RNA binding by the histone methyltransferases Set1 and Set2

Citation for published version:

Digital Object Identifier (DOI):
10.1128/MCB.00165-17

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Molecular and Cellular Biology

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
RNA binding by the histone methyltransferases Set1 and Set2

AUTHORS

Camille Sayou1, Gonzalo Millán-Zambrano2, Helena Santos-Rosa2, Elisabeth Petfalski1, Samuel Robson2, Jonathan Houseley3, Tony Kouzarides2, David Tollervey1*

1 Wellcome Trust Centre for Cell Biology, University of Edinburgh, Michael Swann Building, King’s Buildings, Edinburgh EH9 3BF, Scotland
2 Gurdon Institute and Department of Pathology, Tennis Court Road, Cambridge, CB2 1QN, United Kingdom
3 Epigenetics Programme, The Babraham Institute, Cambridge CB22 3AT, United Kingdom

* Correspondence to:
e-mail: d.tollervey@ed.ac.uk
Tel: +44 131 650 7092
Fax: + 44 131 650 7040

Keywords: Set1, Set2, RNA, chromatin, transcription, yeast, UV crosslinking, RNA-protein interaction
ABSTRACT

Histone methylation at H3K4 and H3K36 is commonly associated with genes actively transcribed by RNA polymerase II (RNAPII) and is catalyzed by yeast Set1 and Set2, respectively. Here we report that both methyltransferases can be UV-crosslinked to RNA in vivo. High-throughput sequencing of the bound RNAs revealed strong Set1 enrichment near the transcription start site, whereas Set2 was distributed along pre-mRNAs. A subset of transcripts showed notably high enrichment for Set1 or Set2 binding relative to RNAPII, suggesting functional post-transcriptional interactions. In particular, Set1 was strongly bound to the SET1 mRNA, Ty1 retrotransposons, and non-coding RNAs from the rDNA intergenic spacers, consistent with its previously reported silencing roles. Set1 lacking RRM2 showed reduced in vivo crosslinking to RNA and reduced chromatin occupancy. In addition, levels of H3K4 trimethylation were decreased whereas di-methylation was increased. We conclude that RNA binding by Set1 contributes to both chromatin association and methyltransferase activity.
A major function of chromatin in eukaryotic cells is the regulation of gene expression in the form of RNA transcripts. It therefore seemed likely that there would be an extensive interplay between the transcriptome and chromatin-associated factors. Consistent with this idea, chromatin proteins were identified by mass-spectrometry following UV-crosslinking and purification of RNA-protein complexes both in yeast and human cells (1–3). Moreover, recent analyses of a panel of chromatin-associated proteins identified 24 protein-RNA interactions that could be recovered through formaldehyde-based crosslinking in human cells (4).

Two prominent modifications in chromatin are the methylation of histone H3 at lysine 4 (H3K4) and lysine 36 (H3K36). In the budding yeast, *Saccharomyces cerevisiae*, these modifications are catalyzed by the Set1 and Set2 methyltransferases respectively. During transcription, the large catalytic subunit of RNA polymerase II (RNAPII), Rpo21 in yeast, undergoes dynamic phosphorylation/dephosphorylation events within the heptad repeats forming the carboxy-terminal domain (CTD). These help coordinate the recruitment of transcription and RNA processing factors to the elongating RNAPII and nascent transcript (reviewed in 5, 6). Set1 functions within the complex of proteins associated with Set1 (COMPASS or Set1C, reviewed in 7), which is brought to RNAPII through interaction with the PAF complex when the CTD is phosphorylated on serine 5 (RNAPII-S5P). Recruitment of Set2 to the elongating RNAPII occurs when serines 2 and 5 are phosphorylated and also requires the PAF complex (reviewed in 8).

H3K4me3 is a characteristic feature of the 5′ regions of actively transcribed genes, and this correlation has often led to an expectation that Set1 functions to stimulate transcription. However, from the earliest analyses in yeast, Set1 was implicated in gene silencing (9, 10). Subsequent analyses implicated Set1 in the repression of many genes (11, 12), with more obvious effects under stress conditions (13). In addition, Set1 has been implicated in transcriptional silencing of retrotransposons in *S. cerevisiae* (14, 15) and in *S. pombe* (16–18).
Set1-dependent silencing of Ty1 retrotransposons is mediated by a non-coding, antisense transcript (15). Set1 is also implicated in silencing of RNAPII transcription from the intergenic spacer (IGS) regions located between the ribosomal DNA (rDNA) genes (19, 9). H3K36 methylation is found throughout protein coding genes and prevents initiation of transcription at cryptic sites via recruitment of the Rpd3S histone deacetylase complex (reviewed in 20). H3K36 methylation is also reported to regulate pre-mRNA splicing (21).

Yeast Set1 has two putative RNA recognition motifs (RRMs) that are implicated in Set1 function, suggesting that it might bind RNA in vivo (22, 23). Set2 does not harbor an evident RNA-binding motif, but was identified in systematic analyses of yeast RNA-interacting proteins (24, 1). However, in vivo targets for these potential RNA-binding activities have not been reported.

To identify potential direct RNA-interactions for Set1 and Set2 we employed UV-crosslinking and analysis of cDNAs (CRAC). This showed that both Set1 and Set2 associate with almost all RNAPII transcripts. However, binding of Set1 and Set2 relative to transcription rates is variable. Transcripts showing high relative binding by Set1 and Set2 are candidates for post-transcriptional regulation. Our results showed that Set1 interactions with RNA are predominately mediated by RRM2, and indicate that contacts with RNA reinforce both chromatin binding and methyltransferase activity.

RESULTS

Set1 and Set2 bind to RNA in vivo

To perform CRAC, the endogenous SET1 gene was tagged with either an N-terminal ProteinA-TEV-His6 (PTH) tag or a C-terminal His6-TEV-ProteinA (HTP) tag. The endogenous SET2 gene was tagged with C-terminal HTP. All constructs were expressed under the control of the endogenous promoter and were the sole form of Set1 or Set2 in the cell (Figure 1A). In strains
expressing only PTH-Set1 or Set2-HTP, global H3K4me3 and H3K36 methylation levels and cell growth were similar to the wild-type (Figure 1B; Figure S1A-C). In contrast, Set1-HTP strains lacked detectable H3K4me3, consistent with previous reports for C-terminal tagged Set1 proteins (25, 22, 16) (Figure S1A), and was slower growing than the wild-type strain (Figure S1C). However, the protein level was unaffected (Figure S1A), in contrast to a previous report that loss of methyltransferase activity results in protein depletion (26). This discrepancy likely reflects structural differences in the alleles used.

To test for in vivo RNA binding, actively growing cells were UV irradiated, the tagged proteins were purified and crosslinked RNAs were labeled and visualized by SDS-PAGE and autoradiography. This showed that PTH-Set1, Set1-HTP and Set2-HTP were all bound to RNAs in vivo (Figure 1C-D; Figure S1D).

Set1 RRM2 was predicted to be a functional RNA binding domain, whereas RRM1 appeared less likely to interact with RNA, and this was supported by in vitro assays (22). Moreover, a deletion overlapping RRM1 reduced Set1 levels, whereas a construct lacking only RRM2 was stable (26). To assess RNA-binding by Set1, we therefore deleted RRM2 (residues 415-494) from the PTH-Set1 strain to obtain PTH-Set1ΔRRM2 (Figure 1A). The abundance of PTH-Set1ΔRRM2 was similar to PTH-Set1, and the deletion did not clearly alter global H3K4me3 levels (Figure 1B) or growth (Figure S1C). In CRAC analyses, PTH-Set1ΔRRM2 greatly reduced, but did not abolish, RNA crosslinking relative to PTH-Set1 (Figure 1C). We therefore conclude that most RNA binding activity in Set1 is attributable to RRM2. Residual binding observed in PTH-Set1ΔRRM2 may result from RRM1.

Set1 and Set2 associate with nascent RNAPII transcripts

RNA fragments purified with Set1 and Set2 (from strains PTH-Set1, PTH-Set1ΔRRM2, Set1-HTP, Set2-HTP) were converted to cDNA and sequenced. RNA was also recovered following mock purification from the untagged strain (BY4741) and represents the background of the
At least 3 replicate datasets were obtained for each of PTH-Set1, PTH-Set1ΔRRM2, Set1-HTP and Set2-HTP (Table S1).

To better estimate the relative in vivo binding to RNA of PTH-Set1 and PTH-Set1ΔRRM2, crosslinked and barcoded samples were mixed prior to SDS-PAGE separation and RT-PCR amplification. Following de-multiplexing, the number of reads recovered for PTH-Set1ΔRRM2 was about 3 fold lower than for PTH-Set1 (Figure S1E), consistent with the reduced binding observed from the autoradiography gels (Figure 1C). Substantially fewer reads were recovered for the BY4741 untagged control.

The distribution of reads among RNA classes showed that both Set1 and Set2 predominately bound mRNAs (Figure 2A). Compared to BY4741, Set1, but not Set2, was also enriched for binding to other non-protein coding RNAs (ncRNAs) transcribed by RNAPII, including SUTs, CUTs, XUTs and intergenic transcripts. For comparison, we also show the distribution of the catalytic subunit of RNAPII (Rpo21-HTP) using previous CRAC data (27).

PTH-Set1, PTH-Set1ΔRRM2 and Set1-HTP showed broadly similar distributions (Figure 2A). PTH-Set1ΔRRM2 samples showed more rRNA and tRNA reads than PTH-Set1 (Figure 2A), although this is likely to reflect a higher background due to reduced RNA binding rather than a difference in endogenous RNA target classes. This indicated that loss of RRM2 greatly reduces affinity for RNA (Figure 1C) but has limited impact on specificity (Figure 2A).

Pre-mRNA splicing is largely co-transcriptional in yeast (28–30). Therefore, the presence of unspliced RNAs in CRAC datasets generally reflects protein interactions with nascent transcripts. To assess whether Set1 and Set2 bind co-transcriptionally, the recovery of spliced and unspliced transcripts from intron-containing genes was calculated as reported (27, 31). The ratio of reads spanning exon-exon junctions (spliced) relative to exon-intron plus intron-exon junctions (unspliced) was below 1 for both Set1 and Set2 (Figure 2B) indicating predominant binding to unspliced, nascent pre-mRNAs. For Set2 the ratio was higher than for Set1, consistent with Set2 binding later during transcription.

Pre-mRNA splicing is largely co-transcriptional in yeast (28–30). Therefore, the presence of unspliced RNAs in CRAC datasets generally reflects protein interactions with nascent transcripts. To assess whether Set1 and Set2 bind co-transcriptionally, the recovery of spliced and unspliced transcripts from intron-containing genes was calculated as reported (27, 31). The ratio of reads spanning exon-exon junctions (spliced) relative to exon-intron plus intron-exon junctions (unspliced) was below 1 for both Set1 and Set2 (Figure 2B) indicating predominant binding to unspliced, nascent pre-mRNAs. For Set2 the ratio was higher than for Set1, consistent with Set2 binding later during transcription.
HTP showed similar ratios. Since both Set1 and Set2 bound a higher proportion of spliced transcripts than RNAPII, we addressed their possible post-transcriptional association with mRNAs by comparing their binding to mRNA stability, determined by RNA-seq following RNAPII inhibition (32). Enrichment of Set1 and Set2 relative to RNAPII did not increase with mRNA half-life (Figure S2A-D), strongly indicating that Set1 and Set2 are not predominantly bound to mature mRNAs. We conclude that Set1 and Set2 are directly associated with nascent RNAPII transcripts, consistent with their function during transcription.

**Set1 binding is enriched near the TSS while Set2 binds across transcripts**

Binding profiles on mRNAs for PTH-Set1 and Set2-HTP were aligned via the transcription start site (TSS) or the poly(A) site (pA) (Figure 2C-E). This showed that PTH-Set1 binding was strongly enriched over the 5’ end of mRNAs, from the TSS to +500 nt. In contrast, Set2-HTP binding was more distributed along transcripts, from +150 nt after the TSS to -150 nt before pA sites. This pattern was also clearly visible on individual mRNAs (Figure S3A-D). We also compared PTH-Set1 binding with the residual binding of PTH-Set1ΔRRM2 and with Set1-HTP. All three proteins showed similar profiles, suggesting that, once Set1 was bound to RNA, RRM2 did not significantly influence its distribution along mRNAs (Figure 2E); and that the lack of methylation activity also did not influence Set1 distribution across mRNAs (Figure 2E).

The RNAPII distribution across transcripts shows higher density over the TSS proximal region (Figure 2E), likely reflecting a substantial level of premature transcription termination in the 5’-proximal region (27, 31, 33–36). To account for this uneven transcript distribution, binding of Set1 and Set2 was expressed relative to RNAPII coverage. Relative coverage was calculated as the log2 (protein coverage / Rpo21-HTP coverage) and plotted along mRNAs (27). Set1 was strongly enriched relative to RNAPII at the 5’ ends of mRNAs (Figures 2F; Figure S3E). Set2 was relatively depleted from the promoter proximal region and progressively rose to peak around +400 nt from the TSS, then remaining elevated along transcripts (Figure 2F; Figure
S3F). These profiles are consistent with previously reported distributions of H3K4me3 and H3K36me3 on chromatin (37).

Set1 was reported to be recruited to chromatin when the CTD is phosphorylated on serine 5, whereas Set2 is recruited when both serine 5 and serine 2 are phosphorylated (8). We therefore compared the relative distributions of Set1 and Set2 to RNAPII with the five types of CTD phosphorylation state (Y1P, S2P, T4P, S5P, S7P), which were recently mapped to RNA using a CRAC-related technique (27). Set1 and RNAPII-S5P both peaked close to the TSS, but their distributions differed significantly. RNAPII-S5P was strongly enriched across the first 130 nt from the TSS and was then sharply depleted. In contrast, the enrichment profile of Set1 extended further 3’ (Figure 2F). The observation of high levels of Set1 binding over regions with low S5P strongly indicates that this is not the only determinant of Set1 distribution over RNAs. The Set2 profile closely resembled both RNAPII-S2P and RNAPII-T4P (Figure 2F), consistent with Set2 recruitment to the CTD modified with S2P and perhaps T4P.

Set1 and, to a lesser extend, Set2 were bound to ncRNAs, including SUTs, CUTs, XUTs, intergenic and antisense transcripts, in addition to mRNAs (Figure 2A). To compare Set1 and Set2 enrichment profiles over mRNAs and ncRNAs, we used an expression-matched subset of mRNAs, SUTs and CUTs, based on their total RNAPII CRAC signal over the first 300 nt (27). Set1 and Set2 were less enriched on SUTs compared to mRNAs, and even less on CUTs (Figure S3G). On the same sets of transcripts, RNAPII-S5P profiles were similar whereas RNAPII-S2P showed decreased enrichment, as previously reported (27). We speculate that the ncRNAs, particularly CUTs, undergo very rapid degradation that occurs immediately following transcription, and may even be partially co-transcriptional, greatly restricting the time available for Set1 or Set2 association.

The Set1 and Set2 RNA binding profiles clearly support predominately co-transcriptional recruitment. However, the correlation between Set1 and RNAPII-S5P indicates that this is not
the sole key recruitment factor, while the ncRNA analyses suggest that RNA association is at least transiently retained with the released transcripts.

**High binding of Set1 and Set2 to specific transcripts suggest functional interactions**

We hypothesized that transcripts with functionally relevant Set1 or Set2 binding would show higher enrichment relative to RNAPII (i.e. transcription rate). Coverage of Set1 or Set2 over genomic features, including mRNAs and non-coding transcripts, was plotted against RNAPII coverage, as determined by crosslinking of Rpo21-HTP (Figure 3A-B). Overall, Set1 and Set2 binding was broadly correlated with RNAPII coverage. There was, however, some heterogeneity, with a subset of transcripts showing high binding despite low levels of transcription.

Set1 showed high relative binding to *SET1* mRNA (Figure 3A and 3C). PTH-Set1 binding along the mRNA was broadly distributed and, in contrast to most mRNAs, did not show a clear 5' peak (Figure 3C), indicating that the interaction is at least not only co-transcriptional. The N-terminal tag is present on the nascent peptide throughout translation, whereas the C-terminal tag is synthesized just before dissociation from the ribosome, and binding to *SET1* mRNA was strongly reduced for Set1-HTP compared to PTH-Set1 (Figure S4A). Those observations are consistent with the previously reported co-translational binding of *SET1* mRNA by Set1 and three other COMPASS components (38). Differences in recovery of *SET1* mRNA between different Set1 strains did not result from altered mRNA abundance (Figure S4C).

Set1 was also enriched on a group of partially overlapping, ncRNA transcripts derived from the rDNA intergenic spacer regions (IGS ncRNAs), and over Ty1 retrotransposons, with strong binding to both mRNAs and antisense transcripts (Figure 3A, 3D and S4B). RT-qPCR analyses showed that those transcripts are unaltered in PTH-Set1, PTH-Set1\textDeltaRRM2 and Set2-HTP strains, relative to Rpo21-HTP or wild-type strain BY4741 (Figure S4D-E). In contrast, Set1-HTP showed increased transcript levels, notably for the rDNA IGS, suggesting the Set1
histone methyltransferase activity, but not RNA binding may be involved in regulating the abundance of these ncRNAs. PTH-Set1ΔRRM2 and PTH-Set1 showed similar enrichment to RNAPII over the different sequence features (Figure S5A-B). We conclude that while RRM2 strongly contributes to the level of RNA association, it is not primarily responsible for the specificity of RNA binding by Set1.

Comparison of Set2 to RNAPII identified only a few mRNAs with high relative Set2 binding (Figure 3B). Set2 was, however, enriched over the rDNA IGS ncRNAs (Figure 3B and 3D) and, most clearly, over a subset of the box C/D class of small nucleolar RNAs (snoRNAs) (Figure S5C). The PAF complex and, less clearly, Set2 were previously implicated in snoRNAs 3' end formation (39, 40), suggesting a possible link between this process and Set2 RNA binding.

To check whether the RNA binding activity could regulate transcript abundance, genes showing differential expression were identified in strains carrying set1Δ (14) or set2Δ (see Materials and Methods). However, for both proteins the differentially expressed genes corresponded to transcripts with low coverage for PTH-Set1, Set2-HTP and RNAPII, indicating their low expression. No clear enrichment was seen for mRNAs showing low or high binding of Set1 and Set2 relative to RNAPII (Figure S5D).

In conclusion, despite co-transcriptional binding to all RNAPII transcription units, Set1 and Set2 were strongly enriched on small numbers of transcripts. For Set1 these largely represent known silencing targets.

RNA binding stabilizes interactions of Set1 with chromatin and regulates the balance between H3K4 di- and tri-methylation

The potential contribution of RNA binding to stabilizing the association of Set1 with chromatin in vivo was assessed by chromatin-immunoprecipitation (ChIP) followed by qPCR. Set1 distribution along PMA1 matched previous reports (26, 41), with stronger crosslinking nearing
the 5' end. PTH-Set1ΔRRM2 binding to chromatin was ~30% reduced compared to PTH-Set1 at the 5' end (primer pairs 1, 2; Figure 4A-B), where Set1 RNA-binding peaked (Figure 2C, 2E; Figure S3A-D). Binding to the 3' end of PMA1 (primer pairs 3, 4) was similar for both proteins (Figure 4A-B). Reductions of ~25 to 30% in binding of PTH-Set1ΔRRM2 was seen for three other genes tested (TEF1, TDH3, ILV5: Figure 4A-B). The data indicate that reduced RNA binding by Set1 leads to weaker interactions with chromatin, but this may be specific for 5' regions that show high Set1-RNA interactions.

We then tested whether the reduced chromatin occupancy caused by the RRM2 deletion affected H3K4 methylation. ChIP-qPCR was performed to assess the levels of H3K4me1, H3K4me2 and H3K4me3 in the strain expressing Set1ΔRRM2 compared to the wild-type strain expressing native Set1. The level of H3K4me3 was reduced by 20 to 30% in the Set1ΔRRM2 strain in the 5' regions of PMA1, TEF1, TDH3 and ILV5, whereas H3K4me3 was unchanged near the 3' end of PMA1 (Figure 4C). In contrast, we observed similarly increased levels of H3K4me2 at all loci tested. In the case of PMA1 this increase was more pronounced at the 5' end (Figure 4C). We observed no change in H3K4me1 (Figure 4C). Significant global change in the three methylation states could not be detected by western-blot, likely due to the lack of sensitivity of the method, compared to ChIP (Figure S6A-B). This shows that RRM2 is required for the normal balance between H3K4me3 and K3K4me2, particularly at the 5' end of genes.

These results demonstrate that RRM2 is functionally important for Set1 targeting at chromatin and for methylation of H3K4. We propose that RNA binding participates to Set1 recruitment and/or stabilization at chromatin, therefore contributing to H3K4 methylation patterns.

**DISCUSSION**

This study presents high resolution, strand-specific, transcriptome-wide mapping of two major histone methyltransferases Set1 and Set2, which are conserved from yeast to human. Both
proteins directly interacted with RNA \textit{in vivo} (Figure 1C-D) and showed preferential interactions with nascent RNAPII transcripts (Figure 2A-B; Figure S2), consistent with their association with transcribing RNAPII. Set1 was enriched at the 5’ end of mRNAs whereas Set2 was distributed along transcripts (Figure 2C-F; Figure S3A-F), matching the distributions of H3K4me3 and H3K36me3 on chromatin, respectively (37).

Binding of Set1 and Set2 was detected for all active RNAPII transcription units. However, some RNAs showed high protein binding relative to their transcription rate, particularly for Set1 (Figure 3A-B; Figure S5C), suggesting post-transcriptional interactions. SET1 mRNA was one of the most enriched transcripts for Set1 binding. The broad distribution of Set1 and the lack of 5’ bias along SET1 mRNA indicates post-transcriptional binding (Figure 3C). This interaction was previously proposed to be co-translational (38) and the reduction in Set1 mRNA binding observed for Set1-HTP for but not PTH-Set1ΔRRM2 would be consistent with this conclusion (Figure S4A).

Ty1 mRNAs and Ty1 antisense transcripts were found to be strongly enriched for Set1 binding. IGS ncRNAs from the rDNA repeats were enriched for both Set1 and Set2 (Figure 3A-B, 3D and S4B). Strikingly, Set1 was previously shown to participate in silencing of retrotransposons (14, 15) and IGS regions (19, 9) supporting the model that functionally important Set1 targets would show preferential binding relative to RNAPII.

Phosphorylation and dephosphorylation of the CTD of the large subunit of RNAPII coordinates the recruitment of numerous factors, including Set1 and Set2 (5, 6). The distribution of Set2 along genes was similar to RNAPII-S2P, consistent with its reported role in recruitment. However, Set2 was also closely matched with RNAPII-T4P (Figure 2F), suggesting the possible involvement of this CTD modification in Set2 recruitment. Surprisingly, the distribution of Set1 was distinct from that of RNAPII-S5P (Figure 2F), strongly indicating that additional parameters help define Set1 localization along transcripts.
The Set1ΔRRM2 protein showed reduced chromatin association in ChIP analyses (Figure 4A-B), indicating that RNA binding functions in the recruitment of Set1 to chromatin and/or stabilizes the association. Consistent with this conclusion, it was previously shown that a truncated version of Set1 containing only the SET domain and most of the N-SET had reduced chromatin occupancy in yeast (26). Notably, analyses at different sites along the PMA1 gene revealed clear differences in chromatin occupancy only in the 5' region. This suggests a potential correlation between stabilization of chromatin association and high RNA binding. Consistent with this model, the absence of RRM2 also led to reduced H3K4me3 and increased H3K4me2, at the 5' end of genes (Figure 4C), demonstrating that RRM2 is required for the correct distribution of H3K4me3 and H3K4me2.

In vivo and in vitro experiments previously showed that the pattern of mono-, di- and tri-methylation deposited by Set1 correlated with interaction time of the COMPASS complex with its nucleosome substrate, monomethylation occurring virtually immediately, followed by dimethylation, and finally trimethylation (42). Other parameters, such as the COMPASS complex subunit composition also directs the distribution of the three methylation states (7, 43). We propose that RNA binding of Set1 via RRM2 near the TSS stabilizes the association of Set1/COMPASS with chromatin, promoting formation of H3K4me3 at the 5’ ends of genes. Due to reduced RNA binding, Set1ΔRRM2-chromatin interaction is weaker or more transient, leading to higher levels of H3K4me2. A major role of H3K4me2 is recruitment of the Set3 histone deacetylase complex, which deacetylate histones in 5’ regions of transcription units and participates in H3K4me2 maintenance (44). This helps regulate overlapping non-coding transcription and contributes to epigenetic transcriptional memory (45, 43). We speculate that the disruption of RNA binding by Set1 adversely affects these processes.

The results reported here contribute to understanding of the crosstalk between RNA synthesis and the modulation of chromatin structure. Recent studies have identified large number of RNA-interacting proteins in eukaryotic cells. Given the key role of chromatin in the
regulation of RNA synthesis, it might be anticipated that functional RNA interactions will be particularly prevalent among the readers, writers and erasers of epigenetic chromatin marks. However, previous analyses have reported comparatively fewer examples of such interactions. In this context, the identification of RNA binding activity by the two major histone methyltransferases in yeast is perhaps not entirely unexpected. Many analyses have revealed substantial functional redundancy among epigenetic regulatory systems in yeast. We anticipate that the importance of RNA interactions by Set1 and Set2 will be more evident in cells that are also deficient in other epigenetic pathways or are undergoing rapid changes in gene expression program, which will be frequent for yeast growing in the natural environment.

METHODOLOGY

Strains

Yeast analyses were performed in strains derived from BY4741, except for the RNA-seq experiment that was done in the W303 background. All strains used are listed in Table S2. Oligonucleotides are listed in Table S3. The PTH-Set1 strain was obtained by integrating a sequence encoding a PTH (2xproteinA-TEV-His) tag at the 5’ end of SET1, resulting in the expression of an N-terminally tagged protein expressed from the endogenous SET1 promoter. Generation of this strain involved inserting a URA3-KAN marker between the SET1 promoter and the SET1 ORF, and then replacing this marker with a sequence encoding the PTH tag. The URA3-KAN marker was amplified from pGSKU (46) using the oligonucleotides oCA164-oCA165. The PTH tag was amplified on a plasmid expressing N-PTH-NPL3 (pRS415-PTH) using the oligonucleotides oCA167-oCA168. The PTH-Set1ΔRRM2 strain was obtained from PTH-Set1. First a URA3-KAN marker was amplified from pGSKU using the oligonucleotides oCA151-oCA152 and integrated in the RRM2 in SET1 ORF. The URA3-KAN marker was removed using oligonucleotides, as described (46). The oCA175-oCA176 oligonucleotides are
homologous to sequences upstream and downstream of RRM2, their insertion resulted in a
deletion from position 243 to 482 on SET1 ORF and residues 415 to 494 on Set1 protein. The
HTP tag with a URA3 marker was amplified from pBS1539-HTP (47) and integrated to obtain
the Set1-HTP and SET2-HTP strains. The Set1ΔRRM2 strain was obtained as described above
for the PTH-Set1ΔRRM2 strain, but starting from BY4741 instead of PTH-Set1. The SET1 ORF
was deleted using a URA3 marker (Δset1:URA:pURA). The URA3 coding sequence and
promoter were inserted antisense relative to the SET1 gene. In the W303 background, the
SET2 ORF was deleted using a KanMX cassette (Δset2:KanMX).

Immunoblotting
For this study, we used the following antibodies: anti-H3 (Abcam Ab1791), anti-H3K4me3
(Upstate 05-745), anti-H3K4me2 (C64G9, cell signaling technology 9725T), anti-H3K4me1
(D1A9, cell signaling technology 5626T), anti-Set1 (Santa Cruz Biotechnology yE-13), anti-Pgk1
(Invitrogen A-6457), anti-H3K36me3 (Abcam Ab9050), anti-H3K36me2 (Abcam Ab9049), anti-
H3K36me1 (Abcam Ab9048), anti-goat (Invitrogen A-21446), anti-mousse (Invitrogen A-21036),
anti-rabbit (Invitrogen A-31537 or Abcam Ab6721 for H3K36me blots).
Cell extracts were prepared using actively growing cells washed with water. Cells were lysed by
vortexing with zirconia beads in TN150 buffer (50 mM Tris-HCl pH 7.8, 150 mM NaCl, 0.1 %
NP-40, 5 mM β-mercaptoethanol, complete EDTA-free protease inhibitor cocktail from Roche).
The lysate was cleared by centrifugation. The protein concentration in the soluble extract was
quantified by Bradford assay. The extract was denatured in NuPAGE sample buffer (Invitrogen)
by incubation at 70°C for 10 min. 15 to 50 µg of protein were resolved on 3-8% Tris-Acetate
NuPAGE gels (Invitrogen), 4–12% Bis-Tris NuPAGE gels (Invitrogen) or 15% SDS-
polyacrylamide gels, for Set1, Pgk1 and H3, respectively. Proteins were transferred to
nitrocellulose membranes, probed with the indicated antibodies and imaged using the Licor
Odyssey system.
**In-vivo RNA crosslinking**

Actively growing cells in SD medium with 2% glucose lacking tryptophan were UV cross-linked at 254nm and processed essentially as described (47, 48).

Tagged proteins were recovered from total lysates by incubation with IgG Sepharose for 2h for Set2 or overnight for Set1, and eluted by TEV cleavage. The eluates were subjected to partial RNase degradation, denatured by the addition of 6M guanidinium–HCl and RNA-protein complexes were bound to nickel columns. The RNAs were labelled using [$^{32}$P] ATP and linkers were added to both ends, on the nickel column. The complexes were eluted with imidazole and resolved on 4–12% Bis-Tris or 3-8% Tris-Acetate NuPAGE gels (Invitrogen), for Set2 and Set1 respectively, transferred to nitrocellulose membranes and detected by autoradiography. Bands corresponding to the size of the protein of interest were excised and incubated with proteinase K to release the bound RNAs. Phenol purified RNAs were reverse transcribed and PCR amplified. Libraries were resolved on agarose gels and fragments with insert sizes from approximately 20 to 80 bp were excised from the gel and sequenced using Illumina HiSeq, 50bp single-end reads (Edinburgh Genomics or Source Bioscience). The reagents used are referenced in (49).

**CRAC data analysis**

The datasets were demultiplexed using pyBarcodeFilter from pyCRAC (50). FLEXBAR (51) was used to remove the 3’ sequencing adapters, trim low-quality positions from the 3’ end of reads and remove reads without a high-quality score (parameters –u 3 –q 30 –m 17 –ao 3). In addition of the barcode, the 5’ linkers contain a random 3 nt sequence, allowing PCR duplicates to be removed by collapsing identical sequences. Reads were filtered to exclude low complexity sequences (with more than 80% of one nucleotide) to avoid potential non-genome-encoded oligo(A) tails to map to A-rich regions of the genome (31). Reads were mapped to the yeast
genome (S. cerevisiae genome version EF4.74, from Ensembl) using novoalign from Novocraft (parameters -s 1 -r Random). To remove PCR duplicates that were not collapsed during preprocessing due to sequencing errors or differential trimming at the 3’ end, any reads with the same random tag in their 5’ linker and with 5’ ends mapping to the same genomic coordinate were collapsed (31).

We used genome annotation from Ensembl (EF4.74), supplemented with non-coding sequences as previously described (31). Distribution of reads across transcript classes was determined using pyReadCounters from pyCRAC. The relative abundance of spliced and unspliced reads was calculated as described (27). The coverage at each position along the genome was calculated and normalized to the library size (reads per million) (27), after exclusion of reads mapping to RNAPI and RNAPIII transcripts (including novel transcripts described in (52) or originating from the mitochondrial genome. Replicate datasets were averaged. The enrichment of Set1, Set2 or phosphorylated RNAPII relative to total RNAPII was calculated as Log2(protein coverage +5/total Rpo21-HTP coverage +5), where the pseudocount of 5 avoids numerical instabilities (27). Coverage around genomic features (metagene analyses and 2D heatmaps) was plotted as in (27). To compare Set1 or Set2 coverage to RNAPII coverage around genomic features, a subset of features with highly reproducible coverage within Set1 or Set2 replicate datasets (features for which the ratio standard deviation to mean was bellow 0.5) and which were confidently bound (RPKM over 30) have been selected (4851, 2867, 4199 features for PTH-Set1, PTH-Set1ΔRRM2, Set1-HTP, Set2-HTP, respectively).

CRAC sequences generated during this work have been deposited with GEO; accession number GSE87919. RNAPII CRAC datasets (27); GEO accession number GSE69676) were reprocessed with pipeline described above.

RT-qPCR
RNA was isolated as described previously (53). Quantity and purity of RNA were analyzed using a NanoDrop 1000. 2µg of total RNA were treated with RQ1 RNase-Free DNase (Promega) and the reaction was stopped by a phenol:chloroform extraction. Single stranded cDNA was generated using gene random primers (Thermo Scientific) and the MuLV reverse transcriptase (Thermo Scientific). The expression level of individual transcripts was determined by quantitative PCR using SYBR Premix Ex Taq II Tli RNase H Plus (Clontech) for detection and using oligonucleotides listed in Table S3. Relative levels were determined by normalization to the ACT1 mRNA in each sample. Using Prism (GraphPad Software, Inc) and assuming normality, an ANOVA followed by a Dunnett’s test were performed to determine whether the relative expression measured in each strain was significantly different from that measured in the with-type BY4741 strain.

**RNA-seq**

Wild-type W303 and otherwise isogenic Δset2 cells (Table S2) were grown in YPDA to OD$_{660}$=0.6. Independent samples of total RNA were prepared from three WT and three Δset2 colonies by hot phenol extraction. RNA was further subjected to DNAse I treatment (E1009-A, Zymo Research) and Ribo-zero treatment (RZY1324, Illumina) following manufacturer instructions. Quantity and purity of RNA was measured using Agilent High sensitivity RNA screen Tape System (Agilent Technologies, cat:5067-5579) and Qubit (Molecular Probes, Invitrogen). Libraries were prepared for sequencing from 200ng of rRNA depleted total RNA using the NEXTflex™ RNA-Seq Kit (Bioo scientific, cat: 5129-02) following the manufacturer’s instructions. Samples were barcoded and combined together at uniform molarity to create a single pool, which was sequenced in a single end 76 bp run on an Illumina NextSeq machine. Multiplexed reads were split based on their NEXTflex barcodes, and 3’ adapter sequences were trimmed using Illumina Basespace software. Trimmed reads were mapped to the sacCer3 genome using tophat (54) with parameters --segment-length 38 --no-coverage-search --max-
multihits 20 --report-secondary-alignments --read-mismatches 2. Mapped reads were filtered to remove reads mapping to more than one unique genomic locus (multihits) by keeping only reads with flag NH:i:1 in the output bam file from tophat. Reads were further filtered to remove reads with mapping quality less than 20 using samtools (55).

Downstream analyses were conducted using the statistical programming language R (R Development Core Team, 2008) and bioconductor packages. Transcriptome annotation was taken from Ensembl (EF4.74), supplemented with non-coding sequences as previously described (31). Read counts within transcriptional units were generated using summarizeOverlaps() from the GenomicAlignments package (56) with parameters mode = "Union", singleEnd = TRUE, inter.feature = TRUE, ignore.strand = TRUE, fragments = FALSE.

Differential expression analysis of Δset2 samples against WT samples was performed using DESeq2 (57). Genes showing significant changes in expression in Δset2 samples were identified based on a fold change greater than 1.5 (up or down) and an adjusted p value (58) less than 0.05. RNA-seq sequences generated during this work have been deposited with GEO; accession number GSE89238.

Set1 ChIP-qPCR

The tagged strains PTH-Set1 and PTH-Set1ΔRRM2, and the untagged BY4741 strain were analysed by ChIP. Actively growing cells in complete minimal media at OD 0.5 were fixed for 15 min with 1% formaldehyde. Crosslinking reaction was quenched by addition of 150mM glycine. Cells were washed in cold PBS, frozen in liquid nitrogen and stored at -80°C. Cell pellet were disrupted in lysis buffer (50 mM HEPES/KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate, 0.1% SDS, Complete Protease inhibitors EDTA free from Roche Applied Science) using a mini bead-beater. Unless stated otherwise, subsequent steps were performed at 4°C. Soluble lysate was discarded after centrifugation and insoluble chromatin was resuspended in lysis buffer. Chromatin was sheared by 20 cycles of sonication, 30s on, 30s off.
a Bioruptor 300 (Diagenode) at high power, leading to fragments from 0.1 to 1 Kb. The
solubilized chromatin was separated from insoluble debris by centrifugation. 1.5 mg of
chromatin were used for IP, 37.5 ug were used as input samples. ChIP was performed by
incubating the lysate with rabbit IgG (Sigma 15006) coupled with Dynabeads M270 Epoxy
(Invitrogen) for 2h. Beads were washed for 15 min with each one of the following buffers: lysis
buffer, 0.5M lysis buffer (as lysis buffer but with 500 mM NaCl), wash buffer (10 mM Tris/HCl pH
8, 0.25 M LiCl, 0.5% NP-40, 0.5% Na-Deoxycholate, 1 mM EDTA), TE (10 mM Tris/HCl pH 8, 1
mM EDTA). Beads were resuspended in elution buffer (50 mM Tris/HCl pH 8, 10 mM EDTA, 1%
SDS) and crosslinking was reverted by overnight incubation at 65°C. Samples were treated with
0.25 mg/mL of proteinase K (Roche) at 55°C for 4h and with 0.2 mg/mL of RNAse A (Thermo
Scientific) at 37°C for 2h. DNA was purified using the Qiaquick kit (Qiagen), elution buffer was
supplemented with 0.2mg/mL of RNAse A and eluted DNA was incubated at 37°C for 2h.
Relative DNA amounts present in input samples and purified fractions were determined by
qPCR using SYBR Premix Ex Taq II (Clonethech). Primer pairs used for amplification are listed
in Table S3. All samples were run at least in triplicate. The mean values and error bars are
derived from three biological replicates. Results, including the BY4741 negative control, are
included in Table S4.

H3 ChIP-qPCR

The wild-type and Set1ΔRRM2 strains were grown and crosslinked as described above for the
Set1 ChIP. Unless stated otherwise, subsequent steps were performed at 4°C. Cells were
disrupted in lysis buffer (20 mM Tris/HCl pH 8, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100,
0.1% SDS, Complete Protease inhibitors EDTA free from Roche Applied Science, 0.5mM
Phenylmethylsulfonyl fluoride) using a Fastprep (MP Biomedicals). Chromatin was sheared by
sonication, 5s on, 5s off, 95% amplitude for 3h a Q800R2 Sonicator (Qsonica). IP buffer (167
mM Tris/HCl pH 8, 167 mM NaCl, 1.2 mM EDTA, 1.1 % Triton-X-100, 0.01 % SDS, 0.5 mM
PMSF, Complete Protease inhibitors from Roche) was added to the solubilized chromatin and incubated for 15 min. 50 µL of chromatin were used as input DNA. ChIP was performed by overnight incubation of 1 mL of chromatin with antibodies (from Abcam) against H3 (ab1791), H3K4me1 (ab8895), H3K4me2 (ab7766), H3K4me3 (ab8580) or GFP as a negative control (ab290), followed by 2h incubation with Dynabeads-protein A (Invitrogen). Beads were washed with TSE-150 buffer (20mM Tris/HCl pH 8, 150mM NaCl, 2mM EDTA, 1% Triton-X-100, 0.1% SDS), TSE-500 (as TSE-150 but with 500 mM NaCl), wash buffer (10 mM Tris/HCl pH 8, 0.25 M LiCl, 1% NP-40, 1% Na-Deoxycholate, 1 mM EDTA), TE (10 mM Tris/HCl, pH 8, 1 mM EDTA). DNA was eluted at 65°C in elution buffer (100 mM NaHCO₃, 1% SDS) and crosslinking was reverted, after addition of 500 mM NaCl, by overnight incubation at 65°C. Samples were treated with 0.5 mg/mL of RNAse A at 37°C for 2h. DNA was purified using the ChIP DNA Clean & Concentrator kit (Zymo Research).

Relative DNA amounts were determined by qPCR using primer pairs listed in Table S3. The mean values and error bars are derived from three biological replicates. Results, including the negative controls, are included in Table S5.

DECLARATIONS

Acknowledgements

We thank H Dunn-Davies and G Kudla for help with the data analysis, ML Winz for critical reading of the MS and A Cook for valuable advice on structural aspects.

Funding

This work was supported by EMBO Long Term Fellowships [ALTF 625-2014] to CS and [ALTF 907-2014] to GMZ, and a Wellcome Trust Fellowship [077248] to DT. Work in the Wellcome Trust Centre for Cell Biology is supported by Wellcome Trust core funding [092076].
Authors' contributions

CS, DT, GMZ, HSR and TK. designed the experiments. CS, GMZ, HSR EP and JH performed the experiments. CS and SR analyzed the sequencing data. CS and DT wrote the paper with contributions from all authors.

Availability of data and material

CRAC and RNA-seq sequences generated during this work have been deposited to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/); accession numbers GSE87919 and GSE89238.

Reviewer links are:
CRAC data
RNA-seq data

Competing interests

The authors declare that they have no competing interests

REFERENCES


Tollervey D, Mattaj IW. 1987. Fungal small nuclear ribonucleoproteins share properties with plant and vertebrate U-snRNPs. EMBO J 6:469.


FIGURES TITLES AND LEGENDS

Figure 1. Set1 and Set2 interact with RNA in vivo.

A. Domain organization of fusion proteins used in this study. RRM: RNA Recognition Motif; H4i: H4 interacting domain; AWS: Associated with SET; PS: Post-SET; WW: Typtophan-rich domain; CC: Coiled Coil domain; SRI: Set2-Rpb1 interaction domain; PTH: ProteinA-TEV-His6 tag; HTP: His6-TEV-ProteinA tag. P_{SET1} and P_{SET2} are SET1 and SET2 promoters, respectively.
B. Western blot showing protein abundance in the samples used in C. Cells were grown in minimal media lacking tryptophan and UV-crosslinked. The input lysate was analyzed with antibodies against H3K4me3, H3 and Pgk1 (loading controls). Molecules eluted from IgG beads using TEV protease were analyzed with anti-Set1 antibodies.

C-D. SDS-PAGE and autoradiography of the 5' $^{32}$P labeled, crosslinked RNAs after purification of the tagged proteins, or after mock purification from the untagged strain (BY4741).

**Figure 2.** Set1 is enriched near the TSS while Set2 binds across nascent RNAPII transcripts.

A. Distribution of reads across transcript classes in the CRAC datasets. Replicates have been averaged. Rpo21-HTP represents RNAPII.

B. Relative recovery of spliced mRNAs versus unspliced pre-mRNAs, expressed as the ratio of RNA fragments spanning exon-exon to intron-exon and exon-intron junctions. Error bars represent standard deviation from the replicates listed in Table S1.

C-D. Distribution of PTH-Set1 (C) and Set2-HTP (D) across individual mRNAs in reads per million of RNAPII transcripts. Transcripts are aligned to the TSS and pA site in the left and right panels, respectively. Distances are indicated in nucleotides. The corresponding total coverages are shown in panel E.

E. Metagene analysis of PTH-Set1, PTH-Set1ΔRRM2, Set1-HTP, Set2-HTP and RNAPII (Rpo21-HTP) across mRNAs, in reads per million of RNAPII transcripts. Transcripts are aligned to the TSS (left) or pA site (right).

F. Metagene analysis of PTH-Set1, RNAPII-S5P, and Set2-HTP, RNAPII-S2P, RNAPII-T4P enrichment relative to total RNAPII, across mRNAs aligned to their TSS (left) or pA site (right). The relative enrichment was calculated as log2(protein coverage/total-Rpo21-HTP coverage). The enrichment across individual mRNAs is shown in Figure S3E-F for PTH-Set1 and Set2-HTP.
Figure 3. Some transcripts show high enrichment for Set1 or Set2 relative to RNAPII.

A-B. PTH-Set1 (A) or Set2-HTP (B) coverage over genomic features (mRNAs, transcripts antisense to mRNAs, intergenic transcripts, SUTs, CUTs, XUTs) is plotted against RNAPII coverage (Rpo21-HTP). The fill color of the points represents the enrichment for Set1 or Set2 relative to RNAPII. Some classes of transcripts have been highlighted, as indicated on the right side. Other RNA classes are shown in Figure S5C.

C-E. Coverage, in reads per million of RNAPII transcripts, at loci where Set1 is enriched over RNAPII.

C. SET1 locus. The transcription unit is represented under the plots with the thicker box corresponding to the coding sequence.

D. A retrotransposon locus. YML045W-A and YML045W are coding for TYA and TYA-TYB, respectively. The LTRs are shadowed on the plots.

E. rDNA intergenic spacers (IGS) region. rRNA genes appear white on the plots.

Figure 4. RNA binding stabilizes interactions of Set1 with chromatin.

A. Schematic representation of the genes analyzed. The transcription unit is represented, with the coding sequences being thicker. Bars indicate PCR products.

B. Set1 ChIP in PTH-Set1 and PTH-Set1ΔRRM2 strains. Associated DNA was analyzed by qPCR, the signal is expressed as percentage of input DNA. Error bars represent the standard deviation from biological triplicates.

C. H3K4me3, H3K4me2 and H3K4me1 ChIP in the wild-type and Set1ΔRRM2 strains. The signal is normalized the total H3 signal. Error bars represent the standard deviation from biological triplicates.
Sayou et al. Fig. 1
Sayou et al. Fig. 2
A  

B  

C  

Sayou et al. Fig. 4
SUPPLEMENTAL MATERIAL

Supplemental tables

Table S1. Number of reads mapping to RNAPII transcripts (any read mapping to the nuclear genome, outside of RNAPI and RNAPII transcribed genes), where PCR duplicates have been removed, in the CRAC datasets used in this study.

<table>
<thead>
<tr>
<th></th>
<th>Sample Name</th>
<th>Number of Reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTH-Set1</td>
<td>PTH-Set1_1</td>
<td>409896</td>
</tr>
<tr>
<td></td>
<td>PTH-Set1_2</td>
<td>207591</td>
</tr>
<tr>
<td></td>
<td>PTH-Set1_3</td>
<td>105599</td>
</tr>
<tr>
<td></td>
<td>PTH-Set1_4</td>
<td>138665</td>
</tr>
<tr>
<td>PTH-Set1ΔRRM2</td>
<td>PTH-Set1dRRM2_1</td>
<td>337413</td>
</tr>
<tr>
<td></td>
<td>PTH-Set1dRRM2_2</td>
<td>68136</td>
</tr>
<tr>
<td></td>
<td>PTH-Set1dRRM2_3</td>
<td>17417</td>
</tr>
<tr>
<td>Set1-HTP</td>
<td>Set1-HTP_1</td>
<td>406822</td>
</tr>
<tr>
<td></td>
<td>Set1-HTP_2</td>
<td>648139</td>
</tr>
<tr>
<td></td>
<td>Set1-HTP_3</td>
<td>75245</td>
</tr>
<tr>
<td></td>
<td>Set1-HTP_4</td>
<td>71919</td>
</tr>
<tr>
<td>Set2-HTP</td>
<td>Set2-HTP_1</td>
<td>182583</td>
</tr>
<tr>
<td></td>
<td>Set2-HTP_2</td>
<td>136682</td>
</tr>
<tr>
<td></td>
<td>Set2-HTP_3</td>
<td>209546</td>
</tr>
<tr>
<td>BY4741</td>
<td>BY4741_1</td>
<td>882</td>
</tr>
<tr>
<td></td>
<td>BY4741_2</td>
<td>1162</td>
</tr>
</tbody>
</table>

Table S2. Yeast strains

<table>
<thead>
<tr>
<th>Name</th>
<th>ID</th>
<th>Genotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>BY4741</td>
<td>BY4741</td>
<td>MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</td>
<td>(59)</td>
</tr>
<tr>
<td>PTH-Set1</td>
<td>yCA14</td>
<td>MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; PTH-SET1</td>
<td>This study</td>
</tr>
<tr>
<td>PTH-Set1ΔRRM2</td>
<td>yCA18</td>
<td>MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; PTH-SET1ΔRRM2 (deletion nucleotides 243-482)</td>
<td>This study</td>
</tr>
<tr>
<td>Set1-HTP</td>
<td>D1202</td>
<td>MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; SET1-HTP-URA3</td>
<td>This study</td>
</tr>
<tr>
<td>Strain construction</td>
<td>Name</td>
<td>Sequence</td>
<td>Description</td>
</tr>
<tr>
<td>---------------------</td>
<td>------</td>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td>Set1ΔRRM2</td>
<td>yCA33</td>
<td>MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; SET1ΔRRM2 (deletion nucleotides 243-482)</td>
<td>This study</td>
</tr>
<tr>
<td>Δset1:URA:pURA</td>
<td>yCA28</td>
<td>MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; Δset1::URA3-pURA3</td>
<td>This study</td>
</tr>
<tr>
<td>Set2-HP</td>
<td>Set2-HP</td>
<td>MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; SET2-HP-URA3</td>
<td>This study</td>
</tr>
<tr>
<td>Rpo21-HP</td>
<td>Rpo21-HP</td>
<td>MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; RPO21-HP-URA3</td>
<td>(27)</td>
</tr>
<tr>
<td>W303</td>
<td>W303</td>
<td>MATa; his3-11,-15; leu2-3,-112; trp1-1; ura3-1; ade2-1; can1-100</td>
<td>R. Rothstein</td>
</tr>
<tr>
<td>Δset2</td>
<td>Δset2</td>
<td>MATa; his3-11,-15; leu2-3,-112; trp1-1; ura3-1; ade2-1; can1-100; Δset2::HIS3</td>
<td>This study</td>
</tr>
</tbody>
</table>

**Table S3. Oligonucleotides**

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>oCA164</td>
<td>TGTCTGTATGAACCAGAAGACGCGTG TGCTCTTCTATAGTAATTTGAGTACCGTGCTTCAGGTCGAC</td>
<td>Amplify a URA-KAN marker from pGSKU and integrate in between pSET1 and SET1 ORF</td>
<td>This study</td>
</tr>
<tr>
<td>oCA165</td>
<td>TAACATTCCTTTTTGGAAATTGTATTTAT AAGAGGTCTCGTCTCGTTTAGAGATAGG GATAACAGGGTAATCCCGAGGCTGGGTCGAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>oCA166</td>
<td>GGTGTCTGTATGAACCAGAAGACGCG TGTCCTGTTCTATAGTAATTTGATGTCGTCGAC</td>
<td>Amplify the PTH tag and replace the URA-KAN marker upstream SET1 ORF</td>
<td>This study</td>
</tr>
<tr>
<td>oCA167</td>
<td>TAACATTCCTTTTTGGAAATTGTATTTAT AAGAGGTCTCGTCTCGTTTAGAGAGGG AGAAATACCATGAAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>oCA168</td>
<td>GGTGTCTGTATGAACCAGAAGACGCG TGTCCTGTTCTATAGTAATTTGATGTCGTCGAC</td>
<td>Amplify a URA-KAN marker from pGSKU and integrate in the RRM2 region in SET1 ORF</td>
<td>This study</td>
</tr>
<tr>
<td>oCA151</td>
<td>GCACATCTGTATCCCTAAGCTTGTAT TTAAGCTCCTCAAGCAGAACCCTTTC GTACGCTGCAGGTCGAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>oCA152</td>
<td>TTAATTACAGACCTTTGTTATCATGCT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This study
<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Target gene(s) or region</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT1-F</td>
<td>GAAATGCAAAACCGCTGCTCA</td>
<td>ACT1</td>
<td>This study</td>
</tr>
<tr>
<td>ACT1-R</td>
<td>TACCCGCGAGATTCCAAAACCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>oCA246</td>
<td>GATGGTACAAGGCGACGCTA</td>
<td>SET1</td>
<td>This study</td>
</tr>
</tbody>
</table>

RT-qPCR
<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Target gene(s) or region</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>oCA247</td>
<td>ATGCCCTCCGACTACTGAT</td>
<td>ORC6</td>
<td>This study</td>
</tr>
<tr>
<td>oCA282</td>
<td>ACCAGGAAACGACGGTTTGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>oCA283</td>
<td>CGCTTCTGCTTTCTTGCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>oCA225</td>
<td>TGAGAGGTCTATCTGGCGAA</td>
<td>All TYA encoding genes</td>
<td>This study</td>
</tr>
<tr>
<td>oCA226</td>
<td>CCGAGCTATAAATTTGGGTTTGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ty-SetC-F</td>
<td>TCACTACACCAGCTGTTCCT</td>
<td>YLR035C-A, YHR214C-B, YER160C</td>
<td>(14)</td>
</tr>
<tr>
<td>Ty-SetC-R</td>
<td>GGAGTGAAGATCAGCGATAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p17-F</td>
<td>GGAAAGCGGGAAGGAAATAAG</td>
<td>IGS1, region of the rDNA repeat between RDN37 and RDN5</td>
<td>(60)</td>
</tr>
<tr>
<td>p17-R</td>
<td>CGATTCAGAAAAATTCGCACT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>oCA235</td>
<td>TGGGGTGTTATAGTCCGCAT</td>
<td>IGS2, region of the rDNA repeat between RDN5 and RDN37</td>
<td>This study</td>
</tr>
<tr>
<td>oCA236</td>
<td>TCGGTGACCGGAATACGCTT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ChIP-qPCR**

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMA1_1up</td>
<td>GGTACCGCTTATGCTCCCCTCCAT</td>
<td>PMA1</td>
</tr>
<tr>
<td>PMA1_1lo w</td>
<td>ATTTTTTTTCTTTTTTTGAATGTGTG</td>
<td></td>
</tr>
<tr>
<td>PMA1_2up</td>
<td>AAGTCGTTCCAGATGTATTTTGCA</td>
<td></td>
</tr>
<tr>
<td>PMA1_2lo w</td>
<td>AACGAAATGGATTTGCACCGGATAGC</td>
<td></td>
</tr>
<tr>
<td>PMA1_3up</td>
<td>CAGAGTGTGTAATCCATTGC</td>
<td></td>
</tr>
<tr>
<td>PMA1_3lo w</td>
<td>TGTCTGAGGTCTTTCAAAGC</td>
<td></td>
</tr>
<tr>
<td>PMA1_4up</td>
<td>TCATCGCTACCATGTTACC</td>
<td></td>
</tr>
<tr>
<td>PMA1_4lo w</td>
<td>CTTCATTGGCTTACCGGTTTCATC</td>
<td></td>
</tr>
<tr>
<td>oCA262</td>
<td>GTTGTCGTTATCGGTATGTCG</td>
<td>TEF1, TEF2</td>
</tr>
<tr>
<td>oCA263</td>
<td>CTTGTCAAAAACCCAAGCGT</td>
<td></td>
</tr>
<tr>
<td>oCA268</td>
<td>CGACCCATTCATACCAACCG</td>
<td>TDH3</td>
</tr>
<tr>
<td>oCA269</td>
<td>ACCCATGGCAAGTATGC</td>
<td></td>
</tr>
<tr>
<td>oCA272</td>
<td>TCTGCAACTCCCGTGTCATC</td>
<td>ILV5</td>
</tr>
</tbody>
</table>
Table S4. Set1 ChIP-qPCR expressed as percentage of input DNA from biological triplicates. Protein A beads were used for ChIP in BY4741 (untagged strain), PTH-Set1 and PTH-Set1ΔRRM2.

<table>
<thead>
<tr>
<th>Primer pairs</th>
<th>Antibody</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>PTH-Set1</td>
<td>PTH-Set1ΔRRM2</td>
</tr>
<tr>
<td>PMA1-1</td>
<td>H3K4me1</td>
<td>0,403</td>
<td>0,0009</td>
</tr>
<tr>
<td></td>
<td>H3K4me2</td>
<td>0,0672</td>
<td>0,0017</td>
</tr>
<tr>
<td></td>
<td>H3K4me3</td>
<td>0,0301</td>
<td>0,0024</td>
</tr>
<tr>
<td></td>
<td>GFP</td>
<td>0,0254</td>
<td>0,0020</td>
</tr>
<tr>
<td>TEF1</td>
<td></td>
<td>0,1009</td>
<td>0,0013</td>
</tr>
<tr>
<td>TDH3</td>
<td></td>
<td>0,1436</td>
<td>0,0014</td>
</tr>
<tr>
<td>ILV5</td>
<td></td>
<td>0,1070</td>
<td>0,0015</td>
</tr>
</tbody>
</table>

Table S5. Methylated H3K4 ChIP in the wild-type (expressing endogenous Set1) and Set1ΔRRM2 strains from biological triplicates. Antibodies against H3K4me3, H3K4me2 and H3K4me1, or GFP (negative control) have been used for ChIP. The signal is normalized the total H3 signal.

<table>
<thead>
<tr>
<th>Primer pairs</th>
<th>Antibody</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Set1</td>
<td>Set1ΔRRM2</td>
</tr>
<tr>
<td>PMA1-1</td>
<td>H3K4me1</td>
<td>0,408</td>
<td>0,395</td>
</tr>
<tr>
<td></td>
<td>H3K4me2</td>
<td>3,024</td>
<td>4,526</td>
</tr>
<tr>
<td></td>
<td>H3K4me3</td>
<td>0,087</td>
<td>1,888</td>
</tr>
<tr>
<td></td>
<td>GFP</td>
<td>5,869E-05</td>
<td>1,040E-04</td>
</tr>
<tr>
<td>PMA1-2</td>
<td>H3K4me1</td>
<td>0,428</td>
<td>0,437</td>
</tr>
<tr>
<td></td>
<td>H3K4me2</td>
<td>0,983</td>
<td>1,230</td>
</tr>
<tr>
<td></td>
<td>H3K4me3</td>
<td>0,489</td>
<td>0,404</td>
</tr>
<tr>
<td></td>
<td>GFP</td>
<td>4,298E-05</td>
<td>9,150E-05</td>
</tr>
<tr>
<td></td>
<td>H3K4me1</td>
<td>H3K4me2</td>
<td>H3K4me3</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td><strong>PMA1-3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3K4me1</td>
<td>0,202</td>
<td>0,210</td>
<td>0,052</td>
</tr>
<tr>
<td>H3K4me2</td>
<td>0,312</td>
<td>0,441</td>
<td>0,063</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>0,118</td>
<td>0,116</td>
<td>0,013</td>
</tr>
<tr>
<td><strong>GFP</strong></td>
<td>8,729E-05</td>
<td>1,252E-04</td>
<td>8,586E-05</td>
</tr>
<tr>
<td><strong>PMA1-4</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3K4me1</td>
<td>0,208</td>
<td>0,218</td>
<td>0,052</td>
</tr>
<tr>
<td>H3K4me2</td>
<td>0,311</td>
<td>0,389</td>
<td>0,072</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>0,122</td>
<td>0,129</td>
<td>0,010</td>
</tr>
<tr>
<td><strong>GFP</strong></td>
<td>8,352E-05</td>
<td>1,065E-04</td>
<td>7,269E-05</td>
</tr>
<tr>
<td><strong>TDH3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3K4me1</td>
<td>0,188</td>
<td>0,193</td>
<td>0,047</td>
</tr>
<tr>
<td>H3K4me2</td>
<td>2,105</td>
<td>3,779</td>
<td>0,726</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>5,135</td>
<td>3,468</td>
<td>0,335</td>
</tr>
<tr>
<td><strong>GFP</strong></td>
<td>2,862E-05</td>
<td>1,771E-05</td>
<td>1,534E-05</td>
</tr>
<tr>
<td><strong>TEF1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3K4me1</td>
<td>0,084</td>
<td>0,087</td>
<td>0,037</td>
</tr>
<tr>
<td>H3K4me2</td>
<td>1,196</td>
<td>2,388</td>
<td>0,376</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>8,341</td>
<td>5,773</td>
<td>1,398</td>
</tr>
<tr>
<td><strong>GFP</strong></td>
<td>2,636E-05</td>
<td>1,512E-05</td>
<td>2,852E-06</td>
</tr>
<tr>
<td><strong>ILV5</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3K4me1</td>
<td>0,113</td>
<td>0,133</td>
<td>0,040</td>
</tr>
<tr>
<td>H3K4me2</td>
<td>1,165</td>
<td>3,084</td>
<td>0,376</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>4,755</td>
<td>3,074</td>
<td>0,131</td>
</tr>
<tr>
<td><strong>GFP</strong></td>
<td>1,437E-04</td>
<td>6,050E-05</td>
<td>1,728E-04</td>
</tr>
</tbody>
</table>
Supplemental Figures and legends

Figure S1. Protein level and crosslinking efficiency in the different strains.
A. Protein abundance and H3K4me3 levels in the Set1 strains used in this study, detected by western-blot. Yeast cells were grown in rich medium. * indicates a non-specific band detected with the anti-Set1 antibodies.
B. H3K36 mono-, di- and tri-methylation levels in the Set2-HTP, the isogenic Δset2 and the untagged (BY4741) strains, detected by western-blot. Yeast cells were grown in minimal media lacking tryptophan. Bellow each H3K36me blot is the H3 blot obtained from the same membrane.
C. Growth curves of the different strains used in this study. Cells were grown in minimal media lacking tryptophan.
D. SDS-PAGE and autoradiography of the 5’ [32P] labeled, crosslinked RNAs after purification of the Set1-HTP protein, or after mock purification from the untagged strain (BY4741).
E. Number of reads recovered from a CRAC experiment where crosslinked and barcoded samples from PTH-Set1, PTH-Set1ΔRRM2 and from the untagged strain (BY4741) were mixed prior to SDS-PAGE separation and RT-PCR amplification.

Figure S2. Set1 and Set2 relative enrichment and mRNA stability.
A-D. PTH-Set1 (A), PTH-Set1ΔRRM2 (B), Set1-HTP (C), Set2-HTP (D) enrichment relative to RNAPII (Rpo21-HTP) across mRNAs is plotted against the mRNA half-lives (32).

Figure S3. Set1, Set2 and RNAPII distribution across transcripts.
A-D. Coverage, in reads per million of RNAPII transcripts, for selected individual protein coding genes: PMA1 (A), RPS13 (B), TEF1 (C), RPL3 (D). Transcription units are represented under the plots with thicker boxes correspond to coding sequences.
E-F. Distribution of PTH-Set1 (B) and Set2-HTP (C) enrichment relative to total RNAPII, across individual mRNAs. Transcripts are aligned to the TSS and pA site in the left and right panels, respectively. Distances are indicated in nucleotides. Blue color indicates depletion, and red color indicates enrichment relative to total RNAPII.

G. Metagene analysis of PTH-Set1, Set2-HTP, RNAPII-S2P and RNAPII-S5P enrichment relative to total RNAPII across mRNAs (left), SUTs (middle) and CUTs (right) with similar expression levels, based on their total RNAPII CRAC signal over the first 300 nt (27). Each subset contains 211 to 213 transcripts.

Figure S4. Set1 binding to SET1 mRNA and transcripts abundance in the different strains.

A. Set1-HP coverage, in reads per million of RNAPII transcripts, at the SET1 locus. The transcription unit is represented under the plot with the thicker box corresponding to the coding sequence. The corresponding plots for PTH-Set1 and RNAPII are shown in Figure 3C.

B. PTH-Set1 and RNAPII coverage, in reads per million of RNAPII transcripts, at a retrotransposon locus. YML045W-A and YML045W are coding for TYA and TYA-TYB, respectively. The LTRs are shadowed on the plots.

C-E. Transcript level measured by RT-qPCR. * indicates a relative expression different from that of BY4741 with a p-value bellow 0.01, calculated with a Dunnett’s test. SET1 mRNA (C), IGS transcripts (D) and retrotransposons (E) have been assessed.

Figure S5. Enrichment for Set1 or Set2 relative to RNAPII on transcripts.

A. PTH-Set1dRRM2 coverage over genomic features is plotted against PTH-Set1 coverage. The different transcript classes have been plotted separately on different panels. Selected classes of transcripts have been highlighted, as indicated below the plots.
B. PTH-Set1dRRM2 coverage over genomic features is plotted against RNAPII coverage (Rpo21-HTP). Note that the fill color of the points represents the enrichment for PTH-Set1 relative to RNAPII.

C. PTH-Set1 (top), Set2-HTP (bottom) coverage over genomic features is plotted against RNAPII coverage. The fill color of the points represents the enrichment for Set1 or Set2 relative to RNAPII (as in Figure 3A-B).

D. PTH-Set1 (left) or Set2-HTP (right) coverage over mRNAs is plotted against RNAPII coverage (Rpo21-HTP). The fill color of the points represents the enrichment for Set1 or Set2 relative to RNAPII. mRNAs found as differentially expressed in set1Δ (14) or set2Δ (this study) compared to wild-type in RNA-seq analysis are highlighted in green.

Figure S6. Global methylated H3K4 levels in Set1 and Set1ΔRRM2.

A. H3K4me3, H3K4me2, H3K4me1, H3 (loading) and Pgk1 (loading) levels in the wild type (Set1), Set1ΔRRM2 and Δset1 stains, detected by western-blot. Yeast cells were grown in minimal medium, each line results from an independent clone.

B. Quantification of the western-blots shown in A, were the signals have been normalized to H3.
Sayou et al. Fig. S2
Sayou et al. Fig. S5
Sayou et al. Fig. S6