An impact of HP1γ on the fidelity of pre-mRNA splicing arises from its ability to bind RNA via intronic repeated sequences.

Christophe Rachez[1*], Rachel Legendre[2], Mickaël Costallat[1], Hugo Varet[2], Jia Yi[1, 3], Etienne Kornobis[2] , Caroline Proux[2] and Christian Muchardt[1*].

[1] Epigenetic Regulation Unit, Institut Pasteur, CNRS UMR 3738, Paris, France.
[2] Transcriptome and Epigenome Platform, Biomics, Citech, and Hub Bioinformatique et Biostatistique, C3BI, Institut Pasteur, Paris, France.
[3] Sorbonne Université, Collège doctoral, F-75005 Paris, France

*email: rachez@pasteur.fr; muchardt@pasteur.fr

ABSTRACT

HP1 proteins are best known as markers of heterochromatin and gene silencing. Yet, they are also RNA-binding proteins and the HP1γ/Cbx3 family member is present on transcribed genes, together with RNA polymerase II, where it regulates co-transcriptional processes such as alternative splicing. To gain insight in the role of the RNA binding activity of HP1γ in transcriptionally active chromatin, we have captured and analyzed RNAs associated with this protein. We find that a prerequisite for RNAs to be bound by HP1γ is the presence of repeated hexameric motifs and coincidentally certain families of transposable elements. HP1γ is part of a mechanism that tethers the nascent RNA to chromatin via these repeated sequences with functional consequences on the outcome of RNA splicing reactions. Our data unveil novel determinants in the relationship between chromatin and co-transcriptional processes.

INTRODUCTION

Transcriptionally active and silenced forms of chromatin can be distinguished in the nucleus by specific subsets of histone modifications. Heterochromatin is characterized mostly by histone modifications associated with repression. Histone 3 trimethylated at Lysines 9 or 27 (H3K9me3 or H3K27me3) mark constitutive or facultative heterochromatin, respectively[1]. The impact of these histone post-translational modifications on the structure and function of chromatin relies in part on their recognition and binding by a number of chromatin-associated proteins. Heterochromatin protein 1 (HP1) is a typical reader of H3K9me3 histone modification, and it is known to be involved in the formation and maintenance of constitutive heterochromatin[2]. The major domains of constitutive heterochromatin are found at centromeres and telomeres, for which the mechanisms of heterochromatin formation are well described and linked to HP1[3,4]. But repeated sequences dispersed thoughout the genome, such as interspersed repeats (SINEs, LINEs) or endogenous retroviruses (ERVs), may also be repressed by heterochromatin-based silencing mechanisms. For example, Alu SINEs are repressed by H3K9trimethylation and HP1[5]. In embryonic stem cells, intact LINEs or endogenous retroviruses, are repressed by heterochromatin silencing[6 7].

There are three mammalian HP1 isoforms, each one having slightly distinct localizations. While HP1α/CBX5 concentrates in foci of pericentromeric heterochromatin, HP1β/CBX1 is more widespread in the nucleus[8,9], and HP1γ/CBX3 is also part of transcriptionally active chromatin[10]. Specific localizations for the different HP1 isoforms suggest that they may have distinct roles in epigenetic silencing. Indeed HP1γ has been found to localize to gene bodies[11], and surprisingly its occupancy positively correlates with gene expression[12-14]. In fact all three HP1 isoforms have been found associated with genomic regions marked with histone modifications that are characteristic of both heterochromatin and active chromatin[15]. We and others have found that, on the locus of an HIV virus integrated in the human genome, two HP1 isoforms have opposite functionalities and recruitment kinetics: while HP1β had a classical silencing effect at the viral promoter, it was, under stress conditions, replaced by HP1γ which localized along the viral genes and behaved as a transcription activator[16]. In human cells, we have also shown that HP1γ favors inclusion of a cluster of alternative exons on the CD44 gene by bridging pre-messenger RNA into chromatin[17]. On a subset of genes, recruitment of HP1γ via repressive chromatin marks contributes to RNA polymerase II pausing and transcriptional termination[18]. In various model systems HP1γ was found to

modulate gene expression programs[19], with consequences on cell lineage specifications[14,20-22]. HP1γ seems more and more to diverge from the classical model of H3K9me3-mediated epigenetic silencing.

HP1 proteins are composed of two globular domains. While the N-terminal chromodomain is involved in H3K9me3 binding, the C-terminal chromo-shadow domain (CSD) is involved in dimerization, and forms as a dimer a binding interface for a large number of PxVxL motif-containing proteins with diverse functionalities[23]. HP1c in *drosophila* interacts with the histone chaperone complex FACT involved in facilitating the recruitment of RNA polymerase II supporting its positive role on transcription[24]. DamID experiments in *drosophila* cells showed only partial overlap between HP1a and H3K9me3[25]. In mouse cells HP1γ localization on transcribed regions is not always correlated with H3K9me3[14]. Even on pericentromeric chromatin which are hot spots of H3K9me3 marks, HP1 requires RNA binding for its localization, via contacts mediated by its hinge (H) domain[26-28], a flexible region separating the two globular domains in HP1. This Hinge domain (H) forms an electrostatic interface involved in mediating contacts with nucleic acids RNA, DNA[29]. Another chromodomain protein LHP1 requires its hinge region for RNA binding and epigenetic repression [30]. This RNA binding property is not unique to HP1 proteins, since it is also found in the PRC2 complex associated with facultative heterochromatin[31].

Earlier experiments identified RNA as a structural component of chromatin[32] and especially RNA from interspersed repetitive elements such as LINE1 repeats may impact chromatin architecture[33]. Conversely, chromatin is thought to function as an RNA-binding matrix thereby contributing to the regulation of co-transcriptional processes[34].

The relationship between HP1γ and RNA on transcriptionally active chromatin is not well understood, particularly concerning the determinants of HP1γ association with RNA and transcription. Here we have analyzed the genome-wide association of HP1γ with RNA by a chromatin-enriched RNA immunoprecipitation (RNAchIP) assay. We have found that HP1γ associates preferentially with intronic regions of pre-messenger RNA in transcribed genes. RNA that was found associated with HP1γ was enriched in specific repetitive motifs. These motifs appeared to be a key feature involved in the HP1γ targeting along pre-messenger RNAs where this protein was found to affect RNA splicing decisions.

RESULTS

HP1γ associates with introns of pre-messenger chromatin-enriched RNA.

To better understand the relationship between HP1γ and RNA, beyond its classical role in heterochromatin formation, we assayed the interactions between HP1γ and RNA on chromatin, genome-wide (RNAchIP; Fig. 1a). For this we used a modification of our previously described strategy of native solubilization of chromatin which leads to a fraction containing chromatin-enriched RNA fragments suitable for immunoprecipitation[17] (Supplementary Fig. 1d). HP1γ -/- (KO cells) mouse embryonic fibroblast (MEF)-derived cell lines, re-complemented with FLAG-tagged HP1γ (HP1γ cells) as previously described were used in these assays[19]. These cells expressed ectopic FLAG- HP1γ at the same level as the endogenous HP1γ in WT MEFs (Supplementary Fig. 1a). As in our previous studies the cells were treated or not with the phorbol ester PMA in order to observe the cellular response to stress on chromatin and transcription. Nuclei were isolated to obtain a chromatin-enriched RNA fraction, (Supplementary Fig. 1a and b). Our procedure was modified to include a limited formaldehyde cross-linking which was used preferentially instead of UV crosslinking to account for the possibility that an HP1γ-interacting protein partner might act as a bridging factor in HP1γ association with RNA. Indeed multiple configurations of HP1 on chromatin may be involved in contacts of HP1γ with RNA, i.e. direct RNA contacts or contacts mediated via one of the numerous HP1 protein partners, including RNA-binding proteins[35]. Nevertheless, stringent immunoprecipitation conditions were used to eliminate the most obvious candidates. None of the HP1γ candidate partners that were tested so far (HP1a, H3, RNA polymerase II) remained detectable in IP (Supplementary Fig. 1b). Our assays revealed by RT-qPCR that chromatin-enriched RNA associates with HP1γ on the Fosl1 locus with a 20- to 25-fold enrichment compared to KO cells. IP in KO cells showed consistently very low to undetectable RNA levels (Supplementary Fig. 1c). Genome-wide analysis on RNAchIP by HP1γ (IP) versus RNA detected in the chromatin fraction (input) was then performed by Illumina sequencing on biological triplicates, then mapped onto the mouse genome. In a first apporach, RNAchIPseq data were computed at the scale of gene bodies by a transcriptome analysis. Normalized read counts per gene bodies (input and IP) revealed that RNA enrichment in HP1γ-IP is correlated with the levels of chromatin-associated RNA, suggesting a positive correlation with gene expression whether or not transcription was stimulated with PMA (Fig. 1b, c). Furthermore, stimulation of cells with PMA increased both input and IP read density on stress-responsive genes such as the *Fosl1* gene (Fig. 1b middle). Our data

complement/corroborate the earlier observation of a correlation between HP1γ recruitment on DNA and gene expression[13]. However, while several genes were highly enriched in HP1γ-IP (see *Gcnt4* or *Fosl1*, Fig. 1b), transcripts from most histone genes were depleted (*Hist1h4a* and *Hist1h3a*, Fig. 1b right), indicating that HP1γ does not equally associate to all transcripts. Since most histone genes are intron-less, we compared the levels of RNA enrichment in HP1γ IP on genes sorted according to their number of exons (Fig. 1d). Transcripts of intron-less genes appeared depleted in HP1γ IP (orange panel 1exon), including intron-less histone genes (red dots), with a 0.6 fold enrichment (Fig. 1e, 1exon), while all intron-containing genes were enriched in HP1γ IP in a similar fashion, independently of the number of exons (blue and dark blue panels, 2-5exons, and 6+-exons, respectively), with an average enrichment of 1.1 fold (Fig. 1e, 2, 3, and 4 exons). Our results indicate that HP1γ may either associate preferentially with RNA of intron-containing genes, or with pre-messenger transcripts. Further analysis of the number of reads over gene features showed their enrichment in RNAchIP over introns and their depletion over exons (Fig. 1f). We then looked specifically at the subset of RNAchIP reads whose sequence alignment is split at the junction between two exons. The number of such split reads was clearly lower in IP relative to input (Fig. 1g). These proportions reveal that HP1γ associates preferentially with pre-messenger RNA transcripts on chromatin, likely prior to the process of splicing. Since, on most genes, splicing is occurring co-transcriptionally on chromatin, our results suggest together that HP1γ associates preferentially with introns of unspliced pre-messenger transcripts, likely prior to maturation by splicing.

HP1γ-associated RNA is enriched in CACACA and GAGAGA motifs.

Within individual genes, the distribution of HP1γ-associated RNA appeared to be not uniform along the gene body (Fig. 1B). To detect regions of local RNA enrichment by HP1-IP, we searched for peaks of IP versus input RNA on all triplicates. We identified 10013 and 2603 peaks, by merging peaks that were conserved in at least two of the three triplicate samples, in unstimulated and in PMA-stimulated samples, respectively (Fig. 2A). Since our crosslinking procedure (i.e. formaldehyde) allows both indirect and direct HP1/RNA contacts, we looked at the chromatin context in which these peaks were detected. For this purpose these peaks were colocalized with several chromatin features obtained by ChIP in MEF cells selected from the ENCODE database, and compared with a list of our peaks whose genomic location was randomized among genes. Interestingly, our RNAchIP peaks colocalized most

significantly with peaks of both RNA polymerase II and H3K4me3, two features associated with active transcription, while no significant colocalization was observed with H3K9me3 and H3K27me3, the two major histone modifications associated with heterochromatin (Fig. 2B). These data clearly link intragenic RNA-associated HP1γ with sites of active transcription. We then wondered if sequence specific motifs could be found among these peaks. Peak-motif analysis via RSAT on stranded sequences of RNA peaks revealed a preferential enrichment in CACACA motifs (e-val. 4.8 e-88), and to a lesser extend in GAGAGA motifs (e-val. 1.8 e-10) on more than half of our peaks (57%; Fig. 2c). Beyond our peaks, RNA was also found globally enriched in RNAchIP around all intragenic CACACA motifs at expressed genes, as seen by the increased average distribution of the reads from IP compared to input RNA around these motifs (Fig. 2d left), and their occurrence among a large number of intragenic loci centered around the motif (Fig. 2d right). Noticeably, we did not find any enrichment versus input over a TGTGTG motif, which corresponds to the CACACA motif in antisense orientation respective to the corresponding gene (Fig. 2d). Despite the fact that HP1 is known for its capacity to associate with both RNA and DNA itself on chromatin[29], our results suggest that the CACACA motifs driving HP1γ association with RNA are carried by the RNA itself and not by the underlying double stranded DNA because only RNA containing motifs matching the orientation of the overlapping transcript are found enriched in IP. As a representative example, an RNA peak within the Ppp3ca gene shows an overlapping distribution of both CACACA and GAGAGA motifs around the peak (Fig. 2e). In order to test whether or not HP1 can establish direct contacts with RNA in this context, the sequence surrounding one of these motifs was used *in vitro* as an RNA probe, and compared to a scrambled RNA devoid of any CA or GA (Fig. 2e, bottom). HP1γ direct interaction with RNA was tested *in vitro* in a gel mobility shift assay between a bacterially expressed, purified GST-HP1γ and each RNA probe. The Hinge domain of HP1 has been identified as an RNA and DNA binding domain ([26,28]; Supplementary Fig. 2a). Since this domain alone has a better capacity to bind RNA *in vitro* than in the context of the full-length protein ([27]; Supplementary Fig. 2c), both HP1γ protein and its Hinge domain were tested. Both fusion-proteins showed a clear binding to the CACACA-containing probe over the control scrambled RNA (Fig. 2f). Similar results have been obtained with an RNA probe containing the GAGAGA motif (Supplementary Fig. 2d). HP1γ has therefore the capacity to directly interact with RNA, suggesting that the RNA enrichment seen in RNAchIP is likely due to direct HP1γ/RNA associations. Together, our results suggest that HP1γ may be targeted onto chromatin, not

only via the H3K9 trimethylation mark, but also by chromatin-associated RNA at specific positions enriched in CACACA and GAGAGA motifs.

HP1γ-associated RNA is enriched in SINE repeat motifs.

Because heterochromatin-based silencing mechanisms may occur within active chromatin on repeated sequences such as interspersed repeats (SINEs, LINEs) or endogenous retroviruses (LTRs) [6,7]. We therefore asked whether HP1γ-associated RNA could be enriched over the subset of repeat elements that are present within gene bodies. We tested the three major classes of mouse repeated elements: SINEs, LINEs, and LTRs. Interestingly, average read profiles from RNAchIP allowed us to detect a modest enrichment in RNA-IP compared to the levels of RNA in input, essentially on the SINE repeats (Fig. 3a). Enrichment was most obvious on the B4 SINE family, and also more pronounced on the B3 and RSINE families, while undetected on the B1 and B2 SINEs (Fig. 3b). The highest RNA enrichment in IP over B3, B4 and RSINEs correlated well with their propensity to contain at least one CACACA motif in their sequence (Fig. 3c). Indeed, average read profiles from RNAchIP identified an increased density in IP when CACACA-containing B4 SINEs were considered instead of all B4 SINEs (Fig. 3d, compare with Fig. 3b B4 SINE panel). Noticeably, only the retroviral repeat loci that were in the same orientation as their overlapping gene were found enriched, not the repeat loci in antisense orientation (Fig. 3a, 3b). It should be noted that the reverse-complement sequence of the B4 SINEs is not enriched in CACACA motifs (Supplementary Fig. 3). Our results are consistent with a binding of HP1γ to the pre-messenger transcript, not the SINE transcript itself which is anyway not expected to be expressed, and this binding is favored by the presence of a CACACA motif within the SINE sequence. Interestingly, the sequence of the RNAchIP peak in the Ppp3ca gene, illustrated in Fig. 2e, also contains a SINE (Fig. 3e). Indeed, similarly to its binding on CACACA-containing RNA, HP1γ has the capacity establish direct contact with an RNA probe designed around the SINE sequence in a Northwestern assay (Fig. 3f). This association was drastically impaired with a probe devoid of its CACACA motifs by deletion of its 3' part (Fig. 3f). Our results suggest that the intragenic SINE repeats, concomitantly to the presence of a CACACA motif, constitute a targeting motif for the association of HP1γ with pre-messenger transcripts.

The density of the CACACA motif affects HP1γ association with transcripts on chromatin.

In order to address the impact of the CACACA motif on HP1γ association with RNA at the scale of the gene body, genes were sorted according to their global CACACA density into two categories of genes with the highest and lowest CACACA density (Fig. 4a and b, and Fig. 4c and d, respectively). These two categories clearly showed opposite profiles of average read density in RNAchIP, genes containing the highest CACACA density being enriched in RNA IP while the genes having the lowest density being depleted, affecting both pre-messenger RNAs (compare Fig. 4a and c), and mature mRNA on chromatin (compare Fig. 4b and d). These two categories highlight the positive correlation between the CACACA motif density and the enrichment in RNA IP over whole gene bodies. The association of RNA with chromatin via a contribution of HP1γ on the highest CACACA-containing genes may therefore be the sum of local RNA contact points with HP1γ distributed thoughout the gene, as in the example of the *Tfap2p* gene (Fig. 4e), where the high CACACA density results in a global RNA IP enrichment, rather than in the appearance of a localized peak. The global levels of RNA enrichment in IP were also clearly lost in the metagene profiles versus pre-mRNA profiles for both categories of genes (compare Fig. 4b and a, or Fig. 4d and c). This clearly suggests that HP1γ is released from the mature transcripts, and it is consistent with the preferential association of HP1γ with introns that we have observed earlier (Fig. 1). Furthermore, intragenic SINE repeats, but not LINEs, are enriched in introns versus exons, and appear to be distributed along the gene body (compare metagene versus meta cDNA profiles, Fig. 4f and g). The parallel that we have established between the localization of CACACA motifs and SINE repeats, and the appearance of peaks in RNA IP, can be extended to the scale of the gene body when looking at the distribution of these motifs along the genes.

HP1γ has an impact on the fidelity of RNA splicing which is correlated with the presence of CACACA and B4 SINE motifs.

We have previously demonstrated the role of HP1γ in the relationship between chromatin and the regulation of alternative splicing[17]. We therefore explored the impact of CACACA and SINE motifs on the functionality of HP1γ in this context. For this purpose we analyzed the density of SINE repeats in exons referenced as alternatively spliced in the mouse genome (cassette exons in the UCSC Alt Events database) and found these exons to contain more SINEs than all exons in mouse RefSeq database (Fig. 5a). In respect to SINE density, the cassette exons tend towards having similar properties as introns. Furthermore the number of RNAchIP reads matching cassette exons give an enrichment profile versus input which is

similar to the one seen on introns, but not on all exons (Fig. 5b). The intron-like properties of cassette exons relative to SINE and RNAchIP read densities, suggest that they are more prone to be targeted by HP1γ for RNA binding. This increases the likelihood of a chromatin-dependent regulation of transcript splicing decisions at these alternative exons. In order to analyze the impact of HP1γ on the outcome of transcripts we performed a transcriptome analysis comparing the HP1γ-expressing (HP1γ) and HP1γ-null (KO) cells lines that we used in our RNAchIPseq. These two cell lines showed only minor differential transcript expression levels (Supplementary Fig. 5a and b), likely due to compensatory effect by the two other HP1 isoforms in accordance with another recent study[12]. We therefore focused on the impact of HP1γ on RNA splicing. For this purpose we compared the transcriptomes of HP1γ and KO cell lines with respect to splicing decisions by calculating the variations in the splicing events between the two cell lines with the MAJIQ algorithm. This approach led us to identify a number of differentially spliced exon-exon junctions between HP1γ and KO cell lines, i.e. RNA splicing junctions being formed at two different locations in the two cell lines. We detected 174 and 81 high confidence differential splicing events in unstimulated or PMA-treated cells, respectively (Supplementary Fig. 5c). Out of 81 high confidence junctions detected in the PMA-treated samples, 41 junctions analyzed in the HP1γ samples were known splicing events that are referenced in ensembl database. These differential events in HP1γ cells were modified into two main types in KO cells (Fig. 5d; Supplementary Fig. 5d, e). First, 16 events were junctions leading to the formation of referenced cassette exons, meaning alternative exons that were included or excluded in the mature transcript (Cassette exons; Fig. 5d; Supplementary Fig. 5d). Second, 21 events were cassette exon junctions in HP1γ cells which were changed, in KO cells, into *de novo* cryptic splice sites, i.e. not previously referenced, that were observed within introns leading to partial intron retention (Cryptic splice sites; Fig. 5d; Supplementary Fig. 5e). These data suggest that HP1γ contributes to both the regulation of alternatively spliced cassette exons, and to the fidelity of splicing as illustrated by the representative example on the *Cnpy3* gene (Fig. 5c). The aberrant character of the *de novo* junctions in KO cells is also highlighted by the significant and systematic shorter length of these junctions in KO versus HP1γ cells, suggesting that HP1γ may mask regulatory splicing motifs which become accessible to regulators of the spliceosomes in KO cells. Counting the enrichment in the number of RNAchIP reads on the introns surrounding the 81 exons that were differentially spliced with HP1γ revealed much higher levels of RNA in both input and IP samples, relative to a matching list of introns chosen randomly (Fig. 5e). This

aspect is reminiscent of a role of HP1γ in stabilizing the association of the pre-mRNA with the chromatin, that we have previously described[17]. In the *de novo* splicing event observed in the absence of HP1γ on the *Cnpy3* gene which is illustrated (Fig. 5d), not only an enrichment in RNAchIP is observed locally on the corresponding intron, but also a high density of CACACA motifs which is concomitant to a high density of endogenous retroviral repeats, including a B4 SINE at the *de novo* splice site. Beyond the *Cnpy3* gene locus, introns surrounding the exons that are aberrantly spliced in the absence of HP1γ contain significantly higher proportions of SINE repeats (in average more than one per intron, Fig. 5f) than on a random matching list of introns. Our data collectively suggest that CACACA motifs and the coincidentally associated B4 SINE repeats drive the targeting of HP1γ onto premature RNA allowing an HP1γ-dependent regulation of splicing events occurring on chromatin.

DISCUSSION

Besides its classical functionality in H3K9me3-mediated silencing of heterochromatin, HP1γ appears to be also associated with transcriptional regulation at actively transcribed genes. Here we have used a modification of our native chromatin immunoprecipitation assay as a strategy to identify chromatin-enriched RNA associated with HP1γ, in an attempt to decipher the mechanisms involved in HP1γ functionality. We have found that the association of HP1γ with pre-messenger RNA on chromatin is based on its capacity to directly bind RNA molecules bearing preferentially a consensus CACACA or GAGAGA motif. The Hinge domain which contributes to RNA binding in HP1 proteins is not classified in any of the common families of RNA binding domains. This Hinge region is a rather unstructured peptidic segment between the two globular domains in HP1, and containing many positively charged basic residues, Lysines and Arginines, which create an electrostatic interface[29]. Considering the recent concept that heterochromatin-mediated gene silencing may occur in part through liquid-liquid phase separation involving association of HP1α with DNA[36,37], these physicochemical properties may very well be at play in the process of HP1γ interaction with RNA via its unstructured domain, creating local foci of chromatin condensation.

The two RNA consensus motifs we have identified were demonstrated to greatly enhance the direct binding of RNA by purified HP1γ in an *in vitro* assay. However an exact match with

both consensus motifs was found in no more than half of the RNAchIP peaks. One reason for these differences is that a subset of peaks contains motifs bearing a sequence that is slightly divergent from the consensus. It suggests that HP1γ may still recognize an imperfect consensus, such as the CACAC motif present in the RNA probe used in our gel mobility shift assay. Another reason is linked with the use of formaldehyde as a crosslinking reagent in our RNAchIP procedure. We have opted for this approach instead UV-crosslinking used in CLIP assay, to allow protein:protein associations and indirect RNA contacts, and take into account the wide spectrum of configurations of HP1γ involved in its functionalities, and the large number of known HP1-associated partners[35]. This way, a number of peaks may have been detected through indirect interactions. These peaks may bear different consensus motifs. Indeed, many factors associated with HP1 are known RNA binding proteins. For example hnRNP-U interacts with HP1α[38]. But in fact, the CACACA motif has also been identified as a direct RNA binding site for hnRNP-L, an abundant nuclear protein involved in the regulation of splicing[39]. Intronic CA-repeat and CA-rich elements were found to be regulatory motifs of mammalian alternative splicing via binding of hnRNP-L[40]. RNA association with hnRNP-L was found preferably in introns and 3' UTR, while we have found HP1γ-associated RNA mostly within introns. These differences do not exclude the possibility of common RNA binding sites for the two proteins. This possibility of common targets on RNA suggests that there could be cooperativity or interference between HP1γ and hnRNP-L proteins, especially on their regulatory role in splicing. Likewise, there is a strong similarity between the GAGAGA consensus motif for HP1γ and a consensus RNA binding motif found for the splicing factor 9G8[41]. Given the fact that HP1γ has two consensus motifs for RNA binding, we could envision that HP1γ may act by interfering or cooperating with multiple positive and negative regulatory splicing factors, via their own RNA targeting motif within chromatin.

In humans, the short interspersed Alu repeats may be the source of novel splicing sites because of their sequence. In this case an intragenic Alu in antisense orientation relative to the transcript introduces a new 5’ splicing site[42]. These Alu become the target of new exon-exon junctions, a process at the origin of the appearance of alternative exons in the genome. The SINE repeats in our model have been found, in many of our examples, to be located in the vicinity of differentially regulated exon-exon junctions but not at the site of a novel junction, in disfavor of the Alu model. Our results fit well with the model we have raised on the human CD44 gene where HP1γ contributes to alternative splicing by maintaining pre-messenger

RNA within chromatin thereby facilitating the occurrence of co-transcriptional splicing[17]. Clearly, a large majority of novel splicing decisions (cryptic splice sites) that we have observed in the absence of HP1γ were consistently shorter junctions than in the presence of HP1γ. This raises the possibility that HP1γ may be masking binding sites of splicing regulators, not splice sites themselves, thereby preventing them from being aberrantly used by the spliceosome, hence reinforcing the accuracy of splicing decisions.

Our results add a new key element to the model of chromatin-based regulation of pre-mRNA splicing by suggesting that HP1 proteins may compete or synergize with splicing regulatory factors such as SR proteins and hnRNPs. B4 SINE repeats, because they are enriched in CACACA motifs appear to be coincidentally good markers of the genome-wide distribution of HP1γ RNA-binding sites in the regulation of splicing decisions. We cannot exclude that under certain conditions HP1 may still contribute to maintaining interspersed repeats silenced, as does another heterochromatin complex, PRC2, on B2 SINEs[43]. The function of HP1 proteins on the heterochromatin-based silencing of the SINE repeat elements within gene bodies has shifted in the case of HP1γ towards regulation of alternative splicing at the locations of SINE repeats. Because of the diversity of HP1γ target genes and chromatin configurations a wide spectrum of functional effects is observed, even from the specific point of view of splicing. The functional consequences of the CACACA-dependent targeting of HP1γ to RNA may have to be explored in specific cellular model systems such as cell lineage specification.

MATERIALS AND METHODS

Cell lines and tissue culture

Mouse embryonic fibroblast (MEF)-derived cell lines were obtained exactly as previously described[19]. Briefly, HP1γ-expressing or KO cells were obtained by stable re-complementation of immortalized MEF HP1γ -/- cells by retroviral transduction using a retroviral vector carrying HP1γ cDNA, or an empty vector, respectively. Cells were grown in DMEM (Invitrogen) supplemented with 10% (v/v) fetal calf serum and 100 U.ml-1 penicillin-streptomycin. PMA-treated samples were obtained by addition of 100nM of Phorbol 12-Myristate 13-Acetate (PMA) in DMSO for 30 min on nearly confluent cells.

Direct *in vitro* protein-RNA binding assay - Gel mobility shift assay

Assay was performed with bacterially expressed, purified GST fusion proteins, as depicted in Supplementary Fig. 2b, prepared as previously described[16]. Between 0.2 and 0.8 nmoles of GST-fusion proteins were incubated with 1 pmole of Cy3-labelled RNA oligonucleotide probes on ice for 20 min at 4°C in EMSA buffer (10mM Tris-HCl (pH 8.0), 50mM NaCl, 10% glycerol (w/v), 0.01% NP40, 0.1mg/ml BSA). The reaction was resolved by gel electrophoresis at 150 V for 20 min at +4°C, on a 5% native polyacrylamide gel (37.5:1) in 0.5x TBE buffer (45mM Tris-borate, 1mM EDTA). The gel was then immediately scanned on a Typhoon FLA 9000 (GE Healthcare), and subsequently stained with Coomassie blue R-250 (Sigma).

Direct *in vitro* protein-RNA binding assay - North-western blot binding assay

Assay was performed essentially as described previously[28]. Briefly, bacterially expressed, purified GST-fusion proteins as above were separated by SDS-PAGE, transferred onto a nitrocellulose membrane (Biorad), renatured in PBS containing 5% of bovine serum albumin, and then hybridized for 1h at room temperature with biotinylated RNA probes that had been *in vitro* transcribed in the presence of biotin-16-UTP (Sigma) with T7 RNA polymerase (NEB) following the manufacturer's instructions. After two washes and hybridization of Cy3-streptavidin (BioLegend, Inc.), membranes were scanned on a Typhoon FLA 9000 (GE Healthcare), and subsequently stained with Ponceau S (Sigma).

Nuclei isolation and crosslinking

40x10+6 cells (2x 10cm^2 plates) treated or not with 100nM PMA for 30 min were washed directly on their tissue culture plate twice with ice cold Phosphate-Buffered Saline (PBS). All subsequent steps were performed at 4°C, unless otherwise specified. Cell were allowed to swell on ice for 5 min in 5ml of ice cold swelling buffer: (10mMTris-HCl pH7.5, 2mM MgCl2, 3mM CaCl2, supplemented before use with: 1x antiprotease (Roche), 0.5mM Na3VO4, 20mM b-Glycerophosphate, 80U/ml RNAsin (Promega), 0.1mM DTT). Cells were removed from the plate with a plastic cell scraper, transferred to a 15ml conical, and pelleted for 5 min at 4°C at 1600rpm. Cells were resuspended in 0.7ml of swelling buffer supplemented with 10% glycerol and 0.5% Igepal CA630 (Sigma), and gently pipetted up and down 15 times using a p1000 tip. Nuclei were pelleted for 5 min at 4°C at 2500rpm, and washed once in 1ml swelling buffer supplemented with 10% glycerol. Nuclei were then fixed

with 0.3% formaldehyde in 300ul swelling buffer supplemented with 10% glycerol for 10 min at room temperature. Crosslinking was quenched during 10 min with 50mM glycine. Nuclei were pelleted for 5 min at 4°C at 3400rpm, and then washed once in 1ml swelling buffer.

Chromatin-enriched RNA immunoprecipitation assay (RNAchIP)

Crosslinked nuclei were extracted in 0.8ml modified RIPA lysis buffer (50mM Tris , 150mM NaCl, 0.5% sodium deoxycholate, 0.2% SDS, 1% NP-40, supplemented with: 1x antiprotease (Roche), RNasin (Promega), 0.5mM DTT). Cell suspension was sonicated using a Diagenode Bioruptor for 4 times 20s cycles at High amplitude. 7µl of Turbo DNase (Ambion, AM2238) and 7µl MgCl2 1M were added to sonicated material, incubated at 37°C for 10 min, and spun down at 12,000rpm for 10 min at 4°C. Ten percent of solubilized chromatin lysate was kept as input. The remaining volume was mixed with 350µl PBS and 200µl anti-FLAG magnetic beads suspension (Sigma, M8823) that were previously blocked in PBS containing 0,1% BSA, 0.5% Triton X-100 and 0.1% polyvinylpyrrolidone-40 (Sigma), and incubated at 4 °C for 2 h on a rotating wheel. Beads were then washed once in low salt wash buffer (1xPBS, 0.1% SDS, 0.5% NP-40), twice in high salt wash buffer (5x PBS, 0.1% SDS, 0.5% NP-40), and once more in low salt wash buffer. Beads were then eluted twice10 min in 250µl elution buffer (150ng/µl 3xFLAG peptide (Sigma)) in low salt wash buffer supplemented with RNasin). Eluates were combined for a total of 500µl. Eluates were then adjusted to 200mM NaCl and 10mM EDTA, and incubated with 10µg of Proteinase K at 50°C for 45 min, then placed at 65°C for 2h to reverse crosslinking. Samples were then subjected to phenol:chloroform extraction under acidic conditions followed with ethanol precipitation with Glycoblue (Ambion) as a carrier. Nucleic acid Pellets were then washed once in 75% ethanol, air-dried briefly, and resuspended in 30µl of RNase-free water for DNase treatment 20 min at 37°C, followed by RT-qPCR or library preparation.

Reverse transcription and qPCR (RT-qPCR)

Reverse transcription was carried out with SuperScript III (Invitrogen) and random hexanucleotides for 1h at 50°C on 1µg RNA, quantified with a nanodrop (Thermo Scientific). Real-time quantitative PCR (qPCR) was carried out on a Stratagene Mx3005p with Brilliant III SYBR Green kits (Stratagene) according to the manufacturer's instructions. Primer sequences are: Fos-in1-F, 5'-TGGAGACCACGAAGTGTTGGGAT-3'; Fos-in1-R, 5'-ATGGACACCTGCAACCTCTCAAGT-3'.

Total RNA preparation and sequencing for transcriptome analysis

Total RNA was prepared out of HP1γ and KO cells by guanidinium thiocyanate-phenol-chloroform extraction according to the method of Chomczynski and Sacchi[44], followed by proteinase K and DNAse treatments as described above. Total RNA library preparation and sequencing were performed on DNase-treated RNA samples by Novogene Co., Ltd, as a lncRNA sequencing service, including lncRNA directional library preparation with rRNA depletion (Ribo-Zero Magnetic Kit), quantitation, pooling and PE 150 sequencing (30G raw data-100M raw reads/sample) on Illumina HiSeq 2500 platform.

RNAchIP Library preparation

RNA quality and yield were assessed by the RNA integrity number (RIN) algorithm, using the 2100 Bioanalyzer. Directional libraries were prepared using the Smarter Stranded Total RNA-Seq kit-Pico Input Mammalian kit following the manufacturer's instructions (Clontech, 635005). The quality of all libraries was verified with the DNA-1000 kit (Agilent) on a 2100 Bioanalyzer and quantification was performed with Quant-It assays on a Qubit 3.0 fluorometer (Invitrogen). Clusters were generated for the resulting libraries, with the Illumina HiSeq SR Cluster Kit v4 reagents. Sequencing was performed using the Illumina HiSeq 2500 system and HiSeq SBS kit v4 reagents. Runs were carried out over 65 cycles, including seven indexing cycles, to obtain 65-bp single-end reads. Sequencing data were then processed with the Illumina Pipeline software, Casava (v.1.9).

Bioinformatics analysis

Bioinformatics analysis of the RNAchIP-seq was performed using the RNA-seq pipeline from Sequana[45]. Reads were cleaned of adapter sequences and low-quality sequences using cutadapt (v.1.11)[46].Only sequences at least 25 nt in length were considered for further analysis. STAR (v.2.5.0a)[47] (parameters: --outFilterMultimapNmax 30 --outSAMmultNmax 1 --outMultimapperOrder Random) was used for alignment on the reference genome (Mus musculus mm9 from UCSC). Genes were counted using featureCounts (v.1.4.6-p3)[48] from Subreads package (parameters: -t CDS -g ID -s 1). MACS2 (v.2.1.0)[49] was used to call HP1-binding peaks on RNA-ChIP data (parameters: --nomodel --extsize=150 -q 0.1). Bamcoverage from Deeptools[50] was used to produce normalized BigWig files to 1X. Finally, Bedtools (v.2.25.0)[51] closedBed (parameters: -D ref -mdb each) was used to annotate each peak from all conditions with related and public ChIP-seq data from Gene Expression

Omnibus database (H3k4me3, RNA pol II, H3K9me3, H3K27me3; GEO sample accessions: GSM769029, GSM918761, GSM2339533, GSM946547, respectively).

Statistical analysis

Each count dataset was analyzed using R (v.3.4.1) and the Bioconductor package DESeq2 (v.1.16.0)[52] to test for the differential gene expression (DGE). The normalization and dispersion estimation were performed with DESeq2 using the default parameters and statistical tests for differential expression were performed applying the independent filtering algorithm. A generalized linear model was set in order to test for the differential expression between the biological conditions. For RNAchIP-seq, triplicates IP samples were compared to input samples in each PMA-treated and unstimulated samples. For RNA-seq, triplicate HP1 samples were compared to KO samples in each PMA-treated and unstimulated cells. For each pairwise comparison, raw p-values were adjusted for multiple testing according to the Benjamini and Hochberg (BH) procedure[53] and genes with an adjusted p-value lower than 0.05 were considered differentially expressed.

Motif discovery by RSAT was performed online on the RSAT peak motif search interface (http://rsat.sb-roscoff.fr/)[54]. First, sequences of the merged RNAchIPseq peaks from PMA-treated and unstimulated conditions were obtained in fasta format with strand orientation based on the Mus_musculus.NCBIM37.67.dna.toplevel.fa reference database. The fasta files were then used as queries for RSAT peak-motif disccovery (parameters: peak-motifs -v 1 -markov auto -disco oligos,positions -nmotifs 5 -minol 6 -maxol 8 -no_merge_lengths -1str -origin center). The e-value associated with each discovered motif represents the expected number of patterns which would be returned at random for a given probability of occurrence (P-value).

Analysis of differential splicing events in the transcriptome by MAJIQ

Alternative splicing events occuring in the transcriptome between different conditions were analyzed by the MAJIQ computational framework (v.2.1-179b437)[55] with default parameters. For this purpose, the transcriptome was aligned with STAR (parameters: --outFilterMismatchNmax 1 --outMultimapperOrder Random --outSAMmultNmax 1 --outFilterMultimapNmax 30) on the mouse GRCm38/mm10 unmasked genome. The definition of referenced and *de novo* junctions were based on the mouse GRCm38/mm10 annotation database. Analyses were performed by pairwise comparisons between

experimental conditions, namely HP1γ versus KO, and HP1γ +PMA versus KO+PMA, for each triplicate. Local splicing variations (LSVs) were detected by the software and their relative abundance (PSI) quantified for each condition, leading to a relative change (dPSI) between HP1γ and KO for each junction involved in the LSV. The default threshold of change of |dPSI| >= 0.2 (20%) between conditions was used.

Data availability

RNAchIP-seq and RNA-seq data have been deposited in the NCBI Gene Expression Omnibus database under GEO accession number GSE133267.

ACKNOWLEDGEMENTS

We are grateful to all members of the Epigenetic Regulation unit for helpful discussions, Madeleine Moscatelli and Cynthia Bezier for helpful preliminary experiments, Catherine Bodin for technical assistance and Edith Ollivier for administrative assistance.
This work was supported by Institut National de la Santé et la Recherche Médicale (Inserm; C.R.), Centre National de la Recherche Scientifique (CNRS; C.M.), with grants from Agence Nationale de la Recherche (ANR-11-BSV8–0013) and REVIVE—Investissement d'Avenir (to E.K. and C.M.). J.Y. is part of the Pasteur - Paris University (PPU) International PhD Program. This program has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 665807

AUTHOR CONTRIBUTIONS

C.R. conceived, carried out and analyzed the experiments and performed some bioinformatics analyses. R.L. performed bioinformatics analyses for the RNAchIPseq and transcriptome. M.C. performed bioinformatics analyses with MAJIQ. H.V. performed statistical analyses. J.Y. carried out the RNA preparations for the transcriptome. E.K. contributed to bioinformatics analysis of RNAchIP peak colocalizations. C.R. and C.M. contributed to the overall orientations of the project. C.R. wrote the manuscript and all authors were involved in revising it critically for important intellectual content.

COMPETING INTERESTS

Authors declare no competing interests.

REFERENCES

1. Zhou, V.W., Goren, A. & Bernstein, B.E. Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* **12**, 7-18 (2011).
2. Allshire, R.C. & Madhani, H.D. Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol* **19**, 229-244 (2018).
3. Almouzni, G. & Probst, A.V. Heterochromatin maintenance and establishment: lessons from the mouse pericentromere. *Nucleus* **2**, 332-8 (2011).
4. Garcia-Cao, M., O'Sullivan, R., Peters, A.H., Jenuwein, T. & Blasco, M.A. Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. *Nat Genet* **36**, 94-9 (2004).
5. Varshney, D. et al. SINE transcription by RNA polymerase III is suppressed by histone methylation but not by DNA methylation. *Nat Commun* **6**, 6569 (2015).
6. Leung, D.C. & Lorincz, M.C. Silencing of endogenous retroviruses: when and why do histone marks predominate? *Trends Biochem Sci* **37**, 127-33 (2012).
7. Bulut-Karslioglu, A. et al. Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol Cell* **55**, 277-90 (2014).
8. Minc, E., Courvalin, J.C. & Buendia, B. HP1gamma associates with euchromatin and heterochromatin in mammalian nuclei and chromosomes. *Cytogenet Cell Genet* **90**, 279-84 (2000).
9. Dialynas, G.K. et al. Plasticity of HP1 proteins in mammalian cells. *J Cell Sci* **120**, 3415-24 (2007).
10. Kwon, S.H. & Workman, J.L. HP1c casts light on dark matter. *Cell Cycle* **10**, 625-30 (2011).
11. Vakoc, C.R., Mandat, S.A., Olenchock, B.A. & Blobel, G.A. Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. *Mol Cell* **19**, 381-91 (2005).

12. Ostapcuk, V. et al. Activity-dependent neuroprotective protein recruits HP1 and CHD4 to control lineage-specifying genes. *Nature* **557**, 739-743 (2018).
13. Smallwood, A. et al. CBX3 regulates efficient RNA processing genome-wide. *Genome Res* **22**, 1426-36 (2012).
14. Sridharan, R. et al. Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1gamma in reprogramming to pluripotency. *Nat Cell Biol* **15**, 872-82 (2013).
15. Ji, X. et al. Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. *Proc Natl Acad Sci U S A* **112**, 3841-6 (2015).
16. Mateescu, B., Bourachot, B., Rachez, C., Ogryzko, V. & Muchardt, C. Regulation of an inducible promoter by an HP1beta-HP1gamma switch. *EMBO Rep* **9**, 267-72 (2008).
17. Saint-Andre, V., Batsche, E., Rachez, C. & Muchardt, C. Histone H3 lysine 9 trimethylation and HP1gamma favor inclusion of alternative exons. *Nat Struct Mol Biol* **18**, 337-44 (2011).
18. Skourti-Stathaki, K., Kamieniarz-Gdula, K. & Proudfoot, N.J. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* **516**, 436-9 (2014).
19. Harouz, H. et al. Shigella flexneri targets the HP1gamma subcode through the phosphothreonine lyase OspF. *EMBO J* **33**, 2606-22 (2014).
20. Casale, A.M., Cappucci, U., Fanti, L. & Piacentini, L. Heterochromatin protein 1 (HP1) is intrinsically required for post-transcriptional regulation of Drosophila Germline Stem Cell (GSC) maintenance. *Sci Rep* **9**, 4372 (2019).
21. Huang, C. et al. Cbx3 maintains lineage specificity during neural differentiation. *Genes Dev* **31**, 241-246 (2017).
22. Sun, M. et al. Cbx3/HP1gamma deficiency confers enhanced tumor-killing capacity on CD8(+) T cells. *Sci Rep* **7**, 42888 (2017).
23. Smothers, J.F. & Henikoff, S. The HP1 chromo shadow domain binds a consensus peptide pentamer. *Curr Biol* **10**, 27-30 (2000).
24. Kwon, S.H. et al. Heterochromatin protein 1 (HP1) connects the FACT histone chaperone complex to the phosphorylated CTD of RNA polymerase II. *Genes Dev* **24**, 2133-45 (2010).
25. Greil, F. et al. Distinct HP1 and Su(var)3-9 complexes bind to sets of developmentally coexpressed genes depending on chromosomal location. *Genes Dev* **17**, 2825-38 (2003).
26. Maison, C. et al. Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nat Genet* **30**, 329-34 (2002).

27. Maison, C. et al. SUMOylation promotes de novo targeting of HP1alpha to pericentric heterochromatin. *Nat Genet* **43**, 220-7 (2011).
28. Muchardt, C. et al. Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1alpha. *EMBO Rep* **3**, 975-81 (2002).
29. Hiragami-Hamada, K. et al. Dynamic and flexible H3K9me3 bridging via HP1beta dimerization establishes a plastic state of condensed chromatin. *Nat Commun* **7**, 11310 (2016).
30. Berry, S., Rosa, S., Howard, M., Buhler, M. & Dean, C. Disruption of an RNA-binding hinge region abolishes LHP1-mediated epigenetic repression. *Genes Dev* **31**, 2115-2120 (2017).
31. Davidovich, C. & Cech, T.R. The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. *RNA* **21**, 2007-22 (2015).
32. Nickerson, J.A., Krochmalnic, G., Wan, K.M. & Penman, S. Chromatin architecture and nuclear RNA. *Proc Natl Acad Sci U S A* **86**, 177-81 (1989).
33. Hall, L.L. et al. Stable C0T-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell* **156**, 907-19 (2014).
34. Allemand, E., Batsche, E. & Muchardt, C. Splicing, transcription, and chromatin: a menage a trois. *Curr Opin Genet Dev* **18**, 145-51 (2008).
35. Hediger, F. & Gasser, S.M. Heterochromatin protein 1: don't judge the book by its cover! *Curr Opin Genet Dev* **16**, 143-50 (2006).
36. Larson, A.G. et al. Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. *Nature* **547**, 236-240 (2017).
37. Strom, A.R. et al. Phase separation drives heterochromatin domain formation. *Nature* **547**, 241-245 (2017).
38. Ameyar-Zazoua, M. et al. Physical and functional interaction between heterochromatin protein 1alpha and the RNA-binding protein heterogeneous nuclear ribonucleoprotein U. *J Biol Chem* **284**, 27974-9 (2009).
39. Rossbach, O. et al. Crosslinking-immunoprecipitation (iCLIP) analysis reveals global regulatory roles of hnRNP L. *RNA Biol* **11**, 146-55 (2014).
40. Hui, J. et al. Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. *EMBO J* **24**, 1988-98 (2005).
41. Cavaloc, Y., Bourgeois, C.F., Kister, L. & Stevenin, J. The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA* **5**, 468-83 (1999).

42. Elbarbary, R.A., Lucas, B.A. & Maquat, L.E. Retrotransposons as regulators of gene expression. *Science* **351**, aac7247 (2016).
43. Zovoilis, A., Cifuentes-Rojas, C., Chu, H.P., Hernandez, A.J. & Lee, J.T. Destabilization of B2 RNA by EZH2 Activates the Stress Response. *Cell* **167**, 1788-1802 e13 (2016).
44. Chomczynski, P. & Sacchi, N. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nat Protoc* **1**, 581-5 (2006).
45. Cokelaer, T., Desvillechabrol, D., Legendre, R. & Cardon, M. 'Sequana': a Set of Snakemake NGS pipelines. *Journal of Open Source Software* **2**, 352 (2017).
46. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3 (2011).
47. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
48. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-30 (2014).
49. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
50. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-5 (2016).
51. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).
52. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
53. Benjamini, Y. & Hochberg, Y. CONTROLLING THE FALSE DISCOVERY RATE - A PRACTICAL AND POWERFUL APPROACH TO MULTIPLE TESTING. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **57**, 289-300 (1995).
54. Thomas-Chollier, M. et al. RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. *Nucleic Acids Res* **40**, e31 (2012).
55. Vaquero-Garcia, J. et al. A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife* **5**, e11752 (2016).

1

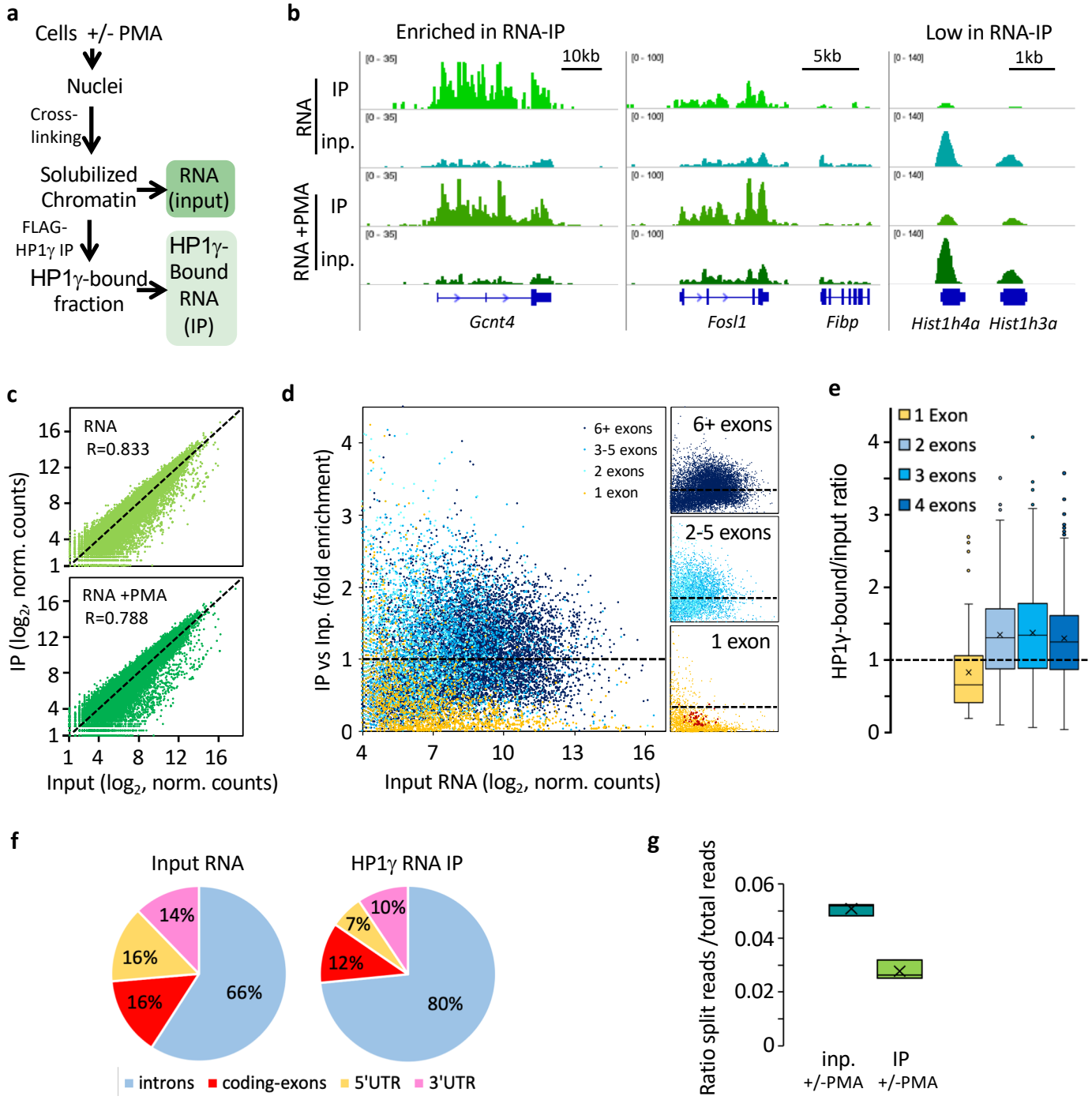


Fig. 1 HP1γ associates with introns of pre-messenger RNA on chromatin. **a,** Scheme of the strategy used to assay HP1γ association with RNA on chromatin. **b**, RNA read density profiles on representative genes enriched (left) or low (right) in HP1γ RNAchIP (IP) and in input RNA (inp.) from HP1γ cells stimulated or not with PMA. **c**, Genome-wide scatter plot of RNAchIP and Input RNA read counts per gene bodies for a representative sample of the triplicate. **d**, Scatter plot of fold enrichment and input represented as in c. Genes were color-coded according to their number of exons (Orange, 1 exon; light to dark blue, 2 to 6 and more exons, respectively), and sorted in individual panels by color. Red dots highlight intron-less histone genes. **e**, Box plot showing the distribution of fold enrichment values per genes sorted by exon number. Box boundaries represent 25th and 75th percentiles; center line represents median; whiskers indicate ±1.5× IQR, and points are values of outliers. **f**, pie chart depicting percentage of RNA sequencing reads from input and RNAchIP counted per gene features, calculated on the basis of all uniquely aligned reads. **g**, Box plot of reads belonging to spliced transcripts versus all reads in all six input or RNAchIP samples calculated as a ratio between split reads and total reads.

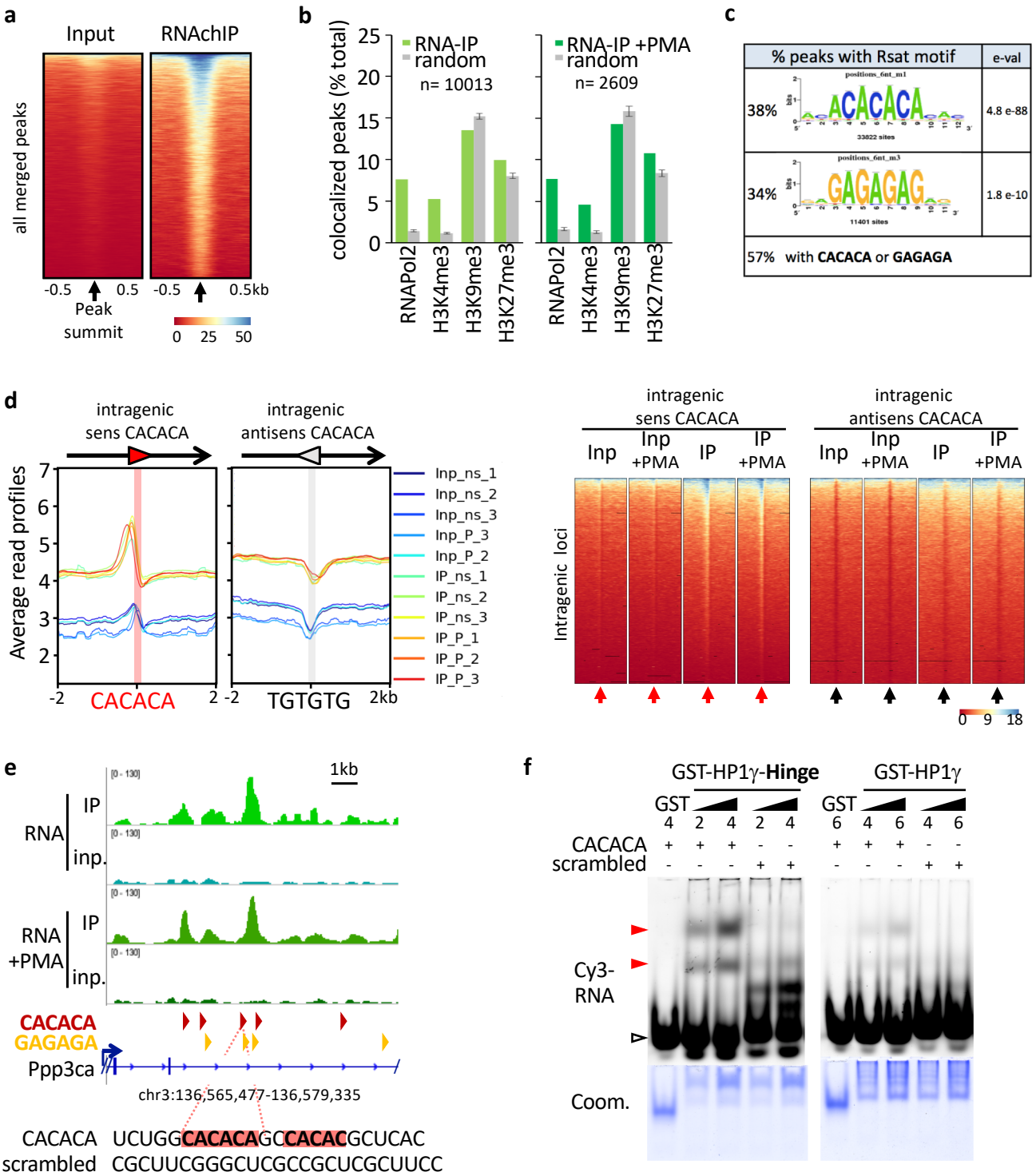
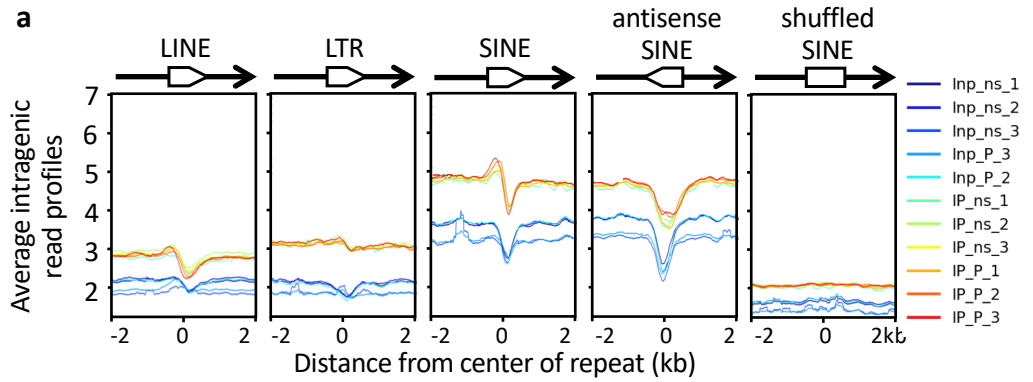
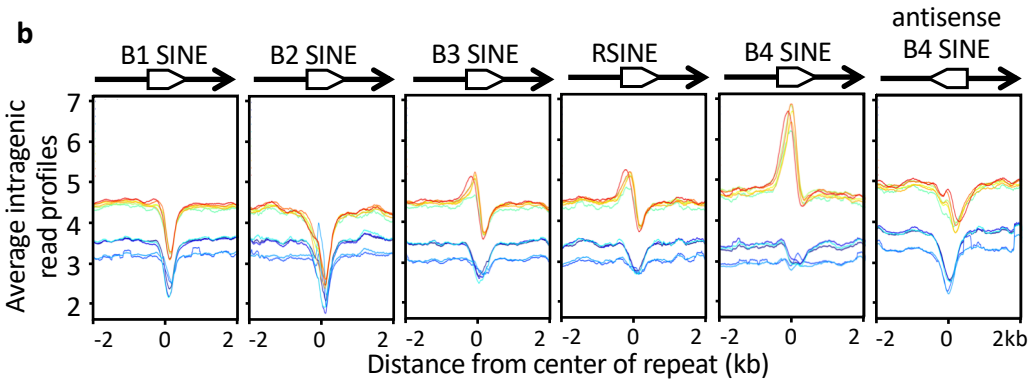
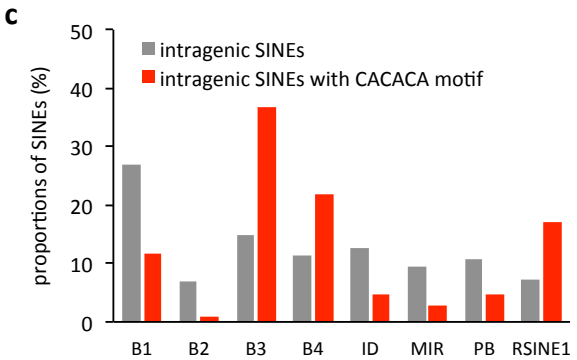

2
a
Input
RNAchIP
all merged peaks
-0.5
0.5
-0.5
0.5kb
Peak summit
0 25 50
b
colocalized peaks (% total)
RNA-IP
random
n= 10013
RNA-IP +PMA
random
n= 2609
RNAPol2
H3K4me3
H3K9me3
H3K27me3
c
% peaks with Rsat motif
e-val
38%
4.8 e-88
34%
1.8 e-10
57% with CACACA or GAGAGA
d
intragenic sens CACACA
intragenic antisens CACACA
Average read profiles
CACACA
TGTGTG
2kb
Inp_ns_1
Inp_ns_2
Inp_ns_3
Inp_P_3
Inp_P_2
IP_ns_1
IP_ns_2
IP_ns_3
IP_P_1
IP_P_2
IP_P_3
Intragenic loci
Inp
Inp +PMA
IP
IP +PMA
0 9 18
e
1kb
RNA
IP
inp.
RNA +PMA
CACACA
GAGAGA
Ppp3ca
chr3:136,565,477-136,579,335
CACACA UCUGGCACACAGCCACACGCUCAC
scrambled CGCUUCGGGCUCGCCGCUCGCUUCC
f
GST-HP1γ-Hinge
GST-HP1γ
GST
CACACA
scrambled
Cy3-RNA
Coom.

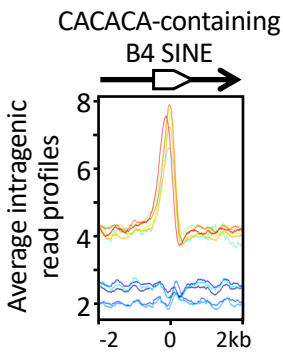
Fig. 2 HP1γ-associated RNA is enriched in CACACA and GAGAGA motifs. **a**, Heat map of RNAchIP and input signal centered on RNAchIP peak summits (n=10013 peaks merged between triplicates of RNAchIP samples). **b**, Percentage of RNAchIP peaks colocalized with indicated chromatin features in MEF samples from ENCODE database (green bars). Colocalization was evaluated by comparison with the list of merged peaks whose genomic location was randomized hundred times among genes (gray bars). Error bars represent mean ± s.d. for the randomized peaks, and n is the number of peaks in each condition. **c**, Consensus motifs discovered among RNAchIP peaks by ChIP-seq peak-motif search with the RSAT pipeline, together with the percentage of peaks containing at least one exact hexameric CACACA or GAGAGA motifs or both. e-val. represent the expected number of patterns which would be returned at random for a given probability. **d**, Left, distribution profiles of average RNAchIP signal in both IP (warm colors) and input (blue colors) over ±2kb centered on intragenic CACACA hexameric motifs oriented in the same orientation (sense) or in opposite orientation (antisense) relative to the overlapping referenced transcript. The antisense CACACA motifs were obtained by querying the transcript sequences with the TGTGTG motif. Right, heat map on intragenic loci in one triplicate sample for each condition illustrated on the left panel. **e**, Top, representative example of a locus on the Ppp3ca gene surrounding an RNAchIP merged peak, showing the RNAchIP signal density as in Fig. 1b. Red and orange arrowheads correspond to oriented CACACA and GAGAGA motifs, respectively. Bottom, sequence of the RNA surrounding a CACACA motif highlighted in red, together with a neighboring imperfect motif. The sequence was used to design a CACACA-containing RNA oligonucleotide probe (CACACA), compared to a control probe (scrambled) devoid of any related motif. **f**, Gel mobility shift assay of bacterially expressed, purified GST proteins fused to HP1γ or HP1γ-Hinge domain, and tested for their direct interaction with the Cy3-labeled RNA oligonucleotide probes depicted in **e**. RNA probes in the gels were detected by their Cy3 fluorescence. Total GST-fusion protein loading was visualized by Coomassie blue staining (Coom.).

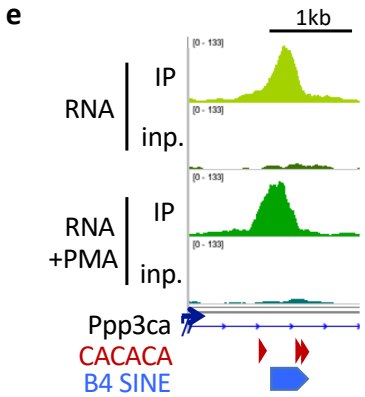
3


a
LINE
LTR
SINE
antisense SINE
shuffled SINE
Average intragenic read profiles
Distance from center of repeat (kb)
Inp_ns_1
Inp_ns_2
Inp_ns_3
Inp_P_3
Inp_P_2
IP_ns_1
IP_ns_2
IP_ns_3
IP_P_1
IP_P_2
IP_P_3


b
B1 SINE
B2 SINE
B3 SINE
RSINE
B4 SINE
antisense B4 SINE
Average intragenic read profiles
Distance from center of repeat (kb)


c
intragenic SINEs
intragenic SINEs with CACACA motif
proportions of SINEs (%)
B1
B2
B3
B4
ID
MIR
PB
RSINE1


d
CACACA-containing B4 SINE
Average intragenic read profiles


e
1kb
RNA
IP
inp.
RNA +PMA
IP
inp.
Ppp3ca
CACACA
B4 SINE

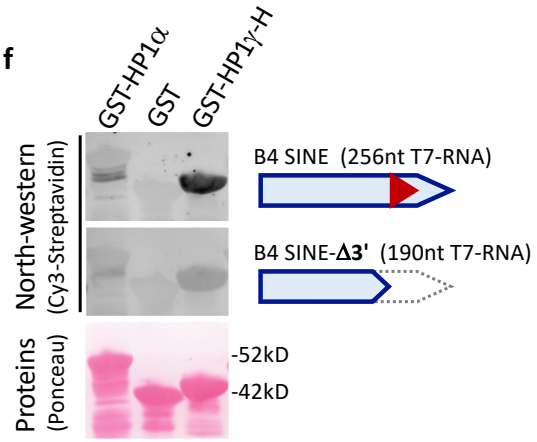

f
GST-HP1α
GST
GST-HP1γ-H
North-western (Cy3-Streptavidin)
B4 SINE (256nt T7-RNA)
B4 SINE-Δ3' (190nt T7-RNA)
Proteins (Ponceau)
-52kD
-42kD

Fig. 3 HP1γ-associated RNA is enriched in oriented SINE repeat motifs. **a**, Profiles of average RNAchIP signal over ±2kb centered on intragenic LINE, LTR, or SINE repeated sequences referenced in Repeat masker database. All repeats are oriented in the same orientation (sense) relative to the overlapping referenced transcript, unless otherwise specified (antisense, shuffled), and as depicted by an oriented white box overlapping the transcript in 5' to 3' orientation (black arrow). **b**, Profiles of average RNAchIP signal, as in **a** centered on the 5 major families of intragenic SINE repeats. **c**, Proportions of the different families of SINE repeats, expressed as a percentage of the total number of intragenic SINEs referenced in the mouse repeat masker database. **d**, Profiles of average RNAchIP signal, as in **b** centered on the subset of intragenic B4 SINE repeats containing an exact CACACA motif. **e**, Genomic position of a B4 SINE repeat and CACACA consensus motifs at the location of the RNAchIP peak depicted in Fig. 2e. **f**, North-western blot assay showing direct association between the indicated bacterially expressed, purified GST-fusion proteins and *in vitro* transcribed, biotinylated RNA probes based on the sequence of the B4 SINE depicted in **e**. Top panels, binding of a 256nt probe corresponding to the CACACA-containing B4 SINE sequence, was compared to an identical B4 SINE deleted of its CACACA by truncation of its 3' portion (B4 SINE-Δ3'). RNA probes hybridized on the membranes were detected by their Cy3 fluorescence. Total GST-fusion protein loading was visualized by Ponceau S staining.

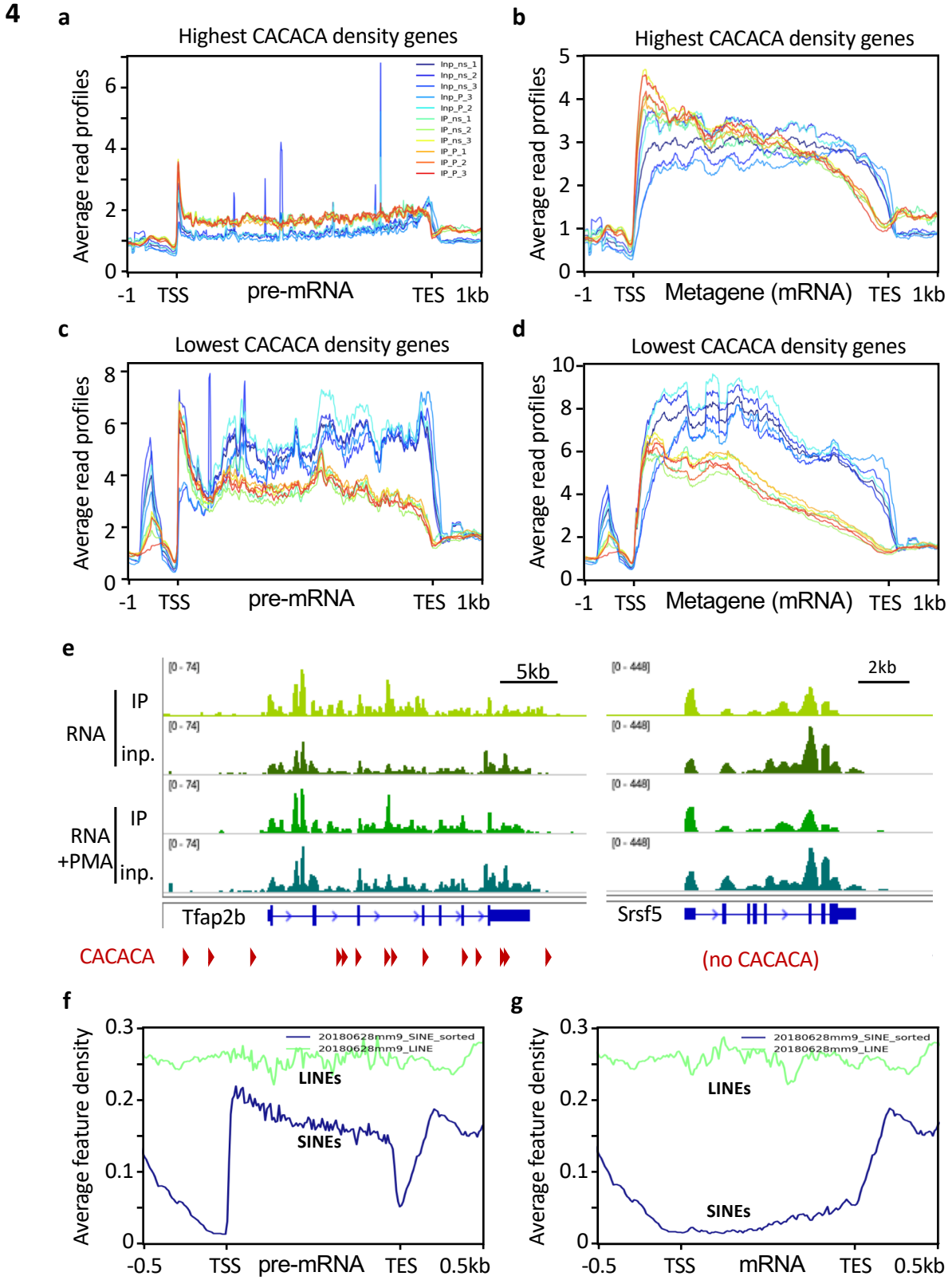


Fig. 4 The density of CACACA motifs affects HP1γ association with transcripts on chromatin. a-d, Average read density profiles computed over entire gene bodies including introns (pre-mRNA; panels **a** and **c**), or as metagenes over exons only (Metagene mRNA; panels **b** and **d**), among two groups of 2,000 RefSeq genes, in the size-range 7 to 12kb, containing the highest or lowest CACACA motif density (panels **a** and **b**, and **c** and **d**, respectively). **e**, Examples of RNAchIP read densities on two representative genes of the highest (left) and lowest (right) categories. **f, g,** Average density of SINE and LINE repeats among 1126 RefSeq genes, in the size-range 7 to 12kb, computed over entire gene bodies including introns (**f**), or as metagenes over exons only (**g**)

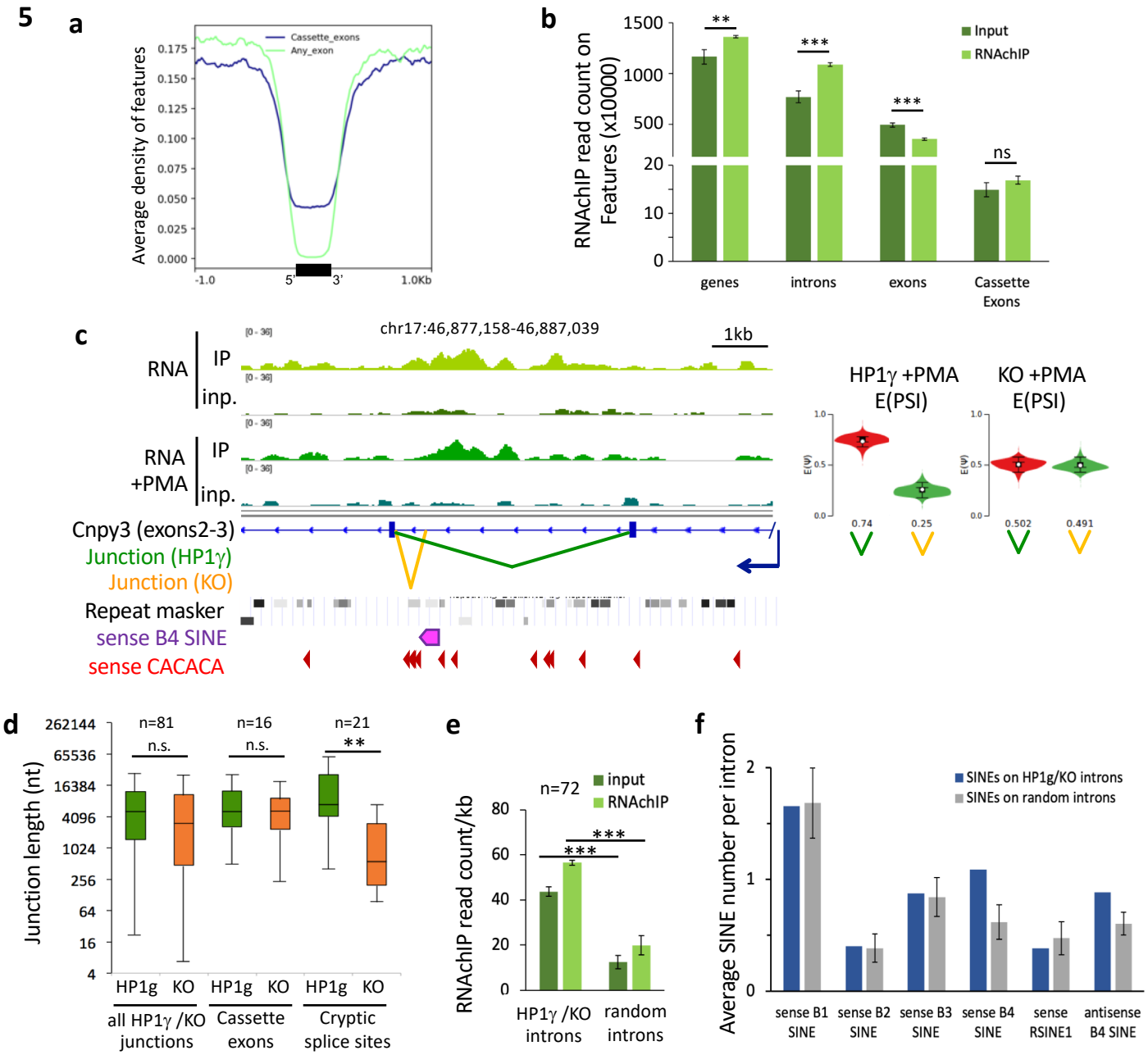


Fig. 5 HP1γ has an impact on RNA splicing which is correlated with CACACA and B4 SINE motifs. **a**, Average density profiles of SINE repeats around cassette exons (blue) and any exon (green), by counting SINEs inside the 5' and 3' exon boundaries (black box), ±1kb around exons. **b**, Number of RNAchIP reads counted on the indicated gene features for input and RNA-IP in combined samples from both untreated and PMA stimulated cells. **c**, HP1γ-dependent alternative splicing events (junctions) were detected with the MAJIQ framework by comparison of the transcriptomes of HP1γ-expressing cells (HP1γ) and HP1γ null cells (KO). Left, genome view around exons 2 and 3 of the *Cnpy3* gene showing the appearance of a *de novo* splicing event in the KO cells, while the 'legitimate', referenced exon-exon junction is detected in the presence of HP1γ. Below, localizations of oriented CACACA and B4 SINE motifs, as well as all repeat masker features are depicted. Right, plot showing the quantification of expected abundance, E(PSI), of the junctions depicted on the left, in HP1γ and KO cells stimulated with PMA. **d**, Box plot depicting the lengths of exon junctions found differentially regulated by MAJIQ between HP1 (green) and KO (orange) cells. Lengths were compared in HP1γ and KO among three categories, all differential junctions, cassette exon junctions, unreferenced *de novo* cryptic splice sites detected in KO only. **e**, RNAchIP read density expressed as read count per kb on the indicated introns features for input and RNAchIP in combined samples from both untreated and PMA stimulated cells. HP1γ/KO introns are all intronic intervals (n=72) between exon junctions differentially detected in HP1γ and KO cells. Random introns are an identical library of randomly chosen intronic intervals. **f**, Average number of each SINE family member per intron as in **e**. All error bars represent mean ± s.d. P-value indicates significantly higher difference between the two conditions (** P < 0.005, ***P < 0.001; two-tailed Student's *t*-test).