The RBM17/SPF45–SAP30BP interaction is essential for splicing in a subset of human short introns

Kazuhiro Fukumura[1]*, Luca Sperotto[2,3], Stefanie Seuß[2,3], Hyun-Seo Kang[2,3], Rei Yoshimoto[4], Michael Sattler[2,3], and Akila Mayeda[1,5]*

[1]Division of Gene Expression Mechanism, Center for Medical Science, Fujita Health University, Toyoake, Aichi 470-1192, Japan
[2]Institute of Structural Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany
[3]Bavarian NMR Center, School of Natural Sciences, Technical University of Munich, 85748 Garching, Germany
[4]Department of Applied Biological Sciences, Faculty of Agriculture, Setsunan University, Hirakata, Osaka 673-0101, Japan

[5]Lead contact
*Correspondences: fukumura@fujita-hu.ac.jp (K.F.), mayeda@fujita-hu.ac.jp (A.M.)

Keywords: Pre-mRNA splicing, short intron, poly-pyrimidine tract, RBM17 (SPF45), SAB30BP, U2AF heterodimer, U2 snRNP, SF3B1 (SF3b155), U2AF-homology motif (UHM), UHM-ligand motif (ULM)

SUMMARY

We recently demonstrated that splicing in a subset of human short introns with a truncated polypyrimidine tract (PPT) depends on RBM17 (SPF45) instead of the known essential splicing factor, U2AF2–U2AF1 (U2AF^{65}–U2AF^{35}). On a truncated PPT, the UHM (U2AF-homology motif) of RBM17 can interact with the ULM (UHM-ligand motif) of SF3B1 (SF3b155) in the U2 snRNP. Fruit fly SAP30BP (HCNGP) is found in spliceosomes assembled on short introns and RBM17 can bind to SAP30BP in yeast two-hybrid assays, suggesting that SAP30BP could be a functional cofactor of RBM17. We show that siRNA-knockdown of SAP30BP specifically represses endogenous, and ectopically expressed, RBM17-dependent splicing, but not conventional U2AF-dependent splicing. Our RNA-Seq data indicate that RBM17 and SAP30BP cooperate for splicing in short introns with a truncated PPT. GST pull-down assays and NMR data reveal the presence of a ULM in SAP30BP that is recognized by the RBM17–UHM. Remarkably, GST pull-down assays with an anti-phospho-SF3B1 antibody suggest that an intermediate interaction between SAP30BP and RBM17 is a prerequisite for RBM17 to be recruited by active phosphorylated SF3B1 in U2 snRNP. We conclude that the RBM17–SAP30BP complex is a constitutive splicing factor and it replaces the U2AF heterodimer for splicing in a subset of human short introns.

INTRODUCTION

The length distribution of human pre-mRNA introns vary extensively, ranging from under fifty to over a million nucleotides (nt) (Piovesan et al., 2015). On the other hand, the known molecular mechanism of pre-mRNA splicing was established using ideal model pre-mRNAs with a single relatively short intron, such as human β-globin pre-mRNA with 158-nt intron and adenovirus 2 major late (AdML) pre-mRNA with 231-nt intron, which are efficiently spliced *in vitro* and *in cellulo* (Mayeda and Krainer, 2012; López-Mejía and Tazi, 2012).

Therefore, the known essential *cis*-acting splicing signals and the binding *trans*-acting splicing factors in the early spliceosomal A complex must belong to a part of splicing system targeting such a definite range of the intron length; i.e., the 5′ splice site, the branch-site sequence, and the poly-pyrimidine tract (PPT) followed by the 3′ splice site that are bound by U1 snRNP, U2 snRNP and U2AF2–U2AF1 (U2AF^{65}–U2AF^{35}), respectively (reviewed in Will and Lührmann, 2011; Kastner et al., 2019). In agreement with this aspect, we recently discovered a distinct splicing mechanism targeting a subset of human short introns with the truncated PPT, which are spliced out by RBM17 (SPF45) instead of the known U2AF heterodimer (Fukumura et al., 2021).

On such a short intron with the truncated PPT, the U2AF-homology motif (UHM) of RBM17 competes out that of U2AF2 for binding to the UHM-ligand motif (ULM) of SF3B1 (SF3b155) in the U2 snRNP (Figure 1A). Our nuclear magnetic resonance (NMR) data revealed that neither the UHM domain nor the G-patch motif of RBM17

does not bind the truncated PPT *in vitro*, while both UHM and G-patch of RBM17 are critical to be located on the truncated PPT and for eventual RBM17-dependent splicing *in cellulo*. Therefore, we postulate that RBM17 is recruit in the functional position on short introns by an interaction with another protein factor ('Unknown factor X' in Figure 1A).

To search for this RBM17-interacting factor, we paid attention to three previous studies. (i) SAP30BP (HCNGP) was detected as one of non-snRNP proteins in homogeneous human spliceosome A complexes (Behzadnia et al., 2007). (ii) The fruit fly homolog of SAP30BP, together with that of RBM17, was detected in spliceosomal complexes formed on short intron (62-nt) but not on long intron (147-nt) (Herold et al., 2009). (iii) Yeast two-hybrid protein interaction analyses showed that RBM17 can interact with SAP30BP (Table S3 in Hegele et al., 2012).

These studies suggested that SAP30BP could be a promising candidate of RBM17-interacting cofactor. Here we demonstrate that SAP30BP is a *bona fide* functional partner for RBM17-dependent splicing. Interestingly, the binding of SAP30BP to RBM17 is important prior to the RBM17 interaction with phosphorylated SF3B1, a component of active U2 snRNP.

RESULTS

siRNA-knockdown of SAP30BP represses endogenous RBM17-dependent splicing

To define the functional role of SAP30BP in RBM17-dependent splicing, we performed siRNA-mediated depletions of SAP30B, RBM17 (positive control), and SAP30 (negative control) in HeLa cells (Figure 1B). The recovered total RNAs were analyzed by RT–PCR to examine RBM17-dependent splicing activity of the endogenous pre-mRNAs targeting the 56-nt (intron 7 of *HNRNPH1*), the 71-nt (intron 17 of *EML3*), the 74-nt (intron 13 of *MUS81*), and the 76-nt (intron 5 of *MTA1*); which were all previously identified as RBM17-dependently spliced pre-mRNAs (Fukumura et al., 2021). These short introns were effectively repressed not only in RBM17-depleted cells but also in SAP30BP-depleted cells, whereas such splicing repression was not observed with the control pre-mRNA (366-nt intron 5 of *DUSP1*; Figure 1C). These data suggest that SAP30BP plays a key role in RBM17-dependent splicing but not in U2AF-dependent conventional splicing.

SAP30BP and RBM17 are critical to splice out many short introns

To test whether SAP30BP is generally required for RBM17-dependent splicing, we performed a whole-transcriptome sequencing (RNA-Seq) analysis in the either RBM17-deficient or SAP30BP-deficient HeLa cells. The sequencing reads were mapped to the human genome reference sequence, and we identified total 2077 events of intron retention, or splicing inhibition, in which 326 events were overlapped by knockdown of RBM17 and SAP30BP (Figure 2A; see Supplementary Table S1 for the list of all 326 introns). Notably, the length distribution of these retained-introns in RBM17-depleted and SAP30BP-depleted cells is strongly biased toward shorter lengths compared to that of the whole set of introns (Figure 2B). Most recently,

SAP30BP was identified as one of the interacting factors that regulate neural alternative splicing events, and remarkably the knockdown of SAP30BP in neuroblastoma (Neuro-2a) cells causes retention of short introns with ~50–100 nt length (J. Roth and B. Blencowe, personal communication), that is fully supporting our results in distinct neuronal cells.

Together, these results demonstrate that RBM17 and SAP30BP cooperatively function in the efficient splicing of a distinct population of pre-mRNAs with short introns irrespective of cell/tissue types.

RBM17 and SAP30BP cooperate in RBM17-dependent splicing

We recently demonstrated that a truncated PPT in short intron is a *cis*-element responsible for RBM17-dependent splicing. Therefore, we examined whether the SAP30BP-dependency is also attributed to the truncated PPT.

First, we confirmed that RBM17-dependent splicing, as well as SAP30BP-dependent splicing, is recapitulated in ectopically expressed mini-genes including short introns as in the corresponding endogenous genes (Figure 1C). The HNRNPH1 pre-mRNA splicing (56-nt intron) and EML3 pre-mRNA splicing (71-nt intron) were repressed by depletion of either RBM17 or SAP30BP, whereas splicing of the control adenovirus 2 major late (AdML) pre-mRNA (231-nt intron) was unaffected (Figure 3A). Then, we proceeded to analyze the association of PPT using ectopically expressed chimeric mini-genes.

Regarding the RBM17-dependency on the chimeric HNRNPH1 pre-mRNAs, we observed results consistent with our recent report (Fukumura et al., 2021); i.e., (i) the replacement of the truncated 13-nt PPT of HNRNPH1 with the conventional 25-nt PPT of AdML switched to the RBM17-independent splicing (AdML PPT25; Figure 3B), (ii) the truncation of this AdML PPT to 13-nt gained RBM17-dependency again (AdML PPT13), and (iii) the expansion of upstream intron region between the 5′ splice site and the branch site with short HNRNPH1 PPT remained RBM17 dependency (AdML 5′MT). Remarkably, the knockdown of not only RBM17 but also SAP30BP can exactly follow the above RBM17-denendency. All these results were recapitulated with the another chimeric EML3 pre-mRNA with truncated 12-nt PPT as well (Figure 3C). Moreover, our global PPT length analysis of the retained introns in either RBM17-knockdown or SAP30BP-knockdown cells showed that PPT lengths of the retained short introns (<75 nt) were significantly shorter than those of the whole RefGene introns (Supplementary Figure S1).

We conclude that the truncated PPTs in short introns are crucial for the RBM17- and SAP30BP-dependency in splicing.

RBM17 binds to SAP30BP through UHM–ULM interaction

The cooperation of RBM17 and SAP30BP in splicing of pre-mRNAs with short introns suggests that these two factors could directly interact. We first examined the binding of RBM17 to SAP30BP *in cellulo*. For this purpose, HeLa cells were co-transfected with the Myc-SAP30BP and FLAG-RBM17 expression plasmids, immunoprecipitated with anti-Flag antibody, and the co-precipitated proteins were detected by the specific antibody. The immunoblot showed that RBM17 bound to SAP30BP and SF3B1

(positive control), but not to SAP30 (negative control), suggesting the formation of RBM17–SAP30BP complex (Figure 4A).

Next, we investigated which regions in RBM17 and SAP30BP are required for complex formation. We prepared *E. coli* recombinant glutathione S-transferase (GST)-fusion proteins of RBM17, SAP30BP, and variants (Figures 4B, 4D; left panel). Our GST pull-down assays show that deletion of the N-terminal region with or without deletion of G-patch motif (RBM17/ΔNΔG, RBM17ΔN) greatly reduced binding to SAP30BP but deletion of only the G-patch motif (RBM17/ΔG) did not affect the SAP30BP interaction (right panel). Deletion of either half part of the N-terminal domain (RBM17/ΔN1, RBM17/ΔN2) abolished binding to SAP30BP indicating that the complete N-terminal domain contributes to the SAP30BP interaction (Supplementary Figure S2). However, the complete or middle part of the N-terminal domain alone without the subsequent C-terminal region in which harbors the UHM domain (RBM17/N, RBM17/N3) show no binding to SAP30BP (Supplementary Figure S2). Consistently, a D319K mutation in the UHM of RBM17 (RBM17/UHMmt), which inhibits ULM binding (Fukumura et al., 2021), greatly reduced the SAP30BP interaction. These data suggest that a ULM–UHM interaction is important for the complex formation. Together, we conclude that the RBM17-UHM is essential to interact with SAP30BP but the N-terminal domain of RBM17 might support the stable binding to SAP30BP.

Since the above data revealed that the RBM17–UHM is critical to bind SAP30BP, we postulated that a ULM in SAP30BP may mediate the binding. Indeed a ULM sequence is present in the C-terminal region of SAP30BP (Figure 4C; Figure 4D, left panel). Deletion of the C-terminal region alone or together with deletion of the N-terminal domain (SAP30BP/ΔNΔC, SAP30BP/ΔC) lost binding to RBM17, while deletion of only the N-terminal domain (SAP30BP/ΔN) had no effect on the binding to RBM17 (right panel). A W257A mutation in the ULM of SAP30BP (SAP30BP/ULMmt), which impairs UHM binding (Pabis et al., 2019), no longer bound to RBM17 (Figure 4D, right panel), confirming that the UHM–ULM interaction is critical in the RBM17–SAP30BP complex.

NMR analysis confirmed the RBM17–SAP30BP interaction

To obtain a residue resolution mapping of the RBM17–SAP30BP interaction, we tested binding of the RBM17-UHM (RBM17/ΔNΔG) to a SAP30BP fragment that harbors the ULM (SAP30BP/ΔNΔH, Figure 5A) in NMR titration experiments. Amide NMR chemical shift changes were monitored upon the addition of SAP30BP/ΔNΔH to ^{15}N-labeled RBM17/ΔNΔG. Significant spectral changes demonstrate the UHM–ULM interaction (Supplementary Figure S3). The NMR chemical shift perturbations were mapped onto the structure of RBM17-UHM domain (Figure 5B). The most affected area includes the two α-helices and residues close to the Arg-Tyr-Phe motif (important for ULM binding) of the UHM, indicating a canonical UHM–ULM interaction between RMB17-UHM and SAP30BP-ULM. Isothermal titration calorimetry (ITC) shows a binding affinity with a dissociation constant K_d = 34.4 μM (Figure 5C), which is somewhat reduced compared to other reported UHM-ULM interactions (Corsini et al., 2007). In particular, the affinity is about one order of magnitude reduced compared to the interaction of the RBM17-UHM domain with SF3B1-ULM5 (Figure 4C). We

compared the crystal structure of RBM17-UHM bound to SF3B1-ULM5 with a homology model where we placed the ligand by the SAP30BP-ULM (Figure 5D), based on the sequence similarity between the two ULM motifs. Analysis shows that the key interactions are preserved, with a few exceptions. For example, Arg337 (preceding the important Trp in the ULM) is substituted by Lys256. While this maintains electrostatic interactions with Asp319 and Glu325 in the RBM17-UHM, the specific hydrogen bonds are likely reduced (Figure 5D). Also, the substitution of Glu340 (in the SF3B1-ULM5) with Ser259 (in the SAP30BP-ULM) removes a negative charge that binds to a positively charged region on the surface of the UHM. Taken together, the amino acid differences may rationalize the reduced binding affinity of the SAP30BP-ULM. This may suggest that the SF3B1-ULM may displace SAP30BP for the binding to the RBM17-UHM.

The GST pull-down assays suggest a contribution of the N-terminal domain of RBM17 for the RBM17–SAP30BP interaction (Figure 4B). We therefore performed additional NMR titration experiments with full-length proteins to test whether there are contact of the two proteins through their N-terminal regions. To this end, we titrated ^{15}N-labeled SAP30BP/ΔNΔH to RBM17/ΔNΔG and ^{15}N-labeled full-length SAP30BP to full-length RBM17, and monitored NMR chemical shift changes for the amide groups using ^{1}H-^{15}N correlation experiments. Notably, the experiments with the full-length constructs show not only the expected effects for residues in the ULM sequence, but also additional spectral changes for residues flanking the ULM sequence and in the N-terminal region of SAP30BP upon binding to RBM17 (Figure 5E). These data, together with the data from the GST pull-down assays, demonstrate that the N-terminal regions of RBM17 and SAP30BP also contribute to the interaction between the two proteins.

Intermediate RBM17–SAP30BP complex plays a role to recruit phosphorylated SF3B1

Since RBM17 has only one UHM, the RBM17–SAP30BP interaction has to be replaced by the SF3B1–ULM interaction (Figure 1A). This fact is consistent with the stronger binding of RBM17 with SF3B1 compared to that with SAP30BP. This raises the question: what is the role of the intermediate RBM17–SAP30BP interaction? Previously, it was shown that SF3B1 is phosphorylated in the active spliceosome *in vitro* (Wang et al., 1998; Bessonov et al., 2010; Agafonov et al., 2011). SF3B1 phosphorylated at Thr313, mostly localizes in the chromatin fraction, whereas free unphosphorylated SF3B1 predominantly exists in the nucleoplasm (Girard et al., 2012). We postulate that SAP30BP prevents RBM17 binding to inactive unphosphorylated SF3B1 and promotes its binding to active phosphorylated SF3B1.

To examine this hypothesis, we take advantage of an anti-phospho-SF3B1 antibody in GST pull-down assay (Figure 6A). We found that the depletion of SAP30BP markedly decreases RBM17 binding to phospho-SF3B1 (lanes 4 and 8), indicating that RBM17 alone binds excess unphosphorylated SF3B1, which results in a decrease of the RBM17 association with active phospho-SF3B1 (Figure 7). This suggests that an intermediate SAP30BP interaction with RBM17 might shield the RBM17–UHM binding surface to prevent binding with inactive free SF3B1–ULM.

To confirm the functional role of an intermediate RBM17–SAP30BP complex, we performed rescue experiments *in vitro* with SAP30BP-depleted HeLa cell extracts supplemented by two recombinant proteins; rSAP30BP and rSAP30BP/ULMmt (Figure 6B). rSAP30BP rescued the binding to phospho-SF3B1 in SAP30BP-depleted extracts, however the rSAP30BP/ULMmt, which cannot bind with RBM17 (Figure 4D), did not (lanes 7 and 8, compare with control 'Buffer' lane 6). We conclude that the intermediate RBM17–SAP30BP complex is important for RBM17 to be recruited into active phosphorylated SF3B1 (Figure 7).

DISCUSSION

SAP30BP was first identified and cloned as HTRP (human transcription regulator protein), which is a cellular co-repressor induced by HSV1 infection and it interacts with SAP30 associated with histone deacetylase complex (HDAC)(Li et al., 2004). The transcriptional co-repressors, SAP30BP and HDAC, are interacted with BAHCC1, that is a molecular basis for the modified histone H3 (H3K27me3)-mediated gene repression (Fan et al., 2020). Here we identified SAP30BP as a functional cofactor of RBM17-dependent pre-mRNA splicing, therefore SAP30BP has been rediscovered with a fully distinct function.

A novel critical UHM–ULM interaction in RBM17-dependent splicing

UHM–ULM-mediated protein–protein interactions; e.g., U2AF2–SF1, U2AF2–SF3B1, and U2AF1–U2AF2, play essential roles for spliceosome assembly in constitutive splicing (Kielkopf et al., 2001; Selenko et al., 2003; Corsini et al., 2007; reviewed in Loerch and Kielkopf, 2016). The UHM–ULM interactions are also involved in alternative splicing regulation; e.g., RBM17–SF3B1 (Corsini et al., 2007), PUF60–SF3B1 (Corsini et al., 2009) and RBM39 (CAPERα)–SF3B1 (Loerch et al., 2014). Recently, we demonstrated that the SF3B1-ULM–RBM17-UHM complex replaces the SF3B1-ULM–U2AF-UHM complex on the truncated PPT in a subset of short introns (Fukumura et al., 2021).

Here, we discovered another novel ULM–UHM interaction in constitutive splicing. We show that the RBM17-UHM binds to a newly identified ULM in SAP30BP prior to the recruitment of RBM17 to active SF3B1 on pre-mRNA. The fact that the RBM17-UHM–SAP30BP-ULM interaction is significantly weaker than other UHM–ULM interactions, as demonstrated by our ITC data (Figure 5), is consistent with a release of SAP30BP from RBM17 upon SF3B1 binding. The SAP30BP-ULM may thus occupy and shield the RBM17-UHM to avoid undesired interactions with other ULMs. And then, SAP30BP-ULM is replaced by SF3B1-ULM, which shows an order of magnitude stronger binding affinity.

A conceivable role of RBM17–SAP30BP binding prior to interaction with U2 snRNP

Why does RBM17 need to bind SAP30 before its interaction with SF3B1 in U2 snRNP? Some data may help to address this question. Since previous *in vitro* spliceosome analyses showed that SF3B1 is phosphorylated during spliceosome activation, typcally

from B to B^{act} complex (Wang et al., 1998; Shi et al., 2006; Bessonov et al., 2010; Agafonov et al., 2011; Hluchý et al., 2022), we paid attention to the phosphorylation status of SF3B1 that interacts with RBM17. Our *in vitro* binding assay revealed that RBM17 can bind to total SF3B1, but without SAP30BP, RBM17 does not bind well to active phosphorylated SF3B1. Importantly, most phosphorylated SF3B1 is present in spliceosomes detected in the chromatin fraction, whereas free unphosphorylated SF3B1 predominantly exists in the nucleoplasmic fraction (Girard et al., 2012). We thus hypothesize that SAP30BP might shield the UHM domain of RBM17 to prevent its association with free unphosphorylated SF3B1 in the nucleoplasm, and thereby ensure that RBM17 can be recruited to active phosphorylated SF3B1 on pre-mRNA (Figure 7).

The composition and assembly of spliceosome in RBM17-dependent splicing have not been studied, and thus it is unclear whether the U2 snRNP (SF3B1)–RBM17 complex formation occurs in A complex, as in U2 snRNP (SF3B1)–U2AF dimer complex formation, or after replacement of RBM17 with U2AF2 in either A complex or the following B complex (Figure 7). In either cases, our model suggests that the phosphorylation of SF3B1, or activation of spliceosome, in RBM17-dependent splicing occurs somewhat earlier than in conventional U2AF-dependent splicing. To determine the timing of SF3B1 phosphorylation during the RBM17-dependent splicing process is underway.

3′ splice site recognition in RBM17-dependent splicing

We have not yet identified the responsible factor of the 3′ splice site recognition in the early spliceosomes of RBM17-dependent short introns, such as U2AF1 in U2AF-dependent conventional introns (‘Unknown factor X’ in Figure 7). An involvement of U2AF1 was ruled out by the observation of active RBM17-dependent splicing upon co-depletion of U2AF2 and U2AF1 (Fukumura et al., 2021). Since the RBM17 G-patch motif is critical for RBM17-dependent splicing (Fukumura et al., 2021), we speculate that such a ‘Factor X’, if any, interacts with RBM17 through the G-patch motif to recognize the 3′ splice site. Ongoing efforts in our lab aim to identify the putative ‘Factor X’.

Alternatively, 3′ splice site recognition may also occur later in the splicing process, deducing from U2AF-dependent splicing in which the point mutation in the 3′ splice site ‘AG’ has no effects on splicing reaction until the first step of splicing in C complex (Gozani et al., 1994; and references therein). In RBM17-dependent splicing, we have not yet identified the recognition timing of the 3′ splice site ‘AG’ during the splicing cycle.

Previous studies uncovered diversity in recognition of the 3′ splice site. Strikingly, U2AF-independent splicing was reported, in which SR protein, SRSF2 (SC35), can functionally substitute for U2AF2 in a substrate-specific manner (MacMillan et al., 1997). Indeed, genome-wide analysis of U2AF–RNA interactions suggests that ~12% of the functional 3′ splice sites do not require U2AF2 binding or even its function for their recognition (Shao et al., 2014). We are confident that the 3′ splice sites of the RBM17-dependent introns must be one category of such U2AF2-independent 3′ splice sites.

METHODS

Construction of expression plasmids

The mini-gene plasmids (pcDNA3-AdML, pcDNA3-HNRNPH1 and pcDNA3-EML3), their chimeric variants (pcDNA3-HNRNPH1/AdML-PPT25, pcDNA3-HNRNPH1/AdML-PPT13, pcDNA3-HNRNPH1/5′AdML, pcDNA3-EML3/AdML-PPT25, pcDNA3-EML3/AdML-PPT13 and pcDNA3-EML3/5′AdML), and the RBM17 expression plasmids (pcDNA3-Flag-RBM17, pGEX6p2-RBM17, pGEX6p2-RBM17/ΔG and pGEX6p2-RBM17/UHMmt) were constructed previously (Fukumura et al., 2021; in which reads 'SPF45' not 'RBM17').

To construct the deleted variants (pGEX6p2-RBM17/ΔN, pGEX6p2-RBM17/ΔN1, pGEX6p2-RBM17/ΔN2, pGEX6p2-RBM17/ΔNΔG pGEX6p2-RBM17/N and pGEX6p2-RBM17/N3), the corresponding ORF regions were PCR-amplified from the parent plasmid pcDNA3-Flag-RBM17 with specific primer sets (Supplementary Table S1) and subcloned into the pGEX6p2 vector (GE Healthcare Life Sciences). To construct the series of SAP30BP and SAP30 expression plasmids (pcDNA3-Myc-SAP30BP, pGEX6p2-SAP30BP, pGEX6p2-SAP30BP/ΔN, pGEX6p2-SAP30BP/ΔC, pGEX6p2-SAP30BP/ΔNΔC and pcDNA3-Myc-SAP30), the corresponding ORF regions were PCR-amplified from HeLa cells cDNA with specific primer sets (Supplementary Table S1) and subcloned into the pcDNA3-Myc or pGEX6p2 vectors (GE Healthcare Life Sciences). Overlap extension PCR was performed to induce the mutation in the ULM motif of pcDNA3-Myc-SAP30BP (pGEX6p2-SAP30BP/ULMmt).

Expression and purification of recombinant proteins

Various recombinant RBM17 and SAP30BP proteins were expressed from pETM11 vectors with His6-ProteinA TEV cleavable tag using *E. coli* BL21(DE3) in LB medium or in M9 medium supplemented with $^{15}NH_4Cl$ or $^{15}NH_4Cl$ and ^{13}C-D-Glucose. Protein expression was induced at OD_{600} around 0.8–1.0 with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG), followed by overnight expression at 20°C. Cells were resuspended in 50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10 mM imidazole with protease inhibitors and lysed using french press/sonicator. After centrifugation, the cleared lysates were purified with Ni-NTA resin column. The protein samples were further purified by ion exchange chromatography, according to their pI, on a 5 mL HiTrap SP/Q HP column (Cytiva) with 20 mM sodium phosphate (pH 6.5) and by size-exclusion chromatography on a HiLoad 16/60 Superdex 75 column (GE Healthcare Life Sciences) with 20 mM sodium phosphate (pH 6.5), 150 mM NaCl. The tag was cleaved with TEV protease and removed by Ni-NTA column before ion exchange chromatography or as final step (SAP30BP/ULM, Figure 5A).

Western blotting analyses

Western blotting was performed as previously described (Fukumura et al., 2021). Briefly, protein samples were boiled with 'NuPAGE LDS Sample Buffer' (Thermo Fisher Scientific), separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). The gel was stained with Ponceau S (Cell Signaling Technology), if necessary, and immuno-reactive protein bands were detected by the ECL system and visualized by

imaging analyzer (ImageQuant LAS 500, GE Healthcare Life Sciences).

The following commercially available antibodies were used to detect targeted proteins: anti-RBM17 (1:1500 dilution; HPA037478, Sigma-Aldrich), anti-SF3B1 (1:3000 dilution; D221-3, MBL Life Science), anti-Phospho-SF3B1(Thr313)(1:1500 dilution; D8D8V, Cell Signaling Technology), anti-U2AF2 (1:1500 dilution; U4758, Sigma-Aldrich), anti-SAP30BP (1:1500 dilution; HPA052943, Sigma-Aldrich), anti-SAP30 (1:1500 dilution; 27679-1-AP, Proteintech), anti-GAPDH (1:3000 dilution; M171-3, MBL Life Science), anti-Flag (1:3000 dilution; anti-DDDDK tag, M185-3L, MBL Life Science) and anti-cMyc 9E10 antibodies (1:1500 dilution; sc-40, Santa Cruz Biotechnology). Immuno-reactive protein bands were detected by the ECL system and visualized by imaging analyzer (ImageQuant LAS 500, GE Healthcare Life Sciences).

siRNA knockdown and splicing assays

siRNA-mediated knockdown in HeLa cells (ATCC) were performed as previously described (Fukumura et al., 2021). The siRNAs targeting RBM17, SAP30BP and SAP30 were purchased (Fasmac; Supplementary Table S1 for the sequences). At 72 h post-transfection, each protein depletion was checked by a Western blotting with indicated antibodies.

Endogenous splicing products derived from the *HNRNPH1*, *EML3*, *MUS81*, MTA1 and *DUSP1* genes were analyzed by RT–PCR with oligo-dT, random primers and specific primer sets (Supplementary Table S1) as previously described (Fukumura et al., 2021).

To analyze splicing products derived from mini-genes, RBM17- and SAP30BP-depleted HeLa cells were transfected at 48 h post-transfection with 0.5 μg of mini-gene plasmid (Supplementary Table S1) using lipofectamine 2000 reagent (Invitrogen–Thermo Fisher Scientific). These cells were incubated for 24 h prior to the extraction of RNAs described previously (Fukumura et al., 2021). To analyze splicing products from mini-genes, RT–PCR was performed with T7 primer and a specific primer set for each mini-gene (Supplementary Table S1).

All the oligonucleotide primers were purchased (Fasmac; Supplementary Table S1), the PCRs were performed with Blend Taq polymerase (Toyobo Life Science), and the PCR products were analyzed by 6% PAGE and quantified as described previously (Fukumura et al., 2021). All the experiments were independently repeated three times and the means and standard errors of the splicing efficiencies were calculated.

High-throughput RNA sequencing (RNA-Seq) analyses

Six independent total RNAs derived from HeLa cells, treated with two control siRNAs (universal negative control siRNA, Nippon gene), with two RBM17-targeted siRNAs and with two SAP30BP-treated siRNAs, were prepared by 'NucleoSpin RNA' kit (Macherey-Nagel). Then enrichment of Poly(A) mRNA was performed with the 'NEBNext Poly(A) mRNA Magnetic Isolation Module' (New England Biolabs). RNA libraries were prepared using the 'NEBNext Ultra Directional RNA Library Prep Kit for Illumina' (New England Biolabs). These samples were sequenced on the high-throughput sequencing platform (NovaSeq6000, Illumina) using a 150 bp paired-end

strategy. cDNA library construction and RNA sequencing were performed by a company (Rhelixa).

The sequencing data was analyzed as previously described (Yoshimoto et al., 2021). Obtained sequence reads were mapped onto the human genome reference sequences (hg19) using the HISAT2 version 2.2.1, the mapped sequence read were assembled using Stringtie version 2.1.7, and the changes of alternative splicing isoforms were analyzed using rMATS version 4.1.1 as previously described (Fukumura et al., 2021). Obtained splicing events were defined as significant changes when the false discovery rate (FDR) was calculated at less than 0.05.

PPT score and length were calculated by SVM-BP finder software as previously described (Yoshimoto et al., 2021). The raw data from the RNA-Seq analysis have been deposited in the SRA database (https://www.ncbi.nlm.nih.gov/sra) under accession number GSE220906.

Immunoprecipitation assays

HeLa cells (in 100-mm dishes) were co-transfected with pcDNA3-Flag-RBM17 and pcDNA3-Myc-SAP30BP (or pcDNA3-Myc-SAP30 as a negative control) using Lipofectamine 2000 reagent (Thermo Fisher Scientific). At 48 h post-transfection, harvested cells were suspended in buffer D [20 mM HEPES-NaOH (pH 7.9), 50 mM KCl, 0.2 mM EDTA, 20% (v/v) glycerol] and sonicated for 20 sec by a ultrasonic disrupter (UR-20P, Tomy Digital Biology). The lysate was centrifuged to remove debris and the supernatant was rocked at 4°C for 2 h with an anti-Flag antibody (anti-DDDDK tag, MBL Life Science) or mouse anti-IgG antibody (Santa Cruz Biotechnology) that was conjugated with 'Dynabeads Protein A' (Veritas) in NET2 buffer [50 mM Tris-HCl (pH7.5), 150 mM NaCl, 0.05% Triton-X100]. The beads were washed five times with NET2 buffer, and the associated proteins were detected by Western blotting with anti-SF3B1, anti-cMyc, and anti-Flag antibodies (described above).

GST pull-down assays

Recombinant GST-tagged proteins (GST-RBM17, GST-RBM17/ΔN, GST-RBM17/ΔN1, GST-RBM17/ΔN2, GST-RBM17/N, GST-RBM17/N3, GST-RBM17/ΔG, GST-RBM17/ΔNΔG, GST-RBM17/UHMmt, GST-SAP30BP, GST-SAP30BP/ΔN, GST-SAP30BP/ΔC, GST-SAP30BP/ΔNΔC or GST-SAP30BP/ULMmt) were expressed in *E. coli* 'BL21-CodonPlus (DE3)-RIPL' competent cells (Stratagene–Agilent) and purified as previously described (Fukumura et al., 2021). To remove the GST-tag from GST-SAP30BP proteins (GST-SAP30BP, GST-SAP30BP/ΔN, GST-SAP30BP/ΔC, GST-SAP30BP/ΔNΔC or GST-SAP30BP/ULMmt), 'PreScission Protease' (GE Healthcare Life Sciences) was used according to the manufacturer's protocol.

For *in vitro* binding assays with the recombinant proteins (Figure 4, Supplementary Figure S2), GST-RBM17 series of proteins (20 μg) were mixed with untagged SAP30BP series of proteins (1 μg) in NET2 buffer supplemented with 0.5 mM dithiothreitol (DTT) and incubated at 4°C for 1 h. Then 20 μL of Glutathione Sepharose 4B (GE Healthcare Life Sciences) was added and incubated at 4°C for another 2 h. The incubated beads were washed five times with NET2 buffer and

boiled with 'NuPAGE LDS Sample Buffer' to analyze SAP30BP binding by Western blotting with anti-SAP30BP antibody (described above).

For *in vitro* binding assays in HeLa cell extracts (Figure 6), HeLa cells (in 100-mm dishes) were transfected with 500 pmol SAP30BP siRNA or negative control siRNA (Nippon gene) using Lipofectamine RNAi max (Invitrogen–Thermo Fisher Scientific) according to the manufacturer's protocol. At 72 h post-transfection, transfected HeLa cells were suspended in Buffer D and sonicated for 20 sec twice and centrifuged to remove debris (described above). Obtained siRNA-treated HeLa whole cell extracts (200 μL) were mixed with recombinant GST-RBM17 proteins (20 μg) and incubated at 30°C for 15 min. The SAP30BP or SAP30BP/ULMmt recombinant proteins (1 μg) are added to the above cell extract mixtures to perform rescue experiments (Figure 6B). After RNase A (xx μg) treatment at xx°C for xx min, NET2 buffer was added to a final volume of 500 μL with 20 μL of Glutathione Sepharose 4B and incubated at 4°C for 3 h. The incubated beads were washed five times with NET2 buffer and boiled with 'NuPAGE LDS Sample Buffer' to analyze SF3B1 binding by Western blotting with indicated antibodies (described above).

NMR spectroscopy

NMR experiments were recorded at 298 K on 500/600/900-MHz Bruker Avance NMR spectrometers equipped with cryogenic triple resonance gradient probes. Backbone ^{1}H, ^{15}N and ^{13}C resonances were assigned with standard triple resonance experiments HNCO, HNCACO, HNCACB, CBCA(CO)NH (Sattler et al., 1999). NMR spectra were processed by TOPSPIN3.5 (Bruker) or NMRPipe (Delaglio et al., 1995) and analyzed using Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). Samples were measured at 0.6–0.8 mM protein concentration for triple resonance experiments and 50–100 μM protein concentration for ^{1}H-^{15}N HSQC in NMR buffer [20 mM sodium phosphate (pH 6.5), 150 mM NaCl, 2 mM DTT] with 10% D_2O added as lock signal.

Isothermal titration calorimetry (ITC) analysis

Binding affinities of RBM17/ΔNΔG with SAP30BP/ULM (Figure 5A) were measured using a MicroCal PEAQ-ITC (Malvern Panalytical) in 20 mM sodium phosphate (pH 6.5), 150 mM NaCl at 25°C. RBM17 protein in the cell (concentration ranges, 60–100 μM) were titrated with SAP30BP/ULM (concentration of 800–900 μM) for the final molar ratios of 1:1.7 to 1:3, adjusted depending the affinity interaction. The standard deviation was determined for 3 replicates.

SUPPLEMENTAL INFORMATION

Supplemental information can be found on the following file.

ACKNOWLEDGMENTS

We thank Dr. B. Blencowe for sharing results prior to publication; H. Shirasaki and T. Kanehisa for technical support, and members of our labs for constructive discussions.

We thank Sam Asami and Gerd Gemmecker for support with NMR experiments and acknowledge access to NMR measurements at the Bavarian NMR center in Garching. K.F. was partly supported by Grants-in-Aid for Scientific Research (C) [Grant number: 18K07304] from the Japan Society for the Promotion of Science (JSPS), a Research Grant from the Hori Sciences and Arts Foundation, and a Research Grant from Nitto Foundation, a Research Grant from Takeda Science foundation, and a Research Grant from Aichi Cancer Research Foundation. A.M. was partly supported by Grants-in-Aid for Scientific Research (B) [Grant number: JP16H04705] and for Scientific Research (C) [Grant number: JP21K06024] from JSPS. L.S. and M.S. acknowledged support by the Marie Skłodowska-Curie Innovative Training Network (MSCA-ITN) RNAct research of the European Union's Horizon 2020 program [Grant number: 813239]. M.S. acknowledged funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) with grants [SPP1935 project number 273941853 and SFB1035 project number 201302640].

AUTHOR CONTRIBUTIONS

K.F. and A.M. conceived and designed the experiments; K.F. performed most of the experiments and organized the data; R.Y. performed bioinformatics analyses of the sequencing data; H.-S.K., S.S, L.S. and M.S. performed NMR and ITC experiments and data analysis; A.M. drafted the manuscript and edited the manuscript together with K.F. and M.S. All authors read, corrected and approved the final manuscript. A.M. coordinated and supervised the whole project.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Agafonov, D.E., Deckert, J., Wolf, E., Odenwalder, P., Bessonov, S., Will, C.L., Urlaub, H., and Lührmann, R. (2011). Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. Mol Cell Biol *31*, 2667-2682.

Behzadnia, N., Golas, M.M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C.L., Urlaub, H., Stark, H.*, et al.* (2007). Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. EMBO J *26*, 1737-1748.

Bessonov, S., Anokhina, M., Krasauskas, A., Golas, M.M., Sander, B., Will, C.L., Urlaub, H., Stark, H., and Luhrmann, R. (2010). Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. RNA *16*, 2384-2403.

Corsini, L., Bonnal, S., Basquin, J., Hothorn, M., Scheffzek, K., Valcárcel, J., and Sattler, M. (2007). U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. Nat Struct Mol Biol *14*, 620-629.

Corsini, L., Hothorn, M., Stier, G., Rybin, V., Scheffzek, K., Gibson, T.J., and Sattler, M. (2009). Dimerization and protein binding specificity of the U2AF homology motif of the splicing factor Puf60. J Biol Chem *284*, 630-639.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR *6*, 277-293.

Fan, H., Lu, J., Guo, Y., Li, D., Zhang, Z.M., Tsai, Y.H., Pi, W.C., Ahn, J.H., Gong, W., Xiang, Y.*, et al.* (2020). BAHCC1 binds H3K27me3 via a conserved BAH module to mediate gene silencing and oncogenesis. Nat Genet *52*, 1384-1396.

Fukumura, K., Yoshimoto, R., Sperotto, L., Kang, H.S., Hirose, T., Inoue, K., Sattler, M., and Mayeda, A. (2021). SPF45/RBM17-dependent, but not U2AF-dependent, splicing in a distinct subset of human short introns. Nat Commun *12*, 4910.

Girard, C., Will, C.L., Peng, J., Makarov, E.M., Kastner, B., Lemm, I., Urlaub, H., Hartmuth, K., and Lührmann, R. (2012). Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. Nat Commun *3*, 994.

Gozani, O., Patton, J.G., and Reed, R. (1994). A novel set of spliceosome-associated proteins and the essential splicing factor PSF bind stably to pre-mRNA prior to catalytic step II of the splicing reaction. EMBO J *13*, 3356-3367.

Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C.L., Pena, V., Lührmann, R., and Stelzl, U. (2012). Dynamic protein-protein interaction wiring of the human spliceosome. Mol Cell *45*, 567-580.

Herold, N., Will, C.L., Wolf, E., Kastner, B., Urlaub, H., and Lührmann, R. (2009). Conservation of the protein composition and electron microscopy structure of *Drosophila melanogaster* and human spliceosomal complexes. Mol Cell Biol *29*, 281-301.

Hluchý, M., Gajdušková, P., Ruiz de Los Mozos, I., Rájecký, M., Kluge, M., Berger, B.T., Slabá, Z., Potěšil, D., Weiß, E., Ule, J.*, et al.* (2022). CDK11 regulates pre-mRNA splicing by phosphorylation of SF3B1. Nature *609*, 829-834.

Kastner, B., Will, C.L., Stark, H., and Lührmann, R. (2019). Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes. Cold Spring Harb Perspect Biol *11*, a032417.

Kielkopf, C.L., Rodionova, N.A., Green, M.R., and Burley, S.K. (2001). A novel peptide recognition mode revealed by the X-ray structure of a core $U2AF^{35}/U2AF^{65}$ heterodimer. Cell *106*, 595-605.

Li, J.F., Liu, L.D., Ma, S.H., Che, Y.C., Wang, L.C., Dong, C.H., Zhao, H.L., Liao, Y., and Li, Q.H. (2004). HTRP–an immediate-early gene product induced by HSV1 infection in human embryo fibroblasts, is involved in cellular co-repressors. J Biochem *136*, 169-176.

Loerch, S., and Kielkopf, C.L. (2016). Unmasking the U2AF homology motif family: a bona fide protein-protein interaction motif in disguise. RNA *22*, 1795-1807.

Loerch, S., Maucuer, A., Manceau, V., Green, M.R., and Kielkopf, C.L. (2014). Cancer-relevant splicing factor CAPERα engages the essential splicing factor SF3b155 in a specific ternary complex. J Biol Chem *289*, 17325-17337.

López-Mejía, I., and Tazi, J. (2012). In vivo analysis of splicing assays. In Alternative pre-mRNA splicing: Theory and protocols, S. Stamm, C. Smith, and R. Lührmann, eds. (Weinheim, Germany: Wiley-Blackwell), pp. 392-399.

MacMillan, A.M., McCaw, P.S., Crispino, J.D., and Sharp, P.A. (1997). SC35-mediated reconstitution of splicing in U2AF-depleted nuclear extract. Proc Natl Acad Sci USA *94*, 133-136.

Mayeda, A., and Krainer, A.R. (2012). In vitro splicing assays. In Alternative pre-mRNA splicing: Theory and protocols, S. Stamm, C. Smith, and R. Lührmann, eds. (Weinheim, Germany: Wiley-Blackwell), pp. 321-329.

Pabis, M., Corsini, L., Vincendeau, M., Tripsianes, K., Gibson, T.J., Brack-Werner, R., and Sattler, M. (2019). Modulation of HIV-1 gene expression by binding of a ULM motif in the Rev protein to UHM-containing splicing factors. Nucleic Acids Res *47*, 4859-4871.

Piovesan, A., Caracausi, M., Ricci, M., Strippoli, P., Vitale, L., and Pelleri, M.C. (2015). Identification of minimal eukaryotic introns through GeneBase, a user-friendly tool for parsing the NCBI Gene databank. DNA Res *22*, 495-503.

Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog Nucl Magn Reson Spectrosc *34*, 93-158.

Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A., and Sattler, M. (2003). Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. Mol Cell *11*, 965-976.

Shao, C., Yang, B., Wu, T., Huang, J., Tang, P., Zhou, Y., Zhou, J., Qiu, J., Jiang, L., Li, H.*, et al.* (2014). Mechanisms for U2AF to define 3' splice sites and regulate alternative splicing in the human genome. Nat Struct Mol Biol *21*, 997-1005.

Shi, Y., Reddy, B., and Manley, J.L. (2006). PP1/PP2A phosphatases are required for the second step of pre-mRNA splicing and target specific snRNP proteins. Mol Cell *23*, 819-829.

Wang, C., Chua, K., Seghezzi, W., Lees, E., Gozani, O., and Reed, R. (1998). Phosphorylation of spliceosomal protein SAP 155 coupled with splicing catalysis. Genes Dev *12*, 1409-1414.

Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. Cold Spring Harb Perspect Biol *3*, a003707.

Yoshimoto, R., Chhipi-Shrestha, J.K., Schneider-Poetsch, T., Furuno, M., Burroughs, A.M., Noma, S., Suzuki, H., Hayashizaki, Y., Mayeda, A., Nakagawa, S.*, et al.* (2021). Spliceostatin A interaction with SF3B limits U1 snRNP availability and causes premature cleavage and polyadenylation. Cell Chem Biol *28*, 1356-1365 e1354.

FIGURES

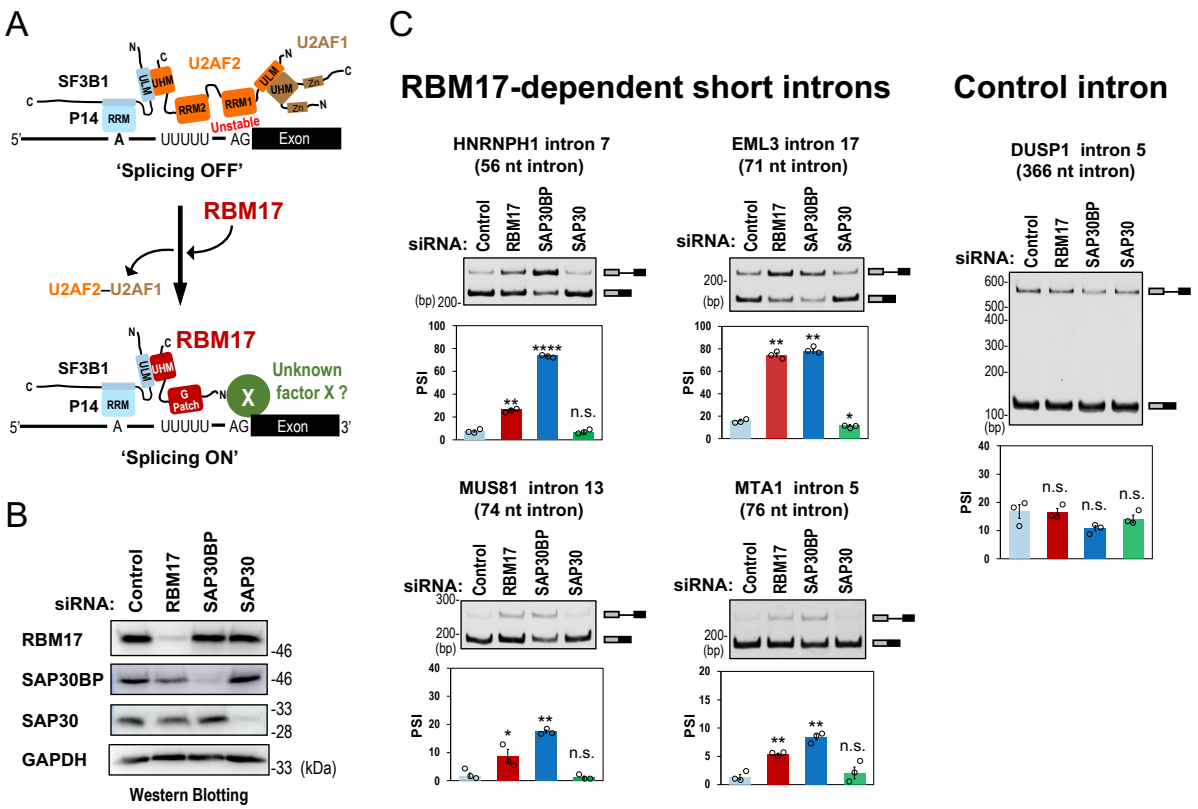

Figure 1. SAP30BP, as well as RBM17, is required for RBM17-dependent splicing of pre-mRNAs including short introns

(A) The model of RBM17-dependent splicing on a short intron with a truncated PPT (indicated by UUUUU). The associated splicing factors with the domain structures and the target sequences of pre-mRNAs are represented schematically. 'Unknown factor X' was postulated in this study.

(B) The depletion of RBM17, SAP30BP, SAP30 proteins by siRNA-knockdown in HeLa cells. The depleted proteins were checked by Western blotting with indicated antibodies.

(C) *In cellulo* splicing assays of the pre-mRNAs including indicated introns. After the indicated siRNA transfection in HeLa cells, endogenous splicing of the four RBM17-dependnt short introns and one RBM17-independent control intron were analyzed by RT–PCR. Means ± SEM are given for three independent experiments and two-tailed paired Student *t*-test values were calculated (**P* < 0.05, ***P* < 0.01, *****P* < 0.0001, $^{n.s.}$ *P* > 0.05).

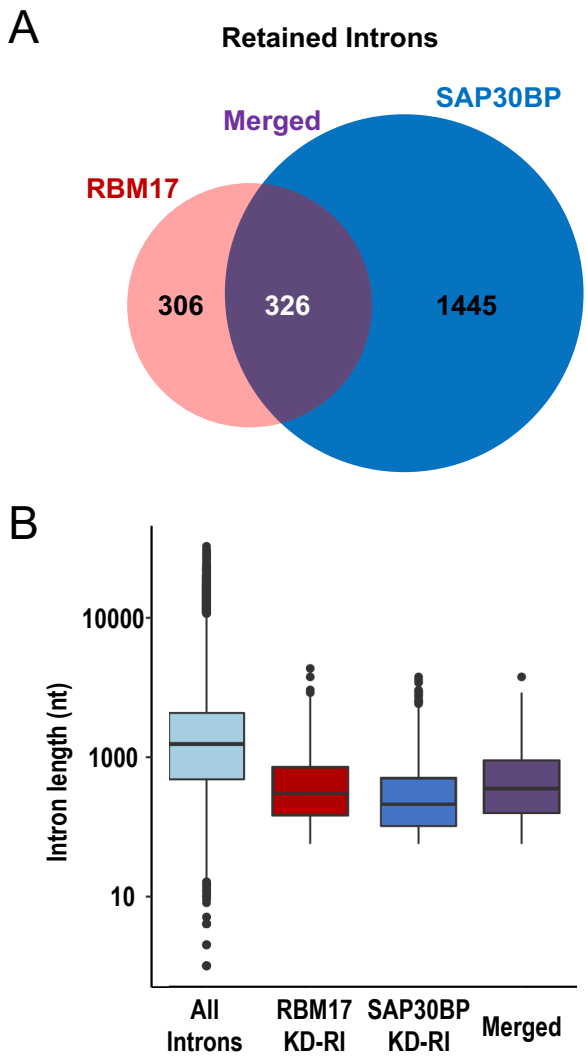

Figure 2. RBM17 and SAP30BP are generally required for the splicing of pre-mRNAs including short introns

(A) Venn diagram for significantly retained introns upon siRNA-knockdown of RBM17 and SAP30BP in HeLa cells.

(B) Box-plots for the intron-length distributions of all introns in human RefGene and the retained introns (RI) in RBM17- and SAP30BP-knockdown (KD) HeLa cells. Box-plots indicate the summary of the dataset; median (middle line), 25–75th percentile (box), and <5th and >95th percentile (whiskers), together with outliers (single points).

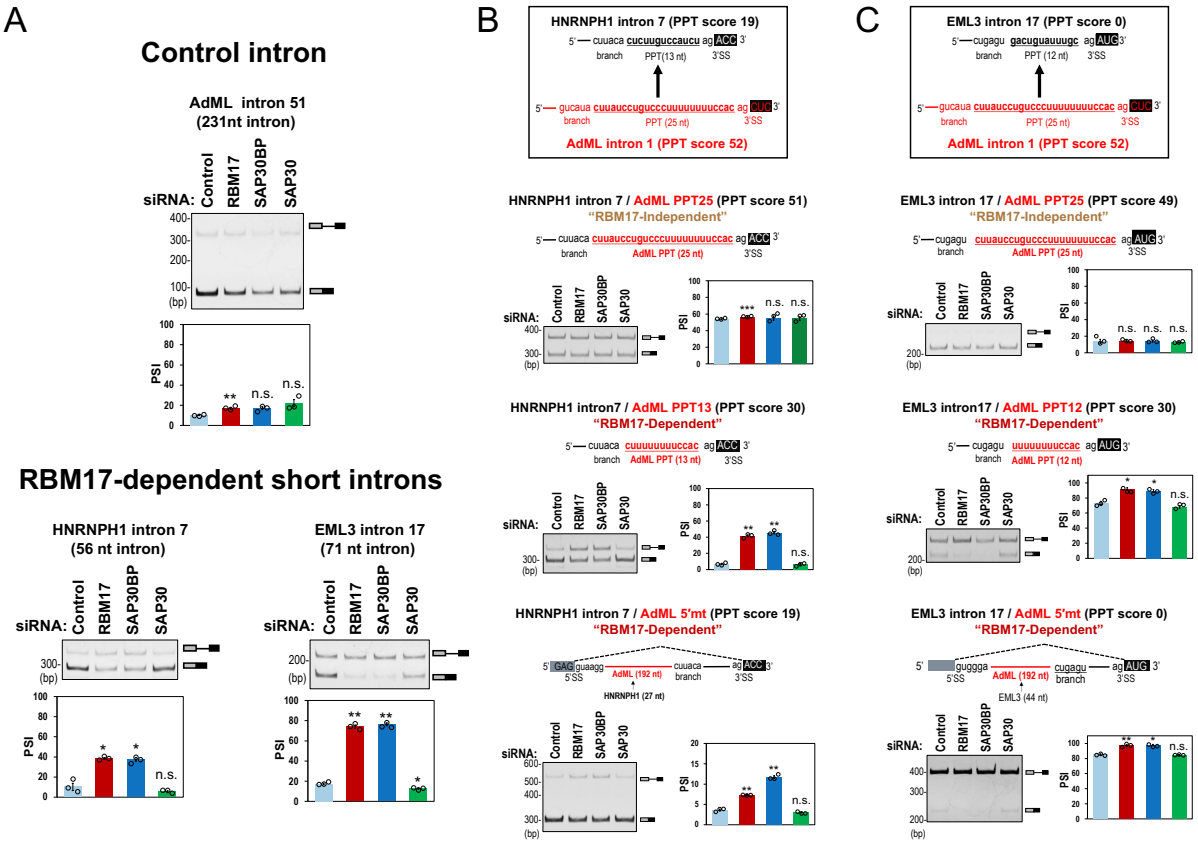

Figure 3. The length of poly-pyrimidine tracts (PPTs) determine the co-dependency of RBM17 and SAP30BP in splicing

(A) Selectively repressed splicing of pre-mRNAs with short introns in RBM17- and SAP30BP-depleted cells. Indicated three pre-mRNAs were expressed from mini-genes in HeLa cells treated with RBM17, SAP30BP or SAP30 (negative control) siRNA and their splicing was assayed by RT–PCR. PAGE images and graphs of PSI (percent spliced in) values are shown. Means ± SEM are given for three independent experiments and two-tailed Student *t*-test values were calculated (**P* < 0.05, ***P* < 0.01, ****P* < 0.001, $^{n.s.}$ *P* > 0.05).

(B, C) Selectively repressed splicing of the chimeric HNRNH1 and EML3 pre-mRNAs with truncated PPT in RBM17- and SAP30BP-depleted cells. These chimeric pre-mRNAs are schematically shown (red color indicates AdML derived sequences). The PPT score (see Methods) is one of the criteria to evaluate effective PPTs. See (A) for *in cellulo* splicing assay and the statistical analysis.

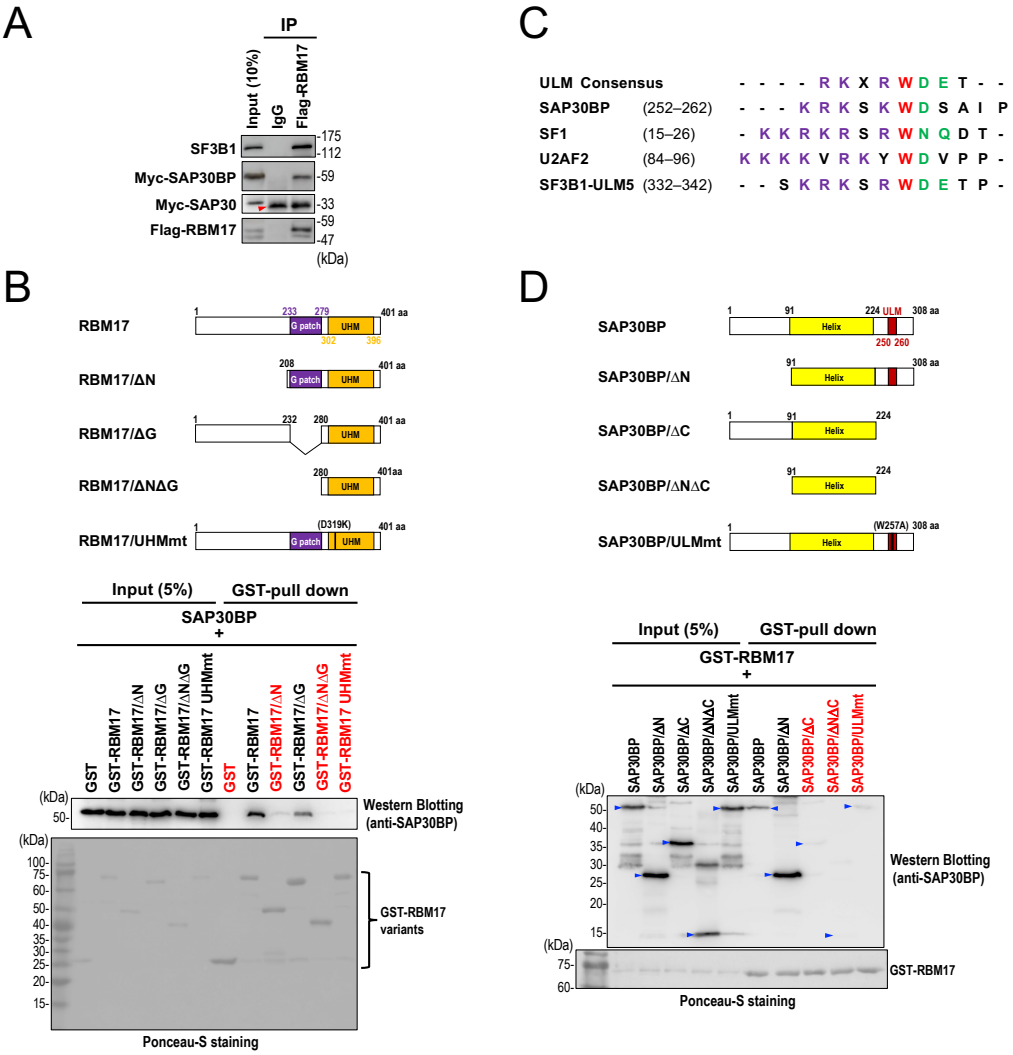

Figure 4. UHM in RBM17 and newly found ULM in SAP30BP are critical for the binding between RBM17 and SAP30BP

(A) Co-immunoprecipitation of RBM17 with SAP30BP and SF3B1. HeLa cells were co-transfected with Myc-SAP30BP, Myc-SAP30 (negative control) and Flag-RBM17 plasmids, these transfected cell extracts were immunoprecipitated using mouse IgG or anti-Flag antibody, and the immunoprecipitated proteins were analyzed by Western blotting. The red arrow head indicates IgG light chain of anti-Flag antibody.

(B) *In vitro* GST-pulldown assay (after RNase A-treatment) to detect the direct binding between indicated RBM17 variant proteins and SAP30BP protein. Red-colored variants indicate the diminished protein binding with SAP30BP. Recombinant SAP30BP protein associated with these GST-fused proteins was detected by Western blotting using anti-SAP30BP antibody. The same membrane was also stained by ponceau S to evaluate the loading amount and the comparable precipitation of the GST or GST fusion proteins.

(C) Alignment of ULM sequences in SAP30BP, SF1, U2AF2 and SF3B1 (ULM5) with the ULM consensus sequence. Conserved amino acid residues are denoted by colors: PURPLE–basic residues preceding the conserved tryptophan (W); RED–conserved W; GREEN–acidic and N/Q-type residues following the conserved W.

(D) *In vitro* GST-pulldown assay (with RNase A-treatment) to detect the direct binding between indicated SAP30BP variants and RBM17. Red-colored variants indicate the diminished protein binding with RBM17. GST-RBM17 protein associated with these SAP30BP variant proteins (indicated by blue arrow heads) was detected by Western blotting using anti-SAP30BP antibody. The same membrane was also stained by ponceau S for loading and precipitation controls.

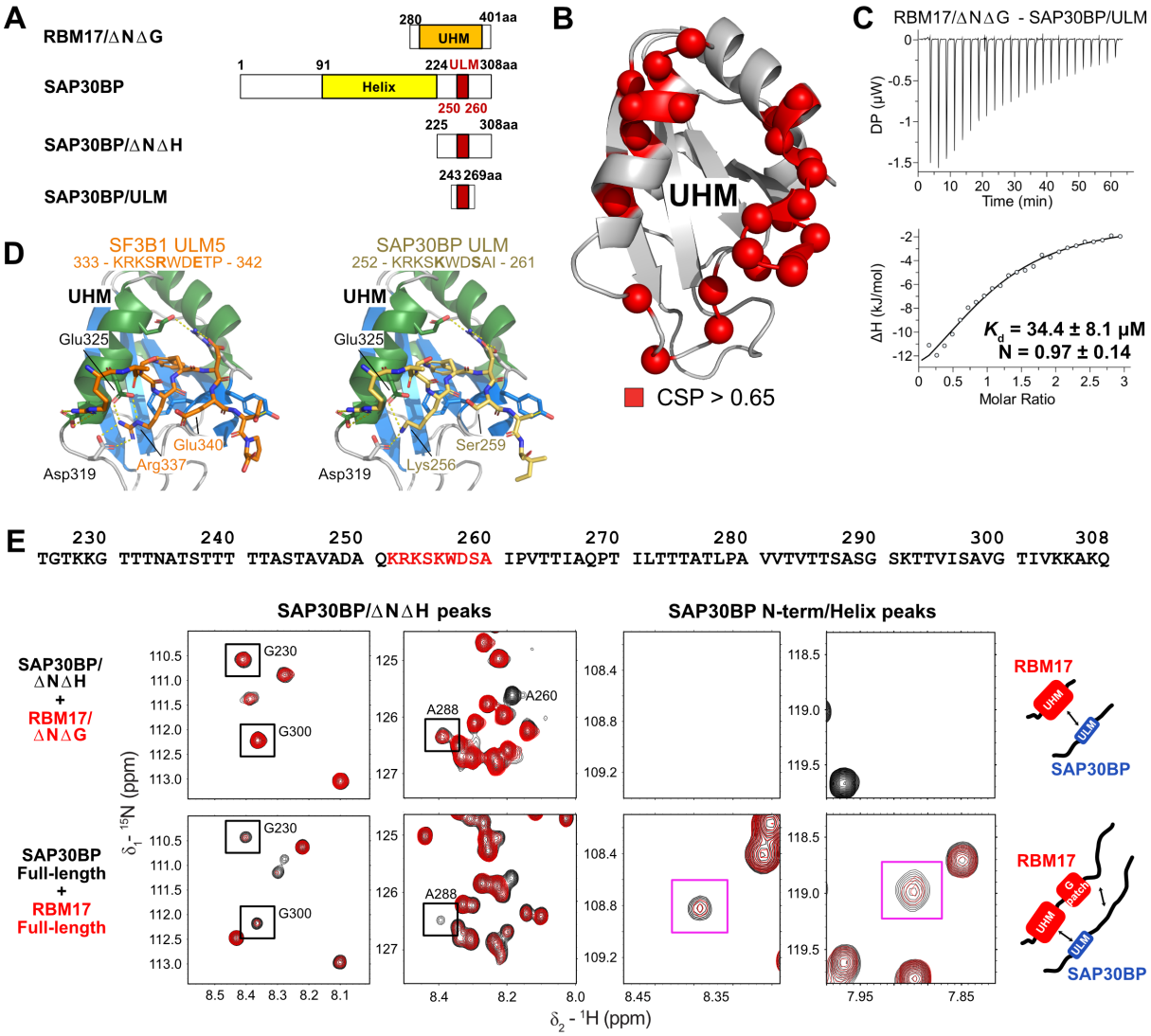

Figure 5. The RBM17–SAP30BP interaction was confirmed by NMR and ITC
(A) RBM17 and SAP30BP protein constructs used for NMR and ICT experiments.
(B) NMR chemical shift perturbation (CSP) upon titration of RBM17/ΔNΔG with 6 molar excess of SAP30BP/ΔNΔH are highlighted on RBM17-UHM structure (red spheres). The binding surface fits with the expected region for a canonical a UHM–ULM type interaction.
(B) ITC analysis of RBM17/ΔNΔG and SAP30BP/ULM.
(D) Crystal structure of SF3B1-ULM5 bound to RBM17-UHM (left, PDB accession code 2peh) and model of the SAP30BP-ULM peptide bound to the RBM17-UHM (right). Amino acid differences between the two ULM motifs may rationalize the significant differences in binding affinity.
(E) Top row: amino acid sequence of SAP30BP C-terminus with the ULM domain highlighted in red. Bottom panel: comparison of ^{1}H-^{15}N HSQC spectra of SAP30BP (in black) upon addition of RBM17 (in red) using indicated minimal constructs and full-length proteins. NMR signals of SAP30BP/ΔNΔH, which are not part of the ULM sequence, show differences in the two titrations highlighted by black boxes. NMR signals from the SAP30BP N-terminal or the flanking helical region that are affected in the titration are highlighted by magenta boxes. The differences highlighted suggest a contribution at RBM17-SAP30BP binding of additional regions outside the UHM-ULM boundaries. Right diagram: the schematic drawing summarizes the RBM17-SAP30BP interactions observed in the NMR titrations.

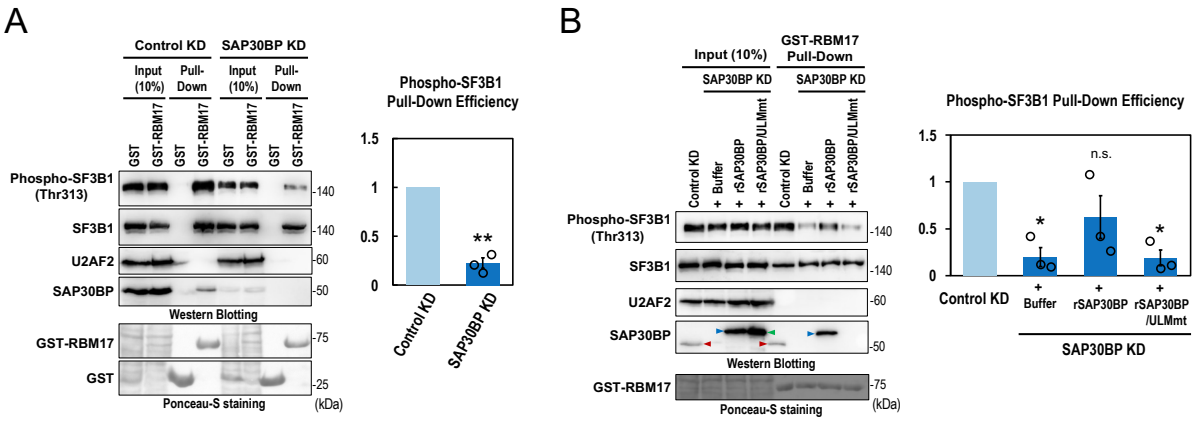

Figure 6. Binding of RBM17 to SAP30BP is critical for RBM17 to be recruited in active phospho-SF3B1

(A) Binding of RBM17 to total SF3B1 and phosphorylated SF3B1 under SAP30bp-depletion. GST pull-down assay with whole cell extracts prepared from SAP30BP-knockdown HeLa cells. Proteins associated with GST-RBM17 protein were detected by Western blotting using the indicated antibodies. Membrane was also stained by ponceau S for loading and precipitation controls. The bar graph shows quantification of the bands on Western blots. Phospho-SF3B1 scanned data were normalized to SF3B1 (Phospho-SF3B1/SF3B1) and plotted. Means ± SEM are given for four independent experiments and two-tailed Student t-test values were calculated (**$P < 0.01$, *$P < 0.05$, n.s. $P > 0.05$).

(B) Rescue experiments of diminished RBM17–phospho-SF3B1 binding under SAP30BP-depletion. GST pull-down assay in whole cell extracts prepared from SAP30BP-knockdown HeLa cells that were supplemented with wild-type and ULM-mutated SAP30BP (SAP30BP/ULMmt) proteins. The same reaction mixtures in panel (A) were supplemented with recombinant SAP30BP or SAP30BP/ULMmt proteins and the associated proteins with GST-RBM17 were analyzed by Western blotting with indicated antibodies. Phospho-SF3B1 scanned data were normalized to SF3B1 (Phospho-SF3B1/SF3B1) and plotted. See (A) for the statistical analysis.

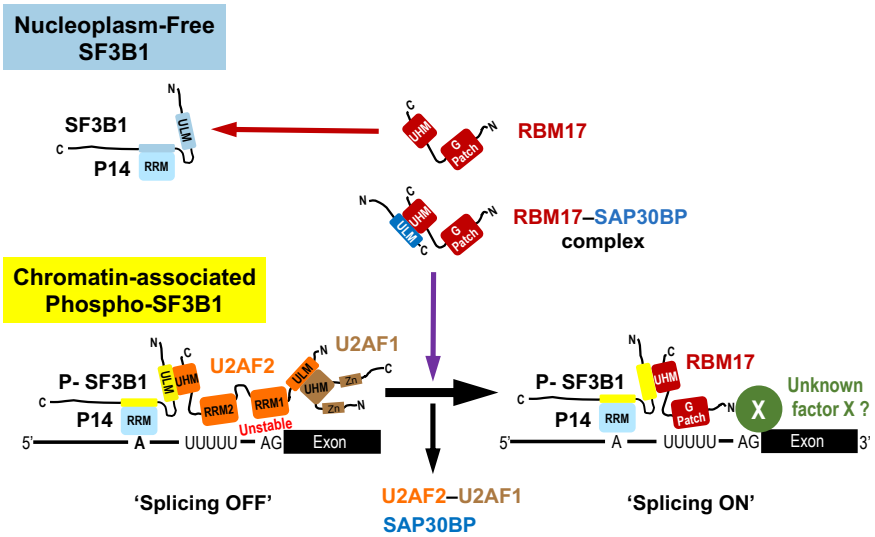

Figure 7. The model of RBM17–SAP30BP complex-dependent splicing on a short intron with a truncated PPT. The associated splicing factors with the domain structures and the target sites of pre-mRNAs are represented schematically.

The RBM17/SPF45–SAP30BP interaction is essential for splicing in a subset of human short introns

Kazuhiro Fukumura[1]*, Luca Sperotto[2,3], Stefanie Seuß[2,3], Hyun-Seo Kang[2,3], Rei Yoshimoto[4], Michael Sattler[2,3], and Akila Mayeda[1,5]*

[1]Division of Gene Expression Mechanism, Center for Medical Science, Fujita Health University, Toyoake, Aichi 470-1192, Japan
[2]Institute of Structural Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany
[3]Bavarian NMR Center, School of Natural Sciences, Technical University of Munich, 85748 Garching, Germany
[4]Department of Applied Biological Sciences, Faculty of Agriculture, Setsunan University, Hirakata, Osaka 673-0101, Japan

[5]Lead contact
*Correspondences: fukumura@fujita-hu.ac.jp (K.F.), mayeda@fujita-hu.ac.jp (A.M.)

Contents

Table S1

Figure S1

Figure S2

Figure S3

Table S1. List of synthetic oligonucleotides used in the experiments

siRNAs (sense sequences) for cellular knockdown (dT: deoxyribonucleotide)		
SPF45-siRNA	5′- GACCCUAUGUUUCCUAAUGdTdT-3′	Knockdown of SPF45
SAP30BP-siRNA	5′- GAGAAAGGCGGAUUGGUAUdTdT-3′	Knockdown of SAP30BP
SAP30-siRNA	5′- GUUCGAAACAGAAGAAAGAdTdT-3′	Knockdown of SAP30
Primer DNAs for plasmid constructions		
SPF45/ΔN infusion F	5′-CAGTGTGCTGGAATTCTGGCCATTCCTCCCCCAGTG-3′	Construction of pGEX6p2-SPF45/ΔN
SPF45 infusion R	5′-TAGATGCATGCTCGAGTCAAACTTGTTCTGCCAAATCC-3′	
SPF45/ΔN1 infusion F	5′-CAGTGTGCTGGAATTCTGGACCCTATGTTTCCTAATGAT-3′	Construction of pGEX6p2-SPF45/ΔN1
SPF45-R-infusion R	5′-TAGATGCATGCTCGAGTCAAACTTGTTCTGCCAAATCC-3′	
SPF45 infusion F	5′-CAGTGTGCTGGAATTCTGATGTCCCTGTACGATGACCT -3′	Construction of pGEX6p2-SPF45/ΔN2
SPF45/ΔN2 R	5′-ATGGCATCATTAGGAAACATAGGGTCATATTCG-3′	
SPF45/ΔN2 F	5′-ATGTTTCCTAATGATGCCATTCCTCCCCCAGTG-3′	
SPF45 infusion R	5′-TAGATGCATGCTCGAGTCAAACTTGTTCTGCCAAATCC-3′	
SPF45 infusion F	5′-CAGTGTGCTGGAATTCTGATGTCCCTGTACGATGACCT-3′	Construction of pGEX6p2-SPF45/N
SPF45/N infusion R	5′-TAGATGCATGCTCGAGTCAAGCGAGGAAGGAGTTG-3′	
SPF45/N3 infusion F	5′-CAGTGTGCTGGAATTCTGCTCGCCCCAGTCATTGA-3′	Construction of pGEX6p2-SPF45/N3
SPF45/N3 infusion R	5′-TAGATGCATGCTCGAGTCATCTCCTTGCAAACCCA-3′	
SPF45/ΔNΔG infusion F	5′-CAGTGTGCTGGAATTCTGGGCGACGCCACAGAGAAA-3′	Construction of pGEX6p2-SPF45/ΔNΔG
SPF45 infusion R	5′-TAGATGCATGCTCGAGTCAAACTTGTTCTGCCAAATCC-3′	
SAP30BP infusion F	5′-CAGTGTGCTGGAATTCTGATGGCGGGGAAGAAGAATGT-3′	Construction of pcDNA3-Myc-SAP30BP& pGEX6p2-SAP30BP
SAP30BP infusion R	5′-TAGATGCATGCTCGAG TCACTGCTTGGCCTTCTTCA-3′	
SAP30BP/ΔN infusion F	5′-CAGTGTGCTGGAATTCTGACAGAAGCAGAAAAGCGAGACC-3′	Construction of pGEX6p2-SAP30BP/ΔN
SAP30BP infusion R	5′-TAGATGCATGCTCGAGTCACTGCTTGGCCTTCTTCA-3′	
SAP30BP infusion F	5′-CAGTGTGCTGGAATTCTGATGGCGGGGAAGAAGAATGT-3′	Construction of pGEX6p2-SAP30BP/ΔC
SAP30BP/ΔC infusion R	5′-TAGATGCATGCTCGAGTCACACAAACTCAATTTTTGTTCGCTC-3′	
SAP30BP/ΔN infusion F	5′-CAGTGTGCTGGAATTCTGACAGAAGCAGAAAAGCGAGACC-3′	Construction of pGEX6p2-SAP30BP/ΔNΔC
SAP30BP/ΔC infusion R	5′-TAGATGCATGCTCGAGTCACACAAACTCAATTTTTGTTCGCTC-3′	
SAP30BP infusion F	5′-CAGTGTGCTGGAATTCTGATGGCGGGGAAGAAGAATGT-3′	Construction of pGEX6p2-SAP30BP/ULMmt
SAP30BP ULMmt-W257A R	5′-GCCGAATCCGCCTTGCTCTTTCTCTTCTGAGC-3′	
SAP30BP ULMmt-W257A F	5′-AGAGCAAGGCGGATTCGGCTATCCCAGTGA-3′	
SAP30BP infusion R	5′-TAGATGCATGCTCGAG TCACTGCTTGGCCTTCTTCA-3′	
SAP30 infusion F	5′-CAGTGTGCTGGAATTCTGATGAACGGCTTCACGCCTG-3′	Construction of pcDNA3-Myc-SAP30
SAP30 infusion R	5′-TAGATGCATGCTCGAGCTAGTGAACACCACTATCAACCTTG-3′	
Primer DNAs for analysis		
HNRNPH1-E7F-EcoRI	5′-GGAATTCCGCTATGGAGGCTATGATGA-3′	Splicing assay of endogenous HNRNPH1 intron 7
HNRNPH1-E8R-XhoI	5′-CCCTCGAGGGCAGTAGCTCTGTAAGGTAAT-3′	
EML3-E17F-EcoRI	5′-GGAATTCCGTGGTTGGTTTTGGACACAG-3′	Splicing assay of endogenous EML3 intron 17
EML3-E18R-XhoI	5′-CCCTCGAGGGCATACAGCGGCCAAAGCGGC-3′	
MUS81-E13F-EcoRI	5′-GGAATTCCCCCTGGGAACCCTGAATCAG-3′	Splicing assay of endogenous MUS81 intron 13
MUS81-E14R-XhoI	5′-CCGCTCGAGGGTGCTGTATCGATCCACCA-3′	
MTA1-E5F-EcoRI	5′-GGAATTCCGAAATGGAGAACCCGGAAATG-3′	Splicing assay of endogenous MTA1 intron 5
MTA1-E6R-XhoI	5′-CCGCTCGAGGCTCCAGGTAGGACTTGAG-3′	
DUSP1-F	5′-TGCAGTACCCCACTCTACGA-3′	Splicing assay of endogenous DUSP1 intron 2
DUSP1-R	5′-GAGACGTTGATCAAGGCAGTG-3′	
T7	5′-AATACGACTCACTATAG-3′	Splicing assay of AdML mini-gene pre-mRNA
AdML-E2R	5′-ACCGCGAAGAGTTTGTCCTCAACC-3′	
T7	5′-AATACGACTCACTATAG-3′	Splicing assay of HNRNPH1 mini-gene pre-mRNA
HNRNPH1-E8R-XhoI	5′-CCCTCGAGGGCAGTAGCTCTGTAAGGTAAT-3′	
T7	5′-AATACGACTCACTATAG-3′	Splicing assay of EML3 mini-gene pre-mRNA
EML3-E1R-XhoI	5′-CCCTCGAGGGCATACAGCGGCCAAAGCGGC-3′	

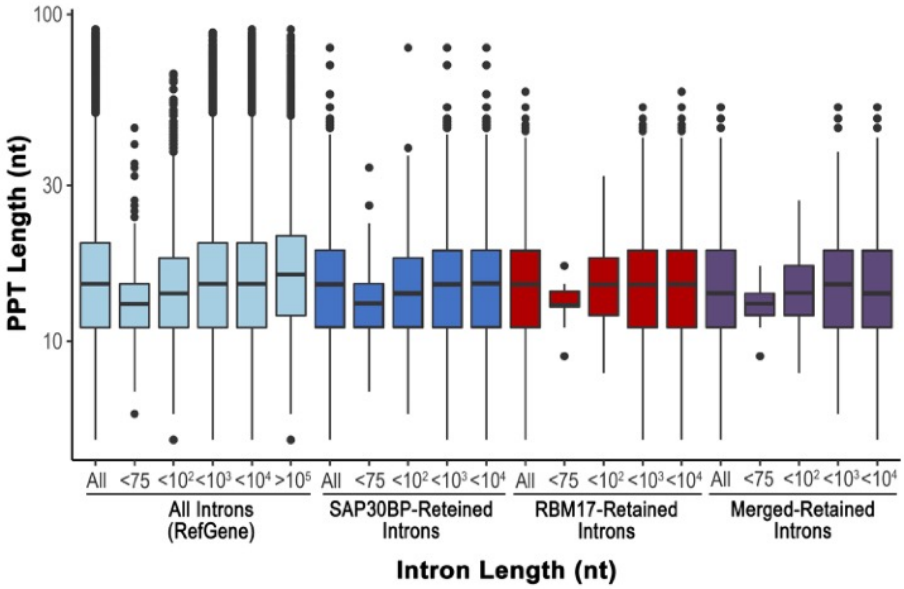

Figure S1. Truncated polypyrimidine tract (PPT) is dominated in introns shorter than 75 nt.
The box plots compare the PPT-length distributions of all introns (in human RefGene) with those of the retained introns in RBM17-knockdown HeLa cells and SAP30BP-knockdown HeLa cells. Box plots show the summary of the data set; median (middle line), 25–75th percentile (box), and <5th and >95th percentile (whiskers), together with outliers (single points). Source data are provided as a Source Data file.

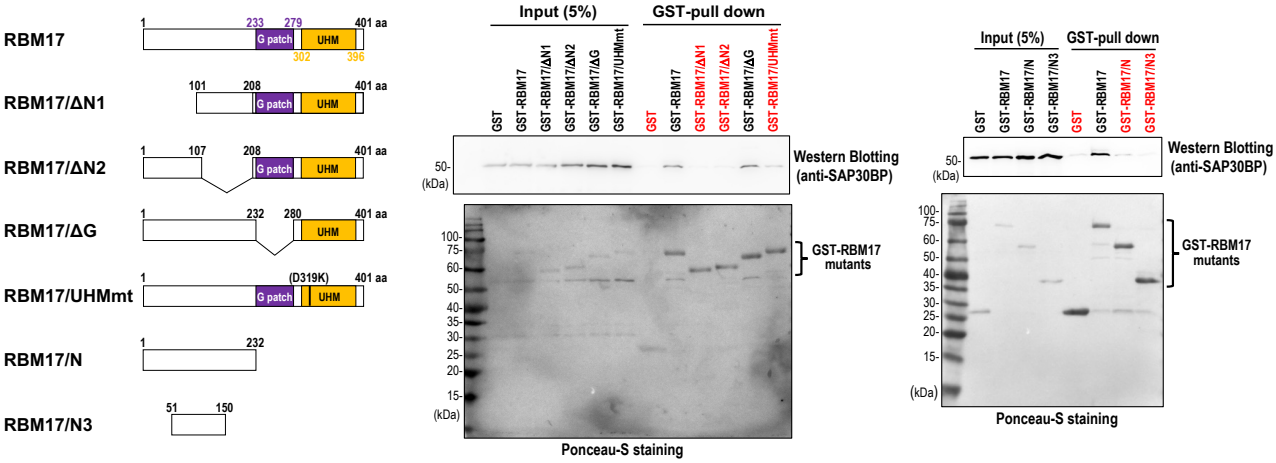

Figure S2. RBM17 requires intact N-terminus and UHM to bind SAP30BP
Direct interaction of RBM17 mutants with SAP30BP, as shown by immunoblot analysis of the *in vitro* GST-pulldown assay for purified GST, GST-RBM17/DG, RBM17/UHMmt,GST-RBM17/DN1, GST-RBM17/DN2 or RBM17/N of RBM17/N3 with purified SAP30BP. Membrane was also stained by ponceau S to show the general loading as well as the comparable precipitation of the GST or GST fusion proteins.

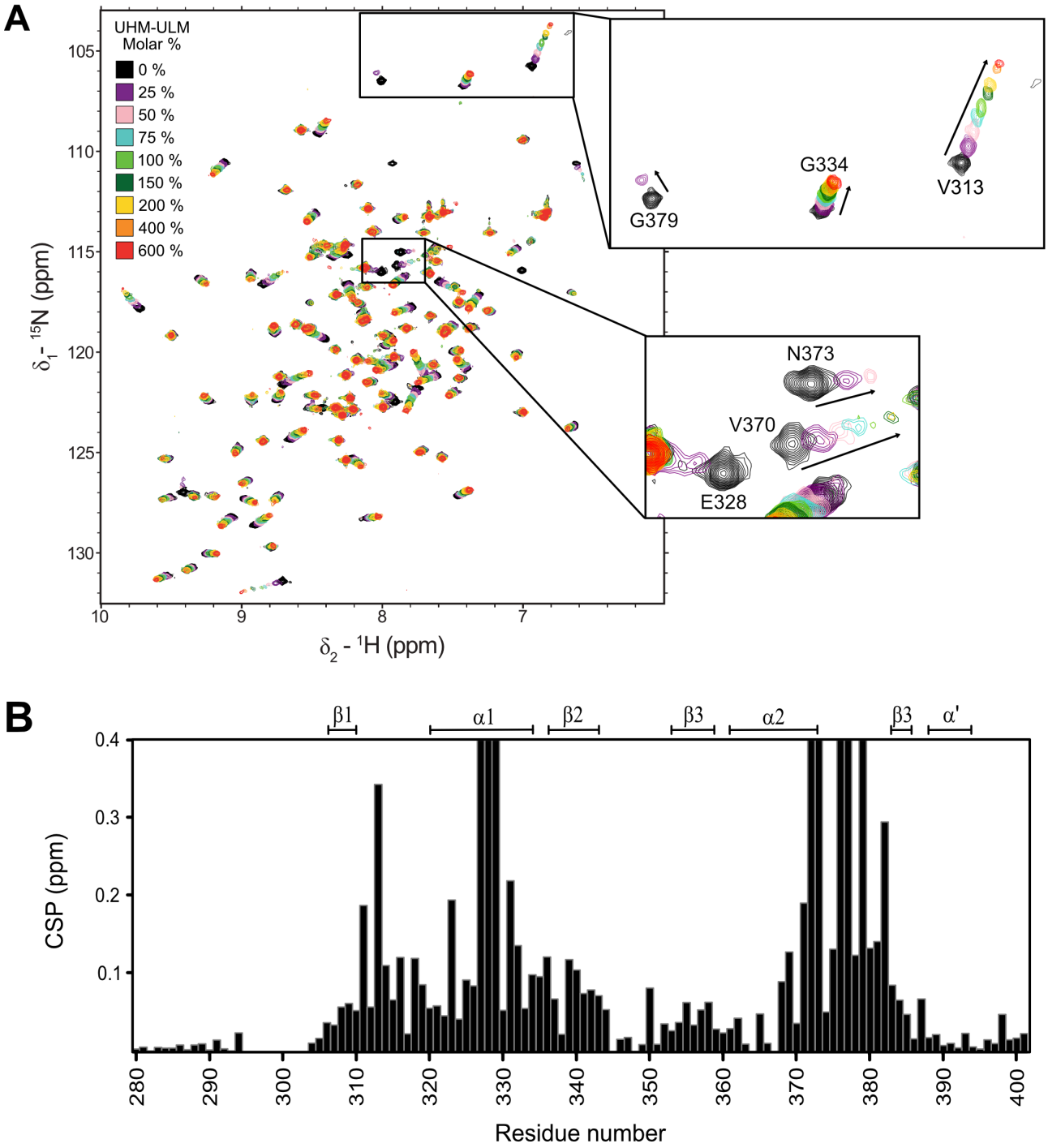

Figure S3. NMR spectral changes reveal the interaction between the RBM17-UHM and SAP30BP-ULM
(A) 1H-15N HSQC spectra of RBM17-UHM domain (RBM17/ΔNΔG) upon titration with SAP30BP-ULM domain (SAP30BP/ΔNΔH).
(B) Bar plot of the calculated chemical shift perturbation observed in the spectrum in panel A. The secondary structure of the RBM17 UHM domain is indicated on top.