TITLE

RNA binding by the histone methyltransferases Set1 and Set2

AUTHORS

Camille Sayou[1], Gonzalo Millán-Zambrano[2], Helena Santos-Rosa[2], Elisabeth Petfalski[1], Samuel Robson[2], Jonathan Houseley[3], Tony Kouzarides[2], David Tollervey[1]*

[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 tri-methylation were decreased whereas di-methylation was increased. We conclude that RNA binding by Set1 contributes to both chromatin association and methyltransferase activity.

INTRODUCTION

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 (1). 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 (2–4). 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 (5).

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 5–7). Set1 functions within the complex of proteins associated with Set1 (COMPASS or Set1C, reviewed in 8, 9), 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 10–12).

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 (14, 15). Subsequent analyses implicated Set1 in the repression of many genes (16, 17), with more obvious effects under stress conditions (18). In addition, Set1 has been implicated in transcriptional silencing of retrotransposons in *S. cerevisiae* (19, 20) and in *S. pombe* (21–23).

Set1-dependent silencing of Ty1 retrotransponsons is mediated by a non-coding, antisense transcript (20). Set1 is also implicated in silencing of RNAPII transcription from the intergenic spacer (IGS) regions located between the ribosomal DNA (rDNA) genes (24, 14). H3K36 methylation is found throughout protein coding genes and prevents initiation of transcription at cryptic sites via recruitment of the Rpd3S histone deacetylase complex (25–27). H3K36 methylation is also reported to regulate pre-mRNA splicing (28).

Yeast Set1 has two putative RNA recognition motifs (RRMs) that are implicated in Set1 function, suggesting that it might bind RNA *in vivo* (29, 30). Set2 does not harbor an evident RNA-binding motif, but was identified in systematic analyses of yeast RNA-interacting proteins (31, 2). 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 (32, 29, 21) (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 (33). 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 (29). Moreover, a deletion overlapping RRM1 reduced Set1 levels, whereas a construct lacking only RRM2 was stable (33). 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

112 experiment. At least 3 replicate datasets were obtained for each of PTH-Set1, PTH-
113 Set1ΔRRM2, Set1-HTP and Set2-HTP (Table S1).
114 To better estimate the relative *in vivo* binding to RNA of PTH-Set1 and PTH-
115 Set1ΔRRM2, crosslinked and barcoded samples were mixed prior to SDS-PAGE separation
116 and RT-PCR amplification. Following de-multiplexing, the number of reads recovered for PTH-
117 Set1ΔRRM2 was about 3 fold lower than for PTH-Set1 (Figure S1E), consistent with the
118 reduced binding observed from the autoradiography gels (Figure 1C). Substantially fewer reads
119 were recovered for the BY4741 untagged control.
120 The distribution of reads among RNA classes showed that both Set1 and Set2
121 predominately bound mRNAs (Figure 2A). Compared to BY4741, Set1, but not Set2, was also
122 enriched for binding to other non-protein coding RNAs (ncRNAs) transcribed by RNAPII,
123 including SUTs, CUTs, XUTs and intergenic transcripts. For comparison, we also show the
124 distribution of the catalytic subunit of RNAPII (Rpo21-HTP) using previous CRAC data (34).
125 PTH-Set1, PTH-Set1ΔRRM2 and Set1-HTP showed broadly similar distributions (Figure 2A).
126 PTH-Set1ΔRRM2 samples showed more rRNA and tRNA reads than PTH-Set1 (Figure 2A),
127 although this is likely to reflect a higher background due to reduced RNA binding rather than a
128 difference in endogenous RNA target classes. This indicated that loss of RRM2 greatly reduces
129 affinity for RNA (Figure 1C) but has limited impact on specificity (Figure 2A).
130 Pre-mRNA splicing is largely co-transcriptional in yeast (35–37). Therefore, the presence
131 of unspliced RNAs in CRAC datasets generally reflects protein interactions with nascent
132 transcripts. To assess whether Set1 and Set2 bind co-transcriptionally, the recovery of spliced
133 and unspliced transcripts from intron-containing genes was calculated as reported (34, 38). The
134 ratio of reads spanning exon-exon junctions (spliced) relative to exon-intron plus intron-exon
135 junctions (unspliced) was below 1 for both Set1 and Set2 (Figure 2B) indicating predominant
136 binding to unspliced, nascent pre-mRNAs. For Set2 the ratio was higher than for Set1,
137 consistent with Set2 binding later during transcription. PTH-Set1, PTH-Set1ΔRRM2 and Set1-

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 (39). 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 (34, 38, 40–43). 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 (34). 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 (44).

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 (13). 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 (34). 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 (34). 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 (34). 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 (45). 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-Set1ΔRRM2 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 (46, 47), 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 *set1*Δ (19) or *set2*Δ (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 transcritpion 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 (33, 48), with stronger crosslinking nearing

241 the 5' end. PTH-Set1ΔRRM2 binding to chromatin was ~30% reduced compared to PTH-Set1
242 at the 5' end (primer pairs 1, 2; Figure 4A-B), where Set1 RNA-binding peaked (Figure 2C, 2E;
243 Figure S3A-D). Binding to the 3' end of *PMA1* (primer pairs 3, 4) was similar for both proteins
244 (Figure 4A-B). Reductions of ~25 to 30% in binding of PTH-Set1ΔRRM2 was seen for three
245 other genes tested (*TEF1*, *TDH3*, *ILV5*: Figure 4A-B). The data indicate that reduced RNA
246 binding by Set1 leads to weaker interactions with chromatin, but this may be specific for 5'
247 regions that show high Set1-RNA interactions.

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

259 These results demonstrate that RRM2 is functionally important for Set1 targeting at
260 chromatin and for methylation of H3K4. We propose that RNA binding participates to Set1
261 recuitment and/or stabilization at chromatin, therefore contributing to H3K4 methylation patterns.
262
263
264 **DISCUSSION**
265 This study presents high resolution, strand-specific, transcriptome-wide mapping of two major
266 histone methyltransferases Set1 and Set2, which are conserved from yeast to human. Both

proteins directly interacted with RNA *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 (44).

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 (45) 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 (19, 20) and IGS regions (24, 14) 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 (6, 8). 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.

292 The Set1ΔRRM2 protein showed reduced chromatin association in ChIP analyses
293 (Figure 4A-B), indicating that RNA binding functions in the recruitment of Set1 to chromatin
294 and/or stabilizes the association. Consistent with this conclusion, it was previously shown that a
295 truncated version of Set1 containing only the SET domain and most of the N-SET had reduced
296 chromatin occupancy in yeast (33). Notably, analyses at different sites along the *PMA1* gene
297 revealed clear differences in chromatin occupancy only in the 5' region. This suggests a
298 potential correlation between stabilization of chromatin association and high RNA binding.
299 Consistent with this model, the absence of RRM2 also led to reduced H3K4me3 and increased
300 H3K4me2, at the 5' end of genes (Figure 4C), demonstrating that RRM2 is required for the
301 correct distribution of H3K4me3 and H3K4me2.
302 *In vivo* and *in vitro* expreriements previously showed that the pattern of mono-, di- and
303 tri-methylation deposited by Set1 correlated with interaction time of the COMPASS complex with
304 its nucleosome substrate, monomethylation occurring virtually imediateley, followed by
305 dimethylation, and finally trimethylation (49). Other parameters, such as the COMPASS
306 complex subunit composition also directs the distribution of the three methylation states (9, 50).
307 We propose that RNA binding of Set1 via RRM2 near the TSS stabilizes the association of
308 Set1/COMPASS with chomatin, promoting formation of H3K4me3 at the 5' ends of genes. Due
309 to reduced RNA binding, Set1ΔRRM2-chromatin interaction is weaker or more transient, leading
310 to higer levels of H3K4me2. A major role of H3K4me2 is recruitment of the Set3 histone
311 deacetylase complex, which deacetylate histones in 5' regions of trabscrption units and
312 participates in H3K4me2 maintenance (51). This helps regulate overlapping non-coding
313 transcription and contributes to epigenetic transcriptional memory (52, 50). We speculate that
314 the disruption of RNA binding by Set1 adversely affects these processes.
315 The results reported here contribute to understanding of the crosstalk between RNA
316 synthesis and the modulation of chromatin structure. Recent studies have identified large
317 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.

METHODS

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 (53) 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 (53). 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 (54) 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.

370

371 ***In-vivo* RNA crosslinking**

372 Actively growing cells in SD medium with 2% glucose lacking tryptophan were UV cross-linked
373 at 254nm and processed essentially as described (54, 55).
374 Tagged proteins were recovered from total lysates by incubation with IgG Sepharose for 2h for
375 Set2 or overnight for Set1, and eluted by TEV cleavage. The eluates were subjected to partial
376 RNase degradation, denatured by the addition of 6M guanidinium–HCl and RNA-protein
377 complexes were bound to nickel columns. The RNAs were labelled using [γ^{32}P] ATP and linkers
378 were added to both ends, on the nickel column. The complexes were eluted with imidazole and
379 resolved on 4–12% Bis-Tris or 3-8% Tris-Acetate NuPAGE gels (Invitrogen), for Set2 and Set1
380 respectively, transferred to nitrocellulose membranes and detected by autoradiography. Bands
381 corresponding to the size of the protein of interest were excised and incubated with proteinase
382 K to release the bound RNAs. Phenol purified RNAs were reverse transcribed and PCR
383 amplified. Libraries were resolved on agarose gels and fragments with insert sizes from
384 approximately 20 to 80 bp were excised from the gel and sequenced using Illumina HiSeq, 50bp
385 single-end reads (Edinburgh Genomics or Source Bioscience). The reagents used are
386 referenced in (56).

387

388 **CRAC data analysis**

389 The datasets were demultiplexed using pyBarcodeFilter from pyCRAC (57). FLEXBAR (58) was
390 used to remove the 3' sequencing adapters, trim low-quality positions from the 3' end of reads
391 and remove reads without a high-quality score (parameters –u 3 –q 30 –m 17 –ao 3). In addition
392 of the barcode, the 5' linkers contain a random 3 nt sequence, allowing PCR duplicates to be
393 removed by collapsing identical sequences. Reads were filtered to exclude low complexity
394 sequences (with more than 80% of one nucleotide) to avoid potential non-genome-encoded
395 oligo(A) tails to map to A-rich regions of the genome (38). 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 (38).

We used genome annotation from Ensembl (EF4.74), supplemented with non-coding sequences as previously described (38). Distribution of reads across transcript classes was determined using pyReadCounters from pyCRAC. The relative abundance of spliced and unspliced reads was calculated as described (34). The coverage at each position along the genome was calculated and normalized to the library size (reads per million) (34), 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 (34). Coverage around genomic features (metagene analyses and 2D heatmaps) was plotted as in (34). 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, 4306, 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 (34); GEO accession number GSE69676) were reprocessed with pipeline described above.

RT-qPCR

RNA was isolated as described previously (60). 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, Invitrogene). 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 (61) 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 (62).

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 (38). Read counts within transcriptional units were generated using summarizeOverlaps() from the GenomicAlignments package (63) 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 (64). 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 (65) 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 4C°. 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. Using Prism (GraphPad Software, Inc) and assuming normality, Student's t-tests were performed to determine a p-value for the differences in percentage of input DNA obtained for PTH-Set1 and PTH-Set1ΔRRM2, for each primer pair. 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 4C°. 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

Molecular and Cellular Biology

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 $NaHCO_3$, 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. Using Prism (GraphPad Software, Inc) and assuming normality, Student's t-tests were performed to determine a p-value for the differences in relative enrichment to total H3 obtained for wild-type and Set1ΔRRM2, for each primer pair. 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

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=apkvegsszryjhyt&acc=GSE87919

RNA-seq data

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=onyfayuepxsnpgx&acc=GSE89238

Competing interests

The authors declare that they have no competing interests

REFERENCES

1. Skalska L, Beltran-Nebot M, Ule J, Jenner RG. 2017. Regulatory feedback from nascent RNA to chromatin and transcription. Nat Rev Mol Cell Biol.

2. Beckmann BM, Horos R, Fischer B, Castello A, Eichelbaum K, Alleaume A-M, Schwarzl T, Curk T, Foehr S, Huber W, Krijgsveld J, Hentze MW. 2015. The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. Nat Commun 6:10127.

3. Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, Krijgsveld J, Hentze MW. 2012. Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. Cell 149:1393–1406.

4. Kramer K, Sachsenberg T, Beckmann BM, Qamar S, Boon K-L, Hentze MW, Kohlbacher O, Urlaub H. 2014. Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. Nat Methods 11:1064–1070.

5. Hendrickson DG, Kelley DR, Tenen D, Bernstein B, Rinn JL. 2016. Widespread RNA binding by chromatin-associated proteins. Genome Biol 17.

6. Zaborowska J, Egloff S, Murphy S. 2016. The pol II CTD: new twists in the tail. Nat Struct Mol Biol 23:771–777.

7. Harlen KM, Churchman LS. 2017. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. Nat Rev Mol Cell Biol 18:263–273.

8. Heidemann M, Hintermair C, Voß K, Eick D. 2013. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. Biochim Biophys Acta BBA - Gene Regul Mech 1829:55–62.

9. Shilatifard A. 2012. The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. Annu Rev Biochem 81:65–95.

10. Piunti A, Shilatifard A. 2016. Epigenetic balance of gene expression by Polycomb and COMPASS families. Science 352:aad9780-aad9780.

11. McDaniel SL, Strahl BD. 2017. Shaping the cellular landscape with Set2/SETD2 methylation. Cell Mol Life Sci.

12. Venkatesh S, Workman JL, Smolle M. 2013. UpSETing chromatin during non-coding RNA production. Epigenetics Chromatin 6:1.

13. Hampsey M, Reinberg D. 2003. Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. Cell 113:429–432.

14. Briggs SD, Bryk M, Strahl BD, Cheung WL, Davie JK, Dent SY, Winston F, Allis CD. 2001. Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. Genes Dev 15:3286–3295.

15. Nislow C, Ray E, Pillus L. 1997. Set1, a yeast member of thetrithorax family, functions in transcriptional silencing and diverse cellular processes. Mol Biol Cell 8:2421–2436.

16. Margaritis T, Oreal V, Brabers N, Maestroni L, Vitaliano-Prunier A, Benschop JJ, van Hooff S, van Leenen D, Dargemont C, Géli V, Holstege FCP. 2012. Two Distinct Repressive Mechanisms for Histone 3 Lysine 4 Methylation through Promoting 3′-End Antisense Transcription. PLoS Genet 8:e1002952.

17. Castelnuovo M, Zaugg JB, Guffanti E, Maffioletti A, Camblong J, Xu Z, Clauder-Munster S, Steinmetz LM, Luscombe NM, Stutz F. 2014. Role of histone modifications and early termination in pervasive transcription and antisense-mediated gene silencing in yeast. Nucleic Acids Res 42:4348–4362.

18. Weiner A, Chen HV, Liu CL, Rahat A, Klien A, Soares L, Gudipati M, Pfeffner J, Regev A, Buratowski S, Pleiss JA, Friedman N, Rando OJ. 2012. Systematic Dissection of Roles for Chromatin Regulators in a Yeast Stress Response. PLoS Biol 10:e1001369.

19. Martín GM, King DA, Green EM, Garcia-Nieto PE, Alexander R, Collins SR, Krogan NJ, Gozani OP, Morrison AJ. 2014. Set5 and Set1 cooperate to repress gene expression at telomeres and retrotransposons. Epigenetics 9:513–522.

20. Berretta J, Pinskaya M, Morillon A. 2008. A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in S. cerevisiae. Genes Dev 22:615–626.

21. Mikheyeva IV, Grady PJR, Tamburini FB, Lorenz DR, Cam HP. 2014. Multifaceted Genome Control by Set1 Dependent and Independent of H3K4 Methylation and the Set1C/COMPASS Complex. PLoS Genet 10:e1004740.

22. Lorenz DR, Mikheyeva IV, Johansen P, Meyer L, Berg A, Grewal SIS, Cam HP. 2012. CENP-B Cooperates with Set1 in Bidirectional Transcriptional Silencing and Genome Organization of Retrotransposons. Mol Cell Biol 32:4215–4225.

23. Lorenz DR, Meyer LF, Grady PJ, Meyer MM, Cam HP. 2014. Heterochromatin assembly and transcriptome repression by Set1 in coordination with a class II histone deacetylase. Elife 3:e04506.

24. Li C, Mueller JE, Bryk M. 2006. Sir2 represses endogenous polymerase II transcription units in the ribosomal DNA nontranscribed spacer. Mol Biol Cell 17:3848–3859.

25. Venkatesh S, Li H, Gogol MM, Workman JL. 2016. Selective suppression of antisense transcription by Set2-mediated H3K36 methylation. Nat Commun 7:13610.

26. Kim JH, Lee BB, Oh YM, Zhu C, Steinmetz LM, Lee Y, Kim WK, Lee SB, Buratowski S, Kim T. 2016. Modulation of mRNA and lncRNA expression dynamics by the Set2–Rpd3S pathway. Nat Commun 7:13534.

27. Hsin J-P, Manley JL. 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev 26:2119–2137.

28. Sorenson MR, Jha DK, Ucles SA, Flood DM, Strahl BD, Stevens SW, Kress TL. 2016. Histone H3K36 methylation regulates pre-mRNA splicing in *Saccharomyces cerevisiae*. RNA Biol 13:412–426.

29. Trésaugues L, Dehé P-M, Guérois R, Rodriguez-Gil A, Varlet I, Salah P, Pamblanco M, Luciano P, Quevillon-Cheruel S, Sollier J, Leulliot N, Couprie J, Tordera V, Zinn-Justin S, Chàvez S, van Tilbeurgh H, Géli V. 2006. Structural Characterization of Set1 RNA Recognition Motifs and their Role in Histone H3 Lysine 4 Methylation. J Mol Biol 359:1170–1181.

30. Schlichter A, Cairns BR. 2005. Histone trimethylation by Set1 is coordinated by the RRM, autoinhibitory, and catalytic domains. EMBO J 24:1222–1231.

31. Klass DM, Scheibe M, Butter F, Hogan GJ, Mann M, Brown PO. 2013. Quantitative proteomic analysis reveals concurrent RNA-protein interactions and identifies new RNA-binding proteins in Saccharomyces cerevisiae. Genome Res 23:1028–1038.

32. Krogan NJ, Dover J, Khorrami S, Greenblatt JF, Schneider J, Johnston M, Shilatifard A. 2002. COMPASS, a Histone H3 (Lysine 4) Methyltransferase Required for Telomeric Silencing of Gene Expression. J Biol Chem 277:10753–10755.

637 33. Soares LM, Radman-Livaja M, Lin SG, Rando OJ, Buratowski S. 2014. Feedback Control
638 of Set1 Protein Levels Is Important for Proper H3K4 Methylation Patterns. Cell Rep 6:961–
639 972.

640 34. Milligan L, Huynh-Thu VA, Delan-Forino C, Tuck A, Petfalski E, Lombraña R, Sanguinetti
641 G, Kudla G, Tollervey D. 2016. Strand-specific, high-resolution mapping of modified RNA
642 polymerase II. Mol Syst Biol 12:874.

643 35. Alexander RD, Innocente SA, Barrass JD, Beggs JD. 2010. Splicing-Dependent RNA
644 Polymerase Pausing in Yeast. Mol Cell 40:582–593.

645 36. Carrillo Oesterreich F, Preibisch S, Neugebauer KM. 2010. Global Analysis of Nascent
646 RNA Reveals Transcriptional Pausing in Terminal Exons. Mol Cell 40:571–581.

647 37. Moehle EA, Braberg H, Krogan NJ, Guthrie C. 2014. Adventures in time and space:
648 Splicing efficiency and RNA polymerase II elongation rate. RNA Biol 11:313–319.

649 38. Tuck AC, Tollervey D. 2013. A transcriptome-wide atlas of RNP composition reveals
650 diverse classes of mRNAs and lncRNAs. Cell 154:996–1009.

651 39. Presnyak V, Alhusaini N, Chen Y-H, Martin S, Morris N, Kline N, Olson S, Weinberg D,
652 Baker KE, Graveley BR, Coller J. 2015. Codon Optimality Is a Major Determinant of mRNA
653 Stability. Cell 160:1111–1124.

654 40. Holmes RK, Tuck AC, Zhu C, Dunn-Davies HR, Kudla G, Clauder-Munster S, Granneman
655 S, Steinmetz LM, Guthrie C, Tollervey D. 2015. Loss of the Yeast SR Protein Npl3 Alters
656 Gene Expression Due to Transcription Readthrough. PLOS Genet 11:e1005735.

657 41. Churchman LS, Weissman JS. 2011. Nascent transcript sequencing visualizes
658 transcription at nucleotide resolution. Nature 469:368–373.

42. Nojima T, Gomes T, Grosso ARF, Kimura H, Dye MJ, Dhir S, Carmo-Fonseca M, Proudfoot NJ. 2015. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell 161:526–540.

43. Mayer A, di Iulio J, Maleri S, Eser U, Vierstra J, Reynolds A, Sandstrom R, Stamatoyannopoulos JA, Churchman LS. 2015. Native Elongating Transcript Sequencing Reveals Human Transcriptional Activity at Nucleotide Resolution. Cell 161:541–554.

44. Chabbert CD, Adjalley SH, Klaus B, Fritsch ES, Gupta I, Pelechano V, Steinmetz LM. 2015. A high-throughput ChIP-Seq for large-scale chromatin studies. Mol Syst Biol 11:777–777.

45. Halbach A, Zhang H, Wengi A, Jablonska Z, Gruber IM, Halbeisen RE, Dehé P-M, Kemmeren P, Holstege F, Géli V, others. 2009. Cotranslational assembly of the yeast SET1C histone methyltransferase complex. EMBO J 28:2959–2970.

46. Tomson BN, Crisucci EM, Heisler LE, Gebbia M, Nislow C, Arndt KM. 2013. Effects of the Paf1 Complex and Histone Modifications on snoRNA 3'-End Formation Reveal Broad and Locus-Specific Regulation. Mol Cell Biol 33:170–182.

47. Sheldon KE, Mauger DM, Arndt KM. 2005. A Requirement for the Saccharomyces cerevisiae Paf1 Complex in snoRNA 3′ End Formation. Mol Cell 20:225–236.

48. Soares LM, Buratowski S. 2012. Yeast Swd2 Is Essential Because of Antagonism between Set1 Histone Methyltransferase Complex and APT (Associated with Pta1) Termination Factor. J Biol Chem 287:15219–15231.

49. Wood A, Shukla A, Schneider J, Lee JS, Stanton JD, Dzuiba T, Swanson SK, Florens L, Washburn MP, Wyrick J, Bhaumik SR, Shilatifard A. 2007. Ctk Complex-Mediated Regulation of Histone Methylation by COMPASS. Mol Cell Biol 27:709–720.

50. D'Urso A, Takahashi Y, Xiong B, Marone J, Coukos R, Randise-Hinchliff C, Wang J-P, Shilatifard A, Brickner JH. 2016. Set1/COMPASS and Mediator are repurposed to promote epigenetic transcriptional memory. Elife 5:e16691.

51. Kim T, Buratowski S. 2009. Dimethylation of H3K4 by Set1 Recruits the Set3 Histone Deacetylase Complex to 5′ Transcribed Regions. Cell 137:259–272.

52. Kim T, Xu Z, Clauder-Münster S, Steinmetz LM, Buratowski S. 2012. Set3 HDAC Mediates Effects of Overlapping Noncoding Transcription on Gene Induction Kinetics. Cell 150:1158–1169.

53. Storici F, Resnick MA. 2006. The Delitto Perfetto Approach to In Vivo Site-Directed Mutagenesis and Chromosome Rearrangements with Synthetic Oligonucleotides in Yeast, p. 329–345. *In* Methods in Enzymology. Elsevier.

54. Granneman S, Kudla G, Petfalski E, Tollervey D. 2009. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. Proc Natl Acad Sci U S A 106:9613–9618.

55. Granneman S, Petfalski E, Tollervey D. 2011. A cluster of ribosome synthesis factors regulate pre-rRNA folding and 5.8 S rRNA maturation by the Rat1 exonuclease. EMBO J 30:4006–4019.

56. Helwak A, Tollervey D. 2014. Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH). Nat Protoc 9:711–728.

57. Webb S, Hector RD, Kudla G, Granneman S. 2014. PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. Genome Biol 15:1.

58. Dodt M, Roehr J, Ahmed R, Dieterich C. 2012. FLEXBAR—Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. Biology 1:895–905.

59. Turowski TW, Leśniewska E, Delan-Forino C, Sayou C, Boguta M, Tollervey D. 2016. Global analysis of transcriptionally engaged yeast RNA polymerase III reveals extended tRNA transcripts. Genome Res 26:933–944.

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

61. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14:1.

62. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

63. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. 2013. Software for Computing and Annotating Genomic Ranges. PLoS Comput Biol 9:e1003118.

64. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15.

65. Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Methodol 57:289–300.

66. Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P, Boeke JD. 1998. Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. YEAST-CHICHESTER- 14:115–132.

67. Huang J, Moazed D. 2003. Association of the RENT complex with nontranscribed and coding regions of rDNA and a regional requirement for the replication fork block protein Fob1 in rDNA silencing. Genes Dev 17:2162–2176.

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).

746 **Figure 2. Set1 is enriched near the TSS while Set2 binds across nascent RNAPII**
747 **transcripts.**
748 A. Distribution of reads across transcript classes in the CRAC datasets. Replicates have been
749 averaged. Rpo21-HTP represents RNAPII.
750 B. Relative recovery of spliced mRNAs versus unspliced pre-mRNAs, expressed as the ratio of
751 RNA fragments spanning exon-exon to intron-exon and exon-intron junctions. Error bars
752 represent standard deviation from the replicates listed in Table S1.
753 C-D. Distribution of PTH-Set1 (C) and Set2-HTP (D) across individual mRNAs in reads per
754 million of RNAPII transcripts. Transcripts are aligned to the TSS and pA site in the left and right
755 panels, respectively. Distances are indicated in nucleotides. The corresponding total coverages
756 are shown in panel E.
757 E. Metagene analysis of PTH-Set1, PTH-Set1ΔRRM2 , Set1-HTP, Set2-HTP and RNAPII
758 (Rpo21-HTP) across mRNAs, in reads per million of RNAPII transcripts. Transcripts are aligned
759 to the TSS (left) or pA site (right).
760 F. Metagene analysis of PTH-Set1, RNAPII-S5P, and Set2-HTP, RNAPII-S2P, RNAPII-T4P
761 enrichment relative to total RNAPII, across mRNAs aligned to their TSS (left) or pA site (right).
762 The relative enrichment was calculated as log2(protein coverage/total-Rpo21-HTP coverage).
763 The enrichment across individual mRNAs is shown in Figure S3E-F for PTH-Set1 and Set2-
764 HTP.
765
766 **Figure 3. Some transcripts show high enrichment for Set1 or Set2 relative to RNAPII.**
767 A-B. PTH-Set1 (A) or Set2-HTP (B) coverage over genomic features (mRNAs, transcripts
768 antisense to mRNAs, intergenic transcripts, SUTs, CUTs, XUTs) is plotted against RNAPII
769 coverage (Rpo21-HTP). The fill color of the points represents the enrichment for Set1 or Set2
770 relative to RNAPII. Some classes of transcripts have been highlighted, as indicated on the right
771 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.

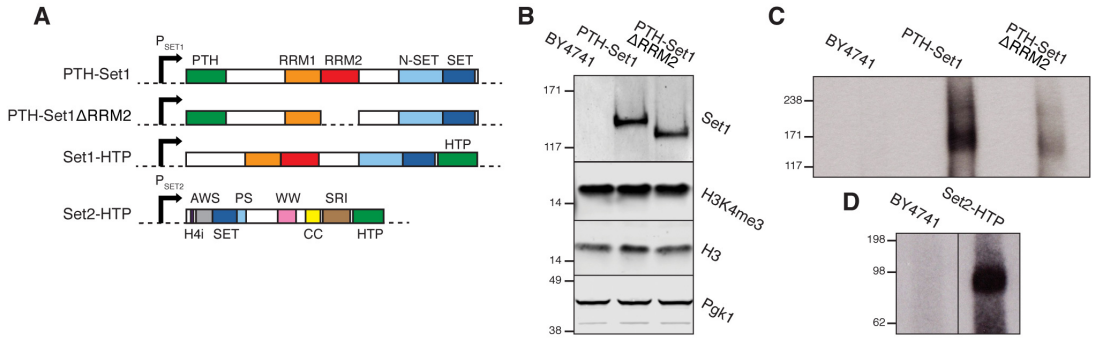
E. rDNA intergenic spacers (IGS) region. rRNA genes appear white on the plots.

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

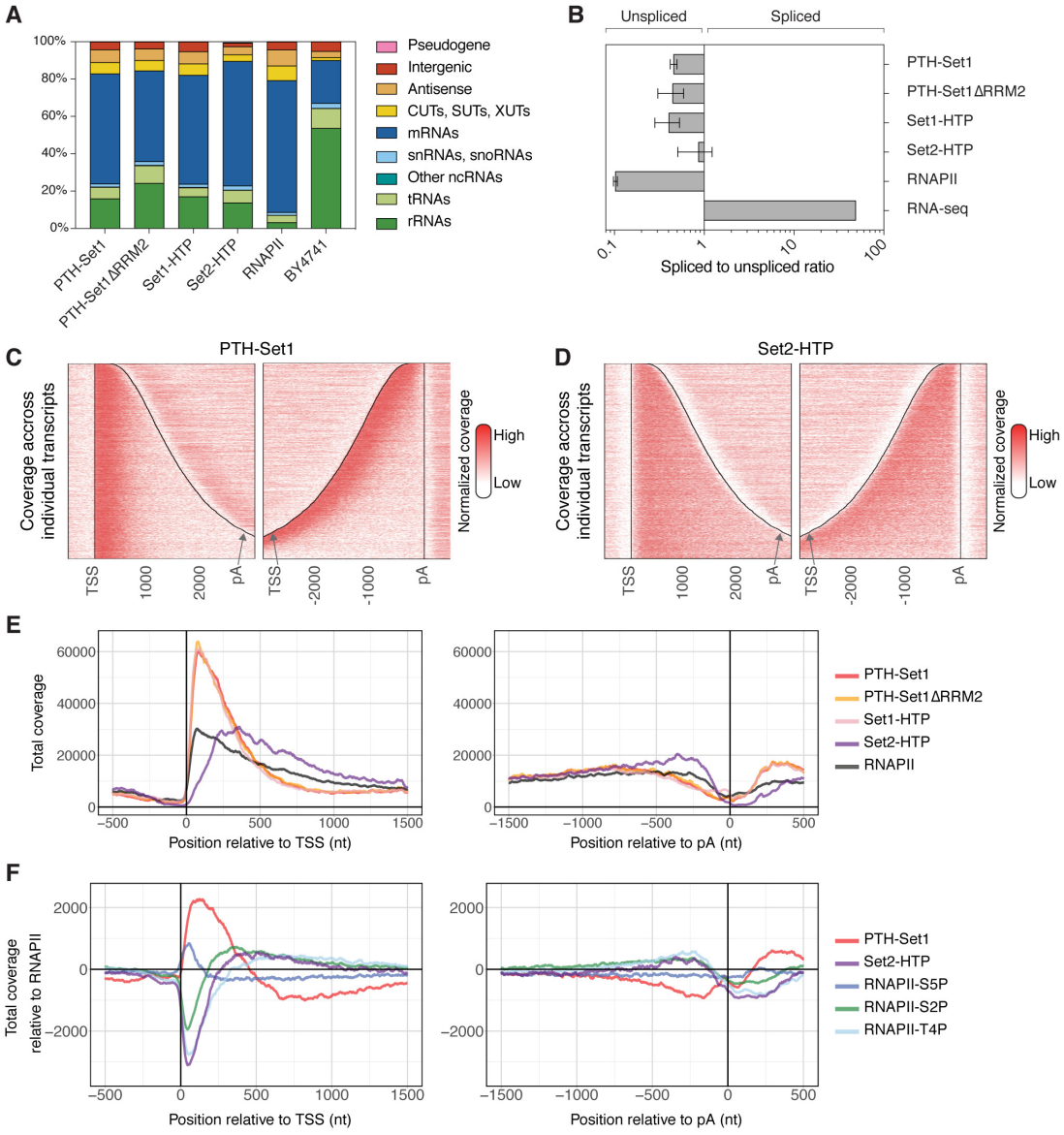
A. Schematic representation of the genes analyzed. The transcription unit is represented, with the coding sequences being thicker. Bars indicate PCR products.

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. * indicates a different signal with a p-value bellow 0.05, calculated with a Student's t-test.

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. * indicates a different signal with a p-value bellow 0.05, calculated with a Student's t-test.

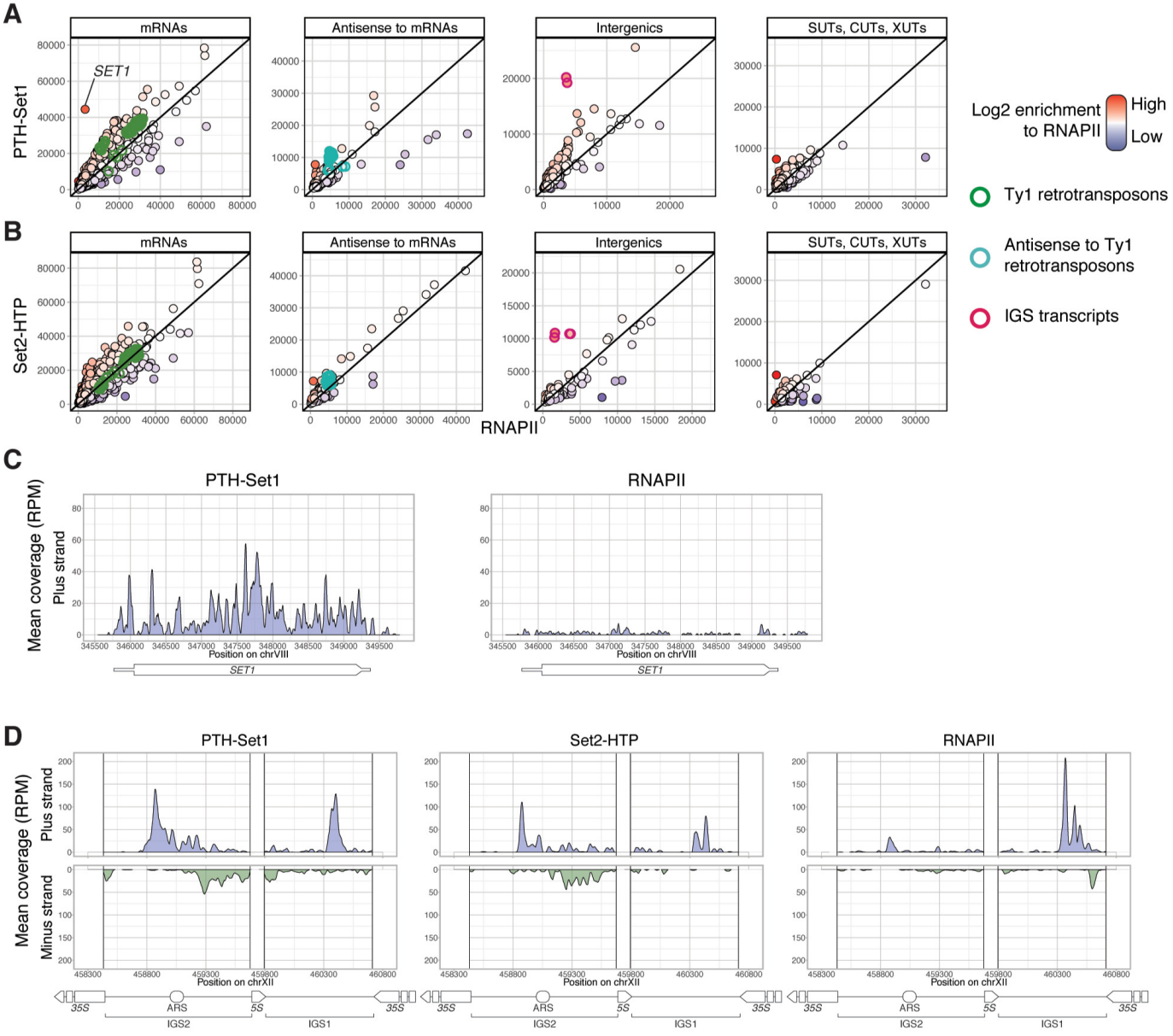


Sayou et al. Fig. 1

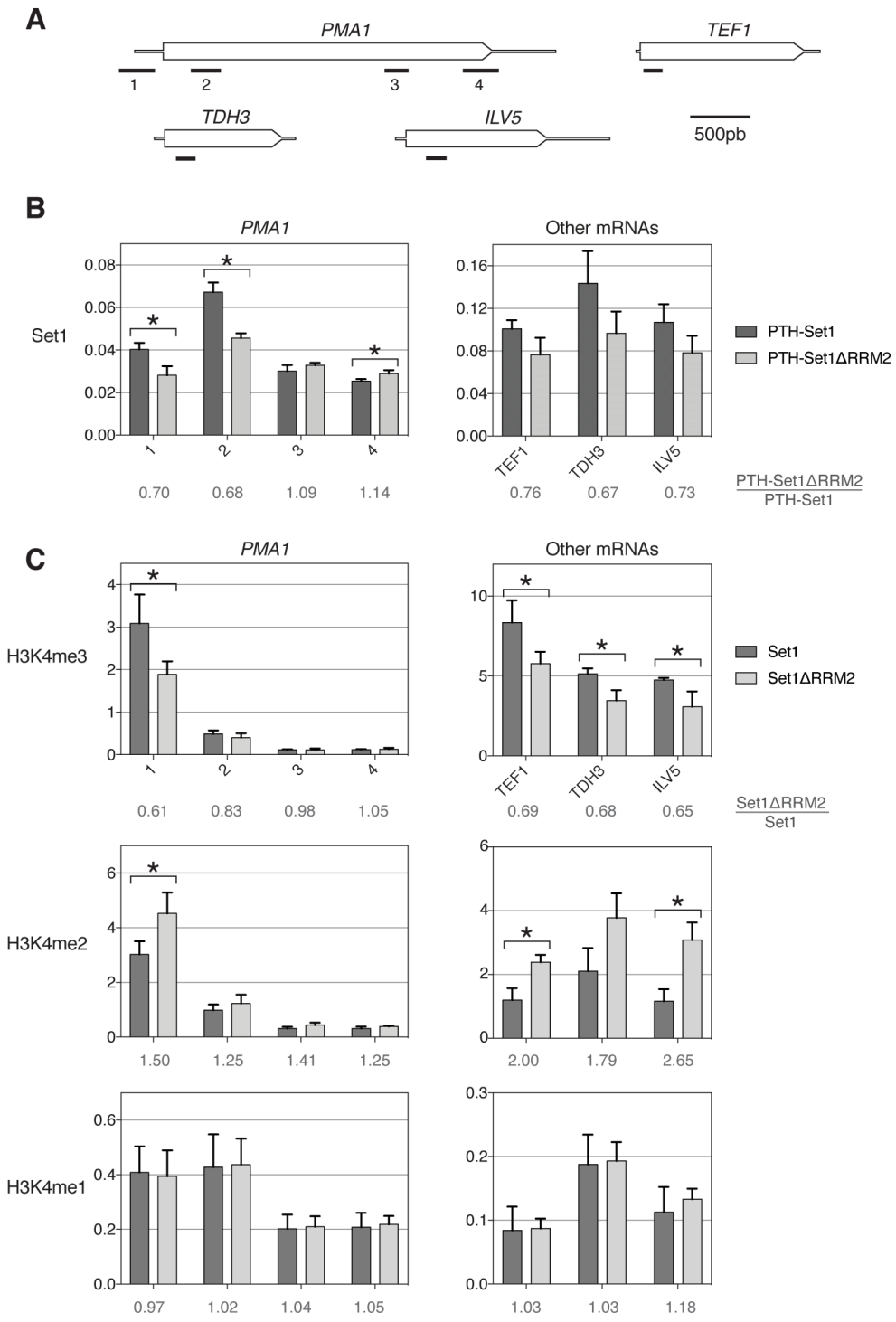


Sayou et al. Fig. 2

Molecular and Cellular Biology



Sayou et al. Fig. 3



Sayou et al Figure 4