Genome-wide analysis of mRNAs bound to the histone stem–loop binding protein

W.H. DAVIN TOWNLEY-TILSON,[1] SARAH A. PENDERGRASS,[1] WILLIAM F. MARZLUFF,[2] and MICHAEL L. WHITFIELD[1,3]

[1]Department of Genetics, Dartmouth Medical School, Hanover, New Hampshire 03755, USA
[2]Program in Molecular Biology and Biotechnology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA
[3]Norris Cotton Cancer Center, Dartmouth Medical School, Hanover, New Hampshire 03755, USA

ABSTRACT

The replication-dependent histone mRNAs are cell-cycle-regulated and expressed only during S phase. In contrast to all other eukaryotic mRNAs, the histone mRNAs end in a highly conserved 16-nucleotide stem–loop rather than a poly(A) tail. The stem–loop is necessary and sufficient for the post-transcriptional regulation of histone mRNA during the cell cycle. The histone mRNA 3′ stem–loop is bound by the stem–loop binding protein (SLBP) that is involved in pre-mRNA processing, translation, and stability of histone mRNA. Immunoprecipitation (IP) of RNA-binding proteins (RBPs) followed by microarray analysis has been used to identify the targets of RNA-binding proteins. This method is sometimes referred to as RIP-Chip (RNA IP followed by microarray analysis). Here we introduce a variation on the RIP-Chip method that uses a recombinant RBP to identify mRNA targets in a pool of total RNA; we call this method recombinant, or rRIP-Chip. Using this method, we show that recombinant SLBP binds exclusively to all five classes of histone mRNA. We also analyze the messages bound to the endogenous SLBP on polyribosomes by immunoprecipitation. We use two different microarray platforms to identify enriched mRNAs. Both platforms demonstrate remarkable specificity and consistency of results. Our data suggest that the replication-dependent histone mRNAs are likely to be the sole target of SLBP.

Keywords: microarray; histone mRNA; cell cycle; gene expression; RIP-Chip; rRIP-Chip; immunoprecipitation

INTRODUCTION

Synthesis of the five classes of histones (H1, H2A, H2B, H3, and H4) is coordinately regulated and their expression is restricted to S phase of the cell division cycle. Most of this regulation is post-transcriptional, occurring at the levels of pre-mRNA processing and mRNA stability. A threefold increase in transcription and a 10-fold increase in the efficiency of pre-mRNA processing leads to a 35–50-fold increase in histone mRNA levels during S phase (Harris et al. 1991). As cells exit S phase or when DNA synthesis is inhibited, histone mRNAs are coordinately destabilized (Sittman et al. 1983).

All of the replication-dependent histone mRNAs end in a conserved stem–loop rather than a poly(A) tail. The histone 3′ end is necessary and sufficient for the cell-cycle regulation of histone mRNA (Pandey and Marzluff 1987).

The protein components that bind the 3′ end include SLBP (also known as the hairpin binding factor; HBF) (Wang et al. 1996; Martin et al. 1997), which binds to the highly conserved 3′ stem–loop, the U7 small nuclear ribonucleoprotein (snRNP), which base pairs with the histone downstream element (HDE), and components of the polyadenylation machinery that are involved in the cleavage reaction (Dominski et al. 2005; Kolev and Steitz 2005). The molecular mechanisms that precisely generate the histone 3′ end involve a single endonucleolytic cleavage that occurs precisely between the two *cis*-acting elements, the stem–loop and HDE (Fig. 1A). In mammalian cells, during pre-mRNA processing in the nucleus, SLBP binds to the 3′ stem–loop and stabilizes the binding of U7 snRNP (Dominski et al. 1999; Pillai et al. 2003). The cleavage reaction that generates the histone 3′ end is performed by the same catalytic protein core used by the polyadenylation machinery and requires CPSF73 (Dominski et al. 2005), the scaffolding protein symplekin (Kolev and Steitz 2005), and possibly other components of the polyadenylation machinery.

SLBP is critical for the regulation of histone mRNA. The protein accompanies the mature mRNA to the cytoplasm as a component of the histone messenger ribonucleoprotein

particle (mRNP) (Whitfield et al. 2004; Erkmann et al. 2005), where it plays a role in translation of the mRNA (Sanchez and Marzluff 2002). SLBP is a major mediator of histone mRNA stability, as the distance between the stem–loop and the stop codon is important for regulated turnover of histone message (Graves et al. 1987; Kaygun and Marzluff 2005b). Although the exact mechanism by which the histone mRNAs are degraded has not been elucidated, it is known to involve the nonsense mediated decay factor UPF1 (Kaygun and Marzluff 2005a) and a 3′–5′ exonuclease (3′hExo) (Dominski et al. 2003).

The SLBP is itself cell-cycle regulated by both translational and post-translational mechanisms (Whitfield et al. 2000). The expression of SLBP mRNA increases twofold during G1/S (Whitfield et al. 2002). In contrast, SLBP protein levels increase more than 20-fold as cells enter S phase (Whitfield et al. 2000), as a result of regulation of translation of SLBP mRNA. SLBP is rapidly degraded by the proteasome at the S/G2 boundary following phosphorylation on two threonine residues (Whitfield et al. 2000; Zheng et al. 2003).

We have previously shown that SLBP is bound to all five classes of replication-dependent histone mRNA as a component of the histone mRNP (Whitfield et al. 2004). Here we systematically examine all the possible SLBP targets in HeLa cells on a genome-wide scale.
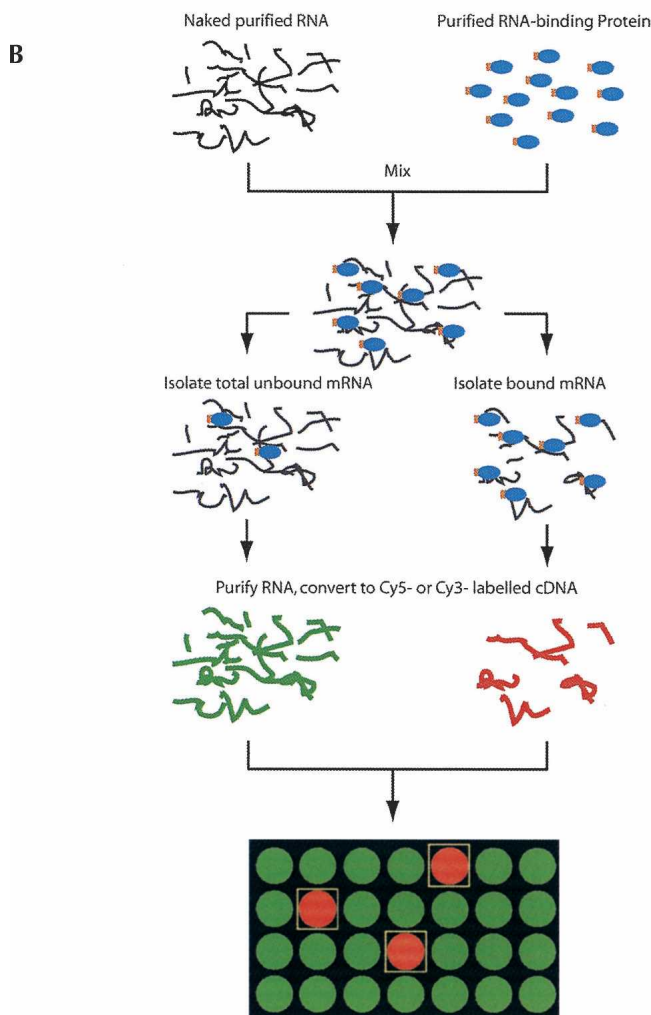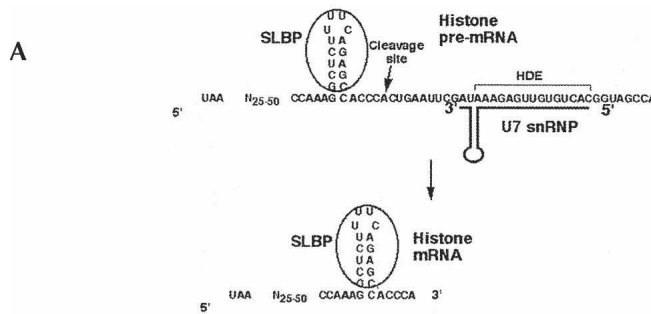
RESULTS

We have previously shown that the SLBP is a component of the histone mRNP by immunoprecipitation (IP) of the protein followed by the detection of specific mRNAs by traditional methods such as S1 nuclease assay and semi-quantitative RT-PCR (Whitfield et al. 2004). Immunoprecipitation of RNA-binding proteins (RBPs) followed by microarray analysis has been used to identify mRNA targets of RBPs in yeast (Shepard et al. 2003; Gerber et al. 2004; Inada and Guthrie 2004; Guisbert et al. 2005) and human cells (Tenenbaum et al. 2000; Eystathioy et al. 2002; Intine et al. 2003; Keene and Lager 2005; Stephens et al. 2005); this method is sometimes referred to as RIP-Chip (RNA ImmunoPrecipitation followed by microarray (Chip) analysis) (Schmitz-Linneweber et al. 2005). In this study we have analyzed the targets of the SLBP in purified total polyribosomal RNA from HeLa cells using a variation on the RIP-Chip method that we have termed rRIP-Chip. We compare these results to those obtained by precipitating the messages associated with endogenous SLBP on polyribosomes by the same method. In each case we find

FIGURE 1. Strategy for identifying SLBP targets using the rRIP-Chip method. (*A*) Histone mRNAs are not polyadenylated but instead end in a conserved stem–loop. The message undergoes a single processing step in the nucleus that involves the SLBP and the U7 snRNP. The mature message ending in a conserved stem–loop structure is then transported to the cytoplasm where it is loaded onto polyribosomes. (*B*) Recombinant purified SLBP was used to identify mRNA targets. Total RNA is purified from either isolated polyribosomes or whole cells. Purified RBP is mixed with the purified RNA and RNA–protein complexes isolated using either an antibody against SLBP or an antibody against a tag in the recombinant protein. RNA is then purified from the bound and unbound fractions. Purified RNA is converted to either Cy3- or Cy5-labeled cDNA by reverse transcription primed with random hexamer primers and hybridized to whole genome microarrays.

evidence to suggest the SLBP associates exclusively with the replication-dependent histone mRNAs.

rRIP-Chip: RNA immunoprecipitation with a recombinant RBP

The experimental approach that we used is outlined in Figure 1B. In this method, purified total RNA is incubated with a purified recombinant RBP. RNP complexes are allowed to form in vitro and subsequently isolated by affinity selection using an antibody to the purified protein or the epitope tag. The RNAs bound to the recombinant protein and the mRNAs in the unbound fraction are purified and converted to labeled cDNA by reverse transcription and the two samples competitively hybridized to DNA microarrays to identify mRNAs bound to the recombinant RBP.

We have tested the rRIP-Chip approach using purified, histidine-tagged SLBP (HIS-SLBP) and purified RNA. HIS-SLBP was produced in the baculovirus system and purified by Ni-affinity chromatography (Zheng et al. 2003). Polyribosomes were isolated from cytoplasmic lysates of HeLa S3 cells by centrifugation through a sucrose cushion, and total polyribosomal RNA was isolated by phenol:chloroform extraction.

Approximately 20 ng of HIS-SLBP were added to 40 μg of purified total RNA (an estimated fivefold molar excess of SLBP relative to total messenger RNA; an estimated 100-fold molar excess relative to the estimated amount of histone mRNA in an asynchronous population of HeLa cells) and SLBP–RNA complexes allowed to form. The resulting complexes were isolated by IP with affinity purified anti-SLBP antibody followed by selection with protein-A agarose beads. We repeated the IPs of these SLBP–RNA complexes four independent times. In every case, a mock IP, with the anti-SLBP preincubated with the antigenic peptide, was performed in parallel as a negative control. As an additional negative control, we performed a single precipitation experiment without any antibody. The mRNA bound in these complexes was purified; mRNA not bound was isolated from the supernatant in parallel. Genes specifically enriched in each IP were identified by microarray hybridization in a two-color hybridization experiment with the precipitated mRNA in the Cy5-labeled channel and the unbound, supernatant mRNA in the Cy3-labeled channel. Extracted RNA was converted to Cy5-labeled or Cy3-labeled cDNA by reverse transcription primed with random hexamer (pdN6) and hybridized to microarrays containing 41,520 elements representing 19,164 UNIGENE clusters (Fig. 1B).

mRNA transcripts enriched in each IP were identified by assigning each element on the microarray a percentile rank based upon the log_2 of the Cy5/Cy3 ratio in each experiment. The percentile rank was calculated based on the enrichment of each RNA precipitated from SLBP–RNA complexes, relative to the unbound RNA in the supernatant. We subsequently used the four percentile rank values from the four repeated experiments to calculate the median

percentile rank for each precipitation. The distribution of intensities (Fig. 2A) in the anti-SLBP experiment shows enrichment of particular mRNAs at the high percentile ranks. The distribution of the median percentile ranks in the five mock IP experiments of SLBP from polyribosomes exhibits a normal distribution (Fig. 2B). Therefore, we see enrichment of a subset of mRNAs in the anti-SLBP experiment that are not observed in the negative control experiments. We have defined SLBP targets as those with a percentile rank above 98.83%. This is similar to the analytical method that has been used to identify targets in Chromatin IP experiments followed by microarray analysis (ChIP-chip) in yeast (Lieb et al. 2001) and the PUF proteins using RNA IPs (Gerber et al. 2004).

Examination of the enriched genes reveals that the replication-dependent histone genes are consistently selected as the most enriched genes on the microarray. In total, 27 genes had a percentile rank above our cutoff. Seventeen of these are replication-dependent histone genes (63% of the total genes selected). This list of genes includes five H2A genes (HIST1H2AL, HIST1H2AC, HIST1H2AX, HIST2H2AA, HIST1H2AM), five H2B genes (HIST1H2BC, HIST1H2BK, HIST2H2BE, HIST1H2BD, HIST1H2BL), two H4 genes (HIST1H4H, HIST1H4C), one H3 gene (HIST1 H3D), and one H1 gene (HIST1H1C). The list of 10 nonhistone genes (which we later show to be false positives) contained an unnamed transcribed locus (H05961), a hypothetical protein (AA936181), and an MHC Class II gene.

Not all replication-dependent histone mRNAs were selected as enriched in our IP experiment, so we examined the distribution of these genes found on the cDNA microarrays. The data for all replication-dependent histone genes on the microarray (see Supplementary Table S1 at http://whitfieldlab.dartmouth.edu/mRNP/) were extracted from the normalized data in which each gene has been centered on the average of the four mock IP experiments. The log_2 of the Cy5/Cy3 ratios has been plotted for all nonhistone genes in the microarray experiment (Fig. 4A), which shows that most of the mRNAs measured in our experiment are distributed around zero. In contrast, when the replication-dependent histone mRNAs are examined, we find that >50% of these genes show 4–16-fold enrichment in the specific IP experiment (Fig. 4A, green). Therefore 17 of the 33 replication-dependent histone genes found on the Stanford cDNA microarray are enriched in our experiment. Two of the genes not identified as enriched, HIST1H4B and HISTH2AC, fall just below our cutoff (and HIST1H2AC is represented by a second clone on the array that was identified as enriched; Supplementary Table S1).

Isolation of messenger ribonucleoprotein complexes formed in vivo

The rRIP-Chip method uses a tagged recombinant RBP and will detect direct RNA–protein interactions in a controlled
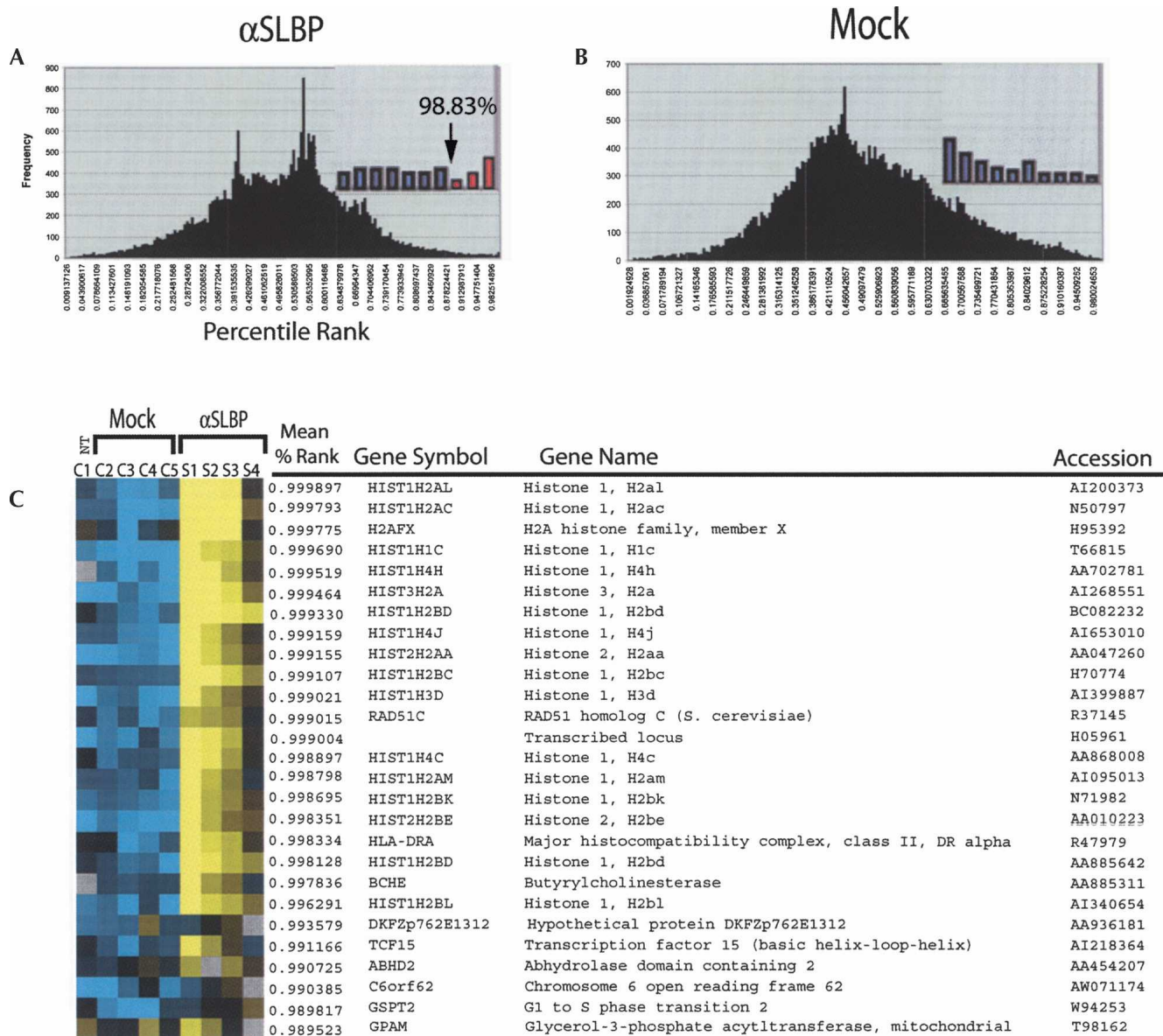
FIGURE 2. The replication-dependent histone mRNAs are bound by recombinant SLBP in vitro as determined by rRIP-Chip. We used purified recombinant, histidine-tagged SLBP and purified polyribosomal RNA to test the rRIP-Chip method described in the text and Figure 1. We performed four independent IP experiments to identify SLBP targets using an anti-SLBP antibody. Enriched mRNAs were identified by assigning each element on the microarray a percentile rank in each of the four experiments. These were subsequently used to calculate the median percentile rank for all four experiments. (*A*) The distribution of the median percentile ranks for the four αSLBP IPs are graphed. The distribution shows a small tail at the high percentile ranks (bin size = 0.005794) and shows a bimodal distribution. We have defined putative SLBP targets as those that fall to the right of this trough with a percentile rank above 98.83% (the distribution of these genes is colored red in the histogram inset). (*B*) The distribution of the median percentile ranks from the four mock IP experiments and the single no antibody IP are shown. There is not an obvious tail evident in the data. (*C*) The data for the genes selected as SLBP targets are displayed ordered by their percentile ranks. The log$_2$ of the Cy5/Cy3 ratio for each gene in each experiment is displayed with the data centered on its median value. Expression values above the median have been color-coded yellow and those below the median have been color-coded blue.

system in vitro. In order to examine the complement of mRNAs to which the SLBP is bound in vivo we have identified the targets of the endogenous SLBP on polyribosomes by microarray analysis. In vivo, the SLBP is bound to the histone mRNA during processing in the nucleus and accompanies the mature message to the cytoplasm. Therefore the complement of messages bound by the SLBP in vivo could be different from the mRNAs bound by the recombinant protein. In order to compare the complement of bound messages in this more physiologically relevant scenario, we have isolated the endogenous histone mRNP complex.
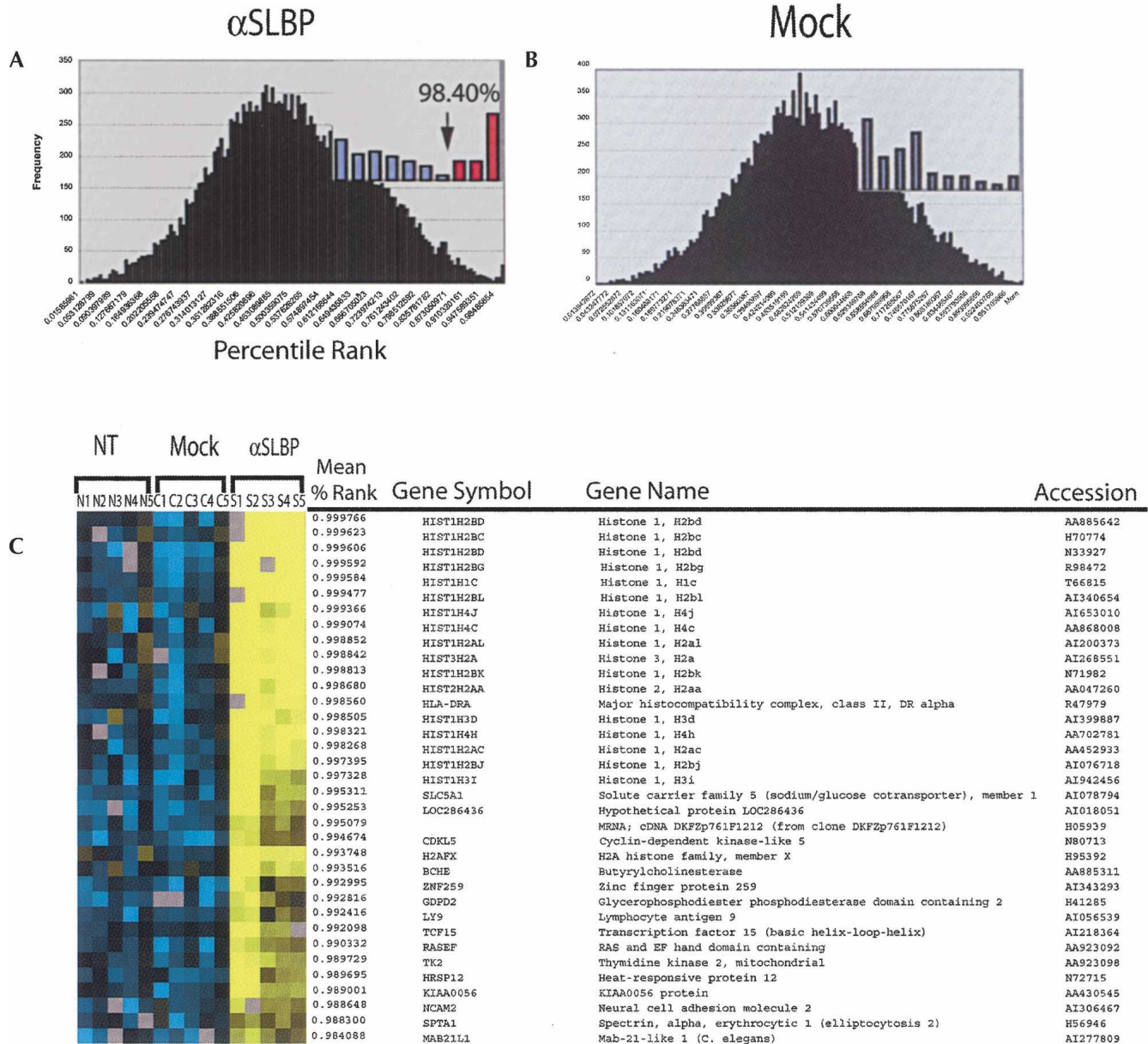
αSLBP — 98.40%

Mock

C

NT	Mock	αSLBP	Mean % Rank	Gene Symbol	Gene Name	Accession
			0.999766	HIST1H2BD	Histone 1, H2bd	AA885642
			0.999623	HIST1H2BC	Histone 1, H2bc	H70774
			0.999606	HIST1H2BD	Histone 1, H2bd	N33927
			0.999592	HIST1H2BG	Histone 1, H2bg	R98472
			0.999584	HIST1H1C	Histone 1, H1c	T66815
			0.999477	HIST1H2BL	Histone 1, H2bl	AI340654
			0.999366	HIST1H4J	Histone 1, H4j	AI653010
			0.999074	HIST1H4C	Histone 1, H4c	AA868008
			0.998852	HIST1H2AL	Histone 1, H2al	AI200373
			0.998842	HIST3H2A	Histone 3, H2a	AI268551
			0.998813	HIST1H2BK	Histone 1, H2bk	N71982
			0.998680	HIST2H2AA	Histone 2, H2aa	AA047260
			0.998560	HLA-DRA	Major histocompatibility complex, class II, DR alpha	R47979
			0.998505	HIST1H3D	Histone 1, H3d	AI399887
			0.998321	HIST1H4H	Histone 1, H4h	AA702781
			0.998268	HIST1H2AC	Histone 1, H2ac	AA452933
			0.997395	HIST1H2BJ	Histone 1, H2bj	AI076718
			0.997328	HIST1H3I	Histone 1, H3i	AI942456
			0.995311	SLC5A1	Solute carrier family 5 (sodium/glucose cotransporter), member 1	AI078794
			0.995253	LOC286436	Hypothetical protein LOC286436	AI018051
			0.995079		MRNA; cDNA DKFZp761F1212 (from clone DKFZp761F1212)	H05939
			0.994674	CDKL5	Cyclin-dependent kinase-like 5	N80713
			0.993748	H2AFX	H2A histone family, member X	H95392
			0.993516	BCHE	Butyrylcholinesterase	AA885311
			0.992995	ZNF259	Zinc finger protein 259	AI343293
			0.992816	GDPD2	Glycerophosphodiester phosphodiesterase domain containing 2	H41285
			0.992416	LY9	Lymphocyte antigen 9	AI056539
			0.992098	TCF15	Transcription factor 15 (basic helix-loop-helix)	AI218364
			0.990332	RASEF	RAS and EF hand domain containing	AA923092
			0.989729	TK2	Thymidine kinase 2, mitochondrial	AA923098
			0.989695	HRSP12	Heat-responsive protein 12	N72715
			0.989001	KIAA0056	KIAA0056 protein	AA430545
			0.988648	NCAM2	Neural cell adhesion molecule 2	AI306467
			0.988300	SPTA1	Spectrin, alpha, erythrocytic 1 (elliptocytosis 2)	H56946
			0.984088	MAB21L1	Mab-21-like 1 (C. elegans)	AI277809

FIGURE 3. The replication-dependent histone mRNAs are bound to the endogenous SLBP in vivo. Purified polyribosomes were incubated with anti-SLBP. We repeated the SLBP IPs independently five different times. In each case, a mock IP was performed in parallel without any SLBP antibody or anti-SLBP preincubated with the antigenic peptide, resulting in a total of 10 negative controls. For each IP, the precipitated mRNAs were assigned a percentile rank. The percentile rank was calculated based on the enrichment of each mRNA precipitated from polyribosomes, relative to the unprecipitated RNA from the polyribosomes. We subsequently used the five percentile rank values to calculate the median percentile rank for each precipitation. (*A*) Shown is the distribution of median percentile ranks in the anti-SLBP experiment. The distribution is skewed toward a single tail of the normal distribution, indicating enrichment of particular mRNAs (bin size = 0.007454). We have defined SLBP targets as those that fall to the right of the trough of the bimodal distribution. This cutoff, which corresponds to a percentile rank of 98.4% (and is colored red in the inset), identifies 36 genes (0.19% of the genes on the array) as significantly enriched. (*B*) The distribution of the median percentile ranks is shown from the 10 mock IP experiments from polyribosomes, which shows a normal distribution and illustrates that very few genes are enriched in the negative control experiments (bin size = 0.007326). (*C*) Genes identified as targets of the SLBP are displayed using Java Treeview, where expression values above the mean have been color-coded yellow and those below the mean have been color-coded blue. The genes are ordered by their median percentile rank in the IP of the endogenous SLBP. The 18 most highly enriched genes are histones. Of those genes selected, a subset of genes (*n* = 15) are not histone genes.

To isolate targets of the SLBP in vivo, polyribosomes were incubated with the anti-SLBP C-terminal antibody. mRNP complexes were isolated by coIP of the SLBP with its bound mRNAs. We performed five IPs from independent preparations of polyribosomes. In each case, a mock IP was performed in parallel with no SLBP antibody, and
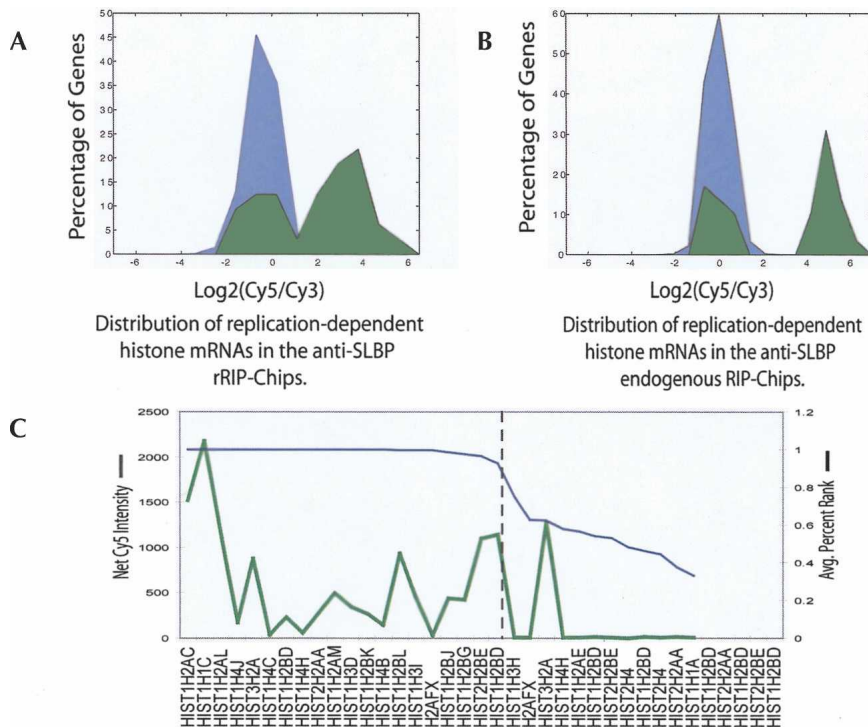
FIGURE 4. Distribution of all replication-dependent histone genes found on the microarray.
(*A*) Shown are histograms showing the distribution of $\log_2(\text{Cy5/Cy3})$ for all genes on the microarray (blue) and for the replication-dependent histone mRNAs (green). In each case, the *y*-axis is the percentage of genes plotted at a given intensity value. The identity of each histone gene, along with accession numbers and the average $\log_2(\text{Cy5/Cy3})$ is given in Supplementary Tables S1 and S2. In panel *A*, the average $\log_2$ of the Cy5/Cy3 ratio distributions across all four αSLBP rRIP-Chip experiments is plotted (bin size = 0.1) as a percentage of all genes (blue; 33,125 genes) or as a percentage of only the replication-dependent histone genes (green; 33 genes). (*B*) The average $\log_2$ (Cy5/Cy3) ratio across the five endogenous SLBP RIP-Chip experiments are plotted (bin size = 0.1) versus the percentage of all genes on the array (blue, $N = 27,957$) and the percentage of only the replication-dependent histone genes (green, $N = 29$). (*C*) Ten nanograms of a ''synthetic histone mRNA pool'' comprised of T7 transcribed histone coding regions (HIST1H1A, HIST1H2AB, HIST1H2BB, HIST1H3A, HIST1H4A) were labeled and hybridized to Stanford cDNA microarrays. The net Cy5 intensity for each replication-dependent histone gene on the array was extracted from the data and plotted along with the average percentile rank for that gene in both the RIP-Chip and rRIP-Chip experiments. A dashed line indicates the cutoff for an enriched gene. These data show that the histone genes identified as enriched in either the rRIP-Chip or the RIP-Chip experiment show significant intensities when hybridized to the positive control synthetic histone mRNA pool. In contrast, histone mRNAs that were not identified as enriched show hybridization signals approaching background. The exception is HIST3H2A, which shows high net Cy5 signal intensity, yet was not identified as enriched.

ranks in the 10 mock IP experiments; the anti-SLBP distribution is skewed toward a single-tail of the normal distribution indicating enrichment of particular mRNAs (Fig. 3A, see inset).

Again, we defined the specific targets of the SLBP as those that fell to the right of the bimodal distribution, which corresponds to a percentile rank of 98.40%. Using this cutoff, 35 genes (0.19% of the genes on the array) are significantly enriched. This list contains 18 replication-dependent histone genes and 15 nonhistone genes (Fig. 3C). Notably, the genes with the highest percentile rank are exclusively histone genes, which are known SLBP targets. Analysis of these data using statistical analysis of microarrays (SAM) identifies a similar list of genes (data not shown).

We have displayed the patterns of gene expression for each of the 33 genes called enriched, ordered by their percentile rank (Fig. 3C). Each of the five classes of core histone genes is represented. At the top of this list are six different clones for histone H2B genes (HIST1H2BC, HIST1H2BD, HIST1H2BG, HIST1H2BJ, HIST1H2BK, and HIST1H2BL). Among the H2A genes, five different genes are represented including HIST1H2AL, HIST1H2AC, HIST2H2AA, and HIST3H2A. The coding regions of the H2A and H2B genes are 44%–48% identical at the sequence level, and therefore they are unlikely to cross hybridize in our assay. The H4 genes are represented by three different clones (HIST1H4C, HIST1H4H, HIST1H4J) while the H3 genes are represented by two clones (HIST1H3D, HIST1H3I). The linker histone H1c

also with anti-SLBP preincubated with the antigenic peptide, resulting in a total of 10 negative controls. mRNP complexes were isolated by affinity selection of the SLBP complexes with protein-A agarose beads. The mRNA bound in these complexes and in the supernatant was purified and hybridized to cDNA microarrays as described above.

Genes enriched in the SLBP IPs were assigned a percentile rank based on the enrichment of each mRNA precipitated from polyribosomes, relative to the unprecipitated RNA in the supernatant. We used the five percentile rank values to calculate the median percentile rank for each precipitation. Figure 3A shows the distribution of the median percentile

(HIST1H1C) measured in our original study of the histone mRNP complex is the only H1c clone included in the list. In addition, the replacement variant H2A family member X (H2AFX) is also present. This histone gene is unique in that it contains a stem–loop and ends in a histone 3′ end when cells are in S phase but ends in a poly(A) tail outside of S phase (Bonner et al. 1993). In previous work, we have shown that both forms of the H2aX mRNA are indeed bound by the SLBP in mouse cells (Whitfield 1999).

We examined the distribution of all of the replication-dependent histone genes that were represented on the microarrays and passed our minimum criteria for spot

quality. The complete list of histone genes on the microarray can be found in Supplementary Table S2. The log$_2$ of the Cy5/Cy3 intensity ratios for all of these replication-dependent histone genes in the specific anti-SLBP IP are graphed in a histogram (Fig. 4B). When the distribution of intensities for the replication-dependent histone mRNAs are specifically examined in the SLBP IP (Fig. 4B, green), there is an 8–64-fold enrichment of these specific RNAs. Of the genes on the array, 99.90% have a distribution around zero in the negative controls, indicating no enrichment.

Hybridization of a synthetic histone mRNA pool

The simplest explanation for why some histone mRNAs are identified as enriched on our arrays and some are not is that some clones are technical failures. This could result from failed PCR reactions during microarray construction, incorrectly annotated clones, or clones that only contain sequence that maps between the stem–loop and the cryptic polyadenylation site, which is sometimes used outside of S phase. In order to test this hypothesis, we created a "synthetic histone mRNA pool" to use as a positive control by in vitro transcribing the coding regions of five different histone genes (HIST1H1A, HIST1H2AB, HIST1H2BB, HIST1H3A, HIST1H4A) using T7 RNA polymerase. mRNA from each clone was mixed in equal amounts to create the synthetic histone mRNA pool. Next, we labeled increasing amounts of this pool (10 pg, 100 pg, 1.0 ng, and 10 ng) and hybridized each to cDNA microarrays. We find our lower limit of detection is ∼1.0 ng of the pool (0.20 ng of each of the individual histone mRNAs; Supplementary Figure S1A).

We have examined the raw Cy5 (channel 2) intensities that result from hybridizing the 10-ng pool of synthetic histone mRNAs to the microarrays. The raw Cy5 intensity for each replication-dependent histone mRNA on the array has been plotted, ordered by their average percentile rank in both the rRIP-Chip and RIP-Chip experiments (Fig. 4C). The absolute intensity varies for each gene because of variable amounts of DNA spotted for each clone. When the absolute amount of hybridization intensity is examined, we find that those histone genes identified as significantly enriched in both the experiments have a higher signal intensity on average then those genes that were not identified as significantly enriched (Fig. 4C). Only three histone genes that were called significantly enriched (HIST1H4C, HIST1H4H, H2AFX) show hybridization signals close to background. In contrast, those replication-dependent histone genes not called significantly enriched (Fig. 4C) have a Cy5 net signal intensity very close to zero when hybridized to the synthetic histone mRNA pool. The sole exception is the clone for HIST3H2A, which has net Cy5 intensity of 1281, yet was not called significantly enriched in either experiment. Five replication-dependent histone genes did not pass basic criteria for spot quality in at least

80% of the arrays analyzed and therefore did not receive percentile ranks (Fig. 4C, last five genes). Therefore, of the 17 replication-dependent histone genes not identified as enriched in the RIP-Chip or rRIP-Chip experiments, we conclude that 16 are technical negatives and one is a true negative. Of the 21 replication-dependent histone genes on the array that showed signal intensity when hybridized to our positive control sample, we identify 20 as significantly enriched and one as not enriched, resulting in a false negative rate of ∼4.8%.

Concordance between mRNP complexes formed in vivo and in vitro

We examined the overlap between the genes identified in the rRIP-Chip experiment and those genes identified as components of the mRNP complex in vivo using the standard RIP-Chip approach. We found only three non-histone genes that were consistently enriched in both in vitro and in vivo IP experiments (Fig. 5). Shown is the color-coded image showing the genes identified as enriched in both data sets. The median percentile rank for each gene in the rRIP-Chip experiment, the RIP-Chip experiment, and the mean percentile rank from the two experiments are shown. Three nonhistone genes, butyrylcholinesterasee (BCHE), Transcription factor 15 (TCF15), and HLA-DRA, are identified as enriched in both experiments.

Using our positive control histone mRNA pool described above, we have analyzed each of the nonhistone mRNA genes to ask if their enrichment could possibly result from cross hybridization to precipitated histone mRNAs. Examination of the raw Cy5 signal intensity after hybridization of 10 ng of the synthetic histone mRNA pool to Stanford cDNA microarrays (Fig. 5B) shows that each of these are likely to be a result of cross hybridization. Shown are the raw Cy5 signal intensities for the top three histone genes identified as enriched in both data sets (genes HIST1H1C, HIST1H2AL, and HIST1H2BC), the genes BCHE, TCF15, and HLA-DRA, and two commonly used control genes, ACTA2 and TUBA3. The histone genes that are consistently identified as the most enriched have net Cy5 intensities ranging from 2174 to 1147 when hybridized to a synthetic pool of histone mRNAs. The signal intensities for the genes HLA-DRA, BCHE, and TCF15 are above background, whereas those for ACTA2 and TUBA3 are very close to background. These data indicate that the nonhistone mRNAs that are identified as enriched likely result from cross hybridization to the precipitated histone mRNAs.

RIP-Chip analysis on an independent microarray platform

We reasoned that using an independent platform would provide sufficient evidence to determine if the nonhistone genes were indeed false positives, or if they might represent
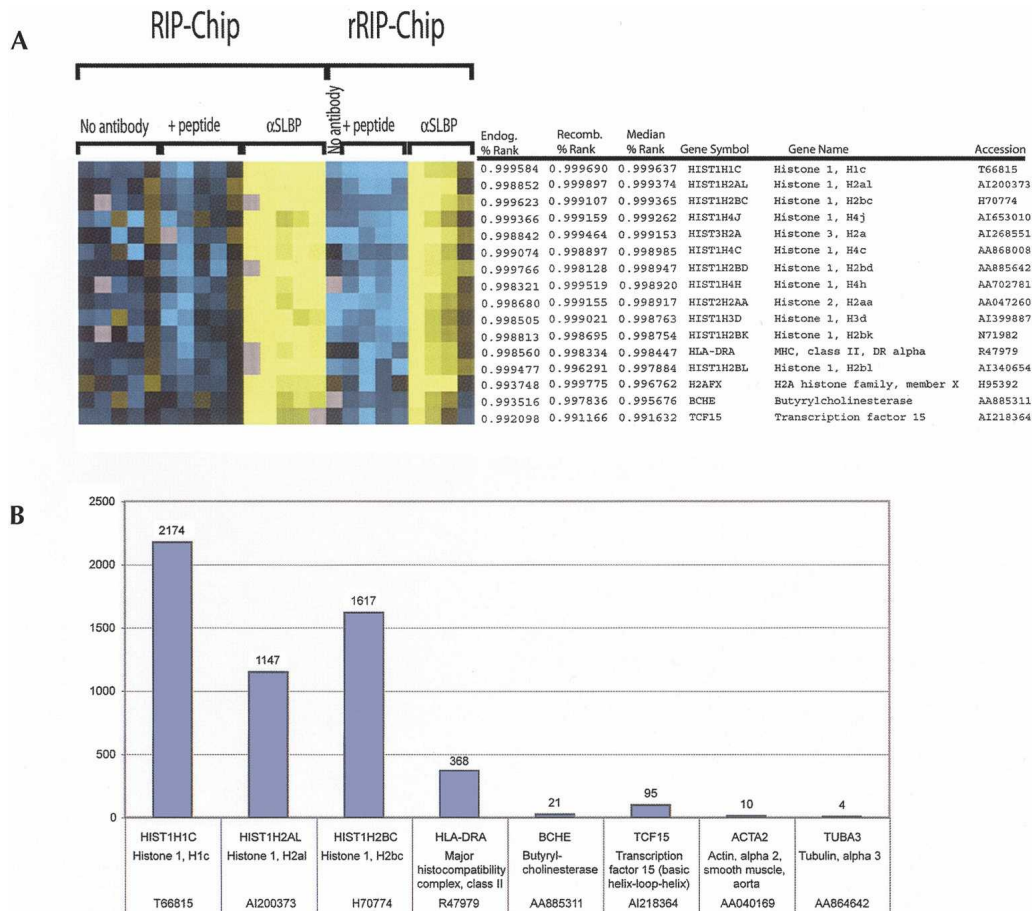
FIGURE 5. Concordance between rRIP-Chip and mRNP complexes formed in vivo. (*A*) Genes identified as enriched in both the rRIP-Chip and RIP-Chip of the endogenous mRNP complexes are shown. Expression values above the mean are color-coded yellow and those below the mean are color-coded blue. The median percentile rank for each gene in the rRIP-Chip experiment and the endogenous RIP-Chip experiment are given, as is the mean of the two measurements. The overlap between the two data sets identifies histone genes as the primary class of mRNAs bound to the SLBP. (*B*) Ten nanograms of a synthetic histone mRNA pool were hybridized to cDNA microarrays to test for cross hybridization. Shown are the net Cy5 signal intensities for the three histone genes with the highest average percentile rank from panel *A* (HIST1H1C, HIST1H2AL, HIST1H2BC), the three nonhistone genes (HLA-DRA, BCHE, TCF15), and two genes not identified as enriched in either the rRIP-Chip or RIP-Chip experiments (ACTA2, TUBA3). The nonhistone genes show signal intensities above background (5–10 units) but lower than that of the replication-dependent histone mRNAs, when hybridized to the synthetic histone mRNA pool containing only the five different replication-dependent histone mRNA coding regions. This suggests their enrichment is likely a result of low-level cross hybridization.

bona fide SLBP targets. We chose to use the Agilent Technologies oligonucleotide microarrays comprised of ~44,000 different 70mers. The SLBP IP was repeated three times, independently, and in each case a negative control with no antibody was performed in parallel. Each of the three SLBP IPs and the three mock IPs were hybridized to the Agilent microarrays (six arrays total) and analyzed by assigning each element on the array a percentile rank as described above.

We selected a cutoff of 99.30% from the distribution of the histogram. Of the 34 genes identified as enriched on the Agilent microarray, 100% were histone genes (Fig. 6). As in the other experiments, only histone genes were found at the upper enrichment values, whereas the nonhistone genes

previously identified using cDNA microarrays were found at much lower percentile ranks. The previous off-target genes identified as significantly enriched in the analysis of our Stanford cDNA microarray data were also present on the Agilent microarray and are not present in the 34 enriched genes selected. Examination of the percentile rank for each of those genes reveals that they are almost all below the 50th percentile on the Agilent microarray. Most notably, BCHE was found at 47th percentile rank; three oligos representing TCF15 were present on the microarray and found at 61st, 27th, and 22nd percentile ranks when analyzed on the Agilent platform. This confirms the interpretation that nonhistone genes are off-target, false positives.
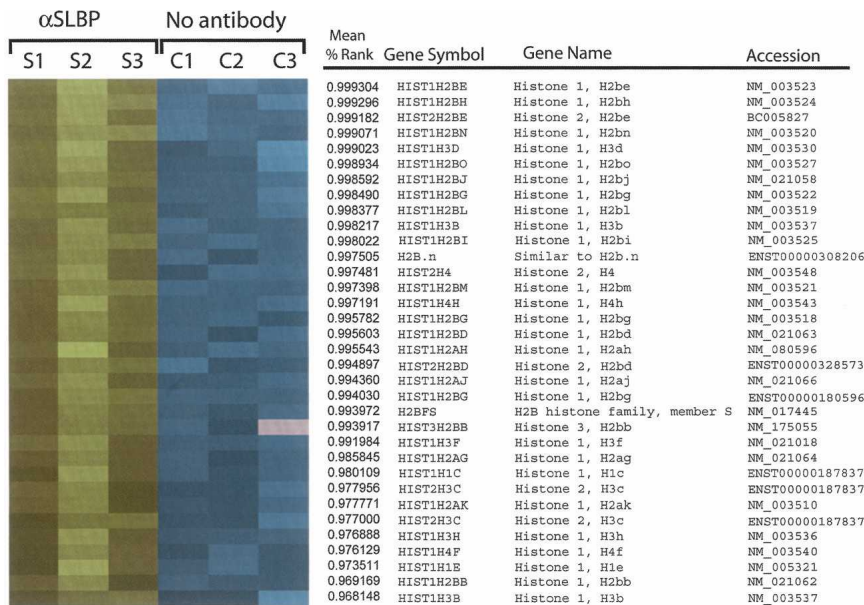
αSLBP			No antibody			Mean % Rank	Gene Symbol	Gene Name	Accession
S1	S2	S3	C1	C2	C3	0.999304	HIST1H2BE	Histone 1, H2be	NM_003523
						0.999296	HIST1H2BH	Histone 1, H2bh	NM_003524
						0.999182	HIST2H2BE	Histone 2, H2be	BC005827
						0.999071	HIST1H2BN	Histone 1, H2bn	NM_003520
						0.999023	HIST1H3D	Histone 1, H3d	NM_003530
						0.998934	HIST1H2BO	Histone 1, H2bo	NM_003527
						0.998592	HIST1H2BJ	Histone 1, H2bj	NM_021058
						0.998490	HIST1H2BG	Histone 1, H2bg	NM_003522
						0.998377	HIST1H2BL	Histone 1, H2bl	NM_003519
						0.998217	HIST1H3B	Histone 1, H3b	NM_003537
						0.998022	HIST1H2BI	Histone 1, H2bi	NM_003525
						0.997505	H2B.n	Similar to H2b.n	ENST00000308206
						0.997481	HIST2H4	Histone 2, H4	NM_003548
						0.997398	HIST1H2BM	Histone 1, H2bm	NM_003521
						0.997191	HIST1H4H	Histone 1, H4h	NM_003543
						0.995782	HIST1H2BG	Histone 1, H2bg	NM_003518
						0.995603	HIST1H2BD	Histone 1, H2bd	NM_021063
						0.995543	HIST1H2AH	Histone 1, H2ah	NM_080596
						0.994897	HIST2H2BD	Histone 2, H2bd	ENST00000328573
						0.994360	HIST1H2AJ	Histone 1, H2aj	NM_021066
						0.994030	HIST1H2BG	Histone 1, H2bg	ENST00000180596
						0.993972	H2BFS	H2B histone family, member S	NM_017445
						0.993917	HIST3H2BB	Histone 3, H2bb	NM_175055
						0.991984	HIST1H3F	Histone 1, H3f	NM_021018
						0.985845	HIST1H2AG	Histone 1, H2ag	NM_021064
						0.980109	HIST1H1C	Histone 1, H1c	ENST00000187837
						0.977956	HIST2H3C	Histone 2, H3c	ENST00000187837
						0.977771	HIST1H2AK	Histone 1, H2ak	NM_003510
						0.977000	HIST2H3C	Histone 2, H3c	ENST00000187837
						0.976888	HIST1H3H	Histone 1, H3h	NM_003536
						0.976129	HIST1H4F	Histone 1, H4f	NM_003540
						0.973511	HIST1H1E	Histone 1, H1e	NM_005321
						0.969169	HIST1H2BB	Histone 1, H2bb	NM_021062
						0.968148	HIST1H3B	Histone 1, H3b	NM_003537

FIGURE 6. RIP-Chip analysis of the SLBP on Agilent oligonucleotide microarrays. We used Agilent 44,000 element oligonucleotide microarrays to analyze the RNAs bound to the endogenous SLBP and to compare the results to those obtained by RIP analysis of SLBP on the cDNA microarrays. The genes are ordered by mean percentile rank in the RIP-Chip experiment determined using the Agilent arrays. Only those genes above a percentile rank of 99.30% are shown. The results obtained on the Agilent platform indicate that histone genes are likely to be the primary target of the SLBP. It is notable that the nonhistone genes identified as significantly enriched in the rRIP-Chip and RIP-Chip treatments are not enriched in this analysis.

Analysis of targets for the histone stem–loop regulatory elements

We further analyzed each precipitated message identified using the cDNA microarrays for the presence of the histone stem–loop, the well-characterized binding site for the SLBP. The identity of each clone identified as enriched was confirmed by comparing the 5′ and 3′ EST sequence reads for each clone on the array to the human genome sequence using BLAT software (BLAST-like Alignment Tool; UCSC Genome Browser) (Kent 2002). The representative mRNA sequence to which each clone corresponded was then analyzed for a degenerate histone stem–loop consensus sequence (AANGGNNNNNNNNNGNGCC). The histone stem–loop sequence was found in 19 of the genes identified as enriched in our screen (Supplementary Table S4). There is a strong consensus among the binding sites found in the precipitated histone mRNAs with four variable positions. These are changes in the flanking sequence of the stem–loop at position 3 (C/A change at positions 13) in the loop of the stem–loop. There are two compensatory changes in the stem—the T/C change at position 7 and A/G at position 16. A graphical representation of this binding site is shown at the bottom on Supplementary Table S4. Of the 15 nonhistone genes identified in the endogenous SLBP IP, none were found to have a conserved stem–loop binding site to

which the SLBP might bind. The lack of binding sites suggests these genes are almost certainly not SLBP targets.

Testing off-target genes by RT-PCR

In order to validate our approach, we analyzed the precipitated and supernatant RNA with RT-PCR primers to the HIST1H1A used in our initial study of the histone mRNP (Fig. 7A; Whitfield et al. 2004). Consistent with our previous data, the HIST1H1A gene is bound by SLBP and specifically precipitates in the IP experiment. As a negative control, we measured levels of another cell-cycle-regulated mRNA that does not contain a binding site for the SLBP (Whitfield et al. 2002; Laoukili et al. 2005), the transcription factor FOXM1. The FOXM1 mRNA is found to be present in the RNA isolated from the supernatant of the IP but is not present in RNA from the SLBP precipitated material.

When we compared the list of genes that were selected in both the IP of the endogenous SLBP and the in vitro reconstituted SLBP (Fig. 5) we found that only two nonhistone genes (BCHE and TCF15) passed our objective criteria for an SLBP target. Semiquantitative RT-PCR was used to further characterize the off-target genes found in our microarray screen. PCR primers were designed for each of the nonhistone genes (Supplementary Table S3) and each gene analyzed by semiquantitative RT-PCR in either total RNA isolated from HeLa cells or the Universal Human Reference RNA (UHRR) that is a pool of RNA isolated from 10 diverse human cell lines (Novoradovskaya et al. 2004). Of 13 nonhistone genes analyzed by RT-PCR, the mRNA for five genes (CDKL5, BCHE, ZNF259, KIAA0056, SPTA1; Fig. 7B, left panel) were found not to be present in total HeLa cell RNA (our source of material in all IP experiments) although they were found to be present in the UHRR, providing a positive control for our primers and PCR reaction. Four genes are expressed in HeLa cells and present in UHRR (GDPD2, LY9, RASEF, HRSP12; Fig. 7B, right panel). Three of these four genes were represented on the Agilent microarrays, and in each case received median percentile ranks in the lower third of the distribution (29th–31st percentile; Supplementary Table S3). A third set of four genes was not detected in either the UHRR or in total Hela cell RNA, suggesting that these genes may not be expressed in either sample analyzed. In one case, TCF15, three independent primer pairs were used in the analysis. Although we cannot distinguish between
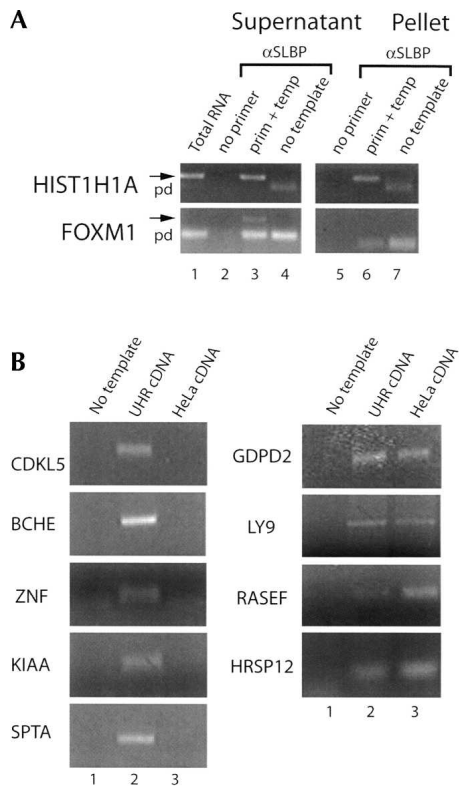
FIGURE 7. Analysis of putative SLBP targets by RT-PCR. Shown are ethidium bromide stained gels analyzing HIST1H1A and putative false-positive genes. (*A*) The RT-PCR results confirming the IP of the HIST1H1A gene. As a negative control for nonspecific IP, we have also measured the expression of an unrelated cell-cycle-regulated gene that does not contain a histone stem–loop, the Forkhead Box M1 (FOXM1) gene. The FOXM1 mRNA is not precipitated in the IP experiment. (*B*) The expression levels of a subset of nonhistone, off-target genes identified as enriched in the endogenous IP experiment were tested by semiquantitative RT-PCR. We measured the expression level of the genes in HeLa total RNA and Stratagene Universal Human Reference RNA by semiquantitative RT-PCR. cDNA was generated by reverse transcription primed with random hexamer (pdN6) and the amount of cDNA for each gene measured by PCR. The *left* panel shows genes that are not expressed in HeLa cells. We find cyclin-dependent kinase-like 5 (CDKL5), Butyrycholinesterase (BCHE), Zinc finger protein 259 (ZNF259), KIAA0056 protein, and Spectrin, alpha, erythrocytic 1 (elliptocytosis 2) (SPTA1) are present in UHRR but not in HeLa cell total RNA, suggesting the genes are not expressed in HeLa cells. Glycerophosphodiester phosphodiesterase domain containing 2 (GDPD2), Lymphocyte antigen 9 (LY9), RAS and EF hand domain containing (RASEF), and Heat-responsive protein 12 (HRSP12) are shown in the *right* panel. Each of these genes was found to be expressed in both UHRR and HeLa total RNA. Despite their presence in the HeLa total RNA, they are found below the 50th percentile when the RIP-Chip experiment is analyzed on Agilent Oligonucleotide microarrays.

failed reactions and absence of such a gene, the information known about each gene's function (http://Source.stanford.edu) suggests that they are expressed in the nervous system and brain tissue or only in early embryonic development; neither neuronal tissue nor early developmental tissues are represented in the UHRR. Two of these genes

were represented on the Agilent microarrays and received relatively low percentile ranks; TCF15 is found in the 61st percentile while SLC5A1 was found in the 48th percentile. Therefore, we conclude that the replication-dependent histone genes are likely to be the only mRNAs bound to the SLBP in HeLa cells, demonstrating the very specific role this RBP plays in the coordinate regulation of histone gene expression.

DISCUSSION

We have addressed the question of how we define all the targets of a specific RBP and how accurate microarrays are for doing this. To explore this question we have used the histone SLBP for which targets are well defined. The SLBP binds to the 3′ end of the replication-dependent histone mRNAs in the nucleus and accompanies the mature mRNA to the cytoplasm as a component of the histone mRNP (Dominski et al. 1999; Whitfield et al. 2004). The SLBP was immunoprecipitated from purified polyribosomes, under conditions that we have previously used to isolate the histone mRNP. In earlier experiments, we assayed individual RNAs, either directly or using RT-PCR to assess whether a given mRNA was bound to SLBP (Whitfield et al. 2004). Here we have broadened our approach using genomic tools to interrogate all mRNAs in the human genome to determine if they are bound by the SLBP. We have tested two different microarray platforms to identify mRNAs that are bound by the SLBP. Histone mRNAs are consistently selected as the most enriched on the array in both the rRIP-Chip and RIP-Chip experiments; the latter on two independent platforms. Some of the replication-dependent histone mRNAs are not enriched in our RIP-Chip experiments. Using a pool of T7 transcribed histone mRNA coding regions, we show that the clones on the array corresponding to these mRNAs are technical failures, showing no signal when hybridized to the positive control pool of synthetic histone mRNAs.

We find no evidence to indicate that SLBP is bound to messages other than the replication-dependent histone mRNAs in HeLa cells. Nonhistone mRNAs are detected in each experiment, although when the overlap between the different histone mRNAs is analyzed, we find that only three of the off-target mRNAs pass our criteria in both set of experiments (Fig. 5). Analysis of the net Cy5 signal intensity for these potential false-positive clones after hybridization to the "synthetic histone mRNA pool" demonstrates cross hybridization of these clones with histone coding regions, suggesting contamination of individual spots on the microarray. Further analysis of these false positives by RT-PCR and using a second microarray platform supports the conclusion that these genes are likely to be false positives. We cannot conclude that histone mRNAs are the only targets of the SLBP, since it could bind to mRNAs not expressed in HeLa cells or to mRNAs that

are represented by clones that are technical failures on our cDNA microarrays. It is also possible that accessory factors that could be lost during purification in the RIP-Chip experiment could assist SLBP in binding to mRNAs other than the histones, which would then not be detected. Nevertheless, all data points to the replication-dependent histone mRNAs as the primary target of the SLBP, demonstrating the remarkable specificity for this RBP. These data provide a controlled study of the targets of a single RNA-binding protein for which the targets are well defined, and raises issues that need to be appropriately addressed for studies of unknown RBPs.

rRIP-Chip method

We have developed a method that uses a recombinant RBP and purified RNA that is a variation on RIP-Chip, which we have termed rRIP-Chip, to analyze the mRNAs that bind to the SLBP in vitro (Figs. 1, 2). This builds upon methods for detecting the targets of RNA-binding proteins pioneered in the yeast *Saccharomyces cerevisiae* (Shepard et al. 2003; Gerber et al. 2004) and in human cells (Tenenbaum et al. 2000; Whitfield et al. 2004). The second method is the standard RIP-Chip method in which we analyze the mRNPs formed in vivo by microarray analysis, which has been used previously to identify the targets of RBPs in yeast (Shepard et al. 2003; Gerber et al. 2004; Inada and Guthrie 2004; Guisbert et al. 2005) and mammalian cells (Tenenbaum et al. 2000, 2003; Brown et al. 2001; Eystathioy et al. 2002; Keene and Tenenbaum 2002; Waggoner and Liebhaber 2003). We have evaluated the two different procedures to identify targets of SLBP.

The advantage of rRIP-chip is that it can be used to identify the targets of an RBP without making a stable cell line expressing an epitope-tagged RBP. Second, it does not require that an antibody be raised against the endogenous protein, often a rate-limiting step in the analysis of unknown proteins. Therefore, if the full-length clone of an unknown RBP is available, its binding specificity could be examined rapidly using only the purified tagged recombinant protein and purified RNA. Second, in the evaluation of an RBP, one is often interested in determining not only the mRNA targets but also the binding site for the protein. Using this method, one can sample all possible binding sites in transcribed RNA from the human genome. Thus, the further development of computational methods to identify RNA binding sites present in different mRNAs will become an important and useful component of the rRIP-Chip procedure.

The rRIP-Chip experiment tests only for RNA–protein interactions rather than interactions that may occur with other proteins bound to the mRNA (for example, via a protein–protein bridge to the mRNA). This has the advantage that an mRNA will only be identified as a target if there is a direct interaction with the RBP. This comple-ments the information obtained from examining the RNAs that are components of mRNPs in vivo. In the analysis of the mRNP in vivo, an RBP could be identified as associated with a particular mRNA because of a protein bridge with another RBP and not by directly recognizing the mRNA. Although this interaction could be physiologically relevant, it is useful to have a technique that will distinguish between direct interactions and indirect interactions.

Coordinating gene expression on a genome-wide scale

The coordinate regulation of genes involved in the same biochemical processes, or that are components of the same macromolecular processes, is now well established. The genome-wide regulation that results in this coordinate expression in many cases has not been well characterized, although it clearly includes both transcriptional and post-transcriptional regulatory mechanisms. There are many examples of genes that are coregulated when steady-state mRNA levels are examined that are too numerous to list here, although the result was clearly observed in early microarray data sets in which hierarchical clustering was applied (Chu et al. 1998; Eisen et al. 1998; Spellman et al. 1998; Iyer et al. 1999; Perou et al. 2000; Whitfield et al. 2002).

The importance of the contribution of post-transcriptional processes, mediated largely by RBPs, has been increasingly studied. For example, the contribution of RNA stability to the coordinate genome-wide regulation of gene expression in yeast has been characterized (Wang et al. 2002; Grigull et al. 2004). One study observed that mRNAs involved in the same biological processes showed similar rates of mRNA degradation (Wang et al. 2002) demonstrating that the coordinate regulation of gene expression extends to the post-transcriptional level.

It has generally been accepted that the post-transcriptional regulation of many RNAs is controlled by changes to the mRNP, either in the complement of RBPs bound or by modification of bound RBPs. This has driven the systematic analysis of the RBPs, which has revealed similar themes. A study of the PUF proteins in yeast has determined each of the five different PUF proteins associated with a subset of messages that had a common theme (Gerber et al. 2004). This same result has been found in the study of other yeast proteins, including the RNA transport proteins She2p, She3p, and Myo4p (Shepard et al. 2003), La-related protein Lhp1p (Inada and Guthrie 2004), the heterogenous nuclear RNP (hnRNP) shuttling proteins Nab2, Npl3, and Nab4/Hrp1 (Guisbert et al. 2005), and the RNA export machinery (Hieronymus and Silver 2003; Hieronymus et al. 2004).

Many studies have also been performed analyzing the complement of mRNA bound to human RBPs. One of the first studies used Atlas arrays containing 597 different cDNAs to analyze the mRNAs bound to the RBP HuB, eIF-4E, and PABP in P19 embryonal carcinoma stem cells and found the complement of mRNAs bound changed

dramatically after induction of neuronal differentiation (Tenenbaum et al. 2000). Targets of the HuR RBP were also identified in colon cancer and HeLa cells by analysis on cDNA arrays containing 9600 genes; a 17–20-nt cis-acting regulatory element was identified in a majority of the enriched transcripts (Lopez de Silanes et al. 2004). In an additional study, it was shown that expression of a tagged RBP in a specific cell type could be used to identify the genes expressed in that cell (Penalva et al. 2004). A study of the α-globin poly-C binding proteins (αCP) used Affymetrix U95A microarrays that represent ∼12,000 different genes and identified 160 mRNAs targets (Waggoner and Liebhaber 2003). These studies have led to a hypothesis of "post-transcriptional operons" that lead to the coordinate regulation of genes (Keene and Tenenbaum 2002; Keene and Lager 2005; Moore 2005).

The combined result of these studies is that coordinating gene expression on a genome-wide scale clearly involves not only a transcriptional component but also a post-transcriptional component. The mechanisms that coordinate histone gene expression have been characterized in detail; it would be informative to perform a similar dissection of the regulatory pathways for a different set of coordinately regulated genes.

Potential pitfalls

The results presented here point to a potential pitfall in RNA-IP experiments. One clear problem is a high potential false-positive rate—the incorrect identification of mRNAs as bound to the RBP. In our case, these appear to be systematic errors, which may have occurred in array manufacture and may be more of a problem with "homemade" microarrays than with more stringently quality-controlled commercial microarray products. For example, each off-target gene identified was consistently enriched in each of the five independent IP experiments. Examination of the spot on the microarray for each of these indicates that it does indeed appear to be enriched on each microarray. The observed hybridization of false-positive genes may be a result of a low level of cross hybridization with the histone mRNA coding region sequences, coupled with the lack of a homologous mRNA sequence to hybridize with. Hybridization of a synthetic histone mRNA pool suggests these systematic positives do indeed result from cross hybridization with histone mRNAs precipitated by the SLBP. Measurement of the levels of each off-target mRNA by RT-PCR suggests that in most cases these genes are not expressed in HeLa cells. Nevertheless, the mRNAs that are known targets of SLBP were consistently identified. Note that use of an oligo microarray rather than a cDNA microarray eliminated these targets, consistent with them being the result of a nonspecific interaction. Alternatively, these spots could be contaminated with a PCR product containing a histone gene sequence during array manufacture.

Conclusions and implications

Here we have interrogated all genes in the human genome to ask if the mRNAs are bound by the SLBP in a HeLa cell line. We find no evidence to indicate that the SLBP is bound to messages other than the replication-dependent histone mRNAs. This data set provides a controlled study of the targets of a single RNA-binding protein for which the targets are well defined, and raises issues that need to be appropriately addressed for studies of unknown genes. This should now provide a resource for the genomics and bioinformatics communities with which to test computational algorithms to select bona fide targets in RIP-Chip experiments.

The rRIP-Chip method documented here will provide a rapid and useful method for the identification of targets of previously uncharacterized RBPs. The methods could be used to identify the targets of the RBPs and to systematically and experimentally identify all RNA-regulatory elements in the transcribed fraction of the human genome. Furthermore, these techniques can be applied to identify the steps in mRNA metabolism in which particular RBPs associate with a message and could ultimately be extended with proteomics to characterize the changes in the mRNP that lead to differential regulation of mRNA processing and stability. All figures and primary data tables are available at http://whitfieldlab.dartmouth.edu/mRNP/.

MATERIALS AND METHODS

Isolation of polyribosomes

HeLa S3 cells were grown in complete Dulbecco's Modified Eagles Media (1× DMEM, 10% FBS, 1% penicillin/streptomycin) to a density of 5×10^5 cells/mL in suspension culture using 1-L spinner flasks (Bellco). Cells were harvested by centrifugation at 1000 r.p.m., washed three times in phosphate-buffered saline (PBS), and resuspended in hypotonic buffer A (10 mM KCl, 1.5 mM MgCl$_2$, 10 mM HEPES at pH 7.9, 0.5 mM DTT, 0.75 mM spermidine, 0.15 mM spermine). Cells were allowed to swell in the hypotonic buffer and lysed using a Dounce homogenizer. Nuclei were removed by centrifugation at 930g in the presence of recovery buffer (50 mM HEPES at pH 7.9, 0.75 mM spermidine, 0.15 mM spermine, 10 mM KCl, 0.2 mM EDTA, 0.1 mM DTT, 67.5% sucrose) to prevent nuclear lysis. Cytoplasmic lysate was separated from cellular components by centrifugation at 10,000g for 10 min. Polyribosomes were purified by layering the cytosol over a 1 M sucrose cushion and centrifuged at 130,000g in a Beckman Coulter Ti60 rotor for 2.5 h. Isolated polyribosomes were resuspended in hypotonic buffer A and stored at −80°C.

RIP-Chip analysis of the endogenous SLBP on polyribosomes

Immunoprecipitations were performed essentially as previously described (Whitfield et al. 2004). The SLBP antibody and affinity purified antibody have been previously described in detail (Wang

et al. 1996; Whitfield et al. 2000). Polyribosomes (30–40 µg of polyribosomal RNA) were diluted into 100 µL of NP-40 lysis buffer (0.1% Tergitol type NP-40, 50 mM Tris-HCl at pH 7.5, 150 mM NaCl, 50 mM NaF, 1 mM DTT, 10 mM EDTA, 40 U RNasin [Promega]). Nonspecific interactions between the poly-ribosomes and protein-A agarose beads were precleared by the addition of 20 µL of the 1:1 protein A bead:NP40LB slurry and incubation at 4°C for 30 min with rotation. Protein A beads were removed by centrifugation at 1500g, and the precleared superna-tant incubated for 1 h at 4°C with either 1.5 µg affinity-purified anti-SLBP, anti-SLBP preincubated with the 13 AA C-terminal peptide, or without any antibody. 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 three times with NP-40 lysis buffer. RNA was prepared from both beads and supernatant as previously described (Whitfield et al. 2004). Urea lysis buffer (7 M urea, 2% SDS, 0.35 M NaCl, 10 mM EDTA, 10 mM Tris-HCl at pH 7.5) was added to an equal volume of the supernatant (100 µL), and 200 µL added to the pellet. RNA from both was extracted in phenol:[chloroform:isoamyl alcohol (24:1)] and precipitated with ethanol.

Isolation of recombinant SLBP–RNA complexes by rRIP-Chip

Approximately 20 ng of HIS-SLBP were incubated with RNA isolated from polyribosomes for 1 h at 4°C. Nonspecific com-plexes were eliminated by preclearing with protein A agarose beads for 1 h at 4°C. The remainder of the procedure was identical to that described above for the endogenous SLBP IP.

cDNA microarray hybridization

RNA isolated from the supernatant and the precipitate of each IP reaction was labeled for microarray analysis. Ten micrograms of supernatant RNA were used for labeling, with an equivalent corresponding fraction from the pellet (∼1/10th the concentra-tion of the total RNA supernatant). The RNA was random primed (pdN6) and incubated in the labeling reaction (1× reverse transcription buffer, 1 mM dATP, 1 mM dGTP, 1 mM dTTP, 0.6 mM dCTP, 20 mM DTT, 0.1 mM Cy3 or Cy5 dCTP, 80 U SuperScript II reverse transcriptase) at 42°C for 2 h. Precipitated RNA was labeled with Cy5-dCTP and the unprecipitated super-natant material labeled with Cy3-dCTP. Input RNA was hydro-lyzed with 0.1 N NaOH at 70°C for 15 min and neutralized with the addition of 0.1 N HCl. Cy-dye labeled cDNA was purified with a Qiagen PCR Purification Kit. The labeled cDNAs were pooled, 20 µg Cot-1 DNA added, and concentrated to a volume <31 µL in TE buffer (8 mM Tris at pH 6.5, 0.8 mM EDTA) using Microcon YM-30 filter columns.

The cDNA mixture was competitively hybridized in 3× SSC, 0.3% SDS along with 10 µg tRNA and 10 µg poly(A) RNA. DNA was denatured by incubation at 100°C for 2 min. Arrays were hybridized for 16 h at 65°C in custom hybridization chambers (DieTech). Microarrays were washed four times in increasingly stringent SSC and SDS (Wash 1: 2× SSC, 0.1% SDS. Wash 2: 2× SSC. Wash 3: 1× SSC. Wash 4: 0.1× SSC) and dried by centrifugation (60g).

Agilent oligonucleotide microarrays

The Agilent Direct Labeling Kit was used to label RNA for hybridization to the Agilent oligonucleotide microarrays. The same amount of input RNA was used here as was used for the Stanford cDNA micoarrays—10 µg of total RNA or the fractional equivalent. Hybridization and washing procedures were consistent with the manufacturer's protocol. Agilent data were centered on their mean values in both the gene and array dimensions. Two arrays (S1 and S2) used for the αSLBP RIP-Chip were sodium hydroxide-stripped according to previously described methods (Hu et al. 2005). We have found that stripped Agilent microarrays perform very similarly to arrays that have not been previously hybridized.

Data acquisition and analysis

Microarrays were scanned using a GenePix 4000B scanner (Axon Instruments) and spots of poor quality removed by visual inspection. Data from each microarray were loaded into the University of North Carolina Microarray Database (Agilent oligonucleotide microarrays) or the Stanford Microarray Database (Stanford cDNA microarrays) (Sherlock et al. 2001). Spots with a Cy3 intensity (Channel 1) over Cy3 background >1.2 (intensity/background ratio at least 20% above background) were excluded; genes with >20% missing data were not considered further. The $\log_2$ (Cy5/Cy3) normalized intensity ratio was retrieved for each gene. Except where indicated, genes were centered on their median expression value.

Each gene's intensity value was assigned a percentile rank based on the ratio of the gene's intensity compared to the intensities for the entire array (the gene with the greatest intensity value would have the highest rank). This rank was converted to a percent relative to the other genes on the array, and each gene assigned a median percentile rank across each set of IP experiments. In the case of the rRIP-Chip experiment that used recombinant SLBP, the median percentile rank was taken for the four specific IP experiments and the five mock IPs separately. The same was done for endogenous IP experiments on Stanford cDNA microarrays (5 specific IPs and 10 mock IPs), and on Agilent oligonucleotide microarrays (three specific IPs and three mock IPs). The mean percentile rank was used when an even number of experiments was analyzed. Using this percentile rank, histograms were gener-ated to segregate individual genes into bin classes (Figs. 2, 3). Analysis of the distribution of percentile ranks typically revealed a bimodal distribution in the specific immunoprecipitation and mock precipitation experiments (Buck and Lieb 2004). In each case, we established a cutoff for an enriched gene as the trough of the bimodal distribution. Genes with a percentile rank above this cutoff were considered possible SLBP targets.

Synthetic histone mRNA pools

Clones containing the coding regions of each of the five classes of histone genes (HIST1H1A, HIST1H2AB, HIST1H2BB, HIS-T1H3A, HIST1H4A) in the vector pcDNA3 (Invitrogen) were a gift of Dr. Alain Verreault. Clones were propagated in *Escherichia coli* and DNA prepared using Qiagen's midi-prep DNA extraction kit. Vectors were linearized by cleavage at the 3′ end of the histone gene with EcoR1 (H1–H3 clones) and BstX1 (H4 clone), and

purified by phenol:chloroform extraction and ethanol precipitation. Linearized DNA for each of the five clonses was transcribed separately in vitro using the MAXIscript T7 kit (Ambion), and free nucleotides were removed using G50 spin columns (Amersham Biosciences). Equal amounts of each of the five different in vitro transcribed histone mRNAs were mixed together and labeled as the "synthetic histone mRNA pool." Dilutions were made to obtain synthetic histone mRNA pools of 10 pg, 100 pg, 1 ng, and 10 ng per microliter. Each synthetic histone mRNA pool was direct labeled with Cy5-dCTP, hybridized to Stanford cDNA microarrays, and scanned on a GenePix 4000B microarray scanner. Values of Cy5 intensity (channel 2) were extracted from the raw data files.

RT-PCR

Total cellular RNA was extracted from HeLa cells with phenol: [chloroform:isoamyl alcohol (24:1)] and recovered by precipitation with ethanol. Four micrograms of total RNA were converted to cDNA by reverse transcription using SuperScript II (Invitrogen Life Technologies) primed with random hexamer (pdN6). Histone and nonhistone gene levels were measured by semiquantitative RT-PCR using this cDNA. Primers were designed using PrimerExpress and Invitrogen software (see Supplementary Table S3 for primer pairs). Each reaction contained 70–100 ng of cDNA template, 10 μmol each of forward and reverse primer, 1 U Taq DNA polymerase (New England Biolabs), 10 mM dNTP mix (Roche), and 1× ThermoPol buffer (New England Biolabs) for 28 cycles (92°C, 30 sec; 52–58°C, 30 sec; 72°C, 30 sec). Annealing temperatures (T_a) for each PCR reaction were chosen based on the T_m calculated for each oligonucleotide. Typically, the T_a was 5°C below the lowest T_m for the primer pair. Products were resolved on a 2% agarose gel in 0.5× TAE buffer, stained with ethidium bromide, and DNA visualized with UV light.

SUPPLEMENTARY MATERIAL

Supplementary material can be found at http://whitfieldlab.dartmouth.edu/mRNP/, a Web site maintained by the authors.

REFERENCES

Bonner, W.M., Mannironi, C., Orr, A., Pilch, D.R., and Hatch, C.L. 1993. Histone H2A.X gene transcription is regulated differently than transcription of other replication-linked histone genes. *Mol. Cell. Biol.* **13:** 984–992.

Brown, V., Jin, P., Ceman, S., Darnell, J.C., O'Donnell, W.T., Tenenbaum, S.A., Jin, X., Feng, Y., Wilkinson, K.D., Keene, J.D., et al. 2001. Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell* **107:** 477–487.

Buck, M.J. and Lieb, J.D. 2004. ChIP-chip: Considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* **83:** 349–360.

Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P.O., and Herskowitz, I. 1998. The transcriptional program of sporulation in budding yeast. *Science* **282:** 699–705.

Dominski, Z., Zheng, L.-X., Sanchez, R., and Marzluff, W.F. 1999. The stem-loop binding protein facilitates 3′ end formation by stabilizing U7 snRNP binding to the histone pre-mRNA. *Mol. Cell. Biol.* **19:** 3561–3570.

Dominski, Z., Yang, X.C., Kaygun, H., Dadlez, M., and Marzluff, W.F. 2003. A 3′ exonuclease that specifically interacts with the 3′ end of histone mRNA. *Mol. Cell* **12:** 295–305.

Dominski, Z., Yang, X.C., and Marzluff, W.F. 2005. The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. *Cell* **123:** 37–48.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci.* **95:** 14863–14868.

Erkmann, J.A., Sanchez, R., Treichel, N., Marzluff, W.F., and 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.

Eystathioy, T., Chan, E.K., Tenenbaum, S.A., Keene, J.D., Griffith, K., and Fritzler, M.J. 2002. A phosphorylated cytoplasmic autoantigen, GW182, associates with a unique population of human mRNAs within novel cytoplasmic speckles. *Mol. Biol. Cell* **13:** 1338–1351.

Gerber, A.P., Herschlag, D., and Brown, P.O. 2004. Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS Biol.* **2:** E79.

Graves, R.A., Pandey, N.B., Chodchoy, N., and Marzluff, W.F. 1987. Translation is required for regulation of histone mRNA degradation. *Cell* **48:** 615–626.

Grigull, J., Mnaimneh, S., Pootoolal, J., Robinson, M.D., and Hughes, T.R. 2004. Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. *Mol. Cell. Biol.* **24:** 5534–5547.

Guisbert, K.K., Duncan, K., Li, H., and Guthrie, C. 2005. Functional specificity of shuttling hnRNPs revealed by genome-wide analysis of their RNA binding profiles. *RNA* **11:** 383–393.

Harris, M.E., Bohni, R., Schneiderman, M.H., Ramamurthy, L., Schümperli, D., and Marzluff, W.F. 1991. Regulation of histone mRNA in the unperturbed cell cycle: Evidence suggesting control at two posttranscriptional steps. *Mol. Cell. Biol.* **11:** 2416–2424.

Hieronymus, H. and Silver, P.A. 2003. Genome-wide analysis of RNA–protein interactions illustrates specificity of the mRNA export machinery. *Nat. Genet.* **33:** 155–161.

Hieronymus, H., Yu, M.C., and Silver, P.A. 2004. Genome-wide mRNA surveillance is coupled to mRNA export. *Genes & Dev.* **18:** 2652–2662.

Hu, Z., Troester, M., and Perou, C.M. 2005. High reproducibility using sodium hydroxide-stripped long oligonucleotide DNA microarrays. *Biotechniques* **38:** 121–124.

Inada, M. and Guthrie, C. 2004. Identification of Lhp1p-associated RNAs by microarray analysis in *Saccharomyces cerevisiae* reveals association with coding and noncoding RNAs. *Proc. Natl. Acad. Sci.* **101:** 434–439.

Intine, R.V., Tenenbaum, S.A., Sakulich, A.L., Keene, J.D., and Maraia, R.J. 2003. Differential phosphorylation and subcellular localization of La RNPs associated with precursor tRNAs and translation-related mRNAs. Mol. *Cell* **12:** 1301–1307.

Iyer, V.R., Eisen, M.B., Ross, D.T., Schuler, G., Moore, T., Lee, J.C., Trent, J.M., Staudt, L.M., Hudson Jr., J., Boguski, M.S., et al. 1999. The transcriptional program in the response of human fibroblasts to serum. *Science* **283:** 83–87.

Kaygun, H. and Marzluff, W.F. 2005a. Regulated degradation of replication-dependent histone mRNAs requires both ATR and Upf1. *Nat. Struct. Mol. Biol.* **12:** 794–800.

———. 2005b. Translation termination is involved in histone mRNA degradation when DNA replication is inhibited. *Mol. Cell. Biol.* **25:** 6879–6888.

Keene, J.D. and Lager, P.J. 2005. Post-transcriptional operons and regulons co-ordinating gene expression. *Chromosome Res.* **13:** 327–337.

Keene, J.D. and Tenenbaum, S.A. 2002. Eukaryotic mRNPs may represent posttranscriptional operons. *Mol. Cell* **9:** 1161–1167.

Kent, W.J. 2002. BLAT—The BLAST-like alignment tool. *Genome Res.* **12:** 656–664.

Kolev, N.G. and Steitz, J.A. 2005. Symplekin and multiple other polyadenylation factors participate in 3′-end maturation of histone mRNAs. *Genes & Dev.* **19:** 2583–2592.

Laoukili, J., Kooistra, M.R., Bras, A., Kauw, J., Kerkhoven, R.M., Morrison, A., Clevers, H., and Medema, R.H. 2005. FoxM1 is required for execution of the mitotic programme and chromosome stability. *Nat. Cell Biol.* **7:** 126–136.

Lieb, J.D., Liu, X., Botstein, D., and Brown, P.O. 2001. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein–DNA association. *Nat. Genet.* **28:** 327–334.

Lopez de Silanes, I., Fan, J., Galban, C.J., Spencer, R.G., Becker, K.G., and Gorospe, M. 2004. Global analysis of HuR-regulated gene expression in colon cancer systems of reducing complexity. *Gene Expr.* **12:** 49–59.

Martin, F., Schaller, A., Eglite, S., Schumperli, D., and Muller, B. 1997. The gene for histone RNA hairpin binding protein is located on human chromosome 4 and encodes a novel type of RNA binding protein. *EMBO J.* **16:** 769–778.

Moore, M.J. 2005. From birth to death: The complex lives of eukaryotic mRNAs. *Science* **309:** 1514–1518.

Novoradovskaya, N., Whitfield, M.L., Basehore, L.S., Novoradovsky, A., Pesich, R., Usary, J., Karaca, M., Wong, W.K., Aprelikova, O., Fero, M., et al. 2004. Universal Reference RNA as a standard for microarray experiments. *BMC Genomics* **5:** 20.

Pandey, N.B. and Marzluff, W.F. 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.

Penalva, L.O., Burdick, M.D., Lin, S.M., Sutterluety, H., and Keene, J.D. 2004. RNA-binding proteins to assess gene expression states of co-cultivated cells in response to tumor cells. *Mol. Cancer* **3:** 24.

Perou, C.M., Sorlie, T., Eisen, M.B., van de, R.M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., et al. 2000. Molecular portraits of human breast tumours. *Nature* **406:** 747–752.

Pillai, R.S., Grimmler, M., Meister, G., Will, C.L., Luhrmann, R., Fischer, U., and Schumperli, D. 2003. Unique Sm core structure of U7 snRNPs: Assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing. *Genes & Dev.* **17:** 2321–2333.

Sanchez, R. and Marzluff, W.F. 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.

Schmitz-Linneweber, C., Williams-Carrier, R., and Barkan, A. 2005. RNA immunoprecipitation and microarray analysis show a chloroplast Pentatricopeptide repeat protein to be associated with the 5′ region of mRNAs whose translation it activates. *Plant Cell* **17:** 2791–2804.

Shepard, K.A., Gerber, A.P., Jambhekar, A., Takizawa, P.A., Brown, P.O., Herschlag, D., DeRisi, J.L., and Vale, R.D. 2003. Widespread cytoplasmic mRNA transport in yeast: Identification of 22 bud-localized transcripts using DNA microarray analysis. *Proc. Natl. Acad. Sci.* **100:** 11429–11434.

Sherlock, G., Hernandez-Boussard, T., Kasarskis, A., Binkley, G., Matese, J.C., Dwight, S.S., Kaloper, M., Weng, S., Jin, H., Ball, C.A., et al. 2001. The Stanford Microarray Database. *Nucleic Acids Res.* **29:** 152–155.

Sittman, D.B., Graves, R.A., and Marzluff, W.F. 1983. Histone mRNA concentrations are regulated at the level of transcription and mRNA degradation. *Proc. Natl. Acad. Sci.* **80:** 1849–1853.

Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D., and Futcher, B. 1998. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9:** 3273–3297.

Stephens, S.B., Dodd, R.D., Brewer, J.W., Lager, P.J., Keene, J.D., and Nicchitta, C.V. 2005. Stable ribosome binding to the endoplasmic reticulum enables compartment-specific regulation of mRNA translation. *Mol. Biol. Cell* **16:** 5819–5831.

Tenenbaum, S.A., Carson, C.C., Lager, P.J., and Keene, J.D. 2000. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc. Natl. Acad. Sci.* **97:** 14085–14090.

Tenenbaum, S.A., Carson, C.C., Atasoy, U., and Keene, J.D. 2003. Genome-wide regulatory analysis using en masse nuclear run-ons and ribonomic profiling with autoimmune sera. *Gene* **317:** 79–87.

Waggoner, S.A. and Liebhaber, S.A. 2003. Identification of mRNAs associated with αCP2-containing RNP complexes. *Mol. Cell. Biol.* **23:** 7055–7067.

Wang, Y., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D., and Brown, P.O. 2002. Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci.* **99:** 5860–5865.

Wang, Z.-F., Whitfield, M.L., Ingledue, T.I., Dominski, Z., and Marzluff, W.F. 1996. The protein which binds the 3′ end of histone mRNA: A novel RNA-binding protein required for histone pre-mRNA processing. *Genes & Dev.* **10:** 3028–3040.

Whitfield, M.L. 1999. The cytoplasmic role and cell cycle regulation of the histone stem–loop binding protein. Biochemistry. Ph.D. thesis, University of North Carolina, Chapel Hill.

Whitfield, M.L., Zheng, L.X., Baldwin, A., Ohta, T., Hurt, M.M., and Marzluff, W.F. 2000. Stem-loop binding protein, the protein that binds the 3′ end of histone mRNA, is cell cycle regulated by both translational and posttranslational mechanisms. *Mol. Cell. Biol.* **20:** 4188–4198.

Whitfield, M.L., Sherlock, G., Saldanha, A.J., Murray, J.I., Ball, C.A., Alexander, K.E., Matese, J.C., Perou, C.M., Hurt, M.M., Brown, P.O., et al. 2002. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell* **13:** 1977–2000.

Whitfield, M.L., Kaygun, H., Erkmann, J.A., Townley-Tilson, W.H., Dominski, Z., and Marzluff, W.F. 2004. SLBP is associated with histone mRNA on polyribosomes as a component of the histone mRNP. *Nucleic Acids Res.* **32:** 4833–4842.

Zheng, L., Dominski, Z., Yang, X.C., Elms, P., Raska, C.S., Borchers, C.H., and Marzluff, W.F. 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.