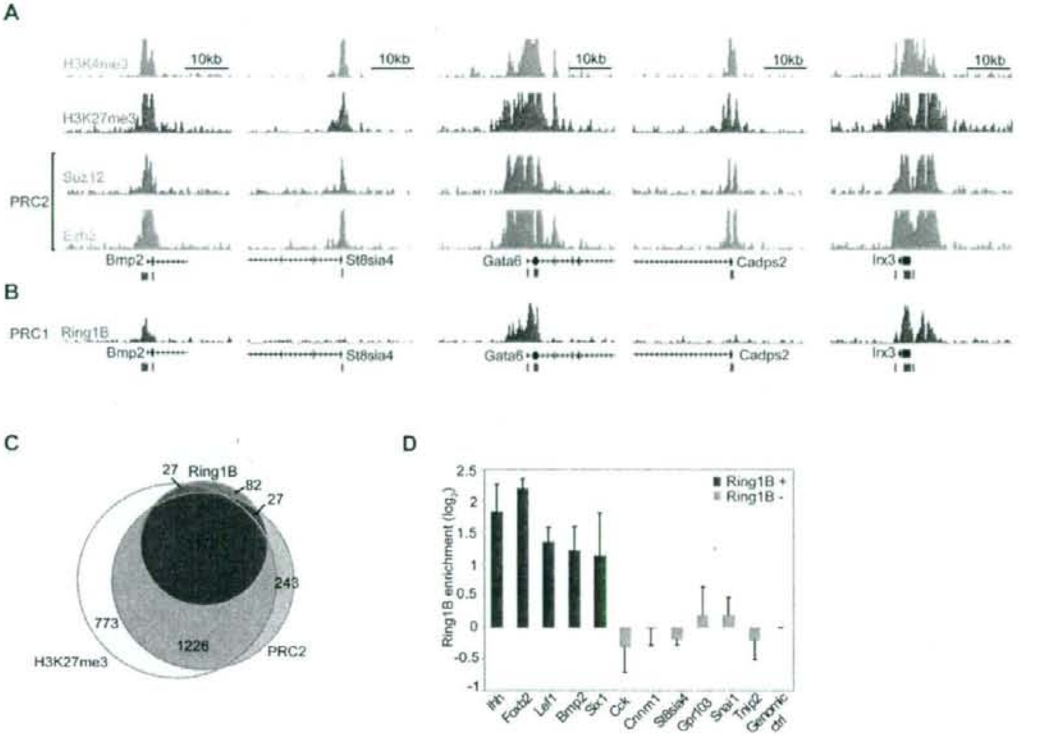


Figure 2. PcG complex occupancy at bivalent domains. (A) ChIP-Seq signals are shown for H3K4me3, H3K27me3 and PRC2 subunits, Suz12 and Ezh2, at a representative panel of bivalent gene promoters. (B) ChIP-Seq signal for the PRC1 subunit Ring1B at these loci. (C) Venn diagram illustrating overlap between promoters marked by H3K27me3, PRC2 and Ring1B. (D) ChIP-qPCR data for Ring1B at bivalent promoters classified by ChIP-Seq as 'Ring1B-positive' or 'Ring1B-negative'. Error bars show standard deviation.
doi:10.1371/journal.pgen.1000242.g002

and, conversely, 78% of bivalent promoters have Ezh2 or Suz12 (Figure 2A,C).

Since PRC2 is the only known complex capable of catalyzing H3K27me3 [2], we considered the minority (22%) of bivalent promoters for which PRC2 was not detected by ChIP-Seq. Many of these promoters show relatively low levels of H3K27me3, and we considered whether PRC2 was simply missed due to sensitivity or thresholding issues. Consistent with this possibility, ChIP with quantitative real-time PCR (qPCR) confirmed modest but significant Ezh2 enrichment at each of these promoters (ratios from 2- to 7-fold; Figure S2A). This suggests that PRC2 is present at essentially all bivalent promoters. Notably, the correspondence between H3K27me3 and PRC2 is not limited to annotated gene promoters, as near-universal PRC2 binding is also evident at the roughly 1000 sites of bivalent chromatin that do not correspond to known genes (see Materials and Methods).

PRC1 Occupies a Conserved Subset of Bivalent Domains. We next turned to examine PRC1 localization, focusing on its catalytic component Ring1B. ChIP-Seq maps reveal roughly 1500 significantly enriched genomic sites in mouse ES cells, including 1308 annotated gene promoters. Nearly all (90%) Ring1B targets correspond to bivalent, PRC2-bound genomic regions. However, just 39% of bivalent promoters are enriched for Ring1B (Figure 2B,C). This occupancy rate is roughly half that observed for Ezh2. As an added measure, we created an Ezh2 ChIP-Seq dataset with exactly the same number of reads as the Ring1B dataset (by randomly selecting reads). Analysis of this truncated dataset reveals Ezh2 binding at 74% of bivalent promoters (compare to 75% for the full Ezh2 ChIP-Seq dataset). Hence, sequencing depth does not account for the difference between Ezh2 and Ring1B occupancy.

Thus, ChIP-Seq analysis suggests that while PRC2 is ubiquitously present at bivalent promoters, PRC1 occupies only a distinct subset. Since PRC2 and PRC1 have generally been described at common genes and loci [9,10], we sought to confirm this unexpected result by orthogonal approaches, as follows:

(i) First, we used ChIP and qPCR to exclude the possibility that the absence of Ring1B at a subset of bivalent promoters reflected a lack of sensitivity of the ChIP-Seq data. This analysis confirmed that Ring1B-negative bivalent promoters also do not show any enrichment by qPCR (Figure 2D).

(ii) Next, to rule out antibody-related bias, we used bioChIP to purify Ring1B-bound chromatin from transgenic ES cells carrying a fusion between Ring1B and biotin ligase recognition peptide (Figure S2B). Ring1B-positive bivalent promoters again showed consistent enrichment, while Ring1B-negative bivalent promoters showed similar enrichment to background controls.

(iii) Third, to test whether the existence of Ring1B-positive and negative bivalent domains is a conserved phenomenon, we examined Ring1B occupancy in human ES cells by ChIP-Seq. We again found that Ring1B occupies only a subset of bivalent domains. The locations of PRC1 show remarkable cross-species conservation: 60% of Ring1B-positive promoters in human are also Ring1B-positive in mouse (Table S4).

(iv) Finally, to confirm that Ring1B status is reflective of PRC1 status, we studied the localization of a distinct PRC1 component, Bmi1. Using an epitope tagged construct in ES cells, we showed that Bmi1 specifically localizes to Ring1B-positive bivalent domains (Figure S2C). This suggests that our findings on Ring1B generally apply to the PRC1 complex. Henceforth, the two sets of bivalent domains are notated as 'PRC1-positive' and 'PRC1-negative'.

PRC1-Bound Bivalent Domains Are Functionally Distinct

The identification of a distinct set of bivalent promoters targeted by Ring1B prompted us to investigate the functional significance of PRC1 occupancy. We made several striking observations relevant to chromatin regulation, epigenetic memory, development and differentiation:

PRC1 Occupancy Correlates with Functional Repression. We first considered whether physical targets of PRC1, as defined above, are also regulated by the complex. Since Ring1B and Ring1A are functionally redundant, we employed a conditional Ring1A/B double-knockout ES cell system in which Ring1B depletion is induced by addition of 4-hydroxy tamoxifen (OHT) [13]. We profiled expression changes after 48 hours of OHT treatment, at which time Ring1B protein levels are markedly depleted while Oct4 levels remain essentially unchanged [8,13]. We found that 32% of PRC1-positive bivalent promoters are up-regulated by at least 50%, compared to just 5% of all genes (Figure 3B). A much smaller proportion of PRC1-negative bivalent promoters are up-regulated at this time point (16%). The difference between the two sets is statistically significant ($p<10^{-10}$), and is not explained by baseline expression levels as bivalent promoters show very low activity, regardless of PRC1 status.

Several factors could contribute to de-repression of this smaller set of PRC1-negative bivalent promoters. The changes may reflect indirect effects as expression is measured after 2 days of OHT treatment. Also, the Ring1 knockout experiment and the location analyses were done in different ES lines, and this could be the basis of some of the discrepancy. Nonetheless, the fact that the PRC1-positive set shows a significantly greater response indicates that PRC1 occupancy correlates with functional repression. As a control, we examined expression changes associated with PRC2 loss. We found that PRC1-positive and PRC1-negative bivalent promoters are de-repressed to roughly equal extents in ES cells lacking the PRC2 component Eed (Figure S4) [13].

PRC1-Positive Bivalent Domains Correspond to Large and Conserved Sites of H3K27me3. Next, we asked whether the patterns of histone modification vary between the two sets of bivalent domains. We observed two significant trends. First, PRC1-positive bivalent domains are associated with much larger regions of H3K27me3 than PRC1-negative bivalent domains (median size of 3.2 kb versus 1.0 kb). The large size is consistent with a proposed role for H3K27me3 in PRC1 recruitment [2,3]. Second, PRC1-positive bivalent domains exhibit greater conservation of chromatin state: bivalent mouse promoters with PRC1 have a bivalent human ortholog in 71% of cases, compared to just 43% of bivalent mouse promoters without PRC1 ($p<10^{-10}$; Figure 3C). Thus, PRC1 occupancy correlates with larger bivalent domains that appear to reflect highly conserved functions.

PRC1-Positive Bivalent Domains Correspond to Developmental Regulator Genes. Next, we examined the gene targets associated with the different classes of bivalent promoters. The PRC1-positive set contains a dramatic enrichment of genes encoding TFs (30%, $p<10^{-20}$), including members of the Hox, Sox, Pax and Pou domain families, or cell signaling and morphogenesis molecules, such as Wnts and Fgfs (Figure S3). In contrast, the PRC1-negative set of bivalent promoters is instead over-represented for genes that encode membrane proteins (50%; $p<10^{-10}$). Remarkably, despite the strong correlation of PcG proteins with developmental TFs, this PRC1-negative (PRC2-only) subset of bivalent domains shows statistically significant depletion of TF genes relative to the genome average (4.1% vs 10.2%; $p<10^{-10}$).

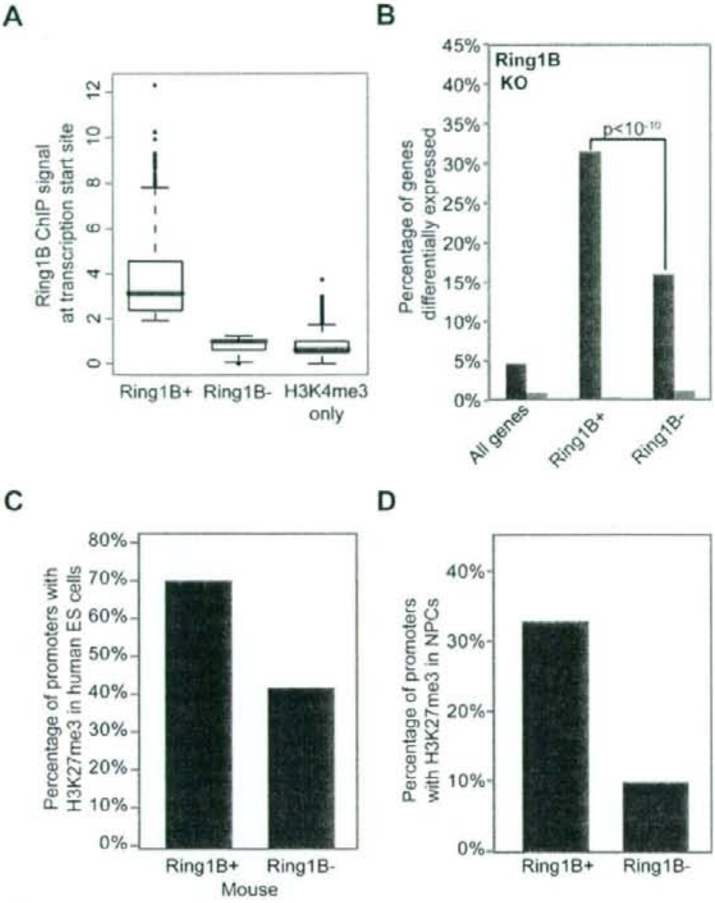
PRC1-Positive Bivalent Domains Efficiently Maintain Repressive Chromatin Environment. Finally, we compared the behavior of PRC1-positive and PRC1-negative bivalent promoters upon ES cell differentiation. We examined ChIP-Seq data for a population of neural progenitors (NPCs) derived from the same ES cell line [16]. Since PRC1 is implicated in the maintenance of a repressive chromatin state, we reasoned that promoters with PRC1 should more efficiently retain H3K27me3 upon differentiation. Consistent with this hypothesis, we found that 33% of PRC1-positive bivalent promoters retain H3K27me3 in the NPCs, compared to just 10% of PRC1-negative bivalent promoters ($p<10^{-10}$) (Figure 3D). Many PRC1-positive bivalent promoters that lose the repressive mark upon differentiation do so in association with transcriptional activation as roughly one-fifth are induced at least 5-fold in the NPCs. Thus, PRC1 occupancy is associated with more stable retention of PcG-associated chromatin marks through differentiation.

We conclude that two distinct sets of bivalent domains can be defined based on PcG complex occupancy in ES cells. Bivalent domains that carry both PRC2 and PRC1 are larger, more conserved and more efficiently retained through differentiation. They account for the vast majority of implicated developmental regulators. By contrast, bivalent domains occupied by PRC2 only are poorly maintained, correspond to distinct non-developmental gene sets, and thus may reflect alternate regulatory processes.

Sequence Elements and Motifs Predict PcG Complex Localization in ES Cells

We next studied the chromatin maps to gain insight into another fundamental unanswered question – namely, the mechanisms that underlie the initial recruitment of PcG complexes and the formation of bivalent domains in ES cells. The extensive epigenetic reprogramming that precedes the pluripotent state suggests that elements in the genomic sequence itself must play central roles in this process [1,27,28]. Yet the identity of these PcG-determining sequence elements has remained elusive.

PRC2 Associates with CG-Rich Sequences Genomewide. To identify sequence elements that could contribute to PcG recruitment, we applied computational sequence analysis and the new ChIP-Seq data. We focused initially on Ezh2, reasoning that this catalytic PRC2 subunit would most closely reflect the initial recruitment mechanisms. Bivalent domains and PcG target sites have been shown previously to correlate with CG-rich DNA; for example, ~50% of Suz12 binding sites in human ES cells correspond to CpG islands [11,16,29]. The ChIP-Seq data for mouse Ezh2 reveal an even higher correspondence, with a full 88% of enriched intervals coinciding with an annotated CpG island. H3K27me3-enriched intervals similarly correlate with CpG islands



E

GO term	% of genes	Ratio	P-value
Ring1B+			
developmental process	40.5%	2.4	$p<10^{-20}$
regulation of gene expression	30.4%	2.2	$p<10^{-20}$
cell-cell signaling	7.7%	3.8	$p<10^{-20}$
embryonic morphogenesis	5.4%	5.0	$p<10^{-20}$
Ring1B-			
membrane	50.1%	1.2	$p<10^{-10}$
cell adhesion	7.9%	2.3	$p<10^{-8}$
transporter activity	12.3%	1.8	$p<10^{-5}$

Figure 3. PRC1-positive bivalent domains are functionally distinct. (A) Box plot shows 25th, 50th and 75th percentile Ring1B ChIP-Seq signals for Ring1B-positive bivalent promoters, Ring1B-negative bivalent promoters, and for H3K4me3 only promoters. (B) Plot illustrates fraction of genes up-regulated (red) or down-regulated (blue) in PRC1-deficient ES cells for the indicated gene sets (see text for details on Ring1A/B dKO ES cell model). De-repression is evident for a significantly greater proportion of PRC1-positive bivalent promoters (p-value by Fisher's exact test). (C) The proportion of bivalent mouse promoters for which the human ortholog also carries H3K27me3 is indicated, contingent on Ring1B status in mouse ES cells. (D) The proportion of bivalent promoters for which H3K27me3 is retained in ES cell-derived neural progenitors ('NPCs'), contingent on Ring1B status in mouse ES cells. (E) Gene Ontology categories over-represented in PRC1-positive or PRC1-negative bivalent gene sets.
doi:10.1371/journal.pgen.1000242.g003

in 79% of cases. Remarkably, the fraction of Ezh2/H3K27me3 sites that coincide with CpG islands is substantially higher than that of H3K4me3 (68%), which has previously been associated with CpG islands [15]. It is also far greater than that of other chromatin structures (Figure S5), including H3K9me3 (1.1%) and H4K20me3 (0.7%).

When we examined the small minority (12%) of Ezh2 binding sites that do not correspond to an annotated CpG island, we found that three-quarters of these sites overlap highly CG-rich sequences that just fall short of the defined threshold for CpG islands (see Materials and Methods). Including those sites, >97% of Ezh2 binding sites in the ES cell genome correspond to annotated CpG islands or other highly CG-rich sequences. These results suggest that such CG-rich sequences, known to be largely un-methylated at the DNA level in ES cells [27], may contribute to the recruitment of PRC2 and the subsequent establishment of H3K27me3 at bivalent domains.

Still, only a minority of CpG islands carries Ezh2 or H3K27me3 in ES cells – that is, are PRC2-positive. Most are enriched for H3K4me3 only and are PRC2-negative (Figure 4A). We thus considered whether additional sequence characteristics distinguish between PRC2-positive and PRC2-negative CpG islands. We collated two sets of CpG islands, one showing clear Ezh2 binding based on ChIP-Seq (n = 2608) and the other lacking any Ezh2 signal (n = 9097). To maximize the power of our analysis, we excluded a subset of CpG islands showing intermediate levels of Ezh2 enrichment (n = 3443).

We considered CpG island length, CG density and the frequency of all possible dinucleotides (Figure S6) as potential characteristics. PRC2-positive CpG islands show a greater median length (721 bp vs 526 bp) and a slightly lower median CpG observed-to-expected ratio (0.88 vs 0.92). However, the overall distributions of length and ratio are largely similar and do not discriminate between PRC2-positive and negative sets.

We also compared the conservation properties of these CpG island sets. Mammalian genomes contain ~200 large regions characterized by striking enrichment for highly conserved non-coding elements [30,31] and exceptionally low CpG divergence rates [32]. These loci contain promoters for many developmental genes, most of which are bivalent in ES cells [33]. Although it has been suggested that conserved elements within these loci contribute to PcG recruitment, we find that only ~10% of Ezh2 binding sites occur within these regions. Overall, we find that PRC2-positive CpG islands show modestly higher sequence conservation relative to PRC2-negative islands, but with overlapping distributions (Materials and Methods). Thus, conservation analysis does not present an obvious explanation for observed PRC2 binding patterns.

PRC2-Positive CpG Islands Can Be Distinguished Based on Motif Content. Because the distinction between PRC2-positive and PRC2-negative CpG islands is not explained by simple sequence composition, we next considered more complex sequence motifs. In *D. melanogaster*, PcG recruitment is mediated by combinations of motifs recognized by specific TFs [4]. We thus explored whether TF motifs could predict PRC2 localization in mammalian ES cells. Since the motifs and TFs implicated in fly show little or no conservation in vertebrates, we broadened our analysis to include all 668 vertebrate DNA binding motifs annotated in the TRANSFAC and Jaspar databases [34,35].

We used the MAST algorithm [36] and position weight matrices (PWMs) from these databases to identify motifs. Taking an unbiased approach, we searched for motifs over-represented in either Ezh2-positive or Ezh2-negative CpG islands. Over-represented motifs were ranked by enrichment ratio, and their significance was confirmed using Fisher's exact test. We also excluded the possibility that enriched motifs simply reflected differences in underlying nucleotide content by repeating each survey with scrambled PWMs. Finally, since there is redundancy among factors and PWMs in the TRANSFAC and Jaspar databases, a clustering algorithm was used to collapse highly similar PWMs to a single representative motif. This analysis yielded a total of 14 motifs enriched between 1.2 and 1.3-fold in the Ezh2-positive CpG islands, and these fall into 10 motif clusters. It also revealed 11 motifs enriched between 2.3 and 6.0-fold in the Ezh2-negative CpG islands, falling into 6 clusters (Figure 4B,C, Figure S8).

We initially focused on the motifs associated with Ezh2-positive CpG islands as these could potentially mediate PRC2 recruitment. Although the enrichment ratios were relatively low, it is conceivable that combinations of factors might be required, as in *Drosophila*. However, most of the corresponding TFs are not actually expressed in ES cells, but rather are expressed in differentiated cells. These include developmental regulators induced along specific differentiation pathways, such as MyoD (myogenesis), Lmo2 (hematopoiesis), Brachyury (paraxial mesoderm) and Pou6F1 (neurogenesis) [37–40]. PRC2 targets include many developmental genes with complex expression patterns which may explain why they are enriched for lineage-specifying TF motifs. Hence, it is unlikely that these non-expressed TFs contribute to PRC2 localization in ES cells.

However, three of the factors identified in the Ezh2-positive islands are expressed in ES cells, and these cases are illustrative (Figure 4D). The most highly-expressed is neuron-restrictive silencing factor (NRSF/REST), a potent transcriptional repressor essential for ES cell pluripotency [41]. Notably, the NRSF motif is among the best characterized and highly predictive binding elements in mammalian genomes [42]. A second expressed factor is Cux1, which also functions as a transcriptional repressor [43]. The third expressed factor is NFκB, a widely studied transcriptional regulator with diverse functions related to immunity, inflammation and differentiation [44]. Although NFκB is clearly expressed, its activity is strongly inhibited in ES cells by the pluripotency factor Nanog [45]. Thus, motifs enriched in Ezh2-positive CpG islands are recognized either by repressors or by TFs that are inactive in ES cells (see Text S2).

Next, we turned to examine motifs enriched in the Ezh2-negative CpG islands. We were immediately struck that these motifs are recognized by several well-characterized classes of transcriptional activators that are highly expressed in ES cells (Figure 4C,D). Some of the implicated factors have key functions in the ES cell regulatory network (e.g., NFY, Myc) while others are constitutive activators with general housekeeping functions (e.g., Ets1; see Text S2) [46–48]. The magnitudes of enrichment observed for these activating motifs are much greater than those observed for motifs identified in Ezh2-positive sequences above. Thus, the strongest sequence correlate of Ezh2 binding at a CpG island appears to be the absence of motifs capable of conferring transcriptional activity.

A simple count of the motif occurrences within a CpG island allows accurate prediction of roughly two-thirds of Ezh2 binding sites (see Materials and Methods; Figure 4E). This compares favorably with the Polycomb response elements predicted in *Drosophila*, which are present at 6 to 27% of experimentally-determined PcG binding sites [4,49–51]. Notably, the motif occurrences we identified in mouse also have considerable predictive value for identifying PcG targets in human ES cells (Figure 4F).

In sum, we find that PRC2-positive CpG islands are characterized by an over-representation of repressor motifs and a strong depletion of transcriptional activator motifs. While it is

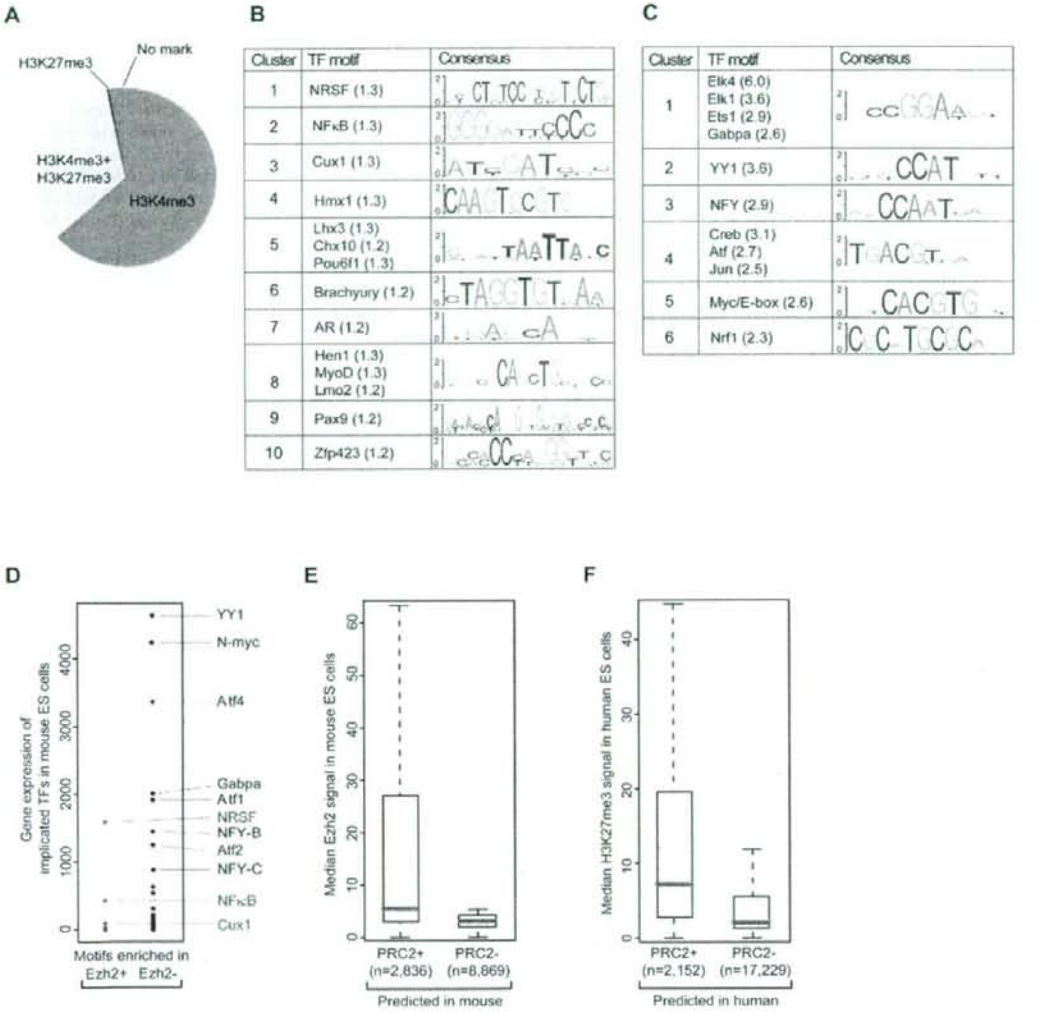


Figure 4. CG-density and DNA motif occurrences predict genomewide PcG complex localization. (A) Proportion of CpG islands with a given chromatin state in mouse ES cells. More than 97% of Ezh2 sites in mouse ES cells correspond to CpG islands or other highly CG-rich sequences. A systematic screen reveals sets of DNA motifs over-represented in (B) Ezh2-positive CpG islands or (C) Ezh2-negative CpG islands (enrichment in parentheses). (D) Expression levels of implicated TFs in mouse ES cells. Motifs enriched in Ezh2-positive CpG islands correspond to repressors or to TFs that are not expressed. Motifs enriched in Ezh2-negative CpG islands correspond to highly expressed activators. (E) Ezh2 ChIP-Seq signals for CpG islands predicted as PRC2-positive or PRC2-negative based on motif occurrences. (F) H3K27me3 ChIP-Seq signals for human ES cells for CpG islands predicted to be PRC2-positive or PRC2-negative based on occurrences of the motifs originally identified in mouse.
doi:10.1371/journal.pgen.1000242.g004

possible that the implicated repressors directly mediate PRC2 recruitment, each has been well-studied and linked to distinct biological processes. Rather, we favor the view that the paucity of activating motifs and, to a lesser extent, the presence of repressive motifs dictate a transcriptionally inactive state in ES cells that is permissive to PRC2 binding. We suggest that CpG islands play a central role in PRC2 recruitment and, in the absence of transcriptional activity, assume a bivalent chromatin state by 'default' in ES cells (see Discussion).

PRC1 Occupies Large PRC2-Positive CpG Islands. Lastly, we considered whether PRC1 association can also be predicted from genome sequence. PRC1 occupies roughly half of all PRC2 sites in ES cells, and is essentially never observed in the absence of this second PcG complex. We collated and compared two sets of Ezh2-positive CpG islands, one with Ring1B (n = 1036) and the other without Ring1B (n = 981) (see Methods). We found no significant differences in nucleotide content (CG-density, dinucleotide frequencies) or in the occurrences of the motifs discussed above.

Rather, the best predictor appears to be the length of CG-rich DNA. PRC1-positive CpG islands are roughly twice as large as those that carry only PRC2 (Figure S9). They are also much more likely to reside in close proximity to other bivalent CpG islands. Consideration of CpG island size and proximity to other bivalent islands enables accurate prediction of PRC1 status for >70% of PRC2-positive CpG islands (see Materials and Methods). Thus, our findings suggest that the genomewide localization of the two main PcG complexes in ES cells may be largely predicted from the location, size and underlying motif content of CpG islands.

Discussion

We have applied ChIP-Seq and computational genomic analysis to study the genomewide distributions of key histone modifications and PcG subunits in mouse and human ES cells, thereby gaining insight into the structure, function and establishment of bivalent domains.

The ChIP-Seq data reveal two distinct sets of bivalent domains in ES cells. One set, defined based on co-occupancy by both PRC1 and PRC2, shows special epigenetic properties, including higher evolutionary conservation of chromatin state and robust retention of repressive chromatin through differentiation. This set is exquisitely enriched for developmental targets in that over one third of the corresponding genes encode TFs, morphogens or cytokines. In striking contrast, a second set of bivalent domains, occupied by PRC2 only, is actually under-represented for TF genes relative to the genome average, and shows weak conservation and retention of the PcG-associated chromatin marks. We suggest that the complete repertoire of PcG machinery is needed for full functionality of bivalent domains and associated chromatin in the epigenetic regulation of key developmental genes.

The data also suggest a potential model for understanding the initial recruitment of PcG complexes for the coordinated establishment of bivalent chromatin. In particular, we find that PRC2 association in ES cells is entirely restricted to sequences with high CpG content, the vast majority being annotated CpG islands. The status of a given CpG island – whether it carries PRC2 and bivalent H3K4me3/H3K27me3 chromatin or only H3K4me3 – correlates with underlying motif content. CpG islands with PRC2 show a striking depletion of transcriptional activator motifs and a modest enrichment of repressor motifs. Thus, PRC2 appears to localize to CpG islands that are transcriptionally silent in ES cells because they lack activating DNA sequence motifs.

CpG islands have been extensively correlated with trxG complexes and H3K4me3; recruitment of the former likely involves CXXC proteins with affinity for un-methylated CpG dinucleotides [15,52,53]. We propose that CpG islands by default similarly mediate PcG recruitment and catalysis of H3K27me3 in mammalian ES cells, except when the default is over-ridden by transcriptional activity. In this model, the extent of PcG/H3K27me3 and trxG/H3K4me3 at any given CpG island is determined by its baseline transcriptional status which is dictated by underlying motif content. The view that transcriptional status is upstream of PcG status in ES cells is consistent with the subtle transcriptional changes evident in PcG-deficient ES cells [9,54]. Although our analyses do not shed light on the underlying mechanisms, PRC2 recruitment may also involve proteins with affinity for un-methylated CpGs or may be mediated indirectly through recognition of other histone modifications such as H3K4me3. In either case, active transcription within a locus would preclude stable PRC2 association and thereby restrict it to inactive CpG islands.

Large PRC2-positive CpG islands tend to also carry PRC1. The expansive regions of H3K27me3 associated with these islands may contribute to PRC1 recruitment via chromodomain proteins [2,3]. As discussed above, bivalent domains that carry both PRC2 and PRC1 appear to have unique epigenetic regulatory properties. We therefore propose that large CpG islands depleted of activating motifs confer epigenetic regulation by recruiting both key PcG complexes in pluripotent cells. Such islands may thereby reflect mammalian memory elements analogous to Polycomb response elements in flies.

The tight correspondence between DNA sequence and PcG localization may have implications for important cellular processes, such as development and epigenetic reprogramming. Induced pluripotent stem (iPS) cells and ES cells exhibit nearly identical chromatin patterns, including the locations of bivalent domains [55,56]. The sequences described above may function as templates for the robust assembly and appropriate positioning of PcG complexes and bivalent domains during pre-implantation development or the artificial reprogramming of somatic cells to iPS cells [1,28].

What then might be the purpose of an initial chromatin state fully encoded by genetic sequence and an associated transcriptional program? Based on existing evidence, we suggest that PcG complexes and associated chromatin buffer the pluripotent ground state by reinforcing the repression of factors that induce differentiation. The initial chromatin architecture also appears poised for the dynamic expression changes that accompany differentiation and for the subsequent engagement of epigenetic controls to maintain lineage-specific transcriptional programs. Our analysis suggests that such epigenetic functions mainly apply to large bivalent CpG islands that also carry PRC1. It remains to be seen whether small PRC1-negative bivalent domains have distinct regulatory functions or are simply byproducts of the mechanisms that have evolved for establishment of the former.

Further studies are needed to determine the precise DNA elements and protein interactions that mediate PcG recruitment. As discussed above, the proposed central role for CG-rich sequences implies the involvement of CXXC domains or other proteins that recognize CG dinucleotides. However, several factors complicate the interpretation of our genomic findings. In particular, CpG islands are at least partly a consequence of reduced CpG deamination rates in regions that lack DNA methylation in the germ line [27]. PcG-occupied regions are largely un-methylated at the DNA level, at least in ES cells [57], and this could favor retention of CG-rich sequences. Thus, it remains possible that evolutionary dynamics and/or the generally high CpG content of target regions are masking other key sequence features.

Finally, it should be emphasized that our findings on the relationships among PRC2 and PRC1 and the sequences that underlie their genomic localizations pertain specifically to ES cells. PcG complexes show remarkable tissue-specificities in terms of their expression levels, stoichiometry and localization [2,3,11,12]. Further study is needed to understand how the genomic localizations and regulatory functions of PcG complexes vary with differentiation, lineage specification, environment, and disease.

Materials and Methods

Cell Culture

Mouse v6.5 (genotype 129SvJae×C57BL6, male, passages 10–15) ES cells were cultured on fibroblast feeders in DMEM (Sigma) with 15% fetal bovine serum (Hyclone), GlutaMax (Invitrogen), MEM non-essential amino acids (Invitrogen), pen/strep (Invitrogen), ESGRO (Chemicon) and 2-mercaptoethanol (Sigma), incubating at 37°C, 5% CO_2 [16]. Prior to harvest, these cells were passaged 2–3 times on feeder-free gelatinized tissue culture plates. A transgenic ES cell line expressing a fusion between Ring1B and biotin ligase recognition peptide from the endogenous Ring1B locus and the BirA biotin ligase from the Rosa26 locus (H.K., unpublished) was cultured as described above.

Human H9 (female, passage 45) ES cells were cultured as described [58] and at http:/www.WiCell.org. Briefly, the human ES cells were cultivated on irradiated MEFs (strain DR4) in Knockout DMEM (Invitrogen) containing 10% Knockout Serum Replacement (Invitrogen), 10% Plasmanate (Bayer Healthcare), GlutaMax (2 mM), pen/strep, MEM non-essential amino acids (0.1 mM), 10 ng/ml β-FGF (Invitrogen) and 2-mercaptoethanol. Cells were incubated at 37°C, 5% CO_2. MEF-free ES cells were used for analysis. MEF-free culture was prepared in the following manner: First, MEFs were depleted at the time of trypsin passaging through brief transfer (thirty minutes) of hES cells onto gelatin-coated plates. MEF-subtracted ES cells were then propagated on plates coated with Matrigel (Invitrogen). ES cells grown on Matrigel were supported with the aforementioned human ES cell medium that had first been conditioned on MEFs for 24 hours. Fresh β-FGF was added to the conditioned medium immediately prior to use.

Generation of Flag-Bmi1 mES Cells

Doxycyclin-inducible Flag-Bmi1 transgenic ES cell line was generated by PCR amplifying a 1× flag tagged Bmi1 ORF (Addgene) with primers that incorporate a 3× flag tag as well as EcoRI and XbaI restriction enzyme sites (5′-GGAATTCCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGATATCGACTACAAGGACG-3′, 5′- GCTCTAGAGCACCAGATGAAGTTGCTGATGACCCATTTAGTGATGATTTT-3′). This was cloned into the pLox vector (pPGK-loxP-neoEGFP) and incorporated into Ainv15 mouse ES cells using a cre recombinase expression vector as previously described [59]. Flag-Bmi1 ES cells were cultured similarly to wild-type mES cells as described above. Prior to harvest, Flag-Bmi1 expression was induced by incubating with 1 μg/ml of Doxycycline for two days on gelatinized culture plates.

Chromatin Immunoprecipitation and Antibodies

ChIP experiments for H3K4me3, H3K27me3 and H3K36me3, Ring1B and Flag-Bmi1 were carried out as described [15,16]. ES cells were crosslinked in 1% formaldehyde, lysed and sonicated with either a Branson 250 Sonifier (mouse ES cells) or a Diagenode bioruptor (human ES cells) to obtain chromatin fragments in a size range between 200 and 700 bp. Solubilized chromatin (whole cell lysate or 'WCE') was diluted in ChIP dilution buffer (1:10) and incubated with antibody overnight at 4°C. Protein A sepharose beads (Sigma) were used to capture the antibody-chromatin complex and washed with low salt, LiCl, as well as TE (pH 8.0) wash buffers. Enriched chromatin fragments were eluted at 65°C for 10 min, subjected to crosslink reversal at 65°C for 5 hrs, and treated with Proteinase K (1 mg/ml), before being extracted by phenol-chloroform-isoamyl alcohol, and ethanol precipitated. ChIP DNA was then quantified by Quant-iT Picogreen dsDNA Assay kit (Invitrogen).

ChIP experiments for Ezh2 and Suz12 were carried out on nuclear preps. Crosslinked ES cells were incubated in swelling buffer (0.1 M Tris pH 7.6, 10 mM KOAc, 15 mM MgOAc, 1% NP40), on ice for twenty minutes, passed through a 16G needle 20 times and centrifuged to collect nuclei [60]. Isolated nuclei were then lysed, sonicated and immunoprecipitated as described above.

BioChIP assays were carried out using transgenic Ring1B-Biotin ligase recognition peptide ES cells (above). Nuclei were isolated, lysed and sonicated as described above. Dynabeads M-280 Streptavidin (Invitrogen 112.05D) were used to capture biotinylated Ring1B-DNA complex. Beads were washed with a 2% SDS buffer and a high salt buffer (50 mM HEPES, pH 7.5, 1 mM EDTA, 500 mM NaCl, 1% Triton X-100, 0.1% Deoxycholate), in addition to the regular washes. Elution and cross-link reversal were done simultaneously by incubating Dynabeads in 300 mM NaCl at 65°C overnight [46]. DNA was isolated as described above.

Antibodies used in this study include anti-H3K4me3 (Abcam ab8580), anti-H3K27me3 (Upstate 07-449), anti-H3K36me3 (Abcam ab9050), anti-Ezh2 (Active Motif 39103), anti-Suz12 (Abcam ab12073), anti-Ring1B [61] and anti-Flag (M2) (Sigma F1804). Details on antibody specificity are provided in Text S1.

Sequencing Library Preparation and Illumina/Solexa Sequencing

Library preparation and ultra high-throughput sequencing were carried out as described [16]. Briefly, one to ten nanograms (ng) of ChIP DNA were end-repaired and 5′phosphorylated using END-It DNA End-Repair Kit (Epicentre). We then followed steps four through seven of Illumina standard sample prep protocol (v1.8) using Genomic DNA Sample Prep Kit (Illumina) with minor modifications. A single Adenine was added to 3′ ends by Klenow (3′→5′ exo$^-$), and double-stranded Illumina Adapters were ligated to the ends of the ChIP fragments. Adapter-ligated ChIP DNA fragments between 275 bp to 700 bp were gel-purified and subjected to 18 cycles of PCR. Prepared libraries were quantified using PicoGreen and sequenced on the Illumina Genome Analyzer per standard operating procedures.

Read Alignment and Generation of Density Maps and Modified Intervals

Sequence reads (36 bases) from each ChIP experiment were compiled, post-processed and aligned to the appropriate reference genome using a general purpose computational pipeline as described previously [16]. Aligned reads are used to estimate the number of end-sequenced ChIP fragments that overlap any given genomic position (at 25-bp resolution). For each position, we counted the number of reads that are oriented towards it and closer than the average length of a library fragment (~300 bp). The result is a high-resolution density map that can be viewed through the UCSC Genome Browser [62] and is used for downstream analyses. Prior comparisons to microarray analysis and quantitative real-time PCR have shown that ChIP-Seq density maps accurately reflect enrichment [16]. ChIP-Seq data can be accessed at http://www.broad.mit.edu/seq_platform/chip/.

We used a Hidden Markov Model (HMM) to demarcate chromosomal segments likely to be enriched for a given chromatin modification or PcG protein [16]. In order to model ChIP-Seq read density variations along the genome, we define four observed states: masked, low density, medium density, and high density. This discretization of the data into the four states was based on the signal intensity in known modified regions versus known unmodified regions as determined in prior ChIP-Seq, microarray and ChIP-PCR analyses [15,16], and adjusted for each sample. The model was then used to discriminate enriched and unenriched intervals genome wide. In order to more properly classify enriched regions containing several short interspersed peaks and facilitate subsequent analyses intervals within 2 kb were merged.

Promoter Classification and Definition of Gene and Transcript Intervals

We defined 17760 mouse and 18522 human promoters for 17442 and 17383 genes, respectively, as the sequences between

−0.5 kb and +2.0 kb of the annotated transcription start site, using the mouse mm8 and human hg18 genome builds. Transcripts were defined for these genes as the range from transcription start to end [62]. To identify regions enriched for histone marks or chromatin-associated proteins, we generated a null-hypothesis background model by dividing the alignable parts of each chromosome into 200 bp bins and randomly redistributing the reads aligned on this chromosome. Based on a histogram of the cumulative distribution of reads per bin, a cutoff threshold was determined. Stability of the calculated background cutoff threshold was confirmed through 1000 independent simulations for each ChIP-Seq track and showed remarkable invariance. For promoters, a 200 bp sliding window was moved across the 2.5 kb promoter region and the ratio of median read density over background was calculated. The maximum enrichment achieved in any window at this promoter site was then used for further analysis. Maximum enrichment cutoff thresholds were determined empirically for all tracks, and promoters were then classified based on the maximum enrichment for the various histone marks and PcG proteins. The same procedure was applied to a pan-H3 (modification-insensitive) ChIP-Seq dataset as control where virtually no significant enrichment over background was found. Ring1B-positive bivalent promoters were defined based on normalized ChIP-Seq signal and comprise 40% of all bivalent promoters. A set of Ring1B-negative bivalent promoters was also defined based on absence of ChIP-Seq enrichment, and includes another 40% of all bivalent promoters. The remaining bivalent promoters (20%) with indeterminate Ring1B ChIP-Seq signals were excluded from this analysis.

For conservation analyses of human and mouse promoter states, we used NCBI HomoloGene (build 58) gene clusters to assign orthologous human promoters and transcripts to the 17442 mouse promoters and transcripts, yielding a set of 13200 orthologous promoters and 13625 orthologous transcripts for which human and mouse chromatin state could be compared (ftp://ftp.ncbi.nih.gov/pub/HomoloGene/). Genes with multiple start sites were excluded from this analysis. Promoters were associated with CpG states as described previously [16].

For comparison of Ezh2 and Ring1B occupancy at target genes, a reduced Ezh2 read set was generated by randomly selecting the same number of reads that were available for Ring1B from the full Ezh2 read pool (~3.5 million). Read mapping to the mouse genome and analysis of promoter state were performed as described above.

Real-Time PCR

PCR primer pairs were designed to amplify designated genomic regions using Primer3 (http://fokker.wi.mit.edu/primer3/input.htm). Real-time PCR assays were carried out on ABI 7000 or 7500 detection systems. We used Quantitect SYBR green PCR mix (Qiagen) with 0.1 ng ChIP or 0.1 ng un-enriched input DNA (WCE) as template. Log_2 enrichment was calculated from geometric means obtained from three independent ChIP experiments, each evaluated by duplicate PCR assays. Background was subtracted by normalizing over negative genomic control.

Gene Expression Analysis

Gene expression data for Ring1A/B-dKO (*Ring1A*$^{-/-}$;*Ring1B*$^{fl/fl}$; *Rosa26::CreERT2*) ES cells (2 day post-tamoxifen treatment and no-treatment control, H. Koseki unpublished data) and Eed KO ES cells (Eed −/− and control Eed+/+ ES) [13], acquired with Affymetrix Mouse Genome 430 2.0 Arrays, were normalized using the Genepattern expression data analysis package (http://www.broad.mit.edu/cancer/software/genepattern). CEL files were processed with RMA, quantile normalization and background correction [63]. For a given comparison (Ring1A/B-dKO vs control; or Eed −/− vs +/+), we only considered probes in which at least one of the experiments had a "P" significance call. Fold changes were calculated for each passing probe. Genes with multiple corresponding probes were assigned the geometric average fold change value. Gene expression data for mouse v6.5 mES and NPCs were obtained from previously published Affymetrix mRNA profiles [16].

Gene Class Enrichment Analysis

Gene ontology (GO) functional annotation for the Ring1B positive and negative sets was done using DAVID analysis tool (http://david.abcc.ncifcrf.gov/home.jsp). P-values were adjusted for multiple hypothesis testing using Bonferroni correction.

CG Content and Motif Enrichment Analysis

The HMM described above was used to define enriched intervals for each modification or chromatin protein from the mouse ES cell ChIP-Seq data. We determined the extent to which Ezh2 intervals (and those for other epitopes) overlap with CG-rich sequences. CpG island coordinates were obtained from the UCSC Genome Browser [62]. We identified all Ezh2 intervals that overlap these CpG island coordinates within 500 bp. Next, the EMBOSS analysis package [64] was used to determine the portion of remaining Ezh2 intervals overlapping a 'mini' CpG island defined as a 100 bp window with at least 50% GC content and an O:E ratio >0.6 (instead of the standard CpG island window of 200 bp).

We next classified CpG islands according to their chromatin state (e.g., Ezh2-positive v. Ezh2-negative, H3K4me3 v. bivalent). This was done by computing the median ChIP-Seq read density across each defined CpG island, and setting thresholds using a null background model of randomized reads. For these analyses we excluded CpG islands that fall within unalignable regions, typically due to low complexity sequence, and thus could not be evaluated by ChIP-Seq (<7% of all CpG islands). To maximize discriminatory power, we excluded intermediate CpG islands with sub-threshold Ezh2 signal.

We computed median values and distributions for length, CG density and observed-to-expected ratio for the different CpG island sets, and also evaluated nucleotide content by calculating the frequencies of all 16 dinucleotide combinations. Conservation scores were determined for each CpG island by aligning the regions between mouse and rat, and performing a dinucleotides level comparison of the conservation between the two species. Both CpG and non-CpG dinucleotides were conserved at slightly higher levels in the Ezh2-bound CpG islands (Figure S7).

We next screened the CpG island sets for TF motif occurrences. 668 position weight matrices (PWMs) were obtained from the Jaspar (Release 3.0 [34]) and TRANSFAC (Release 9.4; [35]) databases, excluding any non-vertebrate factors. We prepared sets of Ezh2-positive and Ezh2-negative sequences by extracting each CpG island along with flanking sequence equal to 50% of its length. The MAST algorithm [36] was then used to search for significant PWM matches ($p<5e-5$) in the Ezh2-positive and negative sets. Occurrences were length-normalized and used to calculate ratios that reflect the enrichment in the Ezh2-positive set relative to the Ezh2-negative set, or vice versa. We identified significantly over-represented motifs using Fisher's exact test with Bonferroni-adjusted p-values. These candidate motifs were then scrambled, re-scored, and excluded if any enrichment was observed in the scramble.

We used a clustering algorithm to collapse similar motifs identified as enriched in one of the sets to a single consensus sequence [65]. This was necessary due to high motif redundancy in the databases.

After clustering, all intra-cluster motif occurrences overlapping by more than 50% were counted as a single instance. Expression values for corresponding DNA binding proteins were determined from previously published Affymetrix mRNA profiles for v6.5 ES cells [16].

A simple count-based model was used to determine the extent to which motif occurrences are predictive of Ezh2 status. The motif content which allowed for maximum discrimination in mouse is as follows: a CpG island was predicted to be Ezh2-positive if it either (i) contained >8 'Ezh2-positive' motifs or (ii) contained >4 'Ezh2-positive' motifs and <2 'Ezh2-negative' motifs. Ezh2 status in human was predicted using the motifs identified in mouse but with the following metric: a CpG island was predicted to be Ezh2-positive if it contained >15 'Ezh2-positive' motifs and <2 'Ezh2-negative' motifs.

In order to quantify Ring1B presence in CpG islands, we considered the distribution of ChIP-Seq reads in control regions. We specifically used all alignable, H3K4me3-only CpG islands as our null hypothesis background model. The distribution of Ring1B ChIP-Seq read densities across these islands was calculated and a threshold was set to minimize the false positive detection rate. We then calculated Ring1B ChIP-Seq read density in sliding 200 bp windows in all Ezh2-positive CpG islands, with a CpG island assigned the maximum enrichment in any of its 200 bp windows. For maximum discriminatory power, we excluded 20% of CpG islands with sub-threshold Ring1B signal. Ring1B status was predicted using the length of CpG-richness in PRC2-positive CpG islands. Islands were predicted to be Ring1B-positive if they were either >1200 bp or within 2 kb of another CpG island.

Supporting Information

Figure S1 Comparison of chromatin states in mouse and human ES cells. (A) Conservation of H3K4me3 for 13,200 transcription start sites between human and mouse. Dashed lines indicate cutoff thresholds used to binarize the data for further analysis. Genes that carry H3K4me3 are likely to be conserved (upper right quadrant), as are those that are not marked (lower left quadrant). Less than 12% of genes are differentially methylated between human and mouse (upper left and lower right quadrants). (B) Conservation of H3K27me3 for the same regions used in (A). Most genes in both mouse and human are not marked with H3K27me3 (bottom left quadrant). Only slightly more than half the genes that carry H3K27me3 in mouse do so in human also. (upper and lower right quadrant). (C) H3K4me3 vs. H3K27me3 plotted for 17,760 mouse genes reveal three prominent marks in ESC: H3K4me3 only, (lower right quadrant), H3K4me3+H3K27me3/bivalent (upper right quadrant) and "no mark" (lower left quadrant). Very few genes are marked with H3K27me3 only (upper left quadrant).
Found at: doi:10.1371/journal.pgen.1000242.s001 (3.85 MB PDF)

Figure S2 Quantitative PCR enrichment for Ezh2 ChIP, Ring1B bioChIP and Flag-Bmi1 ChIP. (A) Plot shows Log2 ChIP-qPCR enrichment of Ezh2 in mouse v6.5 ES cells at bivalent gene promoters. Included are promoters classified as PRC2-bound (orange) or PRC2-unbound (yellow) by ChIP-Seq. (B) Plot shows Log2 enrichment of Ring1B bioChIP-qPCR in transgenic mouse ES cells expressing biotin-tagged Ring1B (mES*) at bivalent promoters classified by ChIP-Seq as PRC1-bound (purple) or PRC1-unbound (blue). H3K4me3 only genes are green. (C) Plot shows fold enrichment of Flag ChIP-qPCR in transgenic mouse ES cells expressing Flag-tagged Bmi1 (mEŜ) at bivalent promoters classified by ChIP-Seq as PRC1-bound (purple) or PRC1-unbound (blue).
Found at: doi:10.1371/journal.pgen.1000242.s002 (0.31 MB PDF)

Figure S3 Chromatin states of species-specific factors from ES cell Pathways. Divergent chromatin states of species-specific factors in transcription and signaling pathways observed in mouse and human ES cells reflect known distinctive biological functions between the two pluripotency models.
Found at: doi:10.1371/journal.pgen.1000242.s003 (0.28 MB PDF)

Figure S4 Expression analysis in PRC2 wild-type (WT) and knock-out (KO) mouse ES cells. Expression changes for all genes, Ring1B-positive bivalent and Ring1B-negative bivalent genes in PRC2 knock-out (Eed−/−) mouse ES cells.
Found at: doi:10.1371/journal.pgen.1000242.s004 (0.15 MB PDF)

Figure S5 Analysis of the CG-richness of HMM-defined intervals of H3K4me3, H3K27me3, H3K36me3, H3K9me3, H3K20me3, and Ezh2. (A) The fraction of intervals that either directly overlap or are within 500 bp of a CpG island. (B) The maximum CpG observed-to-expected ratio in any 200 bp window within the interval. The dashed line marks 0.6, one of the criteria used to define a CpG island.
Found at: doi:10.1371/journal.pgen.1000242.s005 (0.21 MB PDF)

Figure S6 Comparison of Ezh2-positive and Ezh2-negative CpG islands. No marked difference was observed in CpG observed-to-expected ratio (A), percent CpG (B), or percent GC (C), whereas Ezh2-positive CpG islands tend to be longer (median 721 bp vs 526 bp; D).
Found at: doi:10.1371/journal.pgen.1000242.s006 (0.22 MB PDF)

Figure S7 Conservation of Ezh2-bound and Ezh2-unbound dinucleotides between rat and mouse. Aligning regions in rat (rn4) for both classes of CpG island were identified, and a dinucleotide level comparison was performed on the conservation between the two species. Both non-CpG (A) and CpG (B) dinucleotides were conserved at slightly higher levels in the Ezh2-bound CpG islands than in those islands that did not bind Ezh2.
Found at: doi:10.1371/journal.pgen.1000242.s007 (0.70 MB PDF)

Figure S8 Motif clusters and their respective enrichment p-values for Ezh2-positive and Ezh2-negative CpG islands. The top ranking motifs (and their Bonferroni-corrected p-values from Fisher's exact test) for Ezh2-negative (A) and positive (B) CpG islands. The motifs were clustered and collapsed to reduce redundancy.
Found at: doi:10.1371/journal.pgen.1000242.s008 (0.49 MB PDF)

Figure S9 Length of CpG islands in Ring1B-positive and Ring1B-negative bivalent promoters. Ring1B-positive bivalent CpG islands are larger than bivalent CpG islands that are only bound by PRC2.
Found at: doi:10.1371/journal.pgen.1000242.s009 (0.12 MB PDF)

Table S1 List of ChIP-Seq datasets showing numbers of aligned reads.
Found at: doi:10.1371/journal.pgen.1000242.s010 (0.28 MB PDF)

Table S2 Chromatin states of analyzed promoters in mES cells.
Found at: doi:10.1371/journal.pgen.1000242.s011 (3.72 MB XLS)

Table S3 Chromatin states of analyzed promoters in hES cells (Microsoft Excel file).
Found at: doi:10.1371/journal.pgen.1000242.s012 (2.81 MB XLS)

Table S4 Comparison of chromatin states of analyzed promoters between mES and hES cells.
Found at: doi:10.1371/journal.pgen.1000242.s013 (1.69 MB XLS)

Table S5 PCR primers used for Ezh2, Ring1B and Flag-Bmi1 ChIP-qPCR in mouse ES cells.
Found at: doi:10.1371/journal.pgen.1000242.s014 (0.61 MB PDF)

Text S1 Supporting information on the specificity of antibodies. Western blots using mouse ES cell protein extracts demonstrate the specificity of anti-Ring1B and anti-Ezh2 (Active Motif 39103), antibodies used in this study. *Indicates the expected molecular weight. Previous publications that demonstrate the specificity of the antibodies used are listed.

Found at: doi:10.1371/journal.pgen.1000242.s015 (1.83 MB PDF)

Text S2 Relevant references for transcription factors (TFs) that correspond to implicated motifs and are active in ES cells.

Found at: doi:10.1371/journal.pgen.1000242.s016 (0.63 MB PDF)

Acknowledgments

We thank the staff of the Broad Institute Genome Sequencing Platform for assistance with reagents and data generation. We thank Manolis Kellis, Pouya Kheradpour and Alex Meissner for helpful discussions. We acknowledge L. Zagachin and the MGH RT-PCR core for assistance with quantitative PCR. We thank Miguel Vidal for Ring1A KO cells and reagents for generating the Ring1B conditional knock-down and transgenic biotin-Ring1B ES cells. The plox (pPGK-loxP-neoEGFP) plasmid was a kind gift from G. Daley.

Author Contributions

Conceived and designed the experiments: BEB. Performed the experiments: MK RPK ER ME MA. Analyzed the data: MK RPK ER EMM TSM AP XX BEB. Contributed reagents/materials/analysis tools: ME TSM CN ASC SK LMP CAC ESL HK. Wrote the paper: MK RPK ER BEB.

References

1. Jaenisch R, Young R (2008) Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell 132: 567–582.
2. Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G (2007) Genome regulation by polycomb and trithorax proteins. Cell 128: 735–745.
3. Sparmann A, van Lohuizen M (2006) Polycomb silencers control cell fate, development and cancer. Nat Rev Cancer 6: 846–856.
4. Ringrose L, Paro R (2007) Polycomb/Trithorax response elements and epigenetic memory of cell identity. Development 134: 223–232.
5. de Napoles M, Mermoud JE, Wakao R, Tang YA, Endoh M, et al. (2004) Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. Dev Cell 7: 663–676.
6. Wang H, Wang L, Erdjument-Bromage H, Vidal M, Tempst P, et al. (2004) Role of histone H2A ubiquitination in Polycomb silencing. Nature 431: 873–878.
7. Zhou W, Zhu P, Wang J, Pascual G, Ohgi KA, et al. (2008) Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. Mol Cell 29: 69–80.
8. Stock JK, Giadrossi S, Casanova M, Brookes E, Vidal M, et al. (2007) Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat Cell Biol 9: 1428–1435.
9. Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, et al. (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441: 349–353.
10. Bracken AP, Dietrich N, Pasini D, Hansen KH, Helin K (2006) Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev 20: 1123–1136.
11. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, et al. (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. Cell 125: 301–313.
12. Squazzo SL, O'Geen H, Komashko VM, Krig SR, Jin VX, et al. (2006) Suz12 binds to silenced regions of the genome in a cell-type-specific manner. Genome Res 16: 890–900.
13. Endoh M, Endo TA, Endoh T, Fujimura Y, Ohara O, et al. (2008) Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. Development 135: 1513–1524.
14. Azuara V, Perry P, Sauer S, Spivakov M, Jorgensen HF, et al. (2006) Chromatin signatures of pluripotent cell lines. Nat Cell Biol 8: 532–538.
15. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, et al. (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125: 315–326.
16. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, et al. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448: 553–560.
17. Pan G, Tian S, Nie J, Yang C, Ruotti V, et al. (2007) Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell Stem Cell 1: 299–312.
18. Zhao XD, Han X, Chew JL, Liu J, Chiu KP, et al. (2007) Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. Cell Stem Cell 1: 286–298.
19. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, et al. (2007) High-resolution profiling of histone methylations in the human genome. Cell 129: 823–837.
20. Wei CL, Miura T, Robson P, Lim SK, Xu XQ, et al. (2005) Transcriptome profiling of human and murine ESCs identifies divergent paths required to maintain the stem cell state. Stem Cells 23: 166–185.
21. Besser D (2004) Expression of nodal, lefty-a, and lefty-B in undifferentiated human embryonic stem cells requires activation of Smad2/3. J Biol Chem 279: 45076–45084.
22. Xu RH, Peck RM, Li DS, Feng X, Ludwig T, et al. (2005) Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. Nat Methods 2: 185–190.
23. Schuringa JJ, van der Schaaf S, Vellenga E, Eggen BJ, Kruijer W (2002) LIF-induced STAT3 signaling in murine versus human embryonal carcinoma (EC) cells. Exp Cell Res 274: 119–129.
24. Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, et al. (2007) New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature 448: 196–199.
25. Goldin SN, Papaioannou VE (2003) Paracrine action of FGF4 during periimplantation development maintains trophectoderm and primitive endoderm. Genesis 36: 40–47.
26. Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, et al. (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet 38: 431–440.
27. Bernstein BE, Meissner A, Lander ES (2007) The mammalian epigenome. Cell 128: 669–681.
28. Surani MA, Hayashi K, Hajkova P (2007) Genetic and epigenetic regulators of pluripotency. Cell 128: 747–762.
29. Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, et al. (2008) Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol Cell 30: 755–766.
30. Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, et al. (2005) Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol 3: e7.
31. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438: 803–819.
32. Tanay A, O'Donnell AH, Damelin M, Bestor TH (2007) Hyperconserved CpG domains underlie Polycomb-binding sites. Proc Natl Acad Sci U S A 104: 5521–5526.
33. Bernstein E, Duncan EM, Masui O, Gil J, Heard E, et al. (2006) Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol Cell Biol 26: 2560–2569.
34. Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res 32: D91–94.
35. Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, et al. (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res 31: 374–378.
36. Bailey TL, Gribskov M (1998) Combining evidence using p-values: application to sequence homology searches. Bioinformatics 14: 48–54.
37. Weintraub H, Davis R, Tapscott S, Thayer M, Krause M, et al. (1991) The myoD gene family: nodal point during specification of the muscle cell lineage. Science 251: 761–766.
38. Yamada Y, Warren AJ, Dobson C, Forster A, Pannell R, et al. (1998) The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis. Proc Natl Acad Sci U S A 95: 3890–3895.
39. Donahue LM, Reinhart AJ (1998) POU domain genes are differentially expressed in the early stages after lineage commitment of the PNS-derived stem cell line, RT4-AC. Brain Res Dev Brain Res 106: 1–12.
40. Yamaguchi TP, Takada S, Yoshikawa Y, Wu N, McMahon AP (1999) T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. Genes Dev 13: 3185–3190.
41. Singh SK, Kagalwala MN, Parker-Thornburg J, Adams H, Majumder S (2008) REST maintains self-renewal and pluripotency of embryonic stem cells. Nature 453: 223–227.
42. Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. Science 316: 1497–1502.
43. Ellis T, Gambardella L, Horcher M, Tschanz S, Capol J, et al. (2001) The transcriptional repressor CDP (Cutl1) is essential for epithelial cell differentiation of the lung and the hair follicle. Genes Dev 15: 2307–2319.
44. Hayden MS, Ghosh S (2004) Signaling to NF-kappaB. Genes Dev 18: 2195–2224.

45. Torres J, Watt FM (2008) Nanog maintains pluripotency of mouse embryonic stem cells by inhibiting NFkappaB and cooperating with Stat3. Nat Cell Biol 10: 194–201.
46. Kim J, Chu J, Shen X, Wang J, Orkin SH (2008) An extended transcriptional network for pluripotency of embryonic stem cells. Cell 132: 1049–1061.
47. Grskovic M, Chaivorapol C, Gaspar-Maia A, Li H, Ramalho-Santos M (2007) Systematic identification of cis-regulatory sequences active in mouse and human embryonic stem cells. PLoS Genet 3: e145.
48. Hollenhorst PC, Shah AA, Hopkins C, Graves BJ (2007) Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. Genes Dev 21: 1882–1894.
49. Negre N, Hennetin J, Sun LV, Lavrov S, Bellis M, et al. (2006) Chromosomal distribution of PcG proteins during Drosophila development. PLoS Biol 4: e170.
50. Schwartz YB, Kahn TG, Nix DA, Li XY, Bourgon R, et al. (2006) Genome-wide analysis of Polycomb targets in Drosophila melanogaster. Nat Genet 38: 700–705.
51. Tolhuis B, de Wit E, Muijrers I, Teunissen H, Talhout W, et al. (2006) Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster. Nat Genet 38: 694–699.
52. Voo KS, Carlone DL, Jacobsen BM, Flodin A, Skalnik DG (2000) Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. Mol Cell Biol 20: 2108–2121.
53. Birke M, Schreiner S, Garcia-Cuellar MP, Mahr K, Titgemeyer F, et al. (2002) The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. Nucleic Acids Res 30: 958–965.
54. Pasini D, Bracken AP, Hansen JB, Capillo M, Helin K (2007) The polycomb group protein Suz12 is required for embryonic stem cell differentiation. Mol Cell Biol 27: 3769–3779.
55. Wernig M, Meissner A, Foreman R, Brambrink T, Ku M, et al. (2007) In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature.
56. Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, et al. (2007) Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell 1: 55–70.
57. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, et al. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454: 766–770.
58. Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, et al. (1998) Embryonic stem cell lines derived from human blastocysts. Science 282: 1145–1147.
59. Kyba M, Perlingeiro RC, Daley GQ (2002) HoxB4 confers definitive lymphoid-myeloid engraftment potential on embryonic stem cell and yolk sac hematopoietic progenitors. Cell 109: 29–37.
60. Weinmann AS, Bartley SM, Zhang T, Zhang MQ, Farnham PJ (2001) Use of chromatin immunoprecipitation to clone novel E2F target promoters. Mol Cell Biol 21: 6820–6832.
61. Atsuta T, Fujimura S, Moriya H, Vidal M, Akasaka T, et al. (2001) Production of monoclonal antibodies against mammalian Ring1B proteins. Hybridoma 20: 43–46.
62. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, et al. (2002) The human genome browser at UCSC. Genome Res 12: 996–1006.
63. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, et al. (2006) GenePattern 2.0. Nat Genet 38: 500–501.
64. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16: 276–277.
65. Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, et al. (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. Nature 434: 338–345.

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 283, NO. 51, pp. 35805–35814, December 19, 2008

Nuclear Inhibitor of Protein Phosphatase-1 (NIPP1) Directs Protein Phosphatase-1 (PP1) to Dephosphorylate the U2 Small Nuclear Ribonucleoprotein Particle (snRNP) Component, Spliceosome-associated Protein 155 (Sap155)*[S]

Received for publication, July 17, 2008, and in revised form, October 1, 2008 Published, JBC Papers in Press, October 8, 2008, DOI 10.1074/jbc.M805468200

Nobuhiro Tanuma[‡§1], **Sei-Eun Kim**[§2], **Monique Beullens**[¶], **Yao Tsubaki**[§], **Shinya Mitsuhashi**[§‖], **Miyuki Nomura**[‡], **Takeshi Kawamura**[**], **Kyoichi Isono**[‡‡], **Haruhiko Koseki**[‡‡], **Masami Sato**[‡], **Mathieu Bollen**[¶], **Kunimi Kikuchi**[§], **and Hiroshi Shima**[‡3]

From the [‡]*Division of Cancer Chemotherapy, Miyagi Cancer Center Research Institute, Natori 981-1293, Japan,* [§]*Division of Biochemical Oncology and Immunology, Institute for Genetic Medicine, and* [‖]*Division of Applied Bioscience, Research Faculty of Agriculture, Hokkaido University, Sapporo 060-0815, Japan,* [¶]*Laboratory of Biosignaling and Therapeutics, Department of Molecular Cell Biology, Faculty of Medicine, KULeuven, Leuven B-3000, Belgium,* [**]*Laboratory for Systems Biology and Medicine, RCAST, The University of Tokyo, Tokyo 153-8904, Japan, and* [‡‡]*RIKEN Research Center for Allergy and Immunology, Yokohama 230-0045, Japan*

Pre-mRNA splicing entails reversible phosphorylation of spliceosomal proteins. Recent work has revealed essential roles for Ser/Thr phosphatases, such as protein phosphatase-1 (PP1), in splicing, but how these phosphatases are regulated is largely unknown. We show that nuclear inhibitor of PP1 (NIPP1), a major PP1 interactor in the vertebrate nucleus, recruits PP1 to Sap155 (spliceosome-associated protein 155), an essential component of U2 small nuclear ribonucleoprotein particles, and promotes Sap155 dephosphorylation. C-terminally truncated NIPP1 (NIPP1-ΔC) formed a hyper-active holoenzyme with PP1, rendering PP1 minimally phosphorylated on an inhibitory site. Forced expression of NIPP1-WT and -ΔC resulted in slight and severe decreases in Sap155 hyperphosphorylation, respectively, and the latter was accompanied with inhibition of splicing. PP1 overexpression produced similar effects, whereas small interfering RNA-mediated NIPP1 knockdown enhanced Sap155 hyperphosphorylation upon okadaic acid treatment. NIPP1 did not inhibit but rather stimulated Sap155 dephosphorylation by PP1 *in vitro* through facilitating Sap155/PP1 interaction. Further analysis revealed that NIPP1 specifically recognizes hyperphosphorylated Sap155 thorough its Forkhead-associated domain and dissociates from Sap155 after dephosphorylation by associated PP1. Thus NIPP1 works as a molecular sensor for PP1 to recognize phosphorylated Sap155.

Pre-mRNA splicing is an essential step for expression of most genes in metazoans. Intron excision from a nascent transcript is achieved by pre-mRNA splicing catalyzed by the spliceosome, a macromolecular complex consisting of five small nuclear ribonucleoprotein particles (snRNPs)[4] and a large number of non-snRNP proteins. Spliceosome assembly is an ordered process that includes stepwise recruitment of U1, U2, U5, and U4/6 snRNPs on a pre-mRNA and sequential formation of complex E → A/B → B* → C. The activated B* spliceosome catalyzes step I of splicing, whereas the C complex catalyzes step II. During and after splicing, spliceosome components dissociate and are recycled for further rounds of splicing. Spliceosome assembly/disassembly and splicing catalysis are thought to be regulated in part by reversible phosphorylation of spliceosomal proteins (1–3).

U2 snRNP includes U2 snRNA and two heteromeric protein complexes, Sf3a and Sf3b. Sap155, also known as Sf3b1 or Sf3b155, is a component of the Sf3b and becomes hyperphosphorylated concomitant with or just after the first catalytic step of splicing *in vitro* (4). A recent study reveals that Sf3a/b proteins are destabilized and dissociate from the RNP core of the activated spliceosome during the transition from the B to C complex (5). Although Sf3a and Sf3b are essential early in the splicing reaction, they are apparently not required for the second catalytic step. Currently, it is not known what triggers exchange of proteins during spliceosome transitions.

Shi *et al.* (6) reported that the protein Ser/Thr phosphatase (PPase) type 1 (PP1) and/or type 2A (PP2A) are essential for splicing *in vitro*, in particular, at the second catalytic step. They also observed that Sap155 and U5-116k are potential substrates of these two PPases, suggesting the importance of dephosphorylation of spliceosomal proteins in spliceosome structural

---

* This work was supported in part by grants-in-aid for scientific research (B and C) provided by the Japan Society for the Promotion of Science of Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. 1–3.

[1] To whom correspondence may be addressed. Fax: 81-22-3811196; E-mail: tanuma-no872@pref.miyagi.jp.

[2] Present address: Clinical Research Team, Hanmi Pharm. Co. Ltd., Seoul, Korea.

[3] To whom correspondence may be addressed: 47-1 Nodayama, Medeshima-Shiode, Natori 981-1293, Japan. Fax: 81-22-3811196; E-mail: shima-hi632@pref.miyagi.jp.

[4] The abbreviations used are: snRNP, small nuclear ribonucleoprotein particle; RNP, ribonucleoprotein; PP1, protein phosphatase-1; NIPP1, nuclear inhibitor of PP1; siRNA, small interfering RNA; PBS, phosphate-buffered saline; Dox, doxycycline; RNAi, RNA interference; OA, okadaic acid; PPase, phosphatase; HTO, HeLa-Tet-Off; RPA, RNase protection assay; FHA, Forkhead-associated; RT, reverse transcription; FHA, Forkhead-associated.

rearrangement. However, how PP1 and PP2A regulate phosphorylation of spliceosomal proteins is not well understood.

Both PP1 and PP2A are often targeted to specific substrates by binding partners. For example, PP1 is regulated *in vivo* primarily by interaction with various regulatory subunits that dictate subcellular localization, substrate specificity, or enzymatic activity (7). In some cases, regulatory subunits are phosphorylated/dephosphorylated, resulting in activity changes that modulate PP1 activity in response to environmental cues such as hormones and second messengers (8–10). In addition to interacting with regulatory subunits, PP1 is regulated by phosphorylation. Phosphorylation by cyclin-dependent kinases lower PP1 activity *in vitro* and *vivo* in *Schizosaccharomyces pombe* and in human cells (11, 12). The PP1 phosphorylation site is highly conserved among species and has been mapped to the C terminus. Studies in yeast and human cells suggest that PP1 phosphorylation is essential for proper cell cycle progression (13, 14).

A major nuclear interaction partner of PP1 in vertebrate cells is the nuclear inhibitor of PP1 (NIPP1) (7). Targeted disruption of the *Nipp1* gene in mice results in embryonic lethality, indicating an essential role *in vivo* (15). NIPP1 has two PP1 binding regions as follows: a central domain containing a so-called RVXF-motif, which represents a high affinity binding motif and a low affinity binding region mapped to the NIPP1 C terminus (16, 17). The precise function of the latter interaction is not yet understood. Previous studies suggest the importance of NIPP1 in regulating pre-mRNA splicing. NIPP1 localizes in nuclear speckles, domains rich in splicing machinery components. Several splicing-related proteins such as Cdc5L, SAP155, and MELK have been identified as potential ligands of the Forkhead-associated (FHA) domain of NIPP1 (18–20). FHA is a modular phospho-Thr/Ser recognition motif (21, 22). Beullens and Bollen (23) reported that NIPP1 physically associates with spliceosomes through the FHA domain and that a dominant negative form of NIPP1 blocks the B to C transition of spliceosome assembly in the HeLa cell nuclear extracts by an unknown mechanism. Although there is a report that NIPP1 protein lacking the RVXF motif partially compromises splicing *in vitro*, data from splicing assays using HeLa nuclear extracts are contradictory (23, 24). Thus both the mode of NIPP1 action and the potential role of PP1/NIPP1 interaction in pre-mRNA splicing remain obscure.

To analyze function of NIPP1 in pre-mRNA splicing, we reevaluated the physical and functional association among pre-mRNA/RNA, NIPP1, and PP1 *in vivo* using cell-based assays. We show that NIPP1 works as a targeting subunit that directs PP1 to dephosphorylate Sap155, not as an inhibitor of PP1. Our results suggest a regulatory mechanism ensuring proper Sap155 phosphorylation-dephosphorylation cycles during pre-mRNA splicing.

MATERIALS AND METHODS

Antibodies—Anti-U5-116k antibody was kindly provided by Dr. H. Brahms and Dr. R. Lührmann (25). Anti-Sap155 (26), anti-Myc 9e10, anti-Cdc5L, anti-FLAG M2, anti-phospho-Thr-320-PP1, anti-MELK, and anti-pan-PP1 E9 antibody were purchased from MBL (Nagoya, Japan), Roche Applied Science, BD Transduction Laboratories, Sigma, Cell Signaling Technology Inc. (Danvers, MA), Abcam (Cambridge, UK), and Santa Cruz Biotechnology Inc. (Santa Cruz, CA), respectively. Antibodies specific to each PP1 isoform have been described (27). Polyclonal or monoclonal anti-NIPP1 antibodies were generated against a peptide corresponding to the C-terminal 20 amino acids in our laboratory or purchased from BD Transduction Laboratories.

Plasmid Constructs—The pTRE-β-globin reporter has been described (28). The BseRI-BstXI region of pTRE-β-globin, encompassing the β-globin gene from the 3′-half of exon 1 to the 5′-half of exon 3, was replaced with a corresponding BseRI-BstXI fragment of β-globin cDNA derived from pTRE-β-globin; the sequence verified, and the resulting plasmid was designated pTRE-intronless-β-globin. Its nucleotide sequence is identical to that of pTRE-β-globin except intron 1 and 2 sequences are excluded. NIPP1 expression plasmids were constructed by subcloning cDNAs encoding wild-type or mutant forms of rat NIPP1 (29) into pCMV-FLAG2 (Sigma). Myc-PP1 constructs are described elsewhere (30).

Cell Culture and Transfection—HeLa-Tet-Off (HTO) cells (Clontech) were cultured in Dulbecco's modified Eagle's medium containing 10% fetal calf serum and 250 μg/ml geneticin. Transfections were done using FuGENE 6 reagent (Roche Applied Science) according to the manufacturer's recommendation. For stable transfection, cDNAs encoding FLAG-tagged wild-type or mutant NIPP1 were subcloned into the pTRE vector (Clontech). HTO cells in 10-cm dishes were transfected with 10 μg of pTRE-NIPP1-WT or -ΔC with 0.5 μg of pTK-Hyg (Clontech). Drug-resistant clones were selected in 200 μg/ml hygromycin, 250 μg/ml geneticin, and 1 μg/ml Dox. Two weeks later, ~150 independent clones for each construct were screened for expression of exogenous NIPP1 by immunoblotting with anti-FLAG antibody. Several clones for each construct (four for WT NIPP1 and three for NIPP1-ΔC) exhibited low and high levels of FLAG-NIPP1 proteins in the presence and absence of Dox, respectively, and were selected for further experiments. After establishment, stable clones were maintained in medium containing 5–10 ng/ml Dox.

In Vitro and in Vivo Splicing Assays—*In vitro* splicing was performed as described (23). RNP immunoprecipitation assays were performed essentially as described (31) with slight modification. Cells were fixed in 0.75% formaldehyde/PBS at room temperature for 10 min, washed twice in PBS, and harvested. Cells were lysed in 1 ml of RIPA buffer (20 mM Tris-Cl, 150 mM NaCl, 5 mM EDTA, 10% glycerol, 0.1% SDS, 1% Triton X-100, and 0.1% deoxycholate) by sonication (Sonifier 450, Branson, Danbury, CT). 50 μl each of lysate was stored on ice as "Input," and the rest was immunoprecipitated with anti-FLAG-agarose beads at room temperature for 2 h. Beads were collected by centrifugation and washed four times with gentle rotation for 5 min, followed by elution of RNPs with FLAG peptide (Sigma). "Inputs" and eluates were heated to 75 °C for 90 min to reverse cross-links and phenol-extracted twice. RNAs were ethanol-precipitated using yeast tRNA and Glyco-blue co-precipitants (Ambion) as carrier and treated with RNase-free DNase (DNA-free turbo, Ambion). RNA was reverse-transcribed using SuperScript III RTase (Invitrogen) using the RT primer 5′-

GTGCAGCTTGTCACAGTGCAG-3′. PCR was undertaken using standard protocols with Ex Taq-HS polymerase (Takara Bio Inc., Otsu, Japan) and the following primers for β-globin: sense, 5′-ATGGTGCATCTGACTCCTGAG-3′; antisense, 5′-ACTAAAGGCACCGAGCACTTTCTTG-3′. To analyze effects of NIPP1 on splicing rates, HTO cells in 10-cm dishes were transfected with 7.5 μg of NIPP1 expression plasmids with 2.5 μg of pTRE-β-globin. Four hours later, cells were split and further cultured without Dox for 20 h. At indicated time points after Dox addition, cells were harvested, and total RNA was isolated for Northern blot analysis with a ^{32}P-labeled β-globin probe. For RPAs, antisense riboprobes complementary to exon 1/intron 1 or exon 2/intron 2 boundaries of β-globin gene were synthesized by *in vitro* transcription using MAXI script (Ambion, Austin, TX) and T7 polymerase. Hybridization and RNase digestions were performed using the RPA III (Ambion) kit according to the manufacturer's instruction. Protected RNAs were subjected to denaturing PAGE and visualized using the FLA system (Fuji Film Co., Tokyo, Japan).

Immunoprecipitation and Western Blot Analysis—Cells were lysed in RIPA or SDS buffer (125 mM Tris-Cl, pH 6.8, 1% SDS) by sonication using a "Bio-ruptor" sonicator (CosmoBio, Tokyo, Japan). For co-immunoprecipitation, cells were lysed in buffer (50 mM Tris-Cl, pH 7.4, 125 mM NaCl, 0.2% Triton X-100, 2 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, leupeptin, and aprotinin) by sonication. Lysates were incubated with anti-FLAG-agarose or anti-Myc-agarose (Sigma) for 3–4 h at 4 °C. Immunocomplexes were collected by centrifugation, washed three times with lysis buffer without protease inhibitors, and immunoblotted.

Immunohistochemistry—Cells grown on collagen-coated coverslips were fixed in PBS containing 3.7% formaldehyde for 10 min. After permeabilization with 0.2% Triton X-100, cells were incubated in PBS containing 5% (w/v) bovine serum albumin and 0.5 μg/ml anti-FLAG M2 monoclonal antibody overnight at 4 °C. Immunoreactants were further visualized using Alexa Fluor488-conjugated anti-mouse IgG secondary antibody (Invitrogen), and images were acquired using a Pascal confocal laser-scanning microscope (Zeiss).

^{32}P-Orthophosphate Labeling—Stable clones were cultured in the presence or absence of 10 ng/ml Dox for 24 h and then in phosphate-free medium (Invitrogen) containing 10% fetal calf serum (dialyzed three times against 100 volumes of Hepes buffer) and ^{32}P-orthophosphate (PerkinElmer Life Sciences) at 1 mCi/ml for an additional 4 h. Cells were harvested, lysed by sonication using the Bio-ruptor sonicating machine, and subjected to immunoprecipitation. Immunoprecipitates were separated by SDS-PAGE, followed by immunoblot analysis and autoradiography using the FLA system.

In Vitro De-phosphorylation Experiments—HTO cells were lysed in buffer (50 mM Tris-Cl, 100 mM NaCl, 2 mM EDTA, 10% EDTA, 0.1% Triton X-100) by sonication. Lysates were incubated with or without either PP1 (a recombinant PP1α isoform; Calbiochem) or NIPP1-PP1 holoenzyme at 30 °C for 10–60 min. The NIPP1-PP1 holoenzyme was constituted by preincubation of recombinant PP1α with His-NIPP1 at several ratios on ice for 20 min in Tris-buffered saline. His-NIPP1 was expressed in *Escherichia coli* using the pET system (Novagen, San Diego, CA) and purified using TALON beads (Clontech) following the manufacturer's recommendations. Sap155 was immunopurified from lysates of HeLa cells treated with 100 nM okadaic acid for 6–8 h using anti-Sap155 monoclonal antibody-conjugated agarose.

Phosphatase Assay—Immunoprecipitates with anti-FLAG-agarose were eluted twice by incubating beads with 200 μg/ml FLAG peptide. Eluates were subjected to a phosphatase assay and Western blotting. PP1 activity measurement was as described (26).

RNAi Experiments—siRNA duplexes against human NIPP1 (Stealth RNAi) were purchased from Invitrogen (HSS143426 and HSS143427). Stealth RNAi Negative control Medium GC duplex (Invitrogen) served as control. siRNA transfection was undertaken using Lipofectamine RNAiMAX (Invitrogen) according to the manufacturer's instructions at final siRNA concentrations of 5 nM in culture.

Overlay Assay—Sap155 was immunoprecipitated from HeLa cell lysates treated with 100 nM okadaic acid (OA) for 6–8 h, separated by SDS-PAGE, and transferred to a nitrocellulose membrane. The membrane was blocked with 5% bovine serum albumin in PBS and reacted with solutions supplemented with phosphatase inhibitors (Roche Applied Science) at 4 °C for 1 h. Membranes were washed with PBS, 0.1% Tween plus phosphatase inhibitors at 4 °C for 5 min twice and then UV-irradiated to fix complexes.

Statistical Analysis—One-way analysis of variance combined with Tukey's test was used to analyze data with unequal variance between each group. A probability level of 0.05 was considered significant.

RESULTS

Physical and Functional Interaction between NIPP1 and Pre-mRNA Splicing in Intact Cells—To elucidate interaction between NIPP1 and the spliceosome *in vivo*, we performed an RNP immunoprecipitation assay (31). Cells transfected with FLAG-NIPP1 (Fig. 1*A*) together with the β-globin reporter plasmid pTRE-β-globin were treated with formaldehyde, recovered, and lysed. RNPs containing FLAG-NIPP1 were immunoprecipitated using an anti-FLAG antibody. The cross-links were reversed, and the immunoprecipitated RNAs were detected by RT-PCR ("Minus RT" controls are presented as supplemental Fig. 1). As shown in Fig. 1*B*, both unspliced and spliced β-globin RNAs co-immunoprecipitated with FLAG-NIPP1-WT. We next examined interaction between NIPP1 mutants (Fig. 1*A*) and β-globin pre-mRNA/mRNAs. Similar results were obtained with NIPP1-R53A, which has a nonfunctional FHA domain (Fig. 1*C*). NIPP1-V201A/F203A (NIPP1-RATA), which is deficient in PP1 binding (Fig. 1*D*), reproducibly co-immunoprecipitated pre-mRNA more efficiently than did NIPP1-WT. NIPP1-ΔC, which lacks the C-terminal 22 amino acids constituting a second PP1-binding region, interacted almost exclusively with unspliced pre-mRNA. This observation correlated with accumulation of unspliced reporter pre-mRNA in cell lysates, as detected by RT-PCR ("total RNA" in Fig. 1*B*), suggesting inhibition of splicing by that mutant form of NIPP1. Because transcription and "post-transcriptional" RNA processing are functionally coupled (32), the

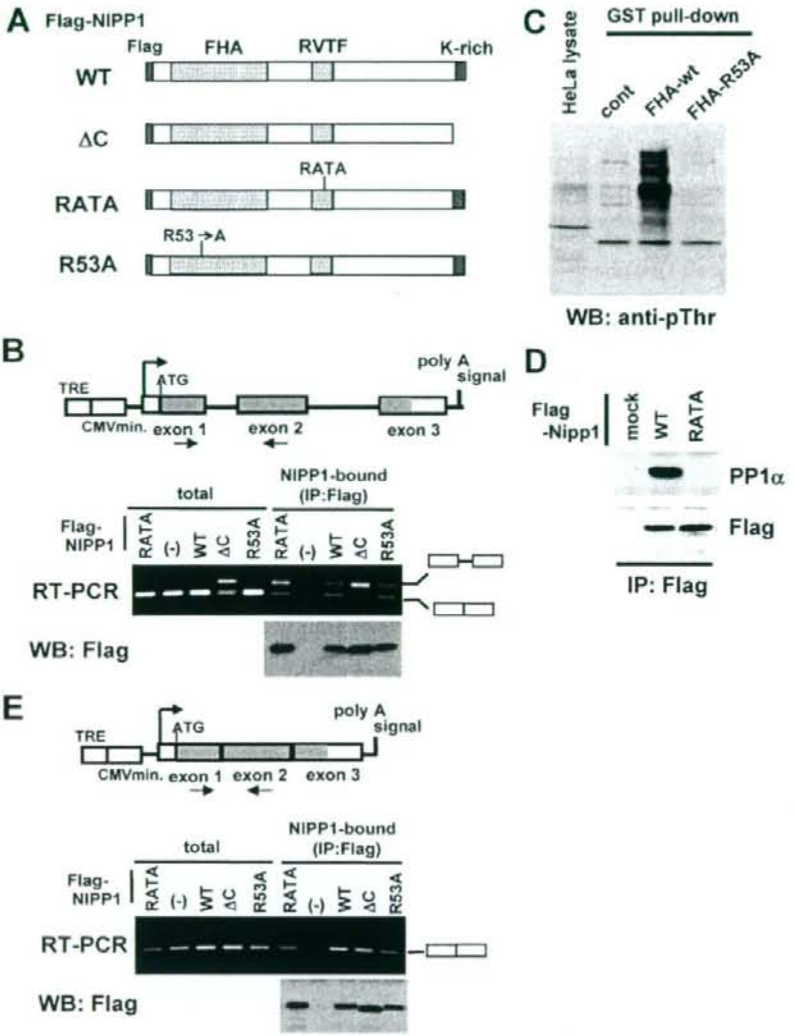


FIGURE 1. **Physical association of NIPP1 and pre-mRNA/mRNA of β-globin reporter genes in intact cells.** *A*, NIPP1 mutant proteins used in this study. Each is N-terminally tagged with the FLAG epitope. The FHA domain and PP1-binding region containing RVTF sequence and Lys (*K*)-rich region of NIPP1 are shown. *B*, *in vivo* interaction between NIPP1 and β-globin reporter pre-mRNA/mRNA. Structure of the human β-globin gene reporter construct, pTRE-β-globin, is shown on the *top*. The open reading frame is represented by *shaded boxes*, and introns and vector stuffer sequences are represented by *lines*. *Arrows* indicate position of primers used in RT-PCR analysis. HTO cells were transiently transfected with FLAG-NIPP1 expression vector with the β-globin reporter plasmid. About 24 h after transfection, cells were subjected to RNP immunoprecipitation assay. Co-immunoprecipitated reporter pre-mRNA/mRNAs with FLAG-NIPP1 were detected by RT-PCR. Positions of amplified DNA fragments derived from unspliced pre-mRNA and spliced mRNA are shown at *right*. Levels of immunoprecipitated FLAG-NIPP1 proteins were also assessed by immunoblotting (*WB*). *C*, pulldown experiments were performed using GST-tagged FHA domains of NIPP1. GST, GST-FHA-WT, or GST-FHA-R53A was mixed with whole cell lysates of HeLa cells. Co-precipitated proteins were subjected to immunoblotting using anti-phospho-Thr antibody. *Cont*, control. *D*, HeLa-TetOff cells were transiently transfected with NIPP1-WT or NIPP1-RATA, lysed, and subjected to immunoprecipitation (*IP*) using an anti-FLAG antibody. Immunoblots were performed using anti-PP1α or anti-FLAG antibodies. *E*, effects of wild-type and mutant forms of NIPP1 on mRNA of an intron-less β-globin reporter gene. Structure of the intron-less human β-globin reporter construct is at the *top*. Interaction between NIPP1 and the intron-less reporter mRNA was analyzed as in *B*.

effects of NIPP1-ΔC on splicing might represent an indirect transcriptional effect. To distinguish between these possibilities, we repeated the assays using a corresponding intron-less reporter gene. As shown in Fig. 1*E*, NIPP1-ΔC and other NIPP1 constructs did not significantly affect mRNA levels of the intron-less reporter, suggesting that the effect of NIPP1-ΔC on splicing is specific for a splicing-related event(s) or for splicing itself. Intriguingly, all NIPP1 variants co-precipitated mRNA derived from the intron-less reporter, consistent with a recent report that the U2 snRNP is recruited to both intron-containing transcripts and transcripts of intron-less genes, such as histone genes, to facilitate 3′-processing (33).

To further delineate effects of mutant NIPP1, we took advantage of the pTRE-β-globin reporter, in which the transcription is regulated by a TRE sequence upstream of the cytomegalovirus promoter (Fig. 1*B*). When introduced into cells expressing the Tet-repressor (Tet-Off cell lines), *de novo* transcription from the reporter is specifically and rapidly suppressed by the stable tetracycline analogue Dox (28). Northern blot analysis showed that expression of NIPP1-ΔC decreased steady-state levels of β-globin mRNA, whereas expression of NIPP1-WT had produced little if no effect (Fig. 2*A*). A band likely representing β-globin pre-mRNA (indicated by *asterisk* in Fig. 2*A*) migrated more slowly than β-globin mRNA and was more abundant in cells transfected with NIPP1-ΔC compared with mock- or NIPP1-WT-transfected cells. To confirm this observation, we performed an RPA (Fig. 2*C*; see supplemental Fig. 2 for assay design) and found that in mock-transfected cells levels of unspliced β-globin pre-mRNA were rapidly decreased after Dox-mediated block of *de novo* transcription of the reporter gene, as seen in Fig. 2*C*. This decrease is likely due to splicing of the reporter pre-mRNA. Strikingly, unspliced pre-mRNA was more abundant in cells transfected with NIPP1-ΔC, whereas mature spliced mRNA levels were decreased (Fig. 2*C*). Additionally, the half-life of unspliced pre-mRNA in NIPP1-ΔC-transfected cells appeared much longer than that seen in mock- and NIPP1-WT-transfected cells

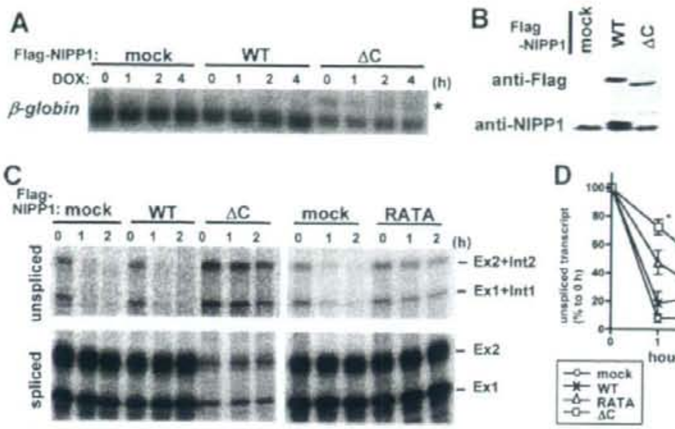


FIGURE 2. **Effects of NIPP1 mutants on splicing *in vivo* and *in vitro*.** *A* and *B*, cells were transiently transfected with FLAG-NIPP1 expression vector and the pTRE-β-globin plasmid. After 24 h, cells were harvested at indicated time points after addition of 10 ng/ml Dox to the medium. Total RNA was isolated and analyzed by Northern blotting with β-globin cDNA (*A*). The lysates of cells at time 0 were analyzed by Western blotting using the anti-FLAG (*upper*) and anti-NIPP1 antibody, which recognizes the Lys-rich region of NIPP1 (*lower*) (*B*). *C* and *D*, decay of unspliced pre-mRNA after the shut-off of *de novo* transcription. Cells were transiently transfected, treated with Dox, and harvested as in *A*. Total RNA was isolated and analyzed by a multiprobe RPA (*C*). Migration of protected fragments derived from unspliced and spliced mRNAs is shown. Band intensities derived from the unspliced transcript were evaluated and are shown as values relative to those at 0 h (*D*). Data represent mean values of 3–4 independent RPAs with S.D. *, $p < 0.001$; **, $p < 0.01$ (ΔC *versus* WT).

(Fig. 2*D*). As described, NIPP1-RATA associated with unspliced pre-mRNA more efficiently than did NIPP1-WT, whereas NIPP1-RATA did not accumulate unspliced pre-mRNA, in contrast to NIPP1-ΔC (Fig. 1*B*). It is noteworthy that decay of unspliced pre-mRNA of NIPP1-RATA-transfected cells was slightly slower than that seen in mock-transfected cells (Fig. 2, *C* and *D*).

Taken together, these results demonstrate physical and functional association between NIPP1 and pre-mRNA splicing and suggest the importance of the NIPP1 C terminus and interaction between NIPP1 and PP1 in NIPP1-mediated regulation of splicing *in vivo*. When we performed an *in vitro* splicing assay in HeLa nuclear extracts, NIPP1-ΔC showed no effect as reported (Ref. 23 and data not shown). Based on these findings, we chose to evaluate NIPP1 function in cell-based rather than *in vitro* experiments.

Characterization of Splicing Inhibition by NIPP1-ΔC—To understand functional differences between NIPP1-ΔC and wild-type protein, we first investigated NIPP1-ΔC subcellular localization and found it indistinguishable from that of NIPP1-WT (Fig. 3*A*). We next prepared NIPP1 constructs containing point mutations in addition to the ΔC deletion to investigate domain requirements for splicing inhibition (Fig. 3*B*). The effects of these mutations on splicing of the β-globin construct were determined by analyzing accumulation of unspliced pre-mRNA. An RPA analysis revealed that NIPP1-ΔC, but not NIPP1-R53A/ΔC or NIPP1-RATA/ΔC, inhibited pre-mRNA splicing of the reporter gene (Fig. 3*C*). These results indicate that inhibition of splicing by NIPP1-ΔC requires both a functional FHA domain and interaction with PP1. Neither the R53A nor RATA mutation alone promoted accumulation of unspliced pre-mRNA (Fig. 1*B* and data not shown).

Because splicing inhibition by NIPP1-ΔC requires interaction with PP1, we compared NIPP1-WT and -ΔC mutant in terms of association with PP1. NIPP1-WT and NIPP1-ΔC were immunoprecipitated, and PP1 activities were evaluated using phosphorylase *a* as a substrate. PP1 activity complexed with NIPP1-ΔC was reproducibly higher than that seen with NIPP1-WT (Fig. 4*A*). Similar amounts of PP1 in immunoprecipitates of NIPP1-WT or -ΔC were confirmed by immunoblotting with anti-PP1 antibody (Fig. 4*B*). We reasoned that differences seen in Fig. 4*A* could be due to potential modification of the catalytic subunit PP1. It is well established that PP1 is phosphorylated by cyclin-dependent kinase(s) on a critical Thr residue in the C terminus (Thr-320 of PP1α) and inactivated (11, 12). NIPP1-WT or NIPP1-ΔC was immunoprecipitated from transfected cells, and the inhibitory phosphorylation of endogenous PP1 associated with NIPP1 was examined and compared. Strikingly, PP1 co-immunoprecipitated with NIPP1-WT was highly phosphorylated on the inhibitory site, whereas PP1 in association with NIPP-ΔC was minimally phosphorylated (Fig. 4*B*). Thus, the higher specific activity of PP1 associated with NIPP1-ΔC could be due to lower levels of inhibitory phosphorylation of PP1, rendering it constitutively active. At present, we cannot differentiate the levels of inhibitory phosphorylation in each PP1 isoforms because the phospho-Thr-320-PP1α antibody may also cross-react with other isoforms phosphorylated on the corresponding residue. Nevertheless, given that NIPP1 preferentially associates with PP1α (see below), it is likely that phospho-PP1 co-immunoprecipitating with NIPP1 is primarily PP1α.

NIPP1-associated PP1 Regulates Sap155 Phosphorylation—To further analyze NIPP1 actions in pre-mRNA splicing, stable cell lines conditionally expressing WT and mutant forms of NIPP1 (HeLa-TetOff (HTO)-NIPP1 clones) were developed. In these clones, FLAG-NIPP1 proteins were detected 8 h after removal of Dox, and levels were maximal at about 24 h (Fig. 5*A*). Previous studies had identified several potential effectors of NIPP1 and/or PP1 such as Cdc5L, Sap155, and U5-116k (6, 18, 19, 34). Levels of phosphorylation of these proteins in HTO-NIPP1 clones were investigated by metabolically labeling cells with ^{32}P-orthophosphate. We found that Sap155 hyperphosphorylation was slightly decreased in cells expressing WT-NIPP1 (Fig. 5*B*). Importantly, this decrease was more apparent in cells expressing NIPP1-ΔC. Induction of NIPP1-ΔC expression by Dox removal caused virtual loss of Sap155 hyperphosphorylation. The slight decrease in Sap155 hyperphosphorylation of HTO-NIPP1-ΔC cells in the presence of Dox was probably due to leaky expression of the construct in this system. In contrast, neither Cdc5L nor U5-116k phosphorylation levels were affected by NIPP1-WT or

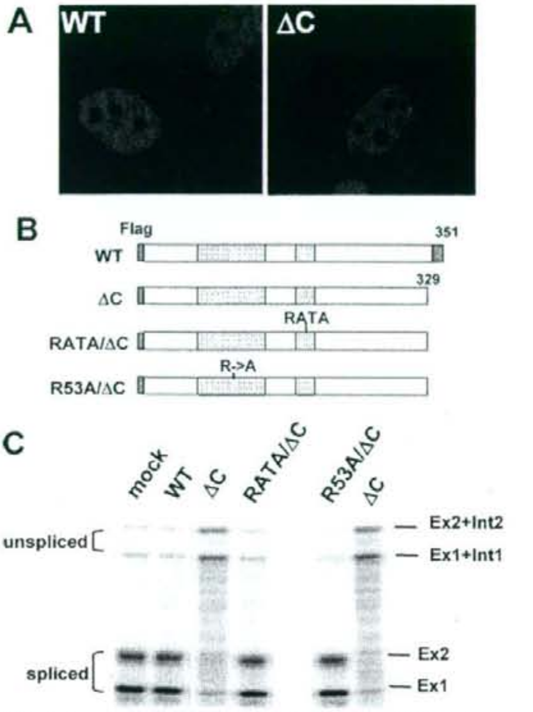


FIGURE 3. **Characterization of splicing inhibition by NIPP1-ΔC.** *A*, normal subcellular localization of NIPP1-ΔC. HTO cells were transiently transfected with NIPP1-WT or NIPP1-ΔC and subjected to immunostaining with the anti-FLAG antibody followed by Alexa Fluor488-conjugated anti-mouse IgG (*green*). F-Actin was co-stained with Alexa Fluor546-phalloidin (*red*). *B*, diagrams of NIPP1-ΔC and deleted forms further mutated in the FHA domain or the PP1-binding motif. *C*, HTO cells were transiently transfected with the β-globin reporter together with NIPP1 constructs shown in *B*. Total RNA was isolated and analyzed by RPA as in Fig. 2.

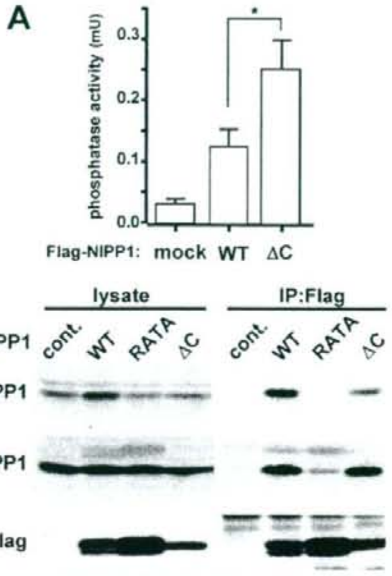


FIGURE 4. **Dysregulation of PP1 by NIPP1-ΔC.** *A*, PP1 activities in association with NIPP1-WT and -ΔC. HTO cells were transiently transfected with NIPP1-WT or NIPP1-ΔC and lysed, and immunoprecipitates with an anti-FLAG antibody were subjected to a phosphatase assay. Data represents mean of four independent experiments with S.D. *, $p < 0.05$. *B*, C-terminal inhibitory phosphorylation of PP1 associating with NIPP1-WT and NIPP1-ΔC. HTO cells were transiently transfected with NIPP1-WT or NIPP1-ΔC. After 24 h, cells were lysed and immunoprecipitated (*IP*) with an anti-FLAG antibody. Immunoblots were performed using anti-phospho-Thr-320-PP1α, anti-pan-PP1 (E9), or anti-FLAG antibodies. Anti-phospho-Thr-320-PP1α antibody also may cross-react with PP1γ and -δ isoforms phosphorylated on the corresponding residues (data from Cell Signaling Technology, Inc.). *Cont*, control.

-ΔC. Taken together, these results demonstrate that Sap155 is a major dephosphorylation target of PP1, which is regulated by NIPP1 *in vivo*.

To test the effects of NIPP1 depletion on Sap155 phosphorylation, parental HTO cells were transfected with NIPP1 siRNA to knockdown NIPP1 (Fig. 5*C*). Little effect of NIPP1 knockdown on Sap155 phosphorylation was observed, as judged by assaying Sap155 mobility shifts on Western blotting. We next treated NIPP1=knockdown cells with the cell-permeable toxin OA, which could preferentially inhibit PP2A class PPases (PP2A, PP4, and PP6) at the adopted concentration of 100 nM. OA treatment of mock- or control siRNA-transfected cells slightly increased levels of hyperphosphorylated Sap155 compared with vehicle-treated cells (Fig. 5*C*). In NIPP1-knockdown cells, OA-induced hyperphosphorylation of Sap155 was more robust than that seen in mock- or control siRNA-transfected cells (Fig. 5*C*). These results indicate the importance of endogenous NIPP1 in regulating Sap155 phosphorylation.

The results in Fig. 5, *B* and *C*, strongly suggest that NIPP1 promotes dephosphorylation of hyperphosphorylated Sap155 by associated PP1. We further tested this hypothesis by *in vitro* dephosphorylation experiments. As shown in Fig. 6*A*, OA treatment of cells greatly induced Thr phosphorylation of numerous cellular proteins (*lower panel*) and also hyperphosphorylation of Sap155 (*upper panel*). Incubation of cell lysates under dephosphorylation conditions (*i.e.* without PPase inhibitors) resulted in time-dependent loss of most phospho-Thr, likely because of the activities of endogenous PPase(s) other than PP1, but did not affect Sap155 phosphorylation. However, addition of recombinant PP1α led to complete loss of hyperphosphorylated Sap155 within 60 min, indicating that Sap155 is a PP1 substrate. Remarkably, further addition of NIPP1 enhanced dephosphorylation of hyperphosphorylated Sap155 (Fig. 6*B*). The stimulating effect of NIPP1 on Sap155 dephosphorylation by PP1 was also observed with immunopurified Sap155, minimizing the possibility that scaffolding proteins other than NIPP1 mediate this effect (Fig. 6*C* and supplemental Fig. 3*B*). Furthermore, Sap155/PP1 interaction was detectable by far-Western analysis only when PP1 was precomplexed with NIPP1 (Fig. 6*D*). Thus, we conclude that NIPP1 functions as a Sap155-targeting subunit for PP1.

Aberrant Sap155 Dephosphorylation Correlates with Inhibition of Pre-mRNA Splicing by Mutant NIPP1—How NIPP1 interacts with Sap155 was further investigated by immunoprecipitation assays. Sap155 was effectively co-immunoprecipitated with NIPP1-WT and -RATA but not with NIPP1-ΔC and -R53A (Fig. 7*A*). Importantly, Sap155 co-immunoprecipitated

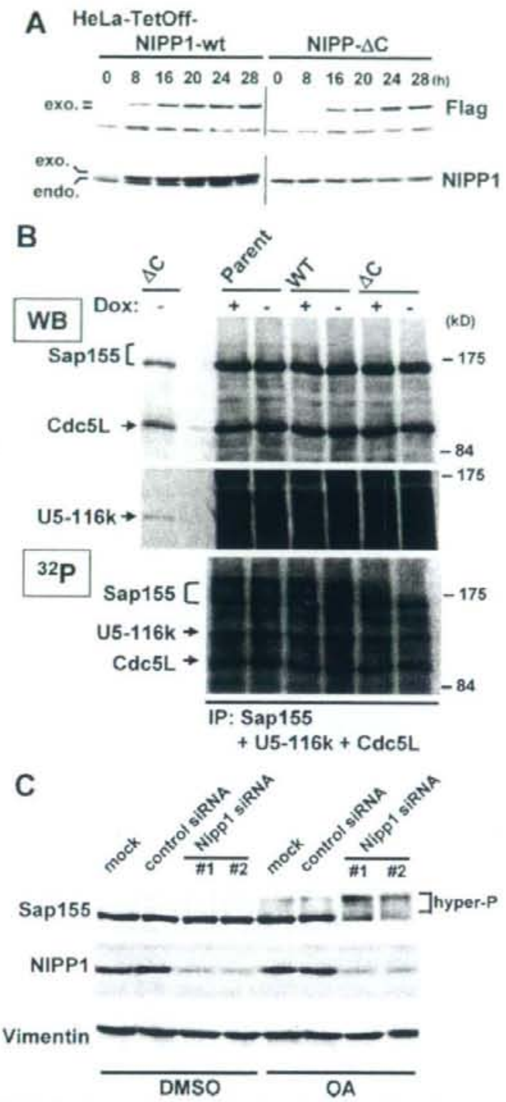


FIGURE 5. **Regulation of Sap155 phosphorylation by NIPP1 in intact cells.** *A*, establishment of HTO clones conditionally expressing exogenous NIPP1. HTO-NIPP1 clones maintained in 5 ng/ml Dox were washed and cultured in the absence of Dox for the times indicated. Cell lysates were subjected to immunoblotting with the anti-FLAG antibody (*upper*) or the anti-NIPP1 antibody, which recognizes the Lys-rich region of NIPP1 (*lower*). The positions of exogenous and endogenous NIPP1 are shown at the left as *exo.* and *endo.*, respectively. *B*, parental HTO and stable clones were cultured in the presence or absence of 10 ng/ml Dox. After 24 h, cells were labeled with ^{32}P-orthophosphate for 4 h. Sap155, U5-116k, and Cdc5L were simultaneously immunoprecipitated by the corresponding antibodies, fractionated by SDS-PAGE, and transferred to a nitrocellulose membrane. The membrane was first immunoblotted with the mixture of mouse monoclonal anti-Sap155 and anti-Cdc5L antibodies, and then with rabbit anti-sera against U5-116k (*WB*). Subsequently, the membrane was subjected to autoradiography (^{32}P). A portion of lysate was also loaded onto the gel to reveal positions of the three proteins shown at the *left*. *C*, HTO cells were transfected with control siRNA or siRNAs against human NIPP1. Forty eight hours later, transfected cells were incubated an additional 4.5 h with 100 nM okadaic acid (*OA*) or vehicle (DMSO). Immunoblots were performed using anti-Sap155, anti-NIPP1, or anti-vimentin antibodies.

with NIPP1-WT was highly phosphorylated. These results are consistent with previous reports that NIPP1/Sap155 interaction requires both the NIPP1 FHA domain and Sap155 phosphorylation (19). Furthermore, interaction between NIPP1 FHA and Sap155 appears to require that Sap155 be "hyper"-phosphorylated because a faster migrating Sap155 species seen on SDS-PAGE was likely phosphorylated at relatively low levels (autoradiography in Fig. 5*B*) and was not significantly co-immunoprecipitated with NIPP1-WT. Notably, NIPP1-RATA, which binds PP1 less efficiently than NIPP1-WT, co-immunoprecipitated hyperphosphorylated Sap155 more efficiently than did NIPP1-WT. In contrast, NIPP1-ΔC again decreased hyperphosphorylation of Sap155 based on its mobility shift ("lysate" in Fig. 7*A*), and hence did not co-immunoprecipitate Sap155 ("IP" in Fig. 7*A*). Collectively, these results indicate that PP1, which is recruited to hyperphosphorylated Sap155 by NIPP1 through the FHA domain, dephosphorylates Sap155 and suggest that its de-regulation by a mutant form of NIPP1, NIPP1-ΔC, results in defects in pre-mRNA splicing. Strikingly, the NIPP1 mutants NIPP1-RATA/ΔC and -R53A/ΔC, which harbor second mutations that no longer inhibit splicing (Fig. 3), also failed to reduce levels of Sap155 hyperphosphorylation (Fig. 7*B*).

PP1 Overexpression Compromises Pre-mRNA Splicing—Three PP1 isoforms exhibiting differing N- and C-terminal sequences (PP1α, PP1γ1, and PP1δ) are expressed in mammalian somatic cells. To determine which isoforms bound to NIPP1, co-immunoprecipitation experiments were performed. Significant amounts of PP1α, the most abundant isoform expressed in HeLa cells, but not PP1γ1 and PP1δ, co-immunoprecipitated with endogenous NIPP1, in the order PP1$\alpha \gg$ PP1γ1 $>$ PP1δ (Fig. 8*A*). Because NIPP1 recruited PP1 to dephosphorylate Sap155 and NIPP1 dysregulation resulted in splicing inhibition, one might predict that excess nuclear PP1 might also inhibit splicing. To test this possibility, the effects of PP1 overexpression on Sap155 phosphorylation and splicing were investigated. Hyperphosphorylated forms of Sap155 were reduced in cells transfected with PP1α-WT and PP1α-T320A but not in cells transfected with the inactive PP1α mutant, PP1α-H125A (Fig. 8, *C* and *D*). As expected, either overexpression of wild-type or the active mutant of PP1α, PP1c-T320A, compromised splicing of the reporter gene (Fig. 8*E*). In contrast, the inactive PP1α mutant, PP1α-H125A had no effect.

DISCUSSION

In this study, we investigated physical and functional association between NIPP1, PP1, and pre-mRNA splicing *in vivo*. Recently, it was reported that PP1 and/or PP2A are required for a catalytic step in splicing, specifically the second step, *in vitro* and that Sap155 and U5-116k are potential targets of these two PPases (6). But how these PPases are recruited to spliceosomal substrates has not been elucidated. Our results demonstrate that NIPP1 targets PP1 to Sap155. Because NIPP1 has been considered a PP1 inhibitor, the stimulating effect of NIPP1 on Sap155 dephosphorylation by PP1 is unanticipated. These observations are reminiscent of Mypt1, another PP1 regulatory protein. Mypt1 inhibits PP1 activity against nonphysiological substrates such as phosphorylase *a* but stimulates it against

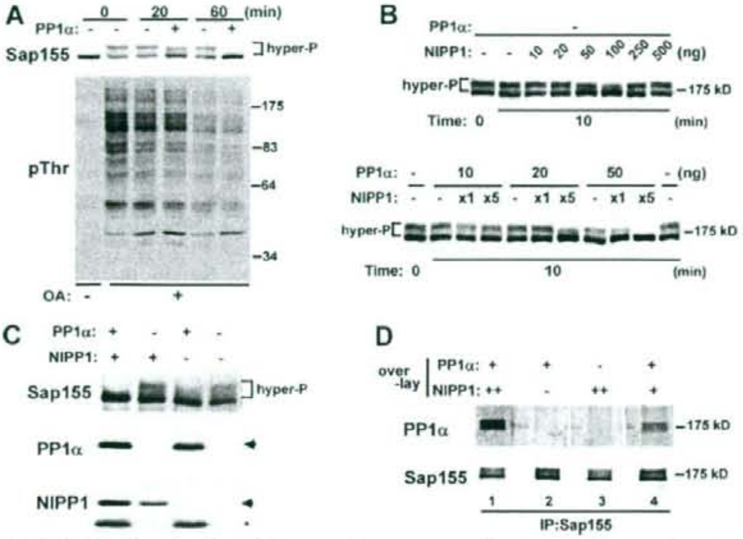


FIGURE 6. **NIPP1 bridges Sap155 and PP1.** *A*, dephosphorylation of Sap155 by PP1 *in vitro*. HTO cells were transfected with siRNA against NIPP1, treated with OA, and lysed. Lysates were incubated with 50 ng of recombinant PP1α at 30 °C for the times indicated. Immunoblots were done with anti-Sap155 or anti-*pThr* antibody. *B*, NIPP1 enhances Sap155 dephosphorylation by PP1 *in vitro*. Lysates of OA-treated HTO cells were incubated with increasing amounts of exogenous NIPP1 (*upper*) or PP1-NIPP1 holoenzyme (*lower*) at 30 °C for 10 min. Immunoblots were done with anti-Sap155 antibody. The holoenzyme was reconstituted using recombinant PP1α and His-NIPP1 expressed in and purified from *E. coli*. *Lower panel*, NIPP1 amounts are shown as a ratio relative to recombinant PP1α. *C*, dephosphorylation experiments using purified Sap155. *In vitro* dephosphorylation experiments were performed, as in *B*, using Sap155 immunopurified from HeLa cells treated with OA. Reactions were done using 50 ng of PP1α and/or His-NIPP1 per reaction and analyzed by immunoblot using anti-Sap155, anti-PP1α, and anti-NIPP1 antibodies. *Asterisk* marks residual signal of anti-PP1α blot. *D*, reconstitution of Sap155-NIPP1-PP1 ternary complex *in vitro*. Physical association between Sap155 and PP1 was analyzed by far-Western. Sap155 was immunoprecipitated, size-fractionated by SDS-PAGE, and transferred to a membrane. Membrane pieces were reacted with 100 ng/ml of PP1α in the presence or absence of recombinant His-NIPP1 (++ and + indicate 300 and 100 ng/ml His-NIPP1, respectively). PP1α overlaid on Sap155 was detected by immunoblot using anti-PP1α antibody. Similar amounts of Sap155 in each lane were confirmed by anti-Sap155 blot (*lower*).

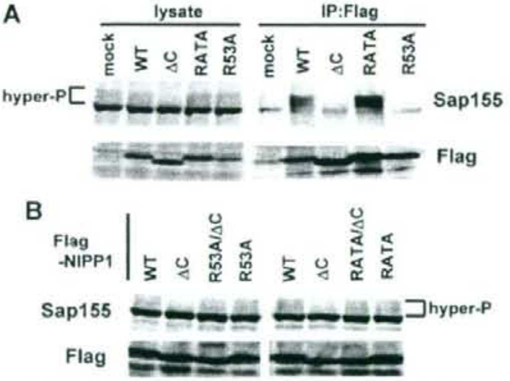


FIGURE 7. **Physical and functional associations between NIPP1 mutant proteins and Sap155.** *A*, ΔC mutation of NIPP1 decreases Sap155 phosphorylation and results in dissociation of the NIPP1-Sap155 complex. HTO cells were transfected with FLAG-NIPP1 plasmids. Twenty eight hours later, cells were lysed and subjected to immunoprecipitation (*IP*) with an anti-FLAG antibody. Immunoprecipitates were analyzed by immunoblot using anti-Sap155 or anti-FLAG antibodies. *Hyper-P* indicates positions of hyperphosphorylated forms of Sap155. *B*, reduction of hyperphosphorylated Sap155 by NIPP1-ΔC requires both a functional FHA domain and interaction with PP1. Cells were transiently transfected with FLAG-NIPP1 plasmids depicted in Fig. 1*A* and Fig. 3*A*, lysed, and immunoblotted with anti-Sap155 and anti-FLAG antibodies.

myosin light chain, a physiological substrate of the Mypt1/PP1 holoenzyme (35). We propose that the inhibitory activity of NIPP1 is also substrate-dependent.

Whereas overexpression of NIPP1-WT slightly reduced levels of hyperphosphorylated Sap155, expression of NIPP1-ΔC, which constitutes a hyper-active PP1 holoenzyme, produced more severe phenotypes, such as virtual loss of Sap155 hyperphosphorylation and defects in pre-mRNA splicing. The robust effect of NIPP1-ΔC is, at least in part, because of defects in regulation of the NIPP1-PP1 holoenzyme by inhibitory phosphorylation of the catalytic subunit. In contrast, PP1 bound to NIPP1-WT is highly phosphorylated at Thr-320. Although Thr-320 phosphorylation is implicated in regulating the cell cycle, our results suggest that regulation of PP1 by inhibitory phosphorylation is important in more general cellular functions, including regulation of pre-mRNA splicing. Recently, it was reported that Sf3a/b proteins, including Sap155, are destabilized and dissociate from the RNP core of the activated spliceosome during the transition from the B to C complex *in vitro* (5). Thus, it is likely that Sf3a and Sf3b are dispensable for the second splicing step while essential in early step(s). Given that PP1/PP2A are required for the second step, dephosphorylation of Sap155 may be critical for dissociation of Sf3a/b from the RNP core, facilitating structural rearrangement of the spliceosome required for further splicing steps. In this scenario, one could imagine that such regulation occurs by modulation of PPase activity through inhibitory phosphorylation of PP1 bound to NIPP1. Currently, why PP1 bound to NIPP1-ΔC is minimally phosphorylated on Thr-320 is not known and is under investigation.

Although our results show that NIPP1 enhances recruitment of PP1 to Sap155 and promotes Sap155 dephosphorylation, detailed mechanisms whereby NIPP1 stimulates Sap155 dephosphorylation remain unknown. It is important to determine how the PP1/NIPP1 holoenzyme can effectively and specifically dephosphorylate Sap155. PP1/NIPP1 interactions may be competed and transiently interrupted by element(s) within Sap155. Interestingly, Sap155 contains an RICF sequence (Arg-1057 to Phe-1060) resembling a consensus PP1-binding motif, although it has not been shown to be functional. Alternatively, NIPP1 may block substrates other than Sap155 from accessing the PP1 active pocket via steric hindrance. Future structural analysis of the PP1/NIPP1 holoenzyme should address these issues. In context, it is still possible that NIPP1 simultaneously

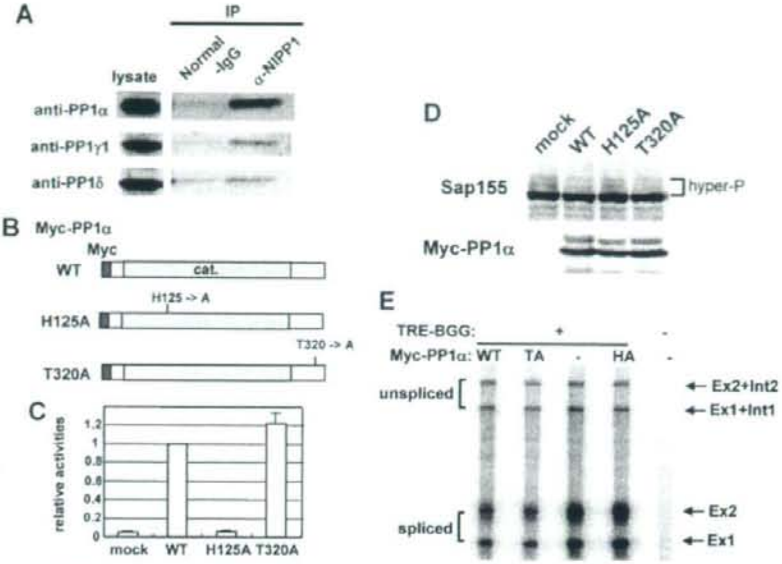


FIGURE 8. **PP1 overexpression affects Sap155 phosphorylation and splicing.** *A*, PP1 isoforms associating with NIPP1. Lysates of HTO cells were incubated with anti-NIPP1 antibody or control normal rabbit IgG. Immunoprecipitates (*IP*) were immunoblotted using antibodies specific to each PP1 isoform. *B*, diagram of wild-type and mutant forms of PP1α. Each is tagged with a Myc epitope at the N terminus. *C*, phosphatase activities of PP1α proteins expressed in HTO cells. Myc-PP1α were transiently expressed in HTO cells and immunoprecipitated with an anti-Myc antibody. Phosphatase activities were determined using phospho-phosphorylase peptide as substrate. Data represent means of three independent experiments with S.D. *D*, effects of PP1α overexpression on Sap155 phosphorylation. HTO cells were transiently transfected with myc-PP1α, lysed, and analyzed by Western blotting using anti-Sap155 and anti-Myc antibodies. *E*, cells were transiently transfected with the β-globin reporter together with myc-PP1α constructs. Total RNA was isolated and analyzed by RPA as in Fig. 2.

suppresses PP1 activity in a way that balances Sap155 phosphorylation activity to ensure proper phosphorylation/dephosphorylation cycle. Our results also do not exclude the possibility that higher activity of PP1-NIPP1ΔC against Sap155 *in vivo* is because of less inhibitory activity of the mutant NIPP1 itself. Currently, we cannot determine whether NIPP1-WT and -ΔC have similar inhibitory activity on hyperphosphorylated Sap155.

In cells expressing NIPP1-WT and -ΔC, Sap155 phosphorylation levels were specifically reduced, although normal phosphorylation levels of U5-116k were seen, suggesting de-regulation of Sap155 phosphorylation alone is sufficient to perturb splicing. But these results do not exclude the possibility that phosphorylation/dephosphorylation of U5-116k is also essential for splicing. Sap155 is the first example of a protein that phosphorylated exclusively at the time of the catalytic step of the splicing reaction (4). Indirect evidence has indicated the importance of Sap155 phosphorylation in splicing, although there has not yet been a direct demonstration. Our results provide additional evidence suggesting an essential role of Sap155 phosphorylation, namely NIPP1-ΔC constituted a hyper-active PP1 holoenzyme, decreasing hyperphosphorylated Sap155 and thereby inhibiting splicing. Sap155 has many potential phosphorylation sites, and several are phosphorylated in *in vitro* splicing reactions and also *in vivo* (36, 37). Identification of Sap155 phosphorylation sites and their functions during splicing is critically important for understanding how Sap155 is regulated byphosphorylation/dephosphorylation. Furthermore, recent reports revealed that Sap155 functions not only in constitutive splicing but also in alternative splicing and in epigenetic gene silencing (26, 38). Roles for Sap155 phosphorylation in these processes are the next questions to be analyzed.

NIPP1 knockdown sensitized Sap155 to hyperphosphorylation upon further stimulation of cells by OA. Because OA preferentially inhibits PP2A family PPases, our results are consistent with a previous report that PP1 and/or PP2A is essential for splicing *in vitro* (6). It is likely that PP1 and PP2A family PPase(s) indeed have overlapping roles in regulating Sap155 phosphorylation in intact cells. According to this notion, the relatively weak effects of mutant NIPP1 lacking the canonical PP1-binding motif (NIPP1-RATA) on splicing seem plausible. NIPP1-RATA was less efficient in binding PP1 and hence co-immunoprecipitated greater levels of hyperphosphorylated Sap155 compared with NIPP1-WT (Fig. 7), indicating that it functions as a dominant negative in some contexts, although its inhibition of splicing was partial (Figs. 1 and 2). One explanation for this observation would be that NIPP1-associated PP1 plays a kinetic role in splicing, affecting only splicing rate. Alternatively, we propose that lower levels of PP1 recruited by NIPP1-RATA could be compensated, for example, by PP2A class PPase(s). The relatively small increase in hyperphosphorylated Sap155 seen in mock- or control siRNA-transfected cells treated with OA (Fig. 5*C*) strengthens this hypothesis, although the precise mechanism by which OA enhances Sap155 hyperphosphorylation is not clear. As seen in Fig. 5*D*, treating cells with OA globally affected phosphorylation levels of several cellular proteins, possibly because of a broad spectrum of targets of PP2A class PPases or to indirect effects. Thus, we cannot exclude the possibility that OA enhances Sap155 hyperphosphorylation by stimulating Sap155 kinase(s). We tried co-transfection of siRNAs targeting PP2A in addition to NIPP, but massive cell death resulting from that treatment prevented analysis of these cells.[5] It is still possible that PP1 and PP2A dephosphorylate Sap155 differentially in terms of target residue or timing during splicing. Furthermore, PP1 and PP2A may be regulated differentially by intra- or extracellular signals through modification of the phosphatases themselves and/or

---

[5] N. Tanuma and M. Nomura, unpublished observations.