read length in the samples that were digested with 0.03 GU/ μ L MNase. However, comparison of total reads in the samples that were digested with 3.0 GU/ μ L MNase showed that the median read length was 4 bp longer in the LMW fraction (median length 27 bp) relative to the HMW fraction (median length 23 bp) (Fig. 2B). This analysis suggests that RNA fragment length did not cause the difference in electrophoretic mobility in the HMW and LMW α -SLBP HITS-CLIP fractions. It is not clear what causes the appearance of HMW fraction, but it may represent crosslinked complexes containing additional proteins in addition to SLBP.

We also examined the read-length distribution for the subset of reads that contain the SL motif (Fig. 2B). The SL motif was identified using a regular expression search for the degenerate 16-nucleotide (nt) core of the SL consensus sequence (Williams and Marzluff 1995). The core SL consensus is 16 nt in length and therefore should be fully captured by our 36-bp sequence reads. This analysis showed a striking difference between the size distribution of reads that contain a SL and those that do not. We found that the median read length for SL-containing reads was between 25 and 30 nt for the SLBP LMW HITS-CLIP data sets, regardless of nuclease concentration (Fig. 2C). Since the SL is at the 3' end of the mRNA, cleavage at the border of the SLBP-RNA complex will result in short fragments about this size. The median length of sequence reads that do not contain the SL was 36 bp, indicating that most of the fragments were >36 bp in length. These reads primarily map to the 3' UTR and 3' end of the coding regions of the histone genes. Some of these reads

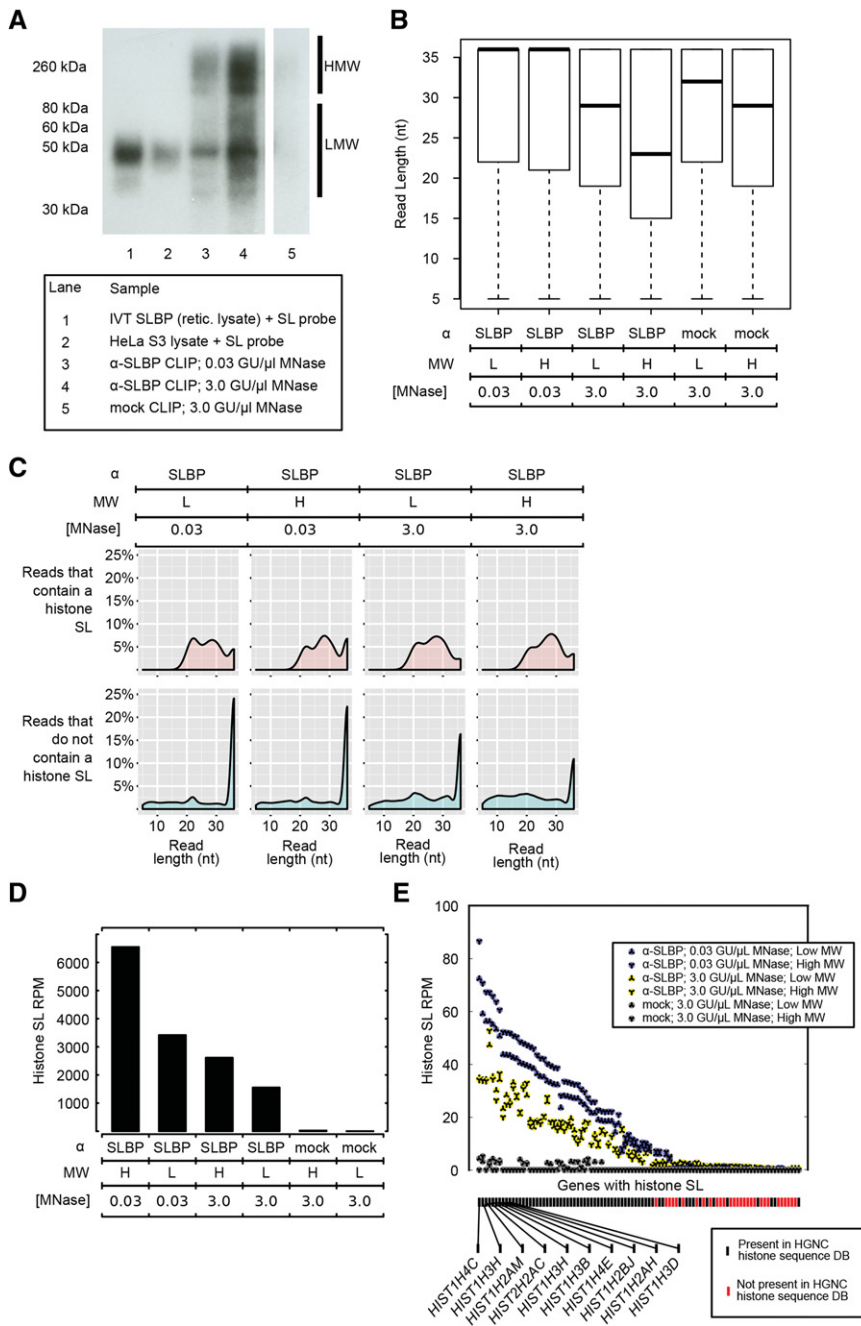


FIGURE 2. Nuclease-resistant histone mRNA stem-loop fragments were prominent in SLBP HITS-CLIP data. (A) Autoradiograph of P^{32} -labeled crosslinked ribonucleoprotein complexes that were isolated during SLBP HITS-CLIP library preparation. Both the experiment and negative controls are on the same gel and are the same exposure, but several intervening lanes were removed. (B) The probability density plots show the sequencing read fragment length distribution of reads that contain a stem-loop motif. (C) Distribution of reads that either contain or do not contain the histone SL in the SLBP HITS-CLIP data. (D) The bar graph shows the number of reads per million (RPM) containing a SL motif sequence for each of the HITS-CLIP libraries that we generated. GFP libraries did not contain an appreciable number of SL-containing reads. (E) Reads that contain SL motifs were mapped to the genome, and the number of reads per million (RPM) is shown for each genomic locus that was detected in each of our HITS-CLIP libraries. Shown are the 95 SL sequences that had at least one read in the HITS-CLIP data. The reads that fall into histone HGNC gene models are shown in black and those not in HGNC gene models are in red. The SL-containing reads primarily mapped to histone genes (shown in color bar).

contain partial SL sequences, suggesting that the 5' side of the SLBP/stem-loop complex is not always cleaved by the MNase. Notably, we found similar read size distributions for the SL-containing reads in the HMW HITS-CLIP data sets, where the SL reads were centered at ~ 27 nt while non-SL fragments were longer than 36 nt. This indicates that SLBP protects ~ 27 nt of the histone SL in the crosslinked RNPs.

SLBP HITS-CLIP enriches for transcripts containing the histone SL consensus

The HITS-CLIPs performed with the α -SLBP antibody resulted in a 60-fold increase in the number of reads containing the histone SL per million reads over the α -GFP HITS-CLIPs (Fig. 2D). The high molecular weight fractions consistently yielded the highest proportion of histone SL-containing reads compared with the low molecular weight fractions from the same nuclease conditions (Fig. 2D). Notably, the lower nuclease concentration (0.03 GU/ μ L MNase) yielded the highest proportion of SL-containing reads in both molecular weight fractions, suggesting that the higher MNase treatment (3.0 GU/ μ L MNase) may have increased the probability of cleavages internal to the SL, resulting in a loss of some sequences containing the SL.

We assessed enrichment of all annotated transcripts as well as unannotated transcribed regions. We generated a custom annotation track that included all known genes in RefSeq, GENCODE, and the UCSC known gene set, as well as mapped sequence read clusters containing >30 mapped sequence reads representing novel transcribed portions of the genome in our data (see Materials and Methods). We then used the edgeR bioconductor package to call significantly enriched transcripts ($P < 0.05$) (Robinson et al. 2010). This analysis identified 41 transcripts as significantly enriched in the HITS-CLIP data (Supplemental Table 4). The transcripts were predominantly comprised of histone mRNAs (Fig. 1B; Supplemental Fig. S1). The nonhistone mRNAs that were called

TABLE 2. BWA alignment results for SLBP HITS-CLIP and RIP-seq data

Method and condition	Total reads	Number of reads removed during 5' adapter clipping	Number of reads removed during 3' adapter clipping	Number of reads remaining after adapter clipping	Uniquely mapped reads	MultiMapped reads	Unmapped reads
HITS-CLIP; α -GFP; 3.0 GU/ μ L Mnase; 80–260 kDa	22,780,067	382,551	2,338,633	20,058,883	4,458,403	14,534,112	1,066,368
HITS-CLIP; α -GFP; 3.0 GU/ μ L Mnase; 30–80 kDa	23,163,618	266,102	956,488	21,941,028	2,635,484	18,447,742	857,802
HITS-CLIP; α -SLBP; 3.0 GU/ μ L Mnase; 80–260 kDa	22,414,650	276,453	3,656,159	18,482,038	3,275,599	14,655,903	550,536
HITS-CLIP; α -SLBP; 3.0 GU/ μ L Mnase; 30–80 kDa	23,394,224	179,503	2,651,984	20,562,737	1,714,177	18,306,286	542,274
HITS-CLIP; α -SLBP; 0.03 GU/ μ L Mnase; 80–260 kDa	21,537,267	621,140	4,963,824	15,952,303	3,013,496	11,685,383	1,253,424
HITS-CLIP; α -SLBP; 0.03 GU/ μ L Mnase; 30–80 kDa	22,589,973	677,054	5,100,659	16,812,260	2,107,569	13,148,862	1,555,829
RIP-seq; α -Flag	8,950,845	342,914	302,518	8,305,413	512,278	7,635,740	157,395
RIP-seq; α -SLBP + peptide	7,018,929	267,480	190,255	6,561,194	417,270	6,071,376	72,548
RIP-seq; no antibody	6,430,041	179,420	560,928	5,689,693	361,231	5,263,029	65,433
RIP-seq; α -SLBP	10,953,180	388,764	318,612	10,245,804	987,377	9,050,954	207,473
RIP-seq; α -SLBP	13,296,202	605,985	441,379	12,248,838	812,773	11,240,939	195,126
RIP-seq; α -SLBP	7,454,272	175,246	311,874	6,967,152	368,174	6,516,782	82,196
Yang poly(A) ⁻ RNA-seq; SRA ID: SRR067391	10,340,237	614,554	208,739	9,516,944	1,772,900	6,341,756	1,402,288
Yang poly(A) ⁻ RNA-seq; SRA ID: SRR067391	20,848,388	722,042	402,551	19,723,795	3,740,079	15,209,282	774,434
Yang poly(A) ⁻ RNA-seq; SRA ID: SRR067391	14,454,805	1,808,525	262,470	12,383,810	2,366,715	9,363,753	653,342

by the analysis did not contain a predicted stem-loop, were largely devoid of CIMS, and were not found as significant in any of our RIP-chip or RIP-seq analyses, indicating they are background. Hundreds of clustered reads were observed that aligned to other regions of the genome, regardless of whether or not they were annotated to transcripts or not, but none were significantly enriched in the RIP-seq or HITS-CLIP EdgeR analysis (Supplemental Tables 3, 4). Thus, we conclude that only RD-histone mRNAs cross-linked to SLBP. The histone transcripts found to show CIMS in the SLBP HITS-CLIP data also correspond to the histone mRNAs expressed in HeLa cells (Table 1).

To determine whether other transcribed regions of the genome contained potential SLBP-binding sites, we determined the genomic coordinates of all SL-containing sequences. We scanned the human genome using the SL consensus regular expression and then counted the number of SLBP HITS-CLIP reads that uniquely aligned to those intervals. We found 95 genomic loci that contained a SL and overlapped with at least one SLBP HITS-CLIP mapped read (Fig. 2E). Sixty-six of these loci are histone genes as defined in the Human Genome Nomenclature Committee's Histone Sequence Database (Marino-Ramirez et al. 2011) (HGNC-HSDB) (<http://www.genenames.org/genefamilies/histones>)

and these correspond to the most abundant fragments in the HITS-CLIP (Fig. 2E, black hashes). This database lists 74 RD-histone genes and 13 pseudogenes; some of the RD-histone genes are not expressed in HeLa cells and none of the pseudogenes are expressed (Supplemental Table 5). Others that were expressed at low levels were not identified as significantly enriched in the edgeR analysis. Only a single nonhistone genomic interval contained a match to our SL search and showed low read coverage (<7 reads per million) in more than one SLBP HITS-CLIP sample (located at nucleotides 24,650,740–24,650,763 of chromosome X in the human genome assembly hg19). Subsequent RT-PCR analysis of HeLa S3 total RNA indicated that it is not expressed in HeLa cells, and is likely to be a sequencing error resulting in a read mapping artifact (data not shown). None of the 28 remaining SL sequences in the genome showed read coverage above background (Fig. 2E, red hashes), and many of these correspond to histone genes not expressed in HeLa cells.

Two distinct HITS-CLIP coverage vector shapes are evident at RD-histone mRNA 3' UTRs

Visual inspection of individual histone gene models showed heterogeneity in HITS-CLIP read coverage shape. Some histone genes had read coverage primarily over the SL in the 3' UTR (e.g., *HIST2H4*, *HIST1H3B*, Fig. 3A). Others also had substantial coverage in the rest of the 3' UTR extending into the 3' end of CDS (e.g., *HIST1H2AG*, Fig. 3A). We developed a method to analyze coverage patterns across gene models and applied it to interrogate HITS-CLIP read coverage across all annotated histone genes. Our technique, termed coverage vector correlation analysis (CVCA), is comprised of three main steps: (i) calculation of coverage vectors for a set of gene models, (ii) calculation of pairwise coverage vector correlation scores, and (iii) clustering of coverage vectors by pairwise correlation scores.

A coverage vector is calculated by counting the number of mapped reads per million mapped reads across all bases in a given gene model. There are 116 gene models in the HGNC-HSDB (87 RD histone genes and 29 non-RD histone genes). Since we sequenced six HITS-CLIP libraries (four experimental and two control), the set of coverage vectors is comprised of 696 coverage vectors. We then processed the coverage vectors for calculation of pairwise correlation scores by removing empty vectors and interpolating all remaining coverage vectors to the same length. Spectral clustering of the pairwise coverage vector correlations (Fig. 3B) revealed three distinct clusters with different coverage vector shapes (Fig. 3C). Cluster A was comprised of coverage vectors with read coverage in the 3' end of the CDS and 3' UTR, including the SL. Cluster B was comprised of coverage vectors with read coverage primarily over the SL sequence. Cluster C had low coverage, indicating transcripts from these genes were not bound (Table 1; Supplemental Table 6). We calculated the proportion of the 116 histone gene models that were

assigned to each cluster (Fig. 3D). In the mock GFP antibody CLIPs the majority of the coverage vectors were in cluster C or empty (Fig. 3D). The histone coverage vectors in the α -SLBP HITS-CLIPs for the RD-histone genes expressed in HeLa cells were in cluster A or B. The remaining α -SLBP HITS-CLIP coverage vectors were empty or assigned to cluster C (22 pseudogenes, 20 replication-independent histone genes, and RD histone genes not expressed in HeLa cells; Fig. 3D). The α -SLBP HITS-CLIPs with lower nuclease concentration show a higher proportion of coverage vectors assigned to cluster A (reads aligning to the CDS and 3' UTR) and α -SLBP CLIPs with higher nuclease concentration showed a higher proportion of coverage vectors in cluster B (reads aligning primarily to the 3' UTR over the known SLBP-binding site) (Fig. 3D). We investigated several factors that could have caused the variation in coverage vector shapes between cluster A and B. These factors included transcript abundance, distance between the stop codon and stem-loop, variations in the stem-loop sequence, variation in the 3' UTR, the histone downstream element (HDE) location, or variation in the HDE sequence. However, none of these factors explained these distinct coverage vector shapes.

Mapping inferred nuclease cleavage sites precisely defines the SLBP-binding site

We reasoned that mapping nuclease cleavage sites would reveal the boundaries of the protected RNA fragments more precisely than the read coverage map. In the HITS-CLIP data, the 5' and 3' termini of all crosslinked RNA fragments result from either MNase cleavage or from transcript termini (for example, the histone mRNA 3' end). Therefore, we are able to map the MNase cleavage sites by mapping the termini of the RNA fragments (Fig. 4A). We identified the 3' terminus of each RNA fragment by searching the reads for the 3' cloning adapter sequence. Only the subsets of reads with an identifiable 3' RNA adapter were mapped (Fig. 4A). The 3' terminus corresponds to the nucleotide immediately preceding the 3' adapter. The 5' terminus of each RNA fragment is the first nucleotide of each read, since all sequence reads initiate from the 5' adapter.

The “cleavage site vectors” could then be calculated by counting the number of mapped read termini per million mapped reads across all bases in a given gene model (Fig. 4B). We applied CVCA analysis to cluster the cleavage site vectors based on their shapes. The cleavage vectors clustered into two coverage shape classes, with most “cleavage” events mapping to the 3' end of the transcript (Fig. 4C,D, clusters A* and B*). The difference between the clusters A* and B* is minor: the average coverage vector for cluster B* has an additional small density of cleavage events internal to the SL sequence (Fig. 4D). The region flanked by the mapped cleavage sites corresponds to the region of the transcripts that are bound and protected by SLBP.

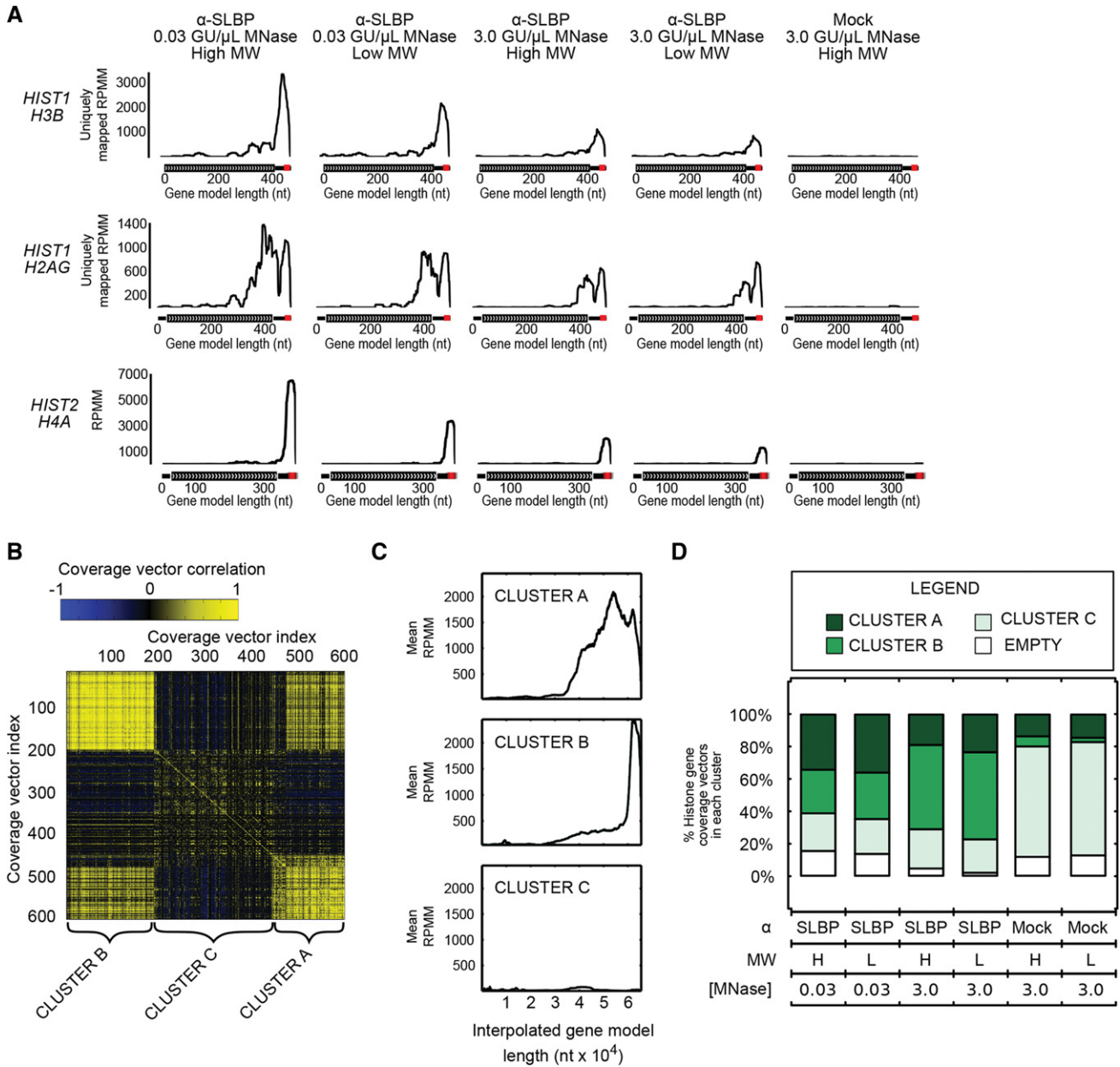


FIGURE 3. HITS-CLIP read coverage maps for representative histone mRNA gene models. (A) Representative coverage plots for select RD histone mRNAs are shown. The uniquely mapped reads for the *HIST1H3B* and *HIST1H2AG* genes are also shown. Multimapped reads are shown for *HIST2H4A*, which is a duplicated gene with reads mapping to both gene copies of the gene in the genome sequence. The coverage distributions are expressed as reads per million mapped (RPM). (B) To identify commonly observed coverage vector shapes across the histone mRNAs, we developed coverage vector correlation analysis (CVCA), which was applied to the 696 coverage vectors derived from 116 histone gene models in the six different conditions. Symmetric correlation map shows each of the 696 coverage vectors plotted against one another and the correlation coefficient is indicated in the heatmap. We find three major clusters that are labeled A, B, and C. (C) The mean coverage vector for each group is displayed. Average coverage vector for the three clusters found by CVCA show two distinct shapes and one comprised of background level reads. The vectors in group C are either histone genes not expressed, non-replication-dependent histone genes, or correspond to vectors from the GFP controls. (D) Proportion of histone gene models found in each cluster from C is shown across HITS-CLIP conditions. This quantifies the proportion of genes that are most similar to a given coverage shape.

We calculated the coverage (Fig. 4E) and cleavage rates (Fig. 4F) at every base across all RD-histone transcripts and found that cleavage events are enriched at the flanks of the conserved SL motif (Fig. 4F). The 3' cleavage peak at the terminal end of the histone mRNA was much sharper than the

cleavage peak 5' of the SL, which resulted in a taller peak. This cleavage site occurs naturally in the cell, and was not due to MNase cleavage. We also detected a modest peak on the 3' side of the SL, which is distinct from the transcript terminus. It is located 5' of the invariant CC nucleotides of the GC

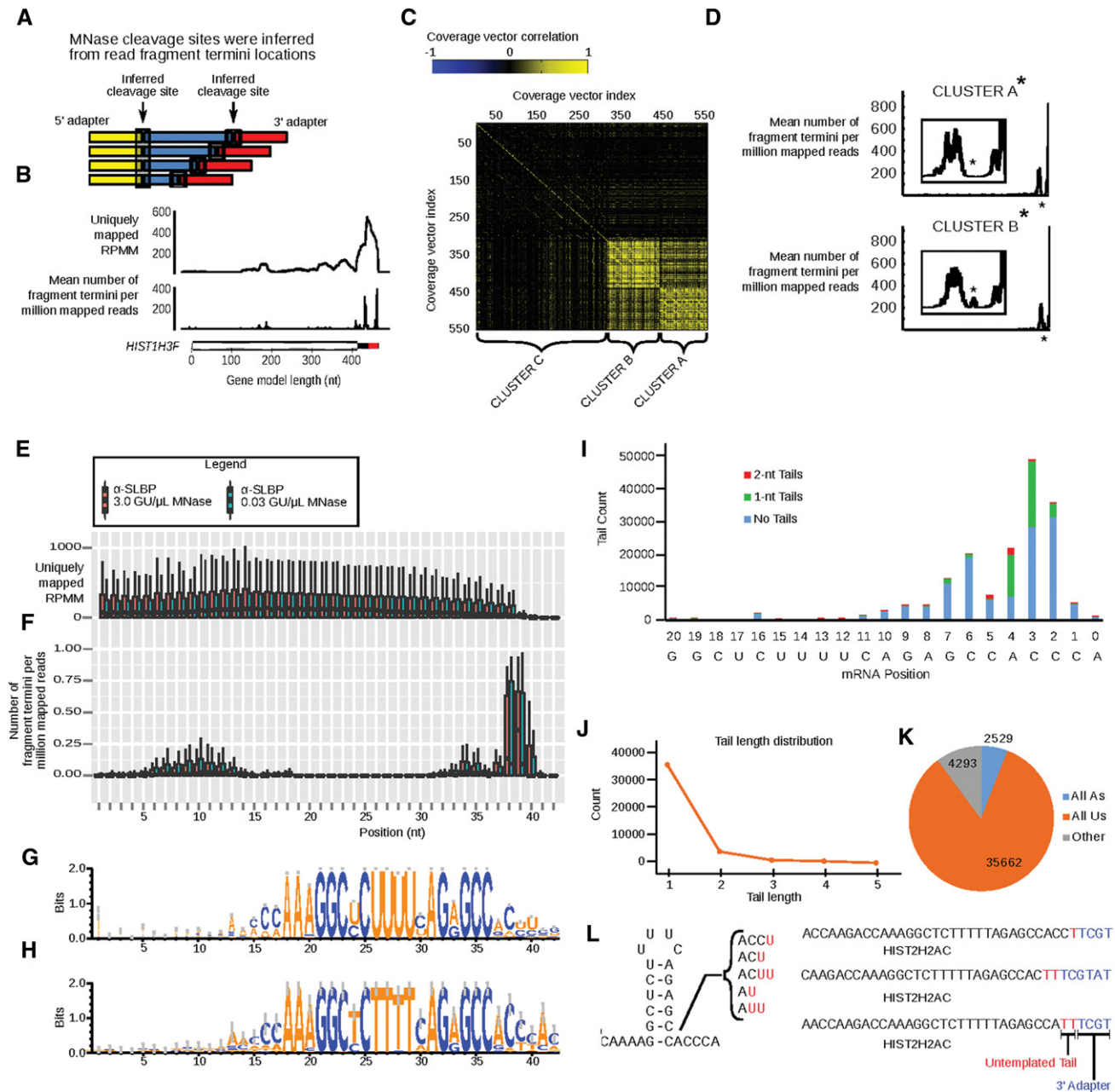


FIGURE 4. Inference of nuclease cleavage sites from HITS-CLIP data maps the boundaries of the SLBP RNP. (A) Cleavage sites were inferred from mapping read termini as depicted. Only termini for which 5' and 3' cleavage sites could be mapped were retained and graphed. The sequencing was 36-bp single end and occurred from the 5' adaptor. (B) A representative coverage plot is shown with coverage expressed as unique reads per million mapped (RPMM) across the *HIST1H3F* locus (upper panel). A plot of the inferred cleavage sites is also shown for comparison, where the number of cleavage sites per million mapped reads is shown. The inferred cleavage sites precisely flank the histone SL, which is indicated in the gene model schematic as a red block. (C) We used CVCA analysis to group histone genes by the similarity of the inferred cleavage site distributions at each locus using the same HGNC gene models shown in Figure 3. (D) The mean cleavage site vectors show that the cleavage peak is most sharply demarcated at the 3' region flanking the SL and that there was only a subtle difference between histone genes in clusters A and B, which is a density of cleavage sites internal to the histone SL (see insets). (E,F) We quantified the number of unique RPMM (E) and cleavage rates (F) for each nucleotide in the SL motifs of 75 histone genes for the low MW band (blue) and HMW weight band (orange). (G,H) Seqlogs (WebLogo 3.3) were generated from a multiple alignment of SLBP HITS-CLIP reads containing the histone SL motif (G) and for the DNA sequence for 3' end of the histone genes (H). Note that the x axis is the same for the boxplots (E,F) and the seqlogs (G,H) to display the sequence composition of the nuclease-resistant histone SL RNA fragments. (I) We used the AppEnD tool (Welch et al. 2015) to identify nontemplated tails on the 3' ends of histone RNA molecules present in HITS-CLIP reads. To avoid calling sequencing errors as tails, only reads that extended at least 4 nt into the 3' adaptor were analyzed. The bar chart shows abundances of unmodified tails (blue), 1-nt tails (green), and 2-nt tails (red) by position from 3' end for all histone mRNAs. (J,K) The length and composition of the nontemplated tails were determined. (L) Examples of types of reads found on the *HIST2H2AC* RNA are shown. The 3' end of the RNA that mapped to the genome (black) is followed by any nontemplated nucleotides (red) and the 3' linker (blue).

base pairs in the SL (Fig. 4F). This modest peak likely represents a site of MNase cleavage and corresponds to the 3' boundary of the portion of the transcript that is bound and protected by SLBP.

We examined the sequence composition of the protected RNA fragments that contained a SL motif and compared it to the sequence composition of the genomically encoded SL motifs (Fig. 4G,H). We noted that an overrepresentation of U's in the nucleotides after the SL was apparent in the consensus sequence derived from the HITS-CLIP RNA fragments (compare Fig. 4G,H). Addition of uridines at the 3' end of histone mRNA is the first step in 3'-5' degradation of histone transcripts (Slevin et al. 2014). Moreover, many histone mRNAs that are not undergoing degradation during S phase have lost 1–2 nt from their 3' end, which have been replaced by one or two uridines to restore the length of the histone mRNA (Welch et al. 2015). These are precisely the positions at which the U's are overrepresented in the HITS-CLIP RNA fragment consensus sequence (positions 39 and 40 in Fig. 4G,H). We used the AppEnD tool (Welch et al. 2015) to define the 3' end of each transcript and directly search for nontemplated 3' additions in the SLBP HITS-CLIP reads. This analysis revealed many nontemplated additions (Fig. 4I–L), the vast majority of which were additions of one or two uridines at the A or AC after the stemloop.

H2AFX mRNA contacts SLBP at two distinct sites

The *H2AFX* gene encodes a histone mRNA that has two distinct isoforms, both of which contain a SL. During S phase, a short isoform (~0.55 kb) that ends at the histone SL predominates whereas outside of S phase, a longer 1.6-kb isoform—which is polyadenylated and replication-independent—is the predominate isoform (Fig. 5A; Mannironi et al. 1989; Bonner et al. 1993). The H2A.X protein, which is encoded by the *H2AFX* gene comprises ~10% of total H2A protein in mammalian cells (West and Bonner 1980); therefore, it must be synthesized in large quantities during S phase. The longer *H2AFX* mRNA isoform predominates in G1 cells and when DNA replication is stopped during genomic replication stress, such as during treatment with the DNA synthesis inhibitor hydroxyurea (HU) (Rogakou et al. 1998; Dickey et al. 2009). The short transcript isoform ends in the SL as a result of cleavage between the SL and the HDE. Both the SL and HDE are retained in the 3' UTR of the long transcript isoform. The two transcript isoforms are clearly distinguishable in a comparison of public RNA-seq data sets from poly(A)⁻ and poly(A)⁺ fractions of HeLa cell transcriptomes (Fig. 5B,C; Yang et al. 2011; Djebali et al. 2012).

Both *H2AFX* isoforms contain the SL and are capable of binding SLBP, but it is not clear whether SLBP is bound to both forms of the *H2AFX* mRNA. SLBP levels are very low outside of S phase (Whitfield et al. 2000), but the polyadenylated *H2AFX* mRNA is stable and some will likely be syn-

thesized in G1 and persist into S phase when SLBP levels are high (Bonner et al. 1993). RNA immunoprecipitation experiments and our α -SLBP RIP-seq data suggest that SLBP binds both isoforms, as sequencing read coverage extends downstream from the HDE into the part of the 3' UTR that is present only in the long isoform (Fig. 5D). SLBP immunoprecipitated both forms of *H2AFX* mRNA from mouse F9 cell polyribosomes (Fig. 5H). These data show that SLBP can bind to both mRNA isoforms and that the poly(A) isoform binds SLBP with at least a similar affinity as the short isoform (as inferred from more efficient recovery of the poly(A) form of the mRNA). However, these results could result from binding of SLBP to the SL after lysis of the cells and before immunoprecipitation (Mili and Steitz 2004). To address this question we interrogated our SLBP HITS-CLIP data for reads that map to *H2AFX* mRNA. There is a coverage peak that corresponds to the histone SL motif (Fig. 5E), and we also observed a second SLBP HITS-CLIP coverage peak near the polyadenylation signal.

The 1D single base pair deletion rate identifies sites of protein crosslinking and indicates that it is a bona fide crosslinked induced mutation (CIM) site (Fig. 5E; Zhang and Darnell 2011). To identify CIMS, we aligned preprocessed reads using BWA (Li and Durbin 2009), which is an open-source indel-aware fast short read alignment program. This allowed us to analyze the distribution of single nucleotide deletion (1D) rates for each position across all bound RNA elements. The 1D rate at this downstream site was similar to the rate at the SL and is indicative of a bona fide CIM site (Fig. 5E). This site maps to the only region other than the SL in the 3' UTR of the *H2AFX* mRNA transcript that is highly conserved across placental mammals (Fig. 5E), suggesting there has been selective pressure to maintain the sequence (Fig. 5G). We cannot conclude that SLBP is bound directly to this sequence. It is likely that the CLIP procedure precipitates SLBP together with any proteins complexed with SLBP, and the crosslinks observed could be due to another protein that interacts with the conserved sequence and SLBP. It is also possible that the 3' end of polyadenylated *H2AFX* is bound in a complex near the ribosome where it is close to SLBP on the SL, and this could result in crosslinking with SLBP.

Crosslink-induced mutation sites (CIMS) precisely identifies nucleic acid contact sites

We analyzed the distribution of CIMS across the RD-histone mRNAs. We observed a preponderance of 1D events mapping to the uridine-rich loop region (Fig. 6A,B). We implemented a bootstrapping technique to test the statistical significance of the observed enrichment of CIMS in the loop region. The resampling method emulates the previously published method for testing significance of putative CIMS (Zhang and Darnell 2011), which involves randomly

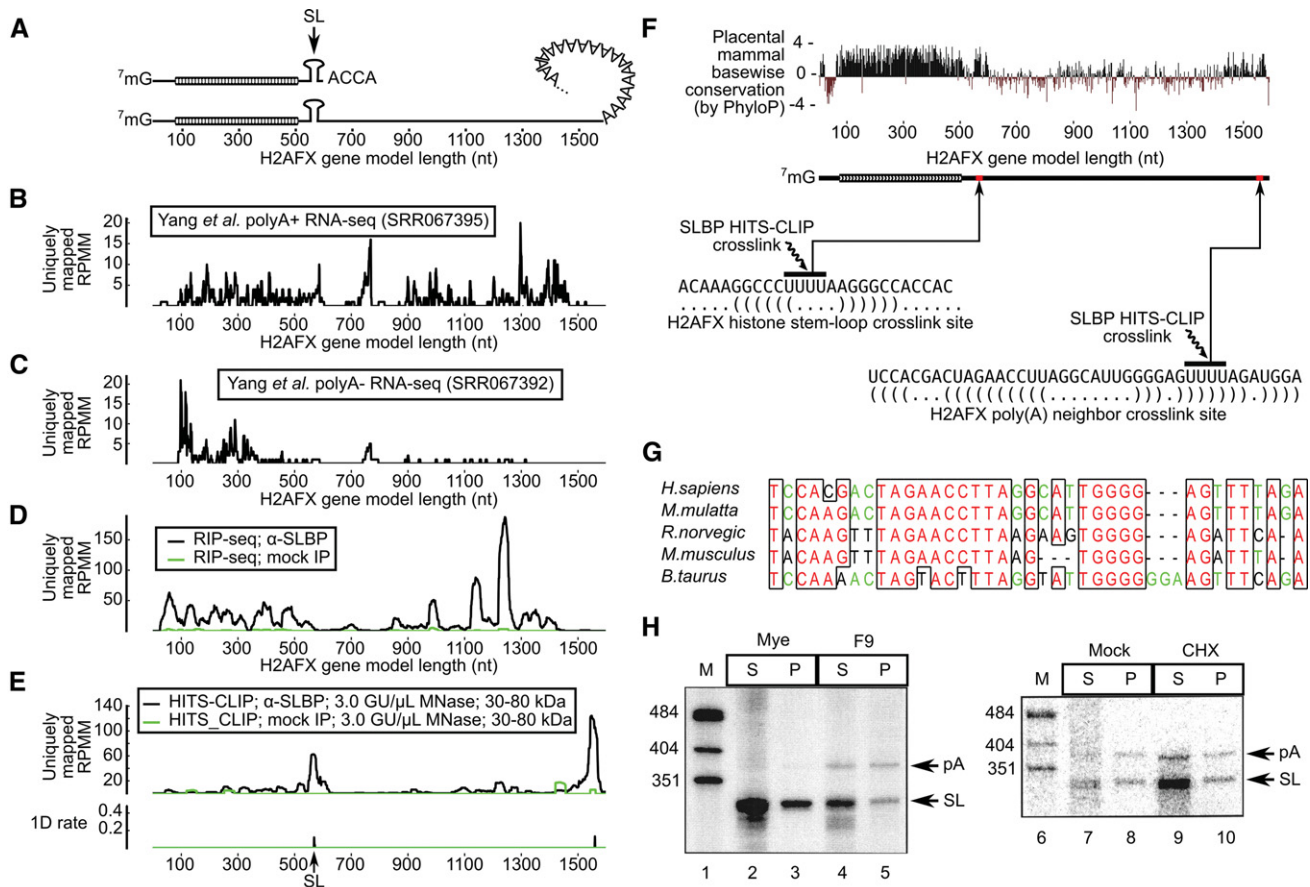


FIGURE 5. Proteins in the SLBP immunoprecipitate interact with H2AFX mRNA at two distinct sites. (A) Schematics of the short and long H2AFX mRNA isoforms are shown. (B,C) The RNA-seq data sets from Yang et al. (2011) produced by sequencing the (B) polyadenylated and (C) nonpolyadenylated fractions of HeLa cell mRNA isolates were aligned and displayed. Reads per million mapped reads (RPMM) is plotted for each nucleotide across the bimorphic H2AFX transcript. (D) Analysis of H2AFX mRNA read coverage in our SLBP RIP-seq data set. (E) Analysis of H2AFX mRNA read coverage in our SLBP HITS-CLIP data set. The *middle* panel shows the HITS-CLIP CIMS (1D rate) with a peak in the histone SL and a peak near the poly(A) site. Shown in the *lower* panel is the placental mammalian conservation which demonstrates conservation of the coding region, histone SL, and the region preceding the poly(A) site. (F) The sites of crosslinking deduced from the sites of indels in the SL and in the sequence adjacent to the poly(A) site are indicated. (G) Comparison of the sequences around the polyadenylation site in the H2AFX gene in different mammals showing the sites of crosslinking defined by the presence of Indels, and the conservation among mammalian species. (H) Total cell RNA from exponentially growing mouse myeloma cells (lanes 2,3) and F9 teratoma cells (lanes 4,5,7,8), or F9 cells treated with cycloheximide for 60 min (lanes 9,10) to freeze the ribosomes and stabilize the stem-loop RNA, were lysed and the histone mRNAs precipitated with anti-SLBP antibody. RNA was prepared from the supernatants (S) and the immunoprecipitates (P) and subjected to S1 nuclease mapping as precisely described (Whitfield et al. 2004). The probe was hybridized with total cell RNA, treated with S1 nuclease, and the protected fragments resolved by electrophoresis on an 8% polyacrylamide-7 M urea gel. Lanes 1 and 6 are markers. The SL form of H2AFX RNA protects a 333-nt fragment, and the polyadenylated form protects a 379 fragment as the probe extends 46 nt past the end of the SL, complementary to the H2AFX gene.

assigning the mutation offset among the reads to preserve any positional bias that is possibly introduced by sequencing errors. We find that 39-RD histone SL motifs have significant enrichment over the null ($P < 0.001$) (Table 1; Fig. 6C). However, since histone SL motifs contain a homopolymer run of four uridines from the UA base pair at the top of the stem into the loop region, it is impossible to discern which of the uridines is deleted (Fig. 6B). The BWA alignment program reports the deletion by assigning it to one of the terminal uridines in the homopolymer run, depending on the strandedness of gene to which the read is being aligned (Fig. 6A).

There are only three histone genes with a cytosine at the second position in the loop, which breaks the homopolymer run and would allow finer mapping of CIMS in the loop. Only one of the three genes, *HIST1H2BI*, is expressed in HeLa cells (Table 1). We compared the transcript abundance of *HIST1H2BI* to the transcript abundance of *HIST1H2BJ*, which is the most similar histone mRNA with a UUUA loop (Supplemental Fig. S1). The two transcripts have nearly identical abundance in the poly(A)⁻ RNA-seq data set (Fig. 6G). The HITS-CLIP data show a 3.75-fold decrease in the magnitude of the read coverage peak for the *HIST1H2BI* transcript with the loop UCUA variant

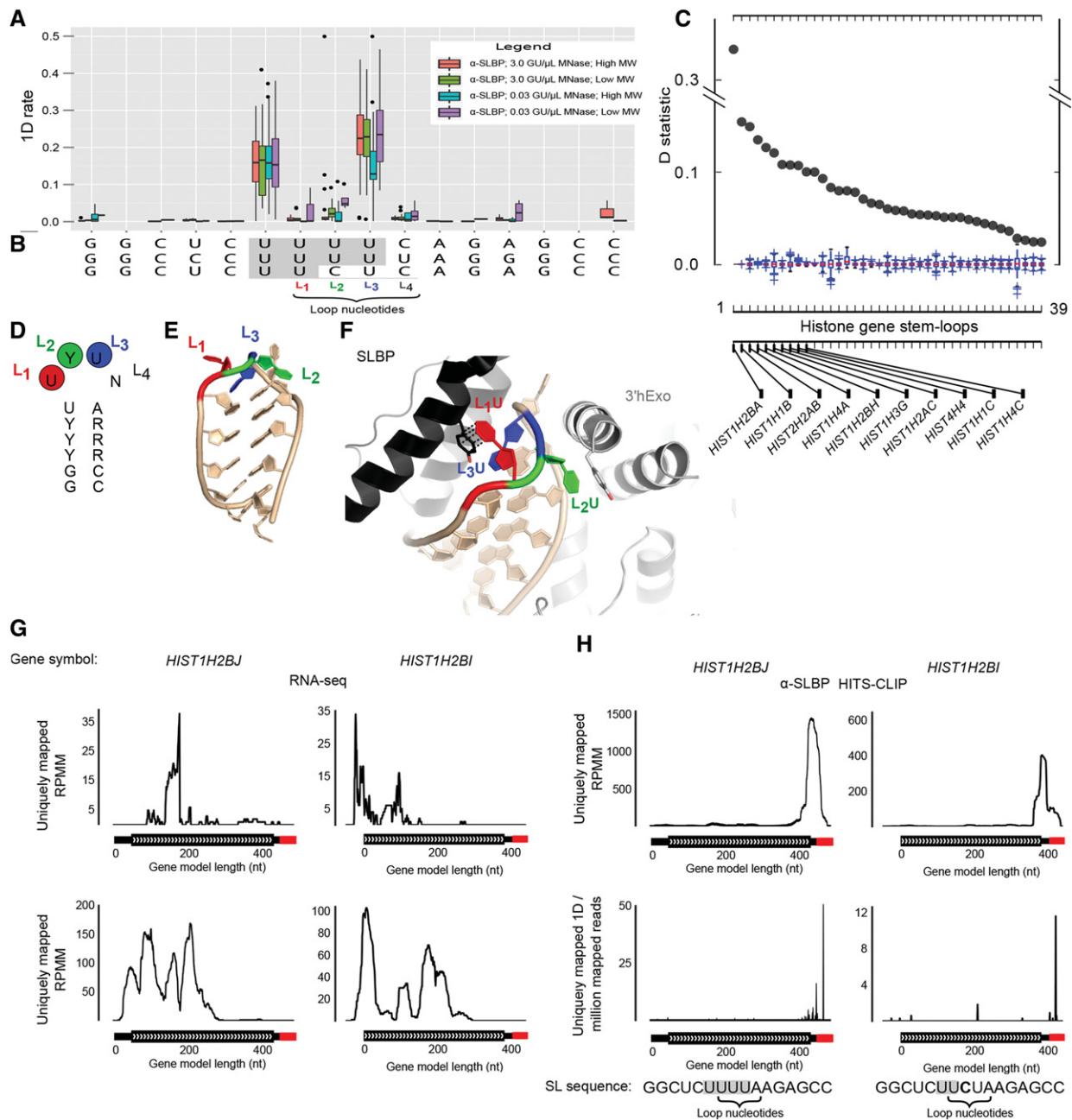


FIGURE 6. Identification of crosslink-induced mutations (CIMS) in histone mRNA stem-loops. (A) Single nucleotide deletion (1D) rates were quantified across 75 histone SL motifs. The distribution of those rates is summarized by a boxplot spanning each nucleotide in the degenerate motif. (B) The number of genomic occurrences of each stem-loop sequence is indicated by the pseudocolor plot alongside each stem-loop. If a 1D maps to a homopolymer run (highlighted in gray in panel B), then the BWA aligner ascribes the 1D to one of the terminal nucleotides in the run depending on whether the gene is on the plus or minus strand in the genome. (C) We implemented a previously published method (Zhang and Darnell 2011) to compute a test statistic (referred to here as “D-statistic”) to assess enrichment of 1Ds in the loop region. The D-statistic computed for each of the histone genes is depicted as a scatter plot. We calculated a P-value by computing a null distribution for the D-statistic at each stem-loop motif using the resampling method from the previously reported method (Zhang and Darnell 2011). The null distributions are displayed as boxplots. Thirty-nine of the histone stem-loops passed the significance cutoff ($P < 0.001$). (D) Highlighted are the three uracils in the stem-loop consensus sequence. (E,F) Those uracils are indicated in the stem-loop RNA fragment (E) that is a component of an SLBP:SL:3’hExo crystal structure (PDB4HXH) (Tan et al. 2013). (G) Coverage vectors for two histone genes with different loop sequences. There are three histone genes which contain a UCUA loop: *HIST1H2BI* (UCUA), *HIST1H4L* (UCUC), and *HIST1H3E* (UCUC). *HIST1H2BI* was expressed at similar levels to the adjacent gene *HIST1H2BJ* with a UUUU loop. In both public data sets (Yang et al. 2011; Djebali et al. 2012) and in our RIP-seq experiment, these genes were expressed at similar relative levels. The *top left* graphs show the coverage vector in the poly(A)-RNA-seq data set (Yang et al. 2011). The *bottom left* row shows the coverage in our SLBP RIP-seq data set. (H) The *top right* shows the coverage in our SLBP HITS-CLIP data set in the 3 GU/ μ L Mnase/High MW fraction. The *bottom right* shows the 1-bp deletions found in each gene model. Each vector is shown relative to the gene model along the x-axis with the 3’ UTR containing the histone SL indicated in red and CDS indicated in black. Quantification of CIMS in the *HIST1H2BJ* (UUUA) and *HIST1H2BI* shows accumulation of CIMS in the homopolymer run of uridines across the loop. In *HIST1H2BJ* the CIMS are located in the stretch of four uridines at the *top* of the stem and the first 3 nt of the loop, and the precise nucleotide crosslinked cannot be assigned. There are fewer CIMS in the *HIST1H2BI* with the majority associated with the UU at the *top* of the stem and first nucleotide in the loop (L_1). This most likely represents crosslinking to L_1 . A small number of CIMS were associated with the naturally occurring L_2C .

compared to HIST1H2BJ, which has a UUUU loop (Fig. 6H). This suggests that the HIST1H2BI sequence variant is cross-linked with lower efficiency. The CIMS detected in the HIST1H2BI loop occur in the UU that encompasses a nucleotide in the stem and the first nucleotide in the loop (L_1) (Fig. 6H). It is likely that this represents crosslinking at nucleotide L_1 in the loop to SLBP based on crystallographic data (Tan et al. 2013), which shows that SLBP may interact with L_1 and L_3 in the loop.

In the cell, SLBP and the 3' to 5' exonuclease ERI1 (also called 3'hExo) are bound simultaneously to the 3' end of mature histone mRNA. 3'hExo trims the processed histone mRNA 3' end, leaving 2–3 nt after the SL, and is necessary for rapid degradation of histone mRNA (Hoefig et al. 2013). The structure of the SLBP/SL/3'hExo complex suggests that both proteins contact uridines in the loop (Fig. 6D–F). SLBP interacts with nucleotides L_1 and L_3 through π -stacking with a tyrosine side chain, and 3'hExo interacts with L_2 of the loop through multiple side-chain interactions (Tan et al. 2013). This suggests that some of the indels observed in UUUC/A loop sequence might result from crosslinking to the 3'hExo at L_2 , while others result from crosslinking of SLBP which would result in a decrease in overall crosslinking to the UCUC/A loop variant. The 3'hExo migrates close to SLBP on PAGE and coimmunoprecipitates with SLBP bound to the SL (Dominski et al. 2003).

In vitro binding analysis identifies the primary sites of crosslinking in histone mRNA

To test this hypothesis, we compared binding and crosslinking of the SL to SLBP and 3'hExo, using a wild-type SL probe (SLWT) and a stem-loop probe that has the stem reversed (SL reverse stem [SLRS]) and does not bind SLBP (Williams and Marzluff 1995). We also analyzed SL probes with U-to-C transitions at U_1 (L_1C), U_2 (L_2C), or U_3 (L_3C) in the loop, and with U-to-C mutations at both U_1 and U_3 ($L_{1,3}C$). Each was analyzed by EMSA (Fig. 7A) and UV-crosslinking assays (Fig. 7B). The L_2C probe has the sequence, UCUC, present in the HIST1H2BI mRNA.

Previous work demonstrated that L_1 and L_3 are important determinants of SLBP binding activity (Williams and Marzluff 1995; Battle and Doudna 2001; Martin et al. 2012). We found that L_2C binds SLBP with a similar affinity as wild-type, but the L_1C , L_3C , and $L_{1,3}C$ mutations reduced SLBP binding. The L_1C probe had reduced affinity for SLBP, and L_3C and $L_{1,3}C$ had greatly impaired binding (Fig. 7A). SLBP crosslinked with the L_1C and L_2C single point mutants with similar efficiency to the SLWT (Fig. 7B,C). The crosslinking efficiency was reduced in the L_3C mutant, as was RNA binding, making it impossible to conclude whether the reduced crosslinking was due to reduced binding. There was no crosslinking to the double loop mutant, $L_{1,3}C$ or the SLRS (Fig. 7B,C), which do not bind the SL RNA.

These experiments did not explain why crosslinking to the UCUC loop variant was reduced by fourfold in the HITS-CLIP experiment. If SLBP crosslinking is not affected in the L_2C variant, we reasoned that perhaps crosslinking of 3'hExo was affected. We tested the affinity of 3'hExo for the SLWT and the L_2C variant. The L_2C variant had reduced affinity for 3'hExo compared to SLWT (Fig. 7D, cf. lanes 1–5 and 11–15). In addition, mutating L_2 to a C greatly reduced UV crosslinking of 3'hExo to the stem-loop (Fig. 7E,F).

To determine whether the reduced crosslinking of 3'hExo was solely a consequence of the lower affinity for the L_2C mutant probe, we took advantage of the fact that SLBP and 3'hExo bind cooperatively to the SL. Single mutations that affect SLBP binding to the SL do not impair formation of the ternary complex (Yang et al. 2006, 2009). SLBP and 3'hExo form a stable ternary complex on both the WT and the L_2C mutant SL (Fig. 7G), allowing us to ask if crosslinking of the 3'hExo is affected in the ternary complex. Crosslinking of SLBP was not affected while crosslinking of 3'hExo was reduced by 10-fold, suggesting that L_2 is the major crosslinking site for 3'hExo (Fig. 7H,I).

Taken together, these data suggest that the loop nucleotides L_1 and L_3 can crosslink to SLBP, and not L_2 , whereas crosslinking of 3'hExo occurs primarily on L_2 . In our HITS-CLIP experiments, it is likely that both 3'hExo and SLBP are present on the replication-dependent histone mRNA and will each crosslink to the mRNA. It is likely that the complex will be immunoprecipitated (Dominski et al. 2003), resulting in both SLBP and 3'hExo crosslinks on histone mRNA. Thus, we have likely mapped crosslinking of both 3'hExo and SLBP to histone mRNAs in the HITS-CLIP experiments. Note that in Figure 8G–I, we used a truncated SLBP in the experiments to clearly resolve SLBP and 3'hExo crosslinks, since full-length 3'hExo and SLBP migrate close to each other on a SDS-PAGE gel. This interpretation explains the reduction in recovery of reads and reduced detection of CIMS in histone SL sequences in the UCUA variant, as a result of reduced crosslinking of the 3'hExo.

UPF1 binds the RD histone 3' end immediately upstream of the SL

We also observed protection of RNA sequences 5' of the SL in the HITS-CLIP experiments (see Fig. 3) on some histone mRNAs. These protected sequences are unlikely to be due to SLBP interacting with that region of histone mRNA. The histone mRNAs are efficiently translated and there is a very short 3' UTR, which likely places the terminating ribosome close to the SLBP. Thus, micrococcal nuclease treatment may not cleave efficiently between the 3' end of the RNA and the terminating ribosome. Additional proteins may also bind to the 3' UTR of some histone mRNAs, including SR-proteins (Änkö et al. 2012).

We identified UPF1 as a likely candidate for binding at this site. UPF1 is a central factor in nonsense-mediated decay

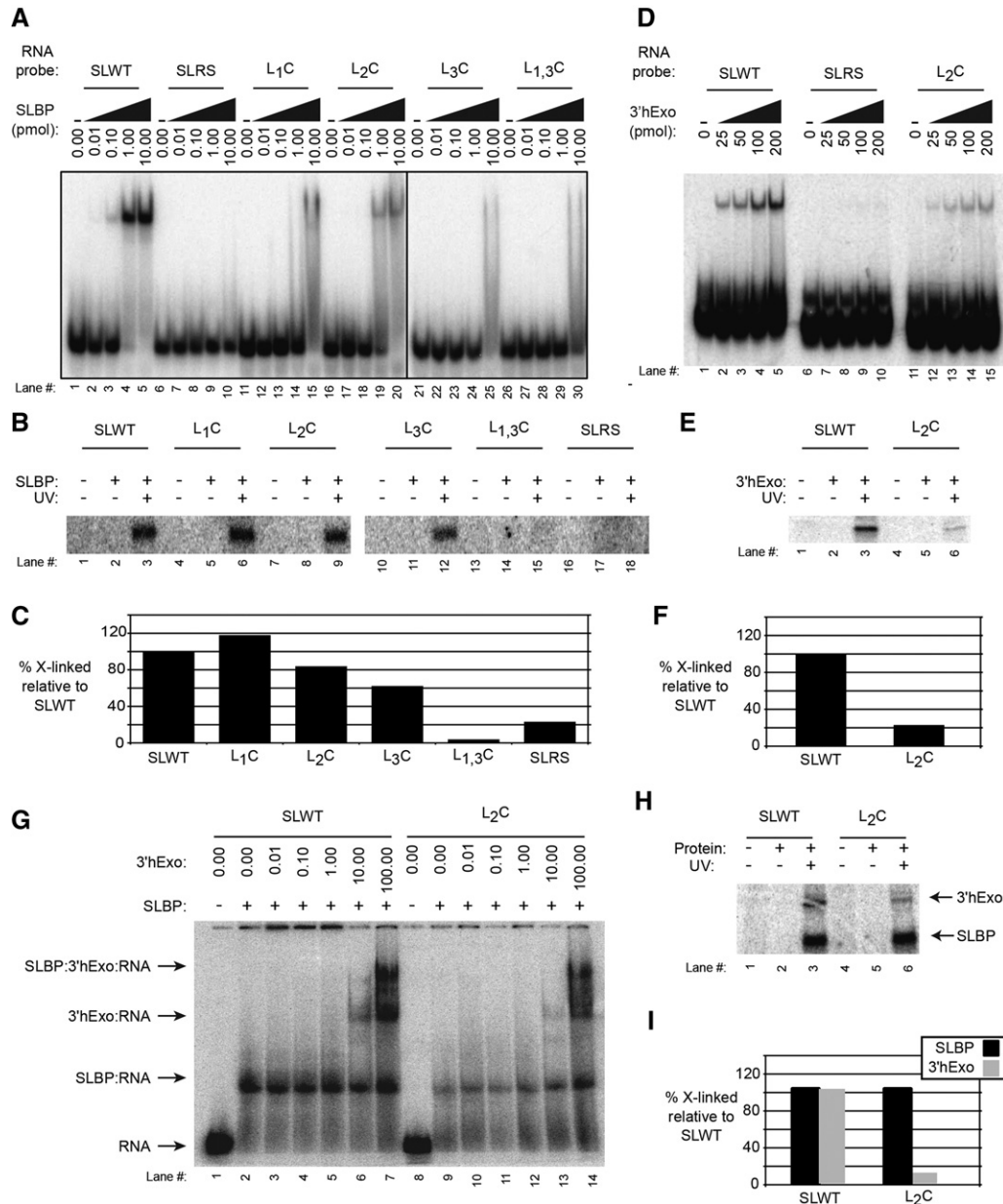


FIGURE 7. SLBP crosslinks to the L_{1,3}U and the 3'hExo crosslinks primarily to L₂U of the loop. (A) Electrophoretic mobility shift assays (EMSA) were performed to ascertain binding of SLBP to mutant SL sequences. Cytosines were substituted for specific uridines in the loop and the effect on SLBP binding determined by EMSA. A stem-loop reverse sequence (SLRS) that does not bind SLBP was used as a negative control. The position(s) of the mutation(s) in the variant SL probes are indicated: L₁C changes the first position in the loop from U to C, L₂C changes the second, L₃C changes the third, and L_{1,3}C changes both the first and third positions from U to C. (B,C) The ability to crosslink recombinant SLBP to these probes was determined by UV irradiation followed by SDS-gel electrophoresis. The relative crosslinking is shown in the graph. (D) The binding of 3'hExo to the SL wild type (SLWT) (lanes 1–5), SLRS (lanes 6–10), and L₂C (lanes 11–15) was measured by EMSA. (E) We carried out *in vitro* UV crosslinking with the SLWT (lanes 1–3) and L₂C (lanes 4–6) probes. (F) The results were quantified by PhosphorImager. (G) We incubated the SLWT (lanes 1–7) and L₂C probes (lanes 8–15) with 10 pmol of the SLBP RNA-processing domain and increasing amounts of 3'hExo. A ternary complex was formed in similar amounts with both probes. (H) The complexes with SLWT (lanes 1–3) and the L₂C probes (lanes 4–6) were crosslinked with UV light, and the two proteins resolved by SDS-gel electrophoresis. (I) The intensity of crosslinking quantified on a PhosphorImager.

(NMD) (Perlick et al. 1996; Sun et al. 1998) and is also involved in degradation of histone mRNAs (Kaygun and Marzluff 2005). It associates with histone mRNAs during histone mRNA degradation, likely by interacting directly with SLBP as well as the ribosome (Kaygun and Marzluff 2005).

Mühlemann and coworkers recently performed an iCLIP analysis to assess the mRNA occupancy preferences of UPF1 to mRNA in puromycin-treated HeLa cells. During puromycin treatment, the mRNAs are released from the ribosomes and additional potential binding sites for UPF1 are

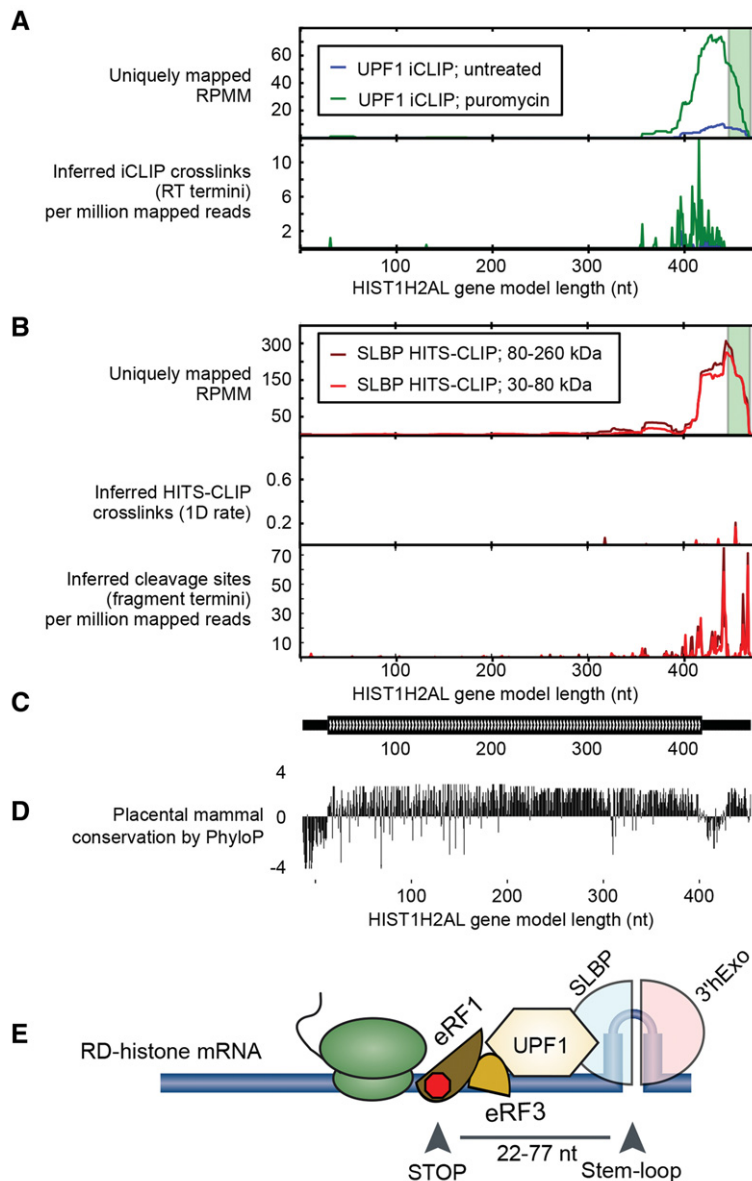


FIGURE 8. Secondary analysis of a public UPF1 iCLIP data set. We analyzed the data sets of UPF1-iCLIP experiments on untreated HeLa cells and cells treated with puromycin (Zünd et al. 2013) to determine whether UPF1 associated with histone mRNAs. (A) A representative histone mRNA, HIST1H2AL, is shown. The RPM across the histone mRNA (*top*) and the reverse transcriptase (RT) termini, which give the site of crosslinking (*bottom*) are plotted. (B) Our SLBP HITS-CLIP data for the HIST1H2AL gene are shown. The RPM (*top*), the single nucleotide deletion (1D) rate (*middle*), and the fragment termini (*bottom*) determined using our cleavage-mapping algorithm, are plotted in the 3.0 GU Mnase SLBP HITS-CLIP data from two specified size ranges. (C) The HIST1H2AL gene model graphic is shown *below* with the CDS and UTRs indicating CDS boundaries. (D) The phyloP (Pollard et al. 2010) placental mammal conservation score (Meyer et al. 2013) is shown for each nucleotide of the mature HIST1H2AL mRNA. (E) A cartoon depicting the potential arrangement of proteins on a generic RD-histone mRNA 3' end (Isken and Maquat 2008).

exposed (Zünd et al. 2013). We analyzed their iCLIP data and found that UPF1 interacts with the 3' UTR of many RD histone mRNAs (Table 1; Zünd et al. 2013), and the binding was enhanced in puromycin-treated cells. An example is shown in Figure 8 with additional examples in Supplemental Figure

S5A–C. The normalized UPF1 iCLIP read coverage is shown in the RPM plots and the UPF1 iCLIP “crosslink signature” is shown in the RT termini plots. These data show that UPF1 contacts RD-histone mRNAs in the 3' UTR before the SL. During puromycin treatment when protein synthesis was inhibited and histone mRNA is dissociated from polyribosomes, there was a substantial increase in UPF1 binding on the 3' UTR. Since SLBP and UPF1 coimmunoprecipitate from HU-treated cells and the binding is not sensitive to RNase (Kaygun and Marzluff 2005), it is possible that once ribosomes (and much of the UPF1) is removed from other mRNAs, UPF1 can then interact with SLBP and the 3' UTR of histone mRNA which is now exposed.

Alignment of the coverage and CIMS graphics from our SLBP HITS-CLIP data set with the UPF1 iCLIP graphs revealed that the two RBPs likely bind adjacent to each other on RD-histone mRNA (Fig. 8B; Supplemental Fig. S5). Since the 3' UTR sequences are not conserved (Fig. 8C,D), and UPF1 is a non-specific RNA-binding protein, it is likely that SLBP is a prime determinant of the UPF1 binding pattern on RD-histone mRNA 3' end, and may help recruit UPF1 to histone mRNAs.

DISCUSSION

SLBP does not interact with mRNAs other than the RD-histone mRNAs and H2AFX

A common result of examining the genome-wide binding of DNA or RNA-binding proteins is that targets that had not been previously described are discovered as a result of the analysis. In the case of SLBP we find that, at the level of our detection, the only targets are the RD-histone mRNAs, which share a highly conserved 26-nt RNA target present only at the 3' end of the RD-histone mRNAs. We describe a novel target, the

H2AFX gene that produces a RD-histone mRNA ending in a SL and a polyadenylated RI-histone mRNA that contains the upstream SL sequence. Neither the RIP-chip, RIP-seq, or HITS-CLIP experiments yielded any novel targets outside of existing histone gene models. SLBP has a novel RNA-

binding domain, which has not been reported in any other RNA-binding protein, and the only homologs found in database searches are SLBPs from other organisms. Homologs of SLBP are only found in organisms with histone genes that contain the histone stem-loop sequence (Davila Lopez and Samuelsson 2008). These include a subset of single-cell eukaryotes and all metazoans. Thus SLBP is an RNA-binding protein with a very limited number of RNA targets, which likely functions only in RD-histone mRNA metabolism.

Detecting conserved mRNP structure and *cis*-acting elements across related gene models

HITS-CLIP data are extremely information rich and here we have extracted a fraction of the information that can be gleaned from these data. The rate-limiting step in fully extracting the biological information from these data is partly the algorithms available to use in the analysis. We used SLBP, which has well-defined targets and a known RNA-binding sequence, as well as a crystal structure, as a test case to develop algorithms that more fully extract the maximum amount of information from HITS-CLIP data.

A major issue we encountered with the analysis of data for histone mRNAs was that of multimapping reads (Supplemental Fig. S3). Since many alignment algorithms deal with multimapping reads differently, the method chosen can impact the results when analyzing closely related gene families. For example, a portion of the HIST2 cluster is duplicated, containing an H2a, H3, and H4 gene (Braastad et al. 2004), and in some RNA-seq studies is annotated as not expressed (Yang et al. 2011). A second complication is that the relatively short reads one gets from HITS-CLIP (partly because of the size of the RNA fragment protected) and the highly conserved SL consensus sequence, makes it difficult to unambiguously assign some of these reads to a particular histone mRNA.

A second complexity is that multiple proteins may bind on an RNA target in a relatively tight complex that may be coimmunoprecipitated after the crosslinking, and in this case, together with SLBP might protect a much larger region than SLBP does by itself. This would result in potential crosslinks to more than one protein bound to RNA in the same immunoprecipitate, each of which would also contain crosslinking CIMS signatures, and/or crosslinking to RNA sequences that are not specifically bound by SLBP but contact it in the immunoprecipitate, as a result of SLBP interacting with another protein bound to RNA. The proteins in the complex could be similar in MW to SLBP (e.g., 3'hExo) or quite different in MW. We observed two broad size ranges of crosslinked RNA complexes when the SLBP immunoprecipitate was resolved by SDS-gel electrophoresis during HITS-CLIP, one in the general size range of SLBP, and the other much larger. Surprisingly, they both contained similar RNA sequences and CIMS in the stemloop. The nature of the different complexes is not clear but others have reported similar observations with other RNA-binding proteins (Yeo et al. 2009).

Analysis of crosslinking data

We developed new tools to help analyze the cross-linking data. These have been organized in a software package we term CLIP-PyL (available at <https://github.com/lb3/CLIP-PyL>) that can be used to analyze HITS-CLIP data.

To accurately compare the coverage vectors across histone genes, we developed the CVCA method to systematically cluster and examine this read coverage in an unbiased manner. The method we developed normalizes all genes to a similar length, a problem that is quite tractable with histone mRNAs that have a narrow size range compared to other mRNAs (400–800 nt). We recognize that this approach may present special challenges when expanded to messages that have a much wider range of sizes and thus will need to be further developed and refined to address the needs of the broader arrays of mRNAs and target sequences present in the human genome.

We also developed a technique for mapping the ends of the fragments as a result of nuclease cleavage to delineate boundaries of nuclease-resistant mRNP complexes. This method allows us to map the cleavage terminus immediately adjacent to the sequencing primer as a free 5' end, and the 3' ends were reads that extended into the 3' linker. Since SLBP binds to the 3' end of the mRNA, we were able to determine the 3' end (protected from nuclease treatment) as well as the 5' end, which delineated cleavage points upstream of the SL. The very 3' end of the consensus binding site determined by CLIP-seq differs from the genomic sequence, with an enrichment of uridine in the last 2 nt. Recently, it was reported that a fraction of all histone mRNAs have the last 2–3' encoded nucleotides replaced with uridines post-transcriptionally (Welch et al. 2015) and we detected these 3' ends present in the cell in our HITS-CLIP experiments.

A limitation of the current data set resulted from the 36-nt single-end sequence reads. We are only able to map the 3' end cleavage sites for fragments that were <27 nt in length, since we could only reliably map those sequences for which we read through the entire RNA fragment and into the adjacent sequencing primer. These 5' end sequences corresponded to the known boundary of the SLBP binding site on histone mRNA (Fig. 4). Although we could determine one cleavage site from the start of the read, we could not determine the other cleavage site for fragments larger than 27 nt in length. Paired-end sequencing would easily allow the mapping of both termini of MNase-resistant fragments. In addition, longer reads would allow the quantification of the CIMS in larger mRNP fragments since one could read a complete fragment of 50–200 nt in length.

SLBP interacts with multiple sites in polyadenylated H2AFX mRNA

A surprising result was the finding that the SLBP immunoprecipitates contains crosslinked proteins to both the SL

and a conserved region near the poly(A) tail in the replication-independent form of the H2AFX mRNA. It seems unlikely that SLBP is bound to the SL in the short form of H2AFX mRNA and only at the poly(A) site in the long form. Rather SLBP is likely bound to the stemloop in the long poly(A)⁺ isoform, and is in contact with a protein bound to H2AFX mRNA which is coimmunoprecipitated with the SLBP antibody or else the mRNA interacts with SLBP as a result of the 3' end being close to the ribosome. The 3' end of the polyadenylated H2AFX mRNA is conserved in mammals, suggesting this interaction has been selected for. It could occur as a result of (or maybe actively promote) the circularization of the mRNP. Since there are CIMS at both positions, a protein clearly crosslinked near the poly(A) site in the conserved region.

UPF1 binds the 3' UTR of histone mRNAs immediately 5' of the SL

Our analyses of published UPF1 iCLIP data from human cells (Zünd et al. 2013) shows that UPF1 binds RD-histone mRNA immediately upstream of the SL in all expressed RD-histone mRNAs. UPF1 is required for histone mRNA degradation (Kaygun and Marzluff 2005) and normally binds to mRNAs near the stop codon as a result of inefficient termination (Amrani et al. 2006). UPF1 also directly interacts with SLBP in vivo and in vitro (Kaygun and Marzluff 2005; S Meaux and WF Marzluff, unpubl.). Since the amount of UPF1 bound to the 3' UTR of histone mRNA is greatly increased when ribosomes are dissociated from the mRNA (Zünd et al. 2013), SLBP may participate in recruiting UPF1 to these histone mRNAs.

Concluding remarks

Consistent with prior results we find that histones are the primary targets of SLBP, confirming its exquisite specificity for histone mRNA. We have extended our knowledge of SLBP:RNA interactions by precisely mapping the cleavage points and CIMS that pinpoint protein–RNA crosslinks in the SLBP HITS data. We also find evidence for additional proteins interacting with SLBP which themselves bind histone mRNA at different sites, and these crosslinked sites were likely recovered because proteins coimmunoprecipitated with SLBP. We have confirmed these results by performing in vitro mobility shift assays and in vitro crosslinking with both WT and mutant SLBP sequences to confirm the CIM predictions. The algorithms we have developed here have been organized in a software package that can be used to analyze HITS-CLIP, PAR-CLIP, and iCLIP data from other RNA-binding proteins and *cis*-acting elements. Furthermore, the data set presented here will be highly valuable for testing new algorithms since the SLBP targets, binding site, and molecular interactions are well characterized.

MATERIALS AND METHODS

The complete raw data from this study are available from NCBI GEO at accession number GSE62154. CLIP-PyL python scripts are available at <https://github.com/lb3/CLIP-PyL>. CLIP-PyL generates coverage metrics per nucleotide to identify crosslinked and protected RNA fragments (e.g., crosslink-induced mutations) from HITS-CLIP experiments. These metrics were subsequently used to determine the number of base mismatches, indels, and aligned read termini at each nucleotide.

SLBP RIP-chip

RIP-chip experiments were performed as previously described (Townley-Tilson et al. 2006) using an antibody to the carboxy-terminus of the SLBP (Wang et al. 1996) (Millipore). A 5 µg portion of the unbound fraction RNA and the volumetric equivalent of the bound fraction was differentially labeled by direct incorporation of CTP:cy3 and CTP:cy5, respectively, by random-primed cDNA synthesis. The labeled cDNA libraries were fragmented by chemical by hydrolysis and hybridized to Agilent 4 × 44k human oligo microarrays. Following hybridization, the microarrays were washed and scanned on the Molecular Devices GenePix 4000b scanner and the image data were saved in the tiff format.

The RIP-chip data set was analyzed using the Statistical Analysis of Microarray software (SAM) (Tusher et al. 2001). The modified unpaired *t*-test provided by SAM returned 82 statistically significant microarray probes (Fig. 1B; Supplemental Fig. S1; Supplemental Table 1), 72 of which are annotated as histone mRNA probes; 10 nonhistone genes did not reproduce in either the Townley-Tilson data, nor our RIP-seq or HITS-CLIP data indicating spurious background and cross-hybridization artifacts that we have detailed previously (Townley-Tilson et al. 2006) and below. We were able to call approximately twice as many statistically significant enriched mRNAs than in our previous RIP-chip analysis (Fig. 1B; Supplemental Fig. S1; Townley-Tilson et al. 2006). This is likely due to the increase in statistical power afforded by the additional experimental replicates included in our current SLBP RIP-chip data set.

We anticipated that cross-hybridization would identify potential false positives in our SLBP RIP-chip data set because of the high degree of sequence similarity across the histone mRNA family. Cross-hybridization was indeed evident as the appearance of *HIST1H2APS1* in the set of significantly enriched mRNAs called by the RIP-chip analysis (Fig. 1B; Supplemental Fig. S1). *HIST1H2APS1* is an unexpressed pseudogene; no expression was detected in our analysis of a publicly available poly(A) RNA-seq data sets from asynchronous HeLa cells (Yang et al. 2011; Djebali et al. 2012) nor is it detected in our SLBP RIP-seq data set. This phenomenon underscores the fact that cross-hybridization artifacts must be considered when interpreting microarray data.

SLBP RIP-seq

A polyclonal αSLBP antibody reagent was prepared by Pacific Immunology by injecting rabbits with an immunogenic peptide comprised of the C-terminal 13 amino acids of SLBP (available from Millipore). Three αSLBP RIPs were performed and three negative

control RIPs were performed. Negative control RIPs were performed with irrelevant antibody, no antibody or by peptide competition. RNA eluates from the RIPs were primed with random hexamers (3 µg/µL, Invitrogen, 48190-011) and reverse transcribed with superscript II (200 U/µL, Invitrogen, 18064-014). RNA–cDNA hybrids were nicked with RNase H (2 U/µL, Invitrogen, 18021-014) and second strand cDNA was polymerized with DNA Pol I (10 U/µL, Invitrogen, 18010-025). The double-stranded cDNA library was sent to the UNC genomics core facility for size selection and subsequent ligation of the Illumina Genomic DNA oligonucleotide Adapters. The libraries were sequenced on a GAI1 to yield 36-bp single-end reads.

SLBP HITS-CLIP

HeLa S3 cells were grown in monolayer to 80% confluency on 100-mm polystyrene culture plates. Prior to UV crosslinking, cells were washed with 10 mL ice-cold PBS. The washed plates were then placed on ice and irradiated with 7400 mJ of 254 nm UV light. Cells were immediately lysed in NP-40 lysis buffer (0.1% NP40, 50 mM Tris–HCl pH 7.5, 150 mM NaCl, 50 mM NaF, 1 mM DTT, 10 mM EDTA, 40U RNasin) and concurrently exposed to micrococcal nuclease and DNase. The nucleolytic digestion was stopped by addition of EGTA. IPs were optimized and performed using conditions previously described that result in specific IP of the SLBP with low background (Whitfield et al. 2004; Townley-Tilson et al. 2006). Nonspecific interactions were precleared by the addition of 20 µL of a 1:1 protein A bead:NP-40 LB slurry and incubation at 4°C for 30 min with rotation. Protein A beads were removed by centrifugation at 1500g, and the precleared supernatant incubated for 1 h at 4°C with either 1.5 mg affinity-purified anti-SLBP. Negative control IPs were performed using irrelevant antibody reagent (e.g., anti-GFP). Antibody–protein complexes were isolated by incubation for 1 h at 4°C with a 20-µL Protein A agarose bead slurry. Beads were recovered by centrifugation at 400g and washed extensively five times with 1 mL of NP-40 lysis buffer. After the IP, SDS sample buffer is added to the pellet and the samples resolved by SDS-PAGE. The samples are then transferred to nitrocellulose under denaturing conditions as described (Ule et al. 2005); only RNA covalently bound to protein adheres to the nitrocellulose while free RNA passes through.

RNA fragments were eluted from the membrane slices by proteinase K digestion, ligated to appropriate RNA-sequencing adapters as previously described (Ule et al. 2003) except illumina sRNA adapters were used. Adaptor ligated RNA was reverse transcribed and PCR-amplified the cDNA libraries. The size distribution of the amplified fragments in the libraries was assessed by PAGE. A ligation reaction comprised of the SL probe and sequencing adapters was prepared to serve as a size standard. RNA isolated from the LMW and HMW regions were used to construct six HITS-CLIP-sequencing libraries. The CLIP libraries were sequenced at the UNC genomics core facility on a GAI1 to yield 36-bp single-end reads.

In vitro transcription of stem–loop substrates

Stem–loops used in electrophoretic mobility shift assays (EMSA) and crosslinking studies were made by in vitro transcription using

oligonucleotides as previously described (Milligan et al. 1987; Pandey et al. 1991). Probes used in EMSAs were labeled with [α - 32 P]-CTP. For stem–loops used in crosslinking studies, the protocol was the same except 10 µL of 3.3 µM [α - 32 P]-3000 Ci/mmol-UTP was used in place of CTP in the transcription reaction.

Expression and purification of recombinant SLBP and 3'hExo from Sf-9 cells

Full-length SLBP, RNA-processing domain (amino acids 125–223) of SLBP and full-length 3'hExo were expressed in baculovirus as previously described (Yang et al. 2006; Dominski et al. 1999).

Electrophoretic mobility shift assays

Ten femtomoles of uniformly labeled RNA was incubated on ice with various amounts of purified recombinant protein in 10 mM HEPES (pH 7.60), 50 mM KCl, 0.1 mM EDTA, 10% glycerol, 1 µg/µL yeast tRNA, 0.1 µg/µL, BSA. For shifts with 3'hExo, yeast tRNA was omitted from reactions. Reactions were directly loaded onto 6% native polyacrylamide gel (acrylamide:bisacrylamide was 29:1 in 1× TBE buffer) without any loading dyes. Gels were fixed in 45% methanol/10% acetic acid, dried and then visualized by autoradiography on phosphor screen and film. Images were analyzed by imageQuant.

In vitro crosslinking assays

For crosslinking assays, in 10 µL, 1 pmol of recombinant protein was incubated with 10 fmol of stem–loop probes uniformly labeled with [α - 32 P]-UTP in 10 mM HEPES (pH 7.60), 50 mM KCl, 0.1 mM EDTA, 10% glycerol, 1 µg/µL yeast tRNA, 0.1 µg/µL BSA for 30 min on ice. Yeast tRNA was omitted for experiments using 3'hExo. RNA:protein complexes were crosslinked for 10 min on ice using stratalinker (Stratagene) with 254-nm UV bulbs. An equal volume of 2× SDS-loading buffer (4% SDS, 10% β -mercaptoethanol, 0.125 M Tris [pH 6.8], 20% glycerol, 0.2% bromophenol blue) was added to the crosslinked proteins and boiled for 10 min before loading onto a 12% SDS-PAGE gel (acrylamide:bis-acrylamide was 37.5:1). Gels were dried and visualized by autoradiography and analyzed using ImageQuant.

S1 nuclease assays for H2AFX

The *H2AFX* gene was cloned from mouse genomic DNA by PCR using a primer complementary to the coding region of the gene at codon 50 containing an EcoRI site and a primer 3' of the histone SL containing an XbaI site. The PCR fragment was digested and cloned into the EcoRI–XbaI sites of pGEM3zf and contains the mouse *H2AFX* gene extending 46 nt past the SL. The plasmid was digested with Ava I and the 5' overhang filled in with Klenow fragment and [α - 32 P]-dCTP. The probe was released by digestion with XbaI which cuts in the plasmid 3' of the insert. A 3' labeled S1 probe protects a 379-bp fragment corresponding to the poly(A)⁺ form of the mRNA and a 333-bp fragment corresponding to the SL form of the mRNA.

RIP-seq and HITS-CLIP bioinformatics methods

Enrichment analysis with EdgeR

For the RIP-seq and HITS-CLIP enrichment analyses, all reads were quality filtered and adapter-trimmed with the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) prior to alignment to the human genome (hg19) using BWA (Li and Durbin 2009). The number of uniquely aligning reads that mapped to each gene model in the UCSC known gene set (Hsu et al. 2006) were counted using the HT-seq python package [<http://www-huber.embl.de/users/anders/HTSeq/doc/index.html>]. The edgeR v2.4.6 (Robinson et al. 2010) package was used to compute normalizing factors, estimate common and tagwise dispersion, and perform an exact test for the negative binomial distribution. The commands that were used are as follows (where x is the data frame object containing my data): $y \leftarrow \text{DGEList}(\text{counts} = x, \text{group} = \text{group})$; $y \leftarrow \text{calcNormFactors}(y)$; $y \leftarrow \text{estimateCommonDisp}(y)$; $y \leftarrow \text{estimateTagwiseDisp}(y)$; $\text{et} \leftarrow \text{exactTest}(y)$; $\text{topTags}(\text{et})$; $\text{de} \leftarrow \text{decideTestsDGE}(\text{et}, P = 0.05, \text{adjust} = \text{"BH"})$; $\text{detags} \leftarrow \text{et}[\text{as.logical}(\text{de})]$; $\text{detags} \leftarrow \text{topTags}(\text{detags}, n = \text{length}(\text{detags}))$.

A P -value threshold of 0.05 was used to identify significantly enriched transcripts enriched by the α -SLBP antibody. This EdgeR analysis was also performed using the custom genome annotation file that we generated. The custom genome annotation was generated by merging multiple public genome annotations including Refseq (Pruitt et al. 2012), GENCODE (Pruitt et al. 2012), and UCSC known gene set. Mapped read clusters from our HITS-CLIP and RIP-seq data sets that were not annotated in the public genome annotations were assigned a generic annotation ID and deposited in the custom genome annotation. The BEDTools software was used to perform the operations (Quinlan and Hall 2010).

We filtered the regular expression matches by selecting sequences with the potential to form a SL structure. This was accomplished using a heuristic that identifies proper base-pairing interactions between apposing nucleotides in the stem. The histone SL was detected in the genome using the following regular expression: GG[C/T][C/T][C/T]TT[C/T]T[A/G/C/T]A[A/G][A/G][A/G]CC and then the positive hits filtered by requiring base-pairing in the stem. This resulted in the final 95 potential SL sequences after filtering for sequences that had at least one read in the HITS-CLIP data.

Coverage vector correlation analysis

Sequence reads derived from six HITS-CLIP libraries were aligned to the human genome (hg19) with the bowtie fast read aligner (Langmead et al. 2009), allowing up to 25 genomic matches. We developed python scripts to calculate basewise coverage vectors for all genomic intervals comprising the HGNC histone sequence database. The coverage vectors were then imported to matlab for all downstream processing steps. A coverage vector is calculated by counting the number of uniquely mapped reads per million mapped reads across all bases in a given gene model. There are 116 gene models in the HGNC-HSDB that includes RD and non-RD histone genes. Since we sequenced six HITS-CLIP libraries, the set of coverage vectors is comprised of 696 coverage vectors. We applied CVCA to the analysis of our SLBP HITS-CLIP data by calculating read coverage per base across each gene model in the HGNC-HSDB. We then processed the coverage vectors for calculation of pairwise correlation scores by removing empty vectors and interpolating all

remaining coverage vectors to the same length. The maximum coverage vector length was determined and we used nearest neighbor interpolation to expand all coverage vectors to the same length. Pairwise correlation scores were calculated for all non-empty vectors and the resultant correlation matrix was sorted organized by spectral clustering to identify distinct coverage shapes.

Cleavage site mapping

We used the FASTX toolkit to select reads that contained 3' adapter sequence. Note that only the reads with 3' adapter sequence were used for this analysis because the 3' adapter sequence was necessary for inference of the 3' end of the fragment. The 3' adapter sequences were then trimmed using the same software and the trimmed reads were aligned to the human genome (hg19) using BWA (Li and Durbin 2009). We developed a python script (CLIP-PyL) that selects the genomic coordinates of the terminal nucleotides from each of the aligned reads. These terminal nucleotides comprise the set of cleavage sites (Fig. 4A). The CLIP-PyL package contains scripts that can parse "crosslink signatures" from preprocessed CLIP-seq data (Supplemental Fig. S5).

Mapping crosslink-induced mutation sites

Sequence reads derived from six HITS-CLIP libraries were aligned to the human genome (hg19) with the indel-aware BWA fast read aligner (Li and Durbin 2009). To quantify the number of single nucleotide deletions (1D) we developed a python script to parse the CIGAR strings present in the alignment file that is generated by BWA. We emulated the previously published method for computing statistical methods for identifying CIMS (Zhang and Darnell 2011). We filtered all sites with >95% of reads containing 1D, as these are likely SNPs. We then tested the null hypothesis that there is no increase in the 1D rate in the loop region of the histone stem-loop. We computed the difference between the 1D rate in the stem region to the 1D rate in the loop region for each histone stem-loop and compared it to the null distribution that was generated by randomizing the assignment of 1D to all reads in the data set (while preserving the offset from the end of the reads, as previously described Zhang and Darnell [2011]). This allowed us to calculate a P -value for 1D enrichment in the loop for each histone stem-loop element (Fig. 6C). Where k_i is the total number of reads mapping to nucleotide i , m_i is the number of mapped 1D of nucleotide i , and n is the total number of nucleotides in the interval of interest. Thus, for a given genomic interval, the average 1D frequency within the interval is computed, such that

$$\bar{X} = \frac{\sum_{i=1}^n \frac{m_i}{k_i}}{n}.$$

For each histone stem-loop, we computed the average 1D frequency for the interval that contains the loop, and for the interval that contains the stem. Our test statistic D is the difference, such that

$$D = \bar{X}_{\text{loop}} - \bar{X}_{\text{stem}}.$$

Computing histone gene abundances from ENCODE HeLa data

RNA-seq reads from ENCODE HeLa data were aligned to hg19 using MapSplice2 with default settings. UCSC gene annotations were

downloaded and used to create transcriptome annotations for RSEM. We computed gene expression levels for UCSC genes using RSEM with the settings `-estimate-rspd` and `-paired-end`. RSEM aligns reads to a reference transcriptome in a way designed to find all possible genes that each read could have originated from and then uses a Bayesian network model to estimate the abundance of all genes simultaneously. This strategy is ideal for dealing with genes that may have many multimapped reads, as is the case with histone genes. For duplicated genes, RSEM computes a separate expression level for each annotated locus. The expression levels that RSEM reports are in units of FPKM (fragments per kilobase per million reads).

Identifying nontemplated 3' additions in HITS-CLIP reads

We used the AppEnD tool (Welch et al. 2015) to identify nontemplated 3' additions in SLBP HITS-CLIP reads. Briefly, AppEnD identifies nontemplated 3' additions by aligning a read to a reference genome, detecting the position within the read where the read sequence stops matching the genome sequence, and removing the portion of this mismatched sequence that comes from a sequencing adapter (if any). The remaining nucleotides from the mismatched sequence represent a nontemplated 3' addition. This process relies heavily on the information provided by the presence of an adapter sequence in the read, indicating the precise 3' terminus of the RNA transcript being sequenced. Therefore, we restricted our analysis to reads containing at least the first four nucleotides of the sequencing adapter added to the 3' end of the RNA. AppEnD located nontemplated additions on many histone mRNAs; we chose to display the transcript positions and counts of these additions on a single histone "metatranscript" to summarize the common pattern of tail addition across the histone mRNAs.

Description of additional files

The following additional data are available with the online version of the paper. Supplemental Table 1 is a tab-delimited text version of Supplemental Figure S1, and contains all histone genes identified as bound to SLBP in HeLa cells from multiple different genomic analyses. Supplemental Table 2 shows the list of microarray probes with statistical significance in the SAM (Tusher et al. 2001) analysis of the RIP-chip experiments. Supplemental Table 3 contains all significant coding regions from the RIP-seq analysis. Supplemental Table 4 contains all significant gene models and clusters with >30 reads from the EdgeR analysis of the HITS-CLIP data. Supplemental Table 5 contains the genomic locations (hg19) and sequence of all the SL motif matches found by our algorithm. Supplemental Table 6 contains all gene vectors from the CVCA analysis and their cluster assignment.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grants R01 HG004499 and R01 CA130795 to M.L.W. and R01 GM29832 to

W.F.M. J.M.M. received support from National Institutes of Health grant R25 CA134286 and the Scleroderma Research Foundation. We thank Jeanine F. Amacher for measuring bond angles and distances in pymol (<http://www.pymol.org>).

Received July 20, 2015; accepted July 23, 2015.

REFERENCES

- Amrani N, Sachs MS, Jacobson A. 2006. Early nonsense: mRNA decay solves a translational problem. *Nat Rev Mol Cell Biol* **7**: 415–425.
- Änkö ML, Müller-McNicoll M, Brandl H, Curk T, Gorup C, Henry I, Ule J, Neugebauer KM. 2012. The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol* **13**: R17.
- Battle DJ, Doudna JA. 2001. The stem-loop binding protein forms a highly stable and specific complex with the 3' stem-loop of histone mRNAs. *RNA* **7**: 123–132.
- Bonner WM, Mannironi C, Orr A, Pilch DR, Hatch CL. 1993. Histone H2A.X gene transcription is regulated differently than transcription of other replication-linked histone genes. *Mol Cell Biol* **13**: 984–992.
- Braastad CD, Hovhannisyan H, van Wijnen AJ, Stein JL, Stein GS. 2004. Functional characterization of a human histone gene cluster duplication. *Gene* **342**: 35–40.
- Cakmakci NG, Lerner RS, Wagner EJ, Zheng L, Marzluff WF. 2008. SLIP1, a factor required for activation of histone mRNA translation by the stem-loop binding protein. *Mol Cell Biol* **28**: 1182–1194.
- Cox MP, Peterson DA, Biggs PJ. 2010. SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinform* **11**: 485.
- Dávila López M, Samuelsson T. 2008. Early evolution of histone mRNA 3' end processing. *RNA* **14**: 1–10.
- Dickey JS, Redon CE, Nakamura AJ, Baird BJ, Sedelnikova OA, Bonner WM. 2009. H2AX: functional roles and potential applications. *Chromosoma* **118**: 683–692.
- Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, et al. 2012. Landscape of transcription in human cells. *Nature* **489**: 101–108.
- Dominski Z, Zheng LX, Sanchez R, Marzluff WF. 1999. Stem-loop binding protein facilitates 3'-end formation by stabilizing U7 snRNP binding to histone pre-mRNA. *Mol Cell Biol* **19**: 3561–3570.
- Dominski Z, Yang X, Kaygun H, Marzluff WF. 2003. A 3' exonuclease that specifically interacts with the 3' end of histone mRNA. *Mol Cell* **12**: 295–305.
- Erkmann JA, Sánchez R, Treichel N, Marzluff WF, Kutay U. 2005. Nuclear export of metazoan replication-dependent histone mRNAs is dependent on RNA length and is mediated by TAP. *RNA* **11**: 45–58.
- Gorgoni B, Andrews S, Schaller A, Schümperli D, Gray NK, Müller B. 2005. The stem-loop binding protein stimulates histone translation at an early step in the initiation pathway. *RNA* **11**: 1030–1042.
- Hanson RJ, Sun J, Willis DG, Marzluff WF. 1996. Efficient extraction and partial purification of the polyribosome-associated stem-loop binding protein bound to the 3' end of histone mRNA. *Biochemistry* **35**: 2146–2156.
- Harris ME, Böhni R, Schneiderman MH, Ramamurthy L, Schümperli D, Marzluff WF. 1991. Regulation of histone mRNA in the unperturbed cell cycle: evidence suggesting control at two posttranscriptional steps. *Mol Cell Biol* **11**: 2416–2424.
- Hoefig KP, Rath N, Heinz GA, Wolf C, Dameris J, Schepers A, Kremmer E, Ansel KM, Heissmeyer V. 2013. Eri1 degrades the stem-loop of oligouridylylated histone mRNAs to induce replication-dependent decay. *Nat Struct Mol Biol* **20**: 73–81.
- Holmes WF, Braastad CD, Mitra P, Hampe C, Doenecke D, Albig W, Stein JL, van Wijnen AJ, Stein GS. 2005. Coordinate control and

- selective expression of the full complement of replication-dependent histone H4 genes in normal and cancer cells. *J Biol Chem* **280**: 37400–37407.
- Hsu F, Kent WJ, Clawson H, Kuhn RM, Diekhans M, Haussler D. 2006. The UCSC known genes. *Bioinformatics* **22**: 1036–1046.
- Isken O, Maquat LE. 2008. The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nat Rev Genet* **9**: 699–712.
- Kaygun H, Marzluff WF. 2005. Regulated degradation of replication-dependent histone mRNAs requires both ATR and Upf1. *Nat Struct Mol Biol* **12**: 794–800.
- Kryzwiniski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* **19**: 1639–1645.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.
- Licalatosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X, et al. 2008. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**: 464–469.
- Mannironi C, Bonner WM, Hatch CL. 1989. H2A.X, a histone isoprotein with a conserved C-terminal sequence, is encoded by a novel mRNA with both DNA replication type and polyA 3' processing signals. *Nucleic Acids Res* **17**: 9113–9126.
- Mariño-Ramirez L, Levine KM, Morales M, Zhang S, Moreland RT, Baxevasis AD, Landsman D. 2011. The Histone Database: an integrated resource for histones and histone fold-containing proteins. *Database (Oxford)* **2011**: bar048.
- Martin L, Meier M, Lyons SM, Sit RV, Marzluff WF, Quake SR, Chang HY. 2012. Systematic reconstruction of RNA functional motifs with high-throughput microfluidics. *Nat Methods* **9**: 1192–1194.
- Marzluff WF, Gongidi P, Woods KR, Jin J, Maltais LJ. 2002. The human and mouse replication-dependent histone genes. *Genomics* **80**: 487–498.
- Marzluff WF, Duronio RJ, Wagner EJ. 2008. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat Rev Genet* **9**: 843–854.
- Meyer LR, Zweig AS, Hinrichs AS, Karolchik D, Kuhn RM, Wong M, Sloan CA, Rosenbloom KR, Roe G, Rhead B, et al. 2013. The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res* **41**: D64–D69.
- Mili S, Steitz JA. 2004. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* **10**: 1692–1694.
- Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* **15**: 8783–8798.
- Pandey NB, Marzluff WF. 1987. The stem-loop structure at the 3' end of histone mRNA is necessary and sufficient for regulation of histone mRNA stability. *Mol Cell Biol* **7**: 4557–4559.
- Pandey NB, Sun J-H, Marzluff WF. 1991. Different complexes are formed on the 3' end of histone mRNA in nuclear and polysomal proteins. *Nucleic Acids Res* **19**: 5653–5659.
- Perlick HA, Medghalchi SM, Spencer FA, Kendzior RJ, Dietz HC. 1996. Mammalian orthologues of a yeast regulator of nonsense transcript stability. *Proc Natl Acad Sci* **93**: 10928–10932.
- Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. 2010. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* **20**: 110–121.
- Pruitt KD, Tatusova T, Brown GR, Maglott DR. 2012. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res* **40**: D130–D135.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.
- Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.
- Rogakou EP, Pilch DR, Orr AH, Ivanova VS, Bonner WM. 1998. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* **273**: 5858–5868.
- Sánchez R, Marzluff WF. 2002. The stem-loop binding protein is required for efficient translation of histone mRNA in vivo and in vitro. *Mol Cell Biol* **22**: 7093–7104.
- Slevin MK, Meaux S, Welch JD, Bigler R, de Marval PLM, Su W, Prins JF, Rhoads RE, Marzluff WF. 2014. Deep sequencing shows multiple oligouridylations are required for 3' to 5' degradation of histone mRNAs on polyribosomes. *Mol Cell* **53**: 1020–1030.
- Sullivan KD, Mullen TE, Marzluff WF, Wagner EJ. 2009. Knockdown of SLBP results in nuclear retention of histone mRNA. *RNA* **15**: 459–472.
- Sun X, Perlick HA, Dietz HC, Maquat LE. 1998. A mutated human homologue to yeast Upf1 protein has a dominant-negative effect on the decay of nonsense-containing mRNAs in mammalian cells. *Proc Natl Acad Sci* **95**: 10009–10014.
- Tan D, Marzluff WF, Dominski Z, Tong L. 2013. Structure of histone mRNA stem-loop, human stem-loop binding protein, and 3'hExo ternary complex. *Science* **339**: 318–321.
- Townley-Tilson WH, Pendergrass SA, Marzluff WF, Whitfield ML. 2006. Genome-wide analysis of mRNAs bound to the histone stem-loop binding protein. *RNA* **12**: 1853–1867.
- Tusher VG, Tibshirani R, Chu G. 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci* **98**: 5116–5121.
- Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. 2003. CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**: 1212–1215.
- Ule J, Jensen K, Mele A, Darnell RB. 2005. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods* **37**: 376–386.
- von Moeller H, Lerner R, Ricciardi A, Basquin C, Marzluff WF, Conti E. 2013. Structural and biochemical studies of SLIP1-SLBP identify DBP5 and eIF3g as SLIP1-binding proteins. *Nucleic Acids Res* **41**: 7960–7971.
- Wang ZF, Whitfield ML, Ingledue TC, Dominski Z, Marzluff WF. 1996. The protein that binds the 3' end of histone mRNA: a novel RNA-binding protein required for histone pre-mRNA processing. *Genes Dev* **10**: 3028–3040.
- Welch JD, Slevin MK, Tatomer DC, Duronio RJ, Prins JF, Marzluff WF. 2015. EnD-Seq and AppEnD: sequencing 3' ends to identify non-templated tails and degradation intermediates. *RNA* **21**: 1375–1389.
- West MHP, Bonner WM. 1980. Histone 2A, a heteromorphous family of eight protein species. *Biochemistry* **19**: 3238–3245.
- Whitfield ML, Zheng LX, Baldwin A, Ohta T, Hurt MM, Marzluff WF. 2000. Stem-loop binding protein, the protein that binds the 3' end of histone mRNA, is cell cycle regulated by both translational and post-translational mechanisms. *Mol Cell Biol* **20**: 4188–4198.
- Whitfield ML, Kaygun H, Erkmann JA, Townley-Tilson WH, Dominski Z, Marzluff WF. 2004. SLBP is associated with histone mRNA on polyribosomes as a component of the histone mRNP. *Nucleic Acids Res* **32**: 4833–4842.
- Williams AS, Marzluff WF. 1995. The sequence of the stem and flanking sequences at the 3' end of histone mRNA are critical determinants for the binding of the stem-loop binding protein. *Nucleic Acids Res* **23**: 654–662.
- Yang XC, Purdy M, Marzluff WF, Dominski Z. 2006. Characterization of 3'hExo, a 3' exonuclease specifically interacting with the 3' end of histone mRNA. *J Biol Chem* **281**: 30447–30454.
- Yang XC, Torres MP, Marzluff WF, Dominski Z. 2009. Three proteins of the U7-specific Sm ring function as the molecular ruler to determine the site of 3'-end processing in mammalian histone pre-mRNA. *Mol Cell Biol* **29**: 4045–4056.

- Yang L, Duff MO, Graveley BR, Carmichael GG, Chen LL. 2011. Genomewide characterization of non-polyadenylated RNAs. *Genome Biol* **12**: R16.
- Yeo GW, Coufal NG, Liang TY, Peng GE, Fu XD, Gage FH. 2009. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol* **16**: 130–137.
- Zhang C, Darnell RB. 2011. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat Biotechnol* **29**: 607–614.
- Zheng L, Dominski Z, Yang X-C, Elms P, Raska CS, Borchers CH, Marzluff WF. 2003. Phosphorylation of stem-loop binding protein (SLBP) on two threonines triggers degradation of SLBP, the sole cell cycle-regulated factor required for regulation of histone mRNA processing, at the end of S phase. *Mol Cell Biol* **23**: 1590–1601.
- Zünd D, Gruber AR, Zavolan M, Mühlemann O. 2013. Translation-dependent displacement of UPF1 from coding sequences causes its enrichment in 3' UTRs. *Nat Struct Mol Biol* **20**: 936–943.