The general splicing factor SPF45 (RBM17) replaces the U2AF heterodimer in a distinct subset of human short introns

Kazuhiro Fukumura[1]*, Rei Yoshimoto[1], Tetsuro Hirose[2], Kunio Inoue[3] and Akila Mayeda[1]*

[1]Division of Gene Expression Mechanism, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Aichi 470-1192, Japan
[2]Institute for Genetic Medicine, Hokkaido University, Sapporo, Hokkaido 060-0815, Japan
[3]Department of Biology, Graduate School of Science, Kobe University, Kobe, Hyogo 657-8501, Japan

*e-mails: fukumura@fujita-hu.ac.jp (K.F.), mayeda@fujita-hu.ac.jp (A.M.)

Keywords: Pre-mRNA splicing, short intron, polypyrimidine tract, SPF45 (RBM17), U2AF heterodimer, U2 snRNP, SF3b155 (SF3b1), U2AF-homology motif (UHM), UHM-ligand motif (ULM)

Human pre-mRNA introns vary in size from under fifty to over a million nucleotides. We searched for specific splicing factors involved in human short introns by screening siRNAs against 154 human nuclear proteins for activity on a model short 56-nucleotide intron-containing *HNRNPH1* pre-mRNA. We identified a known alternative splicing regulator SPF45 (RBM17) as a general splicing factor that is essential to splice out this 56-nt intron. Whole-transcriptome sequencing of SPF45 deficient cells revealed that SPF45 is specifically required for the efficient splicing of many short introns. Our crosslinking and biochemical analyses demonstrate that SPF45 specifically replaces the U2AF heterodimer on the truncated polypyrimidine tracts. To initiate splicing, the U2AF-homology motif (UHM) of the replaced SPF45 interacts with the UHM-ligand motif (ULM) of the U2 snRNP protein SF3b155 (SF3b1). We propose the existence of a distinct subset of SPF45-dependently spliced short introns.

There is a remarkable pattern in the distribution of higher eukaryotic pre-mRNA intron length; most introns fall either within a narrow peak under one hundred nucleotides or in a broad distribution peaking around several thousand nucleotides and extending to over a million nucleotides[1-3]. Pre-mRNA splicing is a ubiquitous and essential part of eukaryotic gene expression, however, little is known about any specific mechanisms for the precise recognition of the degenerated 5′ and 3′ splice site sequences within such extensively varied length of introns.

The basic splicing mechanism was studied and established using model pre-mRNAs with a single relatively short intron of a few hundred nucleotides, which are efficiently spliced in cells and *in vitro*[4,5]. According to such optimal systems, the essential splicing sequences in pre-mRNA, namely the 5′ splice site, the branch-site sequence, and the polypyrimidine tract followed by the 3′ splice site, are simultaneously bound by the splicing factors U1 snRNP, U2 snRNP, and U2AF heterodimer ($U2AF^{65}$/$U2AF^{35}$, U2AF2/U2AF1 as HGNC approved symbol), respectively, leading to early formation of the spliceosomal E and A complexes (reviewed in Ref.[6]). Under the electron microscope, the A-complex is an asymmetric globule is an asymmetric globule (~26 × 20 × 19.5 nm)[7] that fully occupies 79–125 nucleotides (nt) of RNA[8], and recent high-resolution cryo-electron microscopy structure of A-complex is consistent with this notion (reviewed in ref.[9]). Whereas human ultra-short introns under those lengths (43–65 nt) are indeed spliced[10,11]; this raises the question of how ultra-short introns can be recognized and spliced by the 'oversized' A complex without steric hindrance.

To approach this problem, we hypothesized that a general splicing factor must be specifically involved in splicing of pre-mRNAs with short introns.

Results

SPF45 is a novel essential splicing factor for a subset of short introns. To find potential factors involved in splicing of short introns, we screened an siRNA library targeting 154 human nuclear proteins for splicing activity of the *HNRNPH1* pre-mRNA including 56-nt intron 7[10,11]. Many known RNA-binding proteins and splicing factors could be tested with this siRNA library (Table S1).

HeLa cells were transfected with each siRNA and recovered total RNAs were analyzed by RT–PCR to examine splicing activity of the endogenous *HNRNPH1* pre-mRNA with 56-nt intron (Fig. 1, left panel). The strongest splicing repression was markedly caused by knockdown of SPF45 (RBM17 as HGNC approved symbol; right panel) that indeed effectively

depleted SPF45 protein (Fig. 1, middle panel). To test if SPF45 might have a general role in splicing of short introns, we assayed two other endogenous pre-mRNAs targeting the 70-nt (intron 9 of *RFC4*) and the 71-nt (intron 17 of *EML3*). Both introns were also markedly repressed in SPF45-depleted HeLa cells (Fig. 1, right 2 panels).

Splicing inhibition was proportional to SPF45-knockdown efficiency induced by independent siRNAs (Supplementary Fig. S1a). These SPF45 siRNA-induced splicing defects were also observed in HEK293 cells, testifying to the robustness of our results (Supplementary Fig. S1b).

To test our ongoing hypothesis that SPF45 is required to splice out short introns, we performed whole-transcriptome sequencing (RNA-Seq) with RNA from the SPF45-deficient HEK293 cells. The sequencing reads were mapped to the human genome reference sequence. We identified 517 changes in splicing from a total of 47,960 alternative splicing events (Fig. 2a, left panel). The most frequent changes of splicing in SPF45-depleted HEK293 cells were intron retention events (Fig. 2a, right graph; see Table S2 for the list of all 187 introns).

The analysis of these retained introns hinted at a mechanistic insight into how SPF45 might work. Remarkably, the length distribution of the retained-introns in SPF45-depleted cells was strongly biased toward the shorter length than those in the depleted cells of constitutive splicing factors, $U2AF^{65}$ and SF3b155, which were comparing with the distribution of the whole set of introns (Fig. 2b).

We validated these RNA-Seq-based profiles by RT–PCR. Splicing of pre-mRNAs with two control introns were not affected by SPF45-knockdown, while three arbitrarily chosen pre-mRNAs with short introns were repressed (Fig. 2c). These results demonstrate that SPF45 is required for the efficient splicing of a large population of pre-mRNAs with short introns.

SPF45 is required for splicing on intron with truncated polypyrimidine tract. Next we searched for any possible *cis*-element in short introns through which SPF45 might act. From RNA-Seq data of SPF45-depleted cells, we found that strengths of the 5′/3′ splice sites and the branch sites of SPF45-dependent short introns were somewhat weaker than the average in RefGene (Supplementary Fig. S2a). Therefore, we first examined these splicing signal elements by mini-gene splicing assays in SPF45-depleted HeLa cells. As expected, splicing of pre-mRNA containing HNRNPH1 56-nt intron 7 was repressed by depletion of SPF45, whereas splicing of the control adenovirus 2 major late (AdML) pre-mRNA (231-nt intron 1), which was used as a standard splicing substrate previously, was unaffected (Fig. 3; top 2 panels). The SPF45-dependent splicing of the HNRNPH1 pre-mRNA was not altered even after replacement of either the 5′/3′ splice sites or the branch site by those of the AdML pre-mRNA (Supplementary Fig. S2b). These results indicate that the SPF45 does not act through the 5′/3′ splice sites or the branch site.

We then examined whether the SPF45-dependency is attributed to the polypyrimidine tract (PPT). The PPT score (see Experimental Procedures) is one of the criteria to evaluate effective PPTs: 19 for PPT (13 nt) in the HNRNPH1 and 52 for PPT (25 nt) in the AdML (Fig. 3, second panel). Remarkably, splicing of the SPF45-dependent HNRNPH1 pre-mRNA was altered toward that of an SPF45-independent pre-mRNA by replacement of the HNRNPH1-PPT with the conventional AdML-PPT (Fig. 3, ‘AdML PPT25’). To determine whether SPF45 recognizes the quality or the quantity of PPT, we reduced the PPT score of AdML-PPT in two ways: one was transversion mutations in the PPT (C/U→G; score 52→32), and the other was truncation of the PPT (25 nt→13 nt; score 52→30). Notably, the transversion mutations in the PPT did not cause SPF45-dependency (Fig. 3, ‘AdML PPT25mt’) but the truncation of PPT did (Fig. 3, ‘AdML PPT13’). Lastly, we expanded the distance between the 5′ splice site and the branch site in HNRNPH1 intron (27 nt) by replacement with the corresponding

fragment in the AdML intron (192 nt). Interestingly, this chimeric pre-mRNA with the short HNRNPH1 PPT remained SPF45 dependent (Fig. 3, 'AdML 5′MT'). All these results demonstrate that short PPT *per se* in the HNRNPH1 intron 7 is the determinant for the SPF45-dependency in splicing.

All of these results were nicely recapitulated in the distinct SPF45-dependent EML3 pre-mRNA including 71-nt intron (Supplementary Fig. S3a). Moreover, our global PPT length analysis of the retained introns in SPF45-depleted cells showed that PPT lengths of SPF45-dependent short introns (<100 nt) were significantly shorter than those of the whole RefGene introns (Supplementary Fig. S3b). Together, we conclude that it is the truncated PPTs in short introns that are crucial for the SPF45 function in splicing.

SPF45 replaces U2AF^{65} on truncated PPTs in short introns to promote splicing. In an early transition of human spliceosome from the E to A complexes, the branch site, PPT and the 3′ splice site are bound cooperatively by SF1, U2AF^{65} and U2AF^{35}, respectively. A stable U2 snRNP-associated splicing complex is then formed by ATP-dependent displacement of SF1, contact of p14 (U2 snRNP associated factor) with the branch site, and interaction of U2AF^{65} with the U2 snRNP component SF3b155 (reviewed in ref.[6]). Since splicing activity of SPF45 depends on short PPTs, we hypothesized that insufficient U2AF^{65} binding to a truncated PPT would allow its replacement by SPF45.

To examine whether SPF45 associates with short introns, mini-genes encoding HNRNPH1, EML3, MUS81, and control AdML introns, were expressed in HeLa cells. We confirmed that these mini-genes were spliced in an SPF45-dependent manner (Fig. 3 and Supplementary Fig. S4) as we observed in these endogenous genes (Figs. 1 and 2c), so we proceeded to analyze proteins associated with these ectopically expressed RNAs.

We used formaldehyde crosslinking to detect any indirect RNA association of SPF45[12]. HeLa cells were co-transfected with the four mini-genes and Flag-SPF45 expression plasmid, crosslinked with formaldehyde, immunoprecipitated with anti-FLAG antibody, and the co-precipitated RNAs were analyzed by RT–qPCR (Fig. 4a). SPF45 bound to all pre-mRNAs derived from these four mini-genes irrespective of the length of the introns. The binding of SPF45 to control AdML intron (231 nt) is consistent with previous proteomic detection of SPF45 in the human spliceosome[13]. SPF45 was also detected on MINX (131 nt) and PM5 (235 nt) introns[14,15] (reviewed in ref.[6]).

We next examined the binding of U2AF^{65} and SF3b155 to pre-mRNAs by UV crosslinking–immunoprecipitation (CLIP). Whole-cell extracts from crosslinked cells were immunoprecipitated with anti-U2AF^{65} and anti-SF3b155 antibodies, and the precipitated RNAs were analyzed by RT–qPCR (Fig. 4b). SF3b155 bound to all the pre-mRNAs, while significant U2AF^{65} binding was observed only with control AdML pre-mRNA and it was very weak with the three SPF45-dependent short introns.

We biochemically verified the binding of U2AF^{65} and SF3b155 to the HNRNPH1 intron and control AdML intron by affinity pull-down assay using biotinylated pre-mRNAs (Fig. 4c). In agreement with crosslinking experiments, SPF45 bound to both AdML and HNRNPH1 pre-mRNAs; however, importantly, U2AF^{65} strongly bound to AdML pre-mRNA and only bound to HNRNPH1 pre-mRNA when SPF45 was depleted from nuclear extracts. These results together support our proposed hypothesis that SPF45 replaces U2AF^{65} that is unstably bound to truncated PPTs of short introns.

We noticed that endogenous U2AF^{65}-knockdown barely repressed splicing of the SPF45-dependent short intron (Supplementary Table S1, No. 142; Supplementary Fig. S5a). Therefore, we checked splicing efficiencies of these four mini-genes in U2AF^{65}-knockdown HeLa cells (Fig. 5). This depletion of U2AF^{65} also caused effective co-depletion of U2AF^{35}

(Fig. 5, left panel) that is consistent with previous reports[16,17]. In the control AdML mini-gene, spliced mRNA was reduced by the depletion of U2AF^{65}, showing that U2AF heterodimer is essential for conventional AdML pre-mRNA splicing as expected. Remarkably, splicing of SPF45-dependent pre-mRNAs with short introns was rather activated by the depletion of U2AF^{65} (Fig. 5, right 3 panels). In endogenous SPF45-dependent pre-mRNAs (Supplementary Fig. S5a), such marked activation was not observed that could be due to the almost saturated efficiency of splicing (see amounts of unspliced pre-mRNAs). Taken together, we conclude that SPF45 effectively competes out U2AF heterodimer on truncated PPTs and the newly installed SPF45 promotes splicing of pre-mRNAs with short introns.

SF3b155–U2AF^{65}/U2AF^{35} is displaced by SF3b155–SPF45 *via* ULM–UHM binding. The SPF45 protein contains a G-patch motif for the interactions with nucleic acids and proteins[18,19], and C-terminal U2AF-homology motif (UHM) that binds the UHM-Ligand motifs (ULM) of its partner proteins. The binding between the UHM and the ULM plays an essential role in the splicing reactions; e.g., interactions of U2AF^{65}–SF3b155 (SF3B1 as HGNC approved symbols) component of U2 snRNP, U2AF^{65}–U2AF^{35}, and U2AF^{65}–SF1 (Reviewed in ref.[20] and references there in). Remarkably, *in vitro* binding analyses using the purified recombinant proteins showed that SPF45–UHM can bind to SF3b155–ULM, U2AF^{65}–ULM, and SF1–ULM with low micromolar affinity[21]. We therefore postulated that the SF3b155–U2AF^{65}/U2AF^{35} complex is remodeled to the SF3b155–SPF45 complex by switching of their ULM–UHM interactions and truncated pyrimidine tract can be recognized through the G-patch motif of SPF45 (see Fig. 8).

To test our hypothesis, we first examined the binding of SPF45 to SF3b155. We prepared *E. coli* recombinant glutathione S-transferase (GST)-fusion proteins of SPF45, its D319K mutant in the UHM (SPF45/UHMmt) that no longer binds any ULM, and G patch motif-deleted mutant (SPF45/ΔG) that loses potential interaction with nucleic acids and proteins[21] (Supplementary Fig. S6a). Our GST pull-down assays demonstrated that GST-SPF45 bound to SF3b155, but not to SF1, with crude nuclear extracts under physiological conditions even though both SF3b155 and SF1 contain ULM (Fig. 6a). As expected for the SPF45–SF3b155 interaction, the UHM of SPF45 is essential (Fig. 6a, 'GST-SPF45/UHMmt') while the G patch of SPF45 is not required (Fig. 6a, 'GST-SPF45/ΔG'), demonstrating the ULM–UHM interactions in the SF3b155–SPF45 complex. GST-SPF45 also binds to two other previously suggested SPF45-partner proteins that lack ULMs: spliceosomal A complex protein, SF4 (SUGP1 as HGNC approved symbol) and DEAH helicase protein of the U2-related group, hPRP43[22]. However, we confirmed that these two SPF45 interacting factors are not relevant to splicing of short introns (Supplementary Fig. S5b, c).

The binding between U2AF^{65}–ULM and U2AF^{35}–UHM is extremely strong[23]. Remarkably, GST-SPF45 did not pull-down U2AF^{65} and U2AF^{35} in crude nuclear extracts (Fig. 6a). These data together indicate that the U2AF^{65}–ULM cannot interact with the SPF45–UHM and thus the SPF45–UHM and U2AF^{65}–UHM compete for a functional binding toward the SF3b155–ULM (see Fig. 8). Therefore, we next investigated the competitive binding of U2AF^{65} and SPF45 toward SF3b155 by titrating the dose of GST-SPF45 in the immunoprecipitation assays (Fig. 6b). GST-SPF45 interfered with the binding between SF3b155 and U2AF^{65} in a dose-dependent manner, however, GST-SPF45/UHMmt did not disturb this binding. These results indicate that the SPF45–UHM competes with that of U2AF^{65}–UHM for the SF3b155 binding.

Finally, we examined whether the SPF45–SF3b155 interaction and the G-patch of SPF45 are essential for the SPF45-dependent splicing on short introns. We performed functional rescue experiments with SPF45-depleted HeLa cells using three siRNA-resistant proteins; SPF45 (SPF45/siR), SPF45-UHM mutant (SPF45/UHMmt/siR), and a G-patch motif-deleted SPF45

(SPF45/ΔG/siR; Supplementary Fig. S6a). We confirmed that the subcellular localization of these three mutant proteins did not change from that of endogenous SPF45 protein (Supplementary Fig. S6b). Protein expression levels of endogenous SPF45 and ectopically expressed three SPF45 siRNA-resistant mutants were checked by Western blotting in SPF45-depleted HeLa cells (Fig. 7, left panel). We analyzed splicing efficiencies of three mini-genes including short introns by RT–PCR (Fig. 7, right 3 panels). SPF45/siR rescued the splicing defects of all short introns in SPF45-depleted HeLa cells, however the SPF45/UHMmt/siR and SPF45/ΔG/siR did not (compare with control 'vector' lane). Taken together, we conclude that it is SPF45 that competes out U2AF^{65} and settles on the truncated PPT to promote splicing of pre-mRNAs with short introns (Fig. 8).

Discussion

Over a generation ago, two different splicing mechanisms termed 'intron-definition model' for short introns and 'exon-definition model' for long introns were proposed (reviewed in ref.[24]). In the former model, frequent lack of the PPT in vertebrate short introns was remarked and an alternative mechanism that circumvents this problem were assumed. Here we resolve this conjecture by demonstrating that a subset of human short introns, with significantly undersized pyrimidine tracts, is recognized by SPF45 but not by the U2AF heterodimer; implicating that SPF45 as a novel constitutive splicing factor in the spliceosomal complex A. This finding rationally explains why SPF45, which was previously considered just to be an alternative splicing factor, is essential for cell survival and maintenance *in vivo*[25].

We eventually discover the long-sought-after factor responsible for splicing on short introns, and furthermore, we provide mechanistic details of its function. We found that both SPF45 and U2AF^{65} can bind on introns with SF3b155 irrespective of intron size presumably *via* five ULMs in SF3b155, as previously shown in the simultaneous binding of U2AF^{65} and PUF60 to SF3b155[26].

This observation can be explained by the previous mass spectrometry analysis using AdML, MINX and PM5 pre-mRNA with conventional introns; *i.e.*, SPF45 is contained in E, A and B complexes as a U2 snRNP-related protein[13-15,27,28] (reviewed in ref.[29]). Our unique finding was that U2AF^{65} needs to be expelled by SPF45 to promote splicing of a subset of short introns. Previous structural study showed that U2AF^{65} binds to eight or nine nucleotides of pyrimidine tract[30,31]. Here we propose a mechanistic model that the weak and unstable U2AF^{65} binding on the truncated PPT of short intron allows access of SPF45, leading to the replacement of the U2AF heterodimer by SPF45 and its eventual placement on the truncated PPT through the G-patch motif (Fig. 8).

It was demonstrated that SPF45-depleted fruit fly (*Drosophila melanogaster*) S2 cells can be functionally rescued by human SPF45[32]. The SPF45-dependent splicing event on the shorter intron might be conserved in fruit fly. Interestingly, fruit fly spliceosomal B complex formed on *Zeste* pre-mRNA (with 62-nt intron including 14-nt PPT) contains SPF45, but that formed on *Ftz* pre-mRNA (with 147-nt intron including 33-nt PPT) does not[33]. Therefore, SPF45 is located exclusively in short introns in fruit fly, while in human, SPF45 exists in all introns on standby mode for short introns.

Our RNA-Seq analysis in the SPF45-knockdown cells detected the changes in alternative splicing (Fig. 2a). SPF45 was indeed identified and characterized as an alternative splicing regulator. In fruit fly, SPF45 interacts with Sex lethal (Sxl) protein and induces exon 3 skipping of *Sxl* pre-mRNA[32]. In mammals, SPF45 can cause exon 6 skipping in *FAS* pre-mRNA that produces soluble isoform of FAS inducing autoimmune phenotypes in mice[21].

In the SPF45-induced regulation of alternative splicing, there was no competition between SPF45 and U2AF heterodimer on the *Sxl* pre-mRNA[32]. Whereas we found a competitive and mutually exclusive binding of SPF45 and U2AF heterodimer on the truncated PPT to splice out short intron. We speculate that the cooperative interaction of SPF45 and U2AF[65] with SF3b155 may be required for alternative splicing regulation, whereas, exclusive binding of SPF45 with SF3b155, but without the U2AF heterodimer, is critical for short intron-specific constitutive splicing (Fig. 8). Interestingly, the U2-related protein PUF60 and U2AF[65] cooperatively interact with SF3b155 to activate weak 3′ splice sites[34] and the potential binding of both proteins to SF3b155 was demonstrated[26]. SPF45 interacts with other U2-related factors, U2SURP and CHERP, suggesting the role in alternative splicing[35]. However, our siRNA screening showed that knockdown of PUF60 and U2SURP have no effect on the splicing of hnRNPH1 short intron (Table S1, Nos. 89 and 126; unpublished data). Together, we conclude that the critical mechanism of SPF45 as a constitutive splicing factor is distinct from the mechanism of SPF45, with other interactors, as an alternative splicing factor.

Here we have just described one distinct subset of human SPF45-dependent short introns. Most recently, Smu1 and RED proteins were shown to activate spliceosomal B complexes assembled on human short introns[36]. The distance between the 5′ splice site and branch site need to be sufficiently short for Smu1/RED-dependent splicing, whereas in contrast, we clearly showed that this distance *per se* is not responsible for SPF45-dependent splicing (Fig. 3, 'AdML 5′MT'). Therefore, Smu1/Red-dependent and SPF45-dependent splicing mechanisms are essentially different, and thus they target two distinct subsets of human short introns.

The subset of SPF45-dependent short introns was identified by screening our siRNA library based on splicing activity on a 56-nt intron that contains conventional splice sites and branch site. We previously validated a list of ultra-short introns that includes remarkably atypical G-rich introns with completely inefficient splice sites and branch sites, of which the 49-nt intron 12 in the *NDOR1* gene and the 43-nt intron 6 in the *ESRP2* gene were analyzed[10,11]. The mechanism of splicing involved in such atypical G-rich introns is enigmatic. We assume the existence of another exotic subset of human ultra-short G-rich introns.

Methods

Construction of expression plasmids. The mini-gene expression plasmids, pcDNA3-HNRNPH1, pcDNA3-EML3 and pcDNA3-MUS81, were constructed by subcloning the corresponding PCR-amplified fragment into pcDNA3 vector (Invitrogen–Thermo Fisher Scientific). The PCRs were performed using genomic DNA of HeLa cells and specific primer sets (Table S3). For the pcDNA3-AdML, the PCR was performed using the pBS-Ad2 plasmid[37] and specific primer sets (Table S3).

The chimeric expression plasmids, pcDNA3-HNRNPH1/5'SSAdML, pcDNA3-HNRNPH1/branchAdML, pcDNA3-HNRNPH1/3'SSAdML, pcDNA3-HNRNPH1/AdML-PPT25, pcDNA3-HNRNPH1/AdML-PPT25mt, pcDNA3-HNRNPH1/AdML-PPT13, pcDNA3-HNRNPH1/5'AdML, pcDNA3-EML3/AdML-PPT25, pcDNA3-EML3/AdML-PPT13 and pcDNA3-EML3/5'AdML were constructed from the parent plasmids by overlap extension PCR with specific primer sets (Table S3).

To construct expression plasmids, pcDNA3-Flag-SPF45 and pGEX6p2-SPF45, the ORF region was PCR-amplified from HeLa cells cDNA and subcloned into the pcDNA3-Flag and pGEX6p2 vectors (GE Healthcare Life Sciences). In these plasmids, overlap extension PCR was performed to induce the mutation in the UHM motif of SPF45 (pcDNA3-Flag-SPF45/UHMmt and pGEX6p2-SPF45/UHMmt), to delete the G-patch motif (pcDNA3-Flag-SPF45/ΔG and pGEX6p2-SPF45/ΔG), and to make these siRNA-resistant variants (pcDNA3-

Flag-SPF45/siR, pcDNA3-Flag-SPF45/UHMmt/siR and pcDNA3-Flag-SPF45/ΔG/siR).

Western blotting analyses. Protein samples were boiled with NuPAGE LDS sample buffer (Thermo Fisher Scientific), separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE), and the gel was electroblotted onto an Amersham Protran NC Membrane (GE Health Care Life Sciences). The following commercially available antibodies were used to detect targeted proteins: anti-SPF45 (Sigma-Aldrich), anti-SF3b155 (MBL Life Science), anti-U2AF^{65} (Sigma-Aldrich), anti-U2AF^{35} (Proteintech), anti-SF4 (Sigma-Aldrich), anti-U1-70K (Santa Cruz Biotechnology), anti-GAPDH (MBL Life Science) and anti-Flag (anti-DYKDDDDK tag; Wako). The anti-hPRP43 antibody was described previously[38]. Immuno-reactive protein bands were detected by the ECL system and visualized by imaging analyzer (ImageQuant LAS 500, GE Healthcare Life Sciences).

Splicing efficiency screening of siRNA library. HeLa cells were cultured in Dulbecco's modified Eagle's medium (Wako) supplemented with 10% fetal bovine serum. HeLa cells in 35-mm dishes were transfected with 100 pmol of each siRNA in the Stealth siRNA library targeting 154 human nuclear proteins (Invitrogen–Thermo Fisher Scientific) using Lipofectamine RNAiMax (Invitrogen–Thermo Fisher Scientific) according to the manufacturer's protocol.

At 48–96 h post-transfection, total RNAs were isolated from the siRNA-treated HeLa cells and splicing efficiency was analyzed by RT–PCR using a primer set targeting intron 7 of HNRNPH1 (Table S3). The PCR products were separated on 5% PAGE, visualized by imaging analyzer (ImageQuant LAS 500, GE Healthcare Life Sciences), and the unspliced pre-mRNA and spliced mRNA were quantified using NIH Image J software. See below for the detailed procedures.

siRNA knockdown and splicing assays. HeLa cells and HEK293 cells (in 35-mm dishes) were transfected with 100 pmol siRNA using Lipofectamine RNAi max (Invitrogen–Thermo Fisher Scientific) according to manufacturer's protocol. At 72 h post-transfection, total RNAs were isolated from the siRNA-treated cells using a NucleoSpin RNA kit (Macherey-Nagel). To check depletion of the siRNA-targeted proteins, transfected cells were suspended in Buffer D [20 mM HEPES (pH 7.9), 50 mM KCl, 0.2 mM EDTA, 20% glycerol], sonicated for 20 sec, centrifuged to remove debris, and the lysates were subjected to Western blotting (see above). The siRNAs for SPF45 siRNA#1, SPF45 siRNA#2, U2AF^{65} siRNA#1[17], hPRP43 siRNA#1 were purchased from Nippon Gene (Table S3 for the sequences).

To analyze endogenous splicing products derived from the *HNRNPH1*, *RFC4*, *EML3*, *DUSP1*, *NFKBIA*, *MUS81*, *RECQL4* and *MTA1* genes, total RNAs from siRNA-treated cells were reverse transcribed by PrimeScript II reverse transcriptase (Takara Bio) with oligo-dT and random primers, and the obtained cDNAs were analyzed by PCR using the specific primer sets (Table S3). The PCR products were resolved by 6% PAGE. Splicing products were quantified using NIH Image J software. All the experiments were independently repeated three times and the means and standard errors of the splicing efficiencies were calculated.

To analyze splicing products derived from mini-genes, SPF45- and U2AF^{65}-depleted HeLa cells were transfected at 48 h and 68 h post-transfection, respectively, with 0.5 μg of mini-gene plasmid (Table S3) using lipofectamine 2000 reagent (Invitrogen–Thermo Fisher Scientific). These cells were incubated for 24 h and 4 h, respectively, prior to the extraction of RNAs (described above). To analyze splicing products from mini-genes, RT–PCR was performed with T7 primer and a specific primer for each mini-gene (Table S3). All the PCR products were analyzed by 6% PAGE and quantified (described above).

To perform rescue experiments, SPF45-depleted HeLa cells were transfected with 1 μg of pcDNA3-Flag-SPF45/siR, pcDNA3-Flag-SPF45/UHMmt/siR, or pcDNA3-Flag-SPF45/ΔG/siR at 24 h post-transfection. After 48 h culture, total RNA and protein were isolated for RT–PCR and Western blotting, respectively (described above).

In this study, all the oligonucleotide primers were purchased (Fasmac; Table S3) and all the PCRs were performed with Blend Taq polymerase (Toyobo Life Science).

High-throughput RNA sequencing (RNA-Seq) analyses. Six independent total RNAs derived from HEK293 cells, treated with three control siRNAs and three SPF45-targeted siRNAs, were prepared by NucleoSpin RNA kit (Macherey-Nagel). Then rRNA depletion was performed with the RiboMinus Eukaryote System v2 (Invitrogen–Thermo Fisher Scientific). RNA libraries were prepared using the NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs). These samples were sequenced on the high-throughput sequencing platform (HiSeq2500, Illumina) using a 100 bp single-end strategy.

The sequencing data was analyzed as previously described[39]. Obtained sequence reads were mapped onto the human genome reference sequences (hg19) using the TopHat version 2.1.1 (https://ccb.jhu.edu/software/tophat/index.shtml) and the mapped sequence reads, as BAM files, were assembled using Cufflinks version 2.2.1 (http://cufflinks.cbcb.umd.edu). Using the obtained Cufflinks GTF files as a reference, the BAM files were analyzed using rMATS version 3.2.0 (http://rnaseq-mats.sourceforge.net/)[40] to examine the changes of alternative splicing isoforms. Significant changes of splicing events were defined as when the false discovery rate (FDR) was calculated at less than 0.05.

The strengths of the 5′ and 3′ splice sites were calculated using MAXENT (http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html, http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html)[41], and branch point strength, PPT score and PPT length were calculated by SVM-BP finder software (http://regulatorygenomics.upf.edu/Software/SVM_BP/)[42]. 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 GSE135128.

To analyze the sets of retained introns in U2AF^{65}- and SF3b155-depleted HeLa cells, these RNA-Seq data were obtained from the GEO database (Accession numbers GSE65644 and GSE61603)[17,43].

Cellular formaldehyde- and UV-crosslinking followed by immunopreciptation assays. To detect Flag-SPF45 association to a pre-mRNA expressed from a reporter mini-gene, we performed formaldehyde crosslinking followed by immunoprecipitation as previously described[12]. Briefly, HEK293 cells (in 60-mm dishes) were co-transfected with pcDNA3-Flag-SPF45 and a mini-gene plasmid (pcDNA3-hnRNPH1, pcDNA3-EML3, pcDNA3-MUS81 or pcDNA3-AdML) using Lipofectamine 2000 reagent (Invitrogen–Thermo Fisher Scientific). At 48 h post-transfection, cells were harvested, washed with cold PBS buffer, and fixed with 0.2% formaldehyde for 10 min. The fixation was quenched in 0.15 M glycine (pH 7.0) and cells were washed with PBS. Immunoprecipitations were performed using anti-Flag antibody-conjugated beads to analyze pre-mRNA, from the mini-gene, associated with Flag-SPF45.

To detect endogenous U2AF^{65}- and SF3b155-association to a pre-mRNA expressed from a reporter mini-gene, we performed UV crosslinking followed by immunoprecipitation as previously described[44,45]. PBS-washed HEK293 cells were irradiated with 254-nm UV light on ice. The collected cells were lysed and immunoprecipitated with anti-U2AF^{65} and anti-SF3b155 antibodies.

Immunoprecipitated RNAs were extracted with Trizol reagent (Invitrogen–Thermo Fisher Scientific). The isolated RNAs were reverse-transcribed using PrimeScript II reverse transcriptase (Takara Bio) with SP6 primer, and qPCRs were performed using specific primer sets (Table S3).

Biotinylated RNA pull-down assays. Nuclear extracts were prepared from HEK293 cells transfected with control siRNA or SPF45 siRNA according to the small-scale preparation procedure[46]. Biotin-labeled HNRNPH1 and AdML pre-mRNAs were transcribed with a MEGAscript T7 transcription kit (Invitrogen–Thermo Fisher Scientific) according to the manufacturer's instructions.

The biotinylated pre-mRNA (10 pmol) was immobilized with 5 μL of Dynabeads MyOne StreptavidinT1 magnetic beads (Invitrogen–Thermo Fisher Scientific) according to the manufacturer's instruction. The immobilized pre-mRNA beads were incubated at 30°C for 15 min in 30 μL reaction mixture containing 30% nuclear extract, RNase inhibitor (Takara Bio) and nuclease-free water. Then NET2 buffer [50mM Tris (pH7.5), 150 mM NaCl and 0.05% Nonidet P-40] was added to a final volume of 700 μL and incubated at 4°C for 1 h. The incubated beads were washed six times with cold NET2 buffer and boiled in SDS-PAGE sample buffer to analyze by Western blotting (described above).

GST pull-down assays. GST-SPF45, GST-SPF45/UHMmt, or GST-SPF45/ΔG were expressed in *E. coli* BL21(DE3) CodonPlus (DE3) competent cells (Stratagene–Agilent) and the GST-tagged recombinant proteins were checked by SDS-PAGE followed by Coomassie Blue staining. Induction was carried out at 37°C for 3 h. The GST-proteins were purified using Glutathione Sepharose 4B (GE Healthcare Life Sciences) according to the manufacturer's protocol.

The recombinant GST-SPF45 proteins (5 μg) were incubated at 30°C for 15 min in 100 μL mixture containing 30% HeLa cell nuclear extract. After RNase A treatment, NET2 buffer was added to a final volume of 1 mL with 20 μL of Glutathione Sepharose 4B or SF3b155a-antibody conjugated with Protein G Sepharose (GE Healthcare Life Sciences) and incubated at 4°C for 3 h. The incubated beads were washed six times with cold NET2 buffer and boiled in SDS-PAGE sample buffer to analyze by Western blotting (described above).

Immunofluorescence microscopic assays. Immunofluorescence microscopic assays of ectopically expressed Flag-tagged SPF45 proteins were performed as essentially described previously[45].

HeLa cells (in 35-mm dishes) were transfected with 1 μg of pcDNA3-Flag-SPF45/siR, pcDNA3-Flag-SPF45/UHMmt/siR, or pcDNA3-Flag-SPF45/ΔG/siR using lipofectamine 2000 reagent (Invitrogen–Thermo Fisher Scientific). At 48 h post-transfection, cells were fixed with 3% formaldehyde/PBS, permeabilized with 0.1% Triton X-100/PBS, blocked with 5% skimmed milk/PBS and then incubated with primary antibodies in 2% skimmed milk/PBS for 0.5 h. After three washes with PBS, cells were incubated with Alexa Fluor 488 or Alexa Fluor 568 secondary antibody (Invitrogen–Thermo Fisher Scientific) and then washed 5 times with PBS. DNA in cells was counter-stained with 4', 6-diamidino-2-phenylindole (DAPI). The images were analyzed by fluorescence microscope (Olympus).

Author contributions

K.F. and A.M. conceived and designed the experiments; K.F. performed most of the experiments and analyses, organized the data and drafted the manuscript; R.Y. performed

bioinformatics analyses of the sequencing data; T.H. and K.I. contributed toward the success of the screening technology using the siRNA library for human nuclear proteins; K.F. and A.M. revised and edited the manuscript. A.M. coordinated and supervised the whole project. All authors read, corrected and approved the final manuscript.

Acknowledgments

We thank Drs. Hyun-Seo Kang and Michael Sattler for their valuable discussions with shared unpublished data; Dr. Adrian Krainer for helpful suggestions and encouragements; Dr. Julian Venables for critical reading of the manuscript; and our lab members for their constructive discussions. 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) and by a Research Grant from the Hori Sciences and Arts Foundation. A.M. was partly supported by Grants-in-Aid for Scientific Research (B) [Grant number: JP16H04705] and for Challenging Exploratory Research [Grant number: JP16K14659] from JSPS.

Competing interests

The authors declare no competing interests.

Additional information

Supplemental Information is available in separate files.

References

1. Lim, L. P. & Burge, C. B. A computational analysis of sequence features involved in recognition of short introns. *Proc. Natl. Acad. Sci. USA* **98**, 11193-11198 (2001).
2. Yu, J. *et al.* Minimal introns are not "junk". *Genome Res.* **12**, 1185-1189 (2002).
3. Zhu, J. *et al.* A novel role for minimal introns: routing mRNAs to the cytosol. *PLoS One* **5**, e10144 (2010).
4. Mayeda, A. & Krainer, A. R. in *Alternative pre-mRNA splicing: Theory and protocols* (eds S. Stamm, C. Smith, & R. Lührmann) Ch. 30, 321-329 (Wiley-Blackwell, 2012).
5. López-Mejía, I. & Tazi, J. in *Alternative pre-mRNA splicing: Theory and protocols* (eds S. Stamm, C. Smith, & R. Lührmann) Ch. 36, 392-399 (Wiley-Blackwell, 2012).
6. Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701-718 (2009).
7. Behzadnia, N. *et al.* Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *EMBO J.* **26**, 1737-1748 (2007).
8. Glass, J. & Wertz, G. W. Different base per unit length ratios exist in single-stranded RNA and single-stranded DNA. *Nucleic Acids Res.* **8**, 5739-5751 (1980).
9. Zhang, L., Vielle, A., Espinosa, S. & Zhao, R. RNAs in the spliceosome: Insight from cryoEM structures. *WIREs RNA* **10**, e1523 (2019).
10. Sasaki-Haraguchi, N., Shimada, M. K., Taniguchi, I., Ohno, M. & Mayeda, A. Mechanistic insights into human pre-mRNA splicing of human ultra-short introns: potential unusual mechanism identifies G-rich introns. *Biochem. Biophys. Res. Commun.* **423**, 289-294 (2012).
11. Shimada, M. K., Sasaki-Haraguchi, N. & Mayeda, A. Identification and Validation of Evolutionarily Conserved Unusually Short Pre-mRNA Introns in the Human Genome. *Int. J. Mol. Sci.* **16**, 10376-10388 (2015).

12. Yong, J., Kasim, M., Bachorik, J. L., Wan, L. & Dreyfuss, G. Gemin5 delivers snRNA precursors to the SMN complex for snRNP biogenesis. *Mol. Cell* **38**, 551-562 (2010).

13. Zhou, Z., Licklider, L. J., Gygi, S. P. & Reed, R. Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**, 182-185 (2002).

14. Bessonov, S., Anokhina, M., Will, C. L., Urlaub, H. & Lührmann, R. Isolation of an active step I spliceosome and composition of its RNP core. *Nature* **452**, 846-850 (2008).

15. Makarov, E. M., Owen, N., Bottrill, A. & Makarova, O. V. Functional mammalian spliceosomal complex E contains SMN complex proteins in addition to U1 and U2 snRNPs. *Nucleic Acids Res.* **40**, 2639-2652 (2012).

16. Pacheco, T. R., Coelho, M. B., Desterro, J. M., Mollet, I. & Carmo-Fonseca, M. In vivo requirement of the small subunit of U2AF for recognition of a weak 3' splice site. *Mol. Cell. Biol.* **26**, 8183-8190 (2006).

17. Shao, C. *et al.* Mechanisms for U2AF to define 3' splice sites and regulate alternative splicing in the human genome. *Nat. Struct. Mol. Biol.* **21**, 997-1005 (2014).

18. Svec, M., Bauerová, H., Pichová, I., Konvalinka, J. & Strísovský, K. Proteinases of betaretroviruses bind single-stranded nucleic acids through a novel interaction module, the G-patch. *FEBS Lett.* **576**, 271-276 (2004).

19. Silverman, E. J. *et al.* Interaction between a G-patch protein and a spliceosomal DEXD/H-box ATPase that is critical for splicing. *Mol. Cell. Biol.* **24**, 10101-10110 (2004).

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

21. Corsini, L. *et al.* U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. *Nat. Struct. Mol. Biol.* **14**, 620-629 (2007).

22. Hegele, A. *et al.* Dynamic protein-protein interaction wiring of the human spliceosome. *Mol. Cell* **45**, 567-580 (2012).

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

24. Berget, S. M. Exon recognition in vertebrate splicing. *J. Biol. Chem.* **270**, 2411-2414 (1995).

25. Tan, Q. *et al.* Extensive cryptic splicing upon loss of RBM17 and TDP43 in neurodegeneration models. *Hum. Mol. Genet.* **25**, 5083-5093 (2016).

26. Corsini, L. *et al.* Dimerization and protein binding specificity of the U2AF homology motif of the splicing factor Puf60. *J. Biol. Chem.* **284**, 630-639 (2009).

27. Neubauer, G. *et al.* Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat. Genet.* **20**, 46-50 (1998).

28. Agafonov, D. E. *et al.* Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Mol. Cell. Biol.* **31**, 2667-2682 (2011).

29. Wahl, M. C. & Lührmann, R. SnapShot: Spliceosome Dynamics I. *Cell* **161**, 1474-e1471 (2015).

30. Mackereth, C. D. *et al.* Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature* **475**, 408-411 (2011).

31. Agrawal, A. A. *et al.* An extended $U2AF^{65}$-RNA-binding domain recognizes the 3' splice site signal. *Nat. Commun.* **7**, 10950 (2016).

32. Lallena, M. J., Chalmers, K. J., Llamazares, S., Lamond, A. I. & Valcárcel, J. Splicing regulation at the second catalytic step by Sex-lethal involves 3' splice site recognition by SPF45. *Cell* **109**, 285-296 (2002).

33. Herold, N. *et al.* Conservation of the protein composition and electron microscopy structure of Drosophila melanogaster and human spliceosomal complexes. *Mol. Cell. Biol.* **29**, 281-301 (2009).

34. Hastings, M. L., Allemand, E., Duelli, D. M., Myers, M. P. & Krainer, A. R. Control of pre-mRNA splicing by the general splicing factors PUF60 and U2AF^{65}. *PLoS One* **2**, e538 (2007).

35. De Maio, A. *et al.* RBM17 Interacts with U2SURP and CHERP to Regulate Expression and Splicing of RNA-Processing Proteins. *Cell reports* **25**, 726-736 e727 (2018).

36. Keiper, S. *et al.* Smu1 and RED are required for activation of spliceosomal B complexes assembled on short introns. *Nat. Commun.* **10**, 3639 (2019).

37. Fukumura, K., Taniguchi, I., Sakamoto, H., Ohno, M. & Inoue, K. U1-independent pre-mRNA splicing contributes to the regulation of alternative splicing. *Nucleic Acids Res.* **37**, 1907-1914 (2009).

38. Yoshimoto, R., Kataoka, N., Okawa, K. & Ohno, M. Isolation and characterization of post-splicing lariat-intron complexes. *Nucleic Acids Res.* **37**, 891-902 (2009).

39. Yoshimoto, R. *et al.* Global analysis of pre-mRNA subcellular localization following splicing inhibition by spliceostatin A. *RNA* **23**, 47-57 (2017).

40. Shen, H. *et al.* Reversal of multidrug resistance of gastric cancer cells by downregulation of TSG101 with TSG101siRNA. *Cancer Biol. Ther.* **3**, 561-565 (2004).

41. Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11**, 377-394 (2004).

42. Corvelo, A., Hallegger, M., Smith, C. W. J. & Eyras, E. Genome-wide association between branch point properties and alternative splicing. *PLoS Comput. Biol.* **6**, e1001016 (2010).

43. Kfir, N. *et al.* SF3B1 association with chromatin determines splicing outcomes. *Cell reports* **11**, 618-629 (2015).

44. Sasaki, Y. T., Ideue, T., Sano, M., Mituyama, T. & Hirose, T. MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc. Natl. Acad. Sci. U S A* **106**, 2525-2530 (2009).

45. Fukumura, K., Inoue, K. & Mayeda, A. Splicing activator RNPS1 suppresses errors in pre-mRNA splicing: A key factor for mRNA quality control. *Biochem. Biophys. Res. Commun.* **496**, 921-926 (2018).

46. Lee, K. A. W. & Green, M. R. Small-scale preparation of extracts from radiolabeled cells efficient in pre-mRNA splicing. *Methods Enzymol.* **181**, 20-30 (1990).

FIGURES

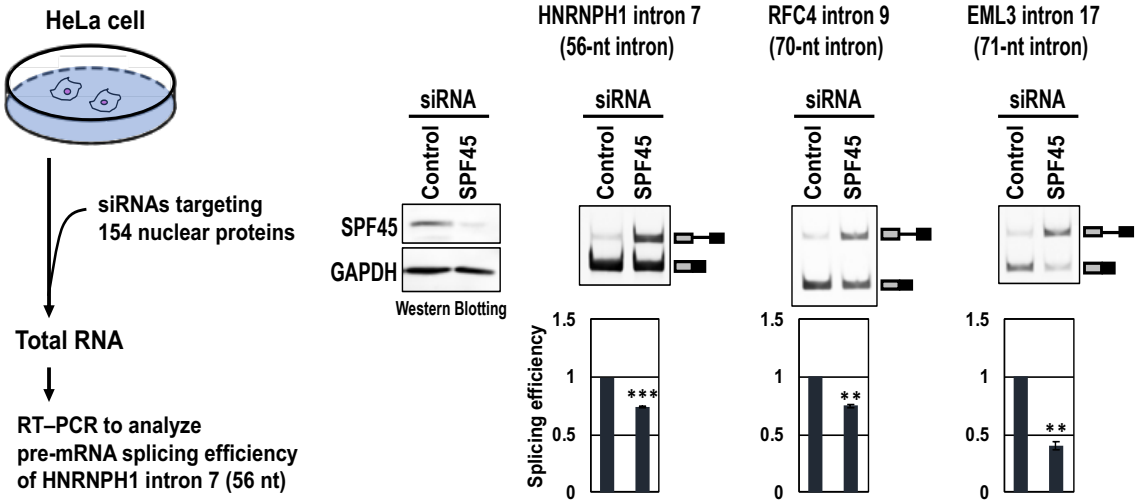

Fig. 1 | SPF45 was identified by siRNA screening for HNRNPH1 pre-mRNA (with 56-nt intron) splicing. The siRNA screening procedure to search for a specific splicing factor for short introns (left panel). The SPF45 protein depletion by siRNA-knockdown in HeLa cells was checked by a Western blotting (middle panel). After the siRNA transfection, endogenous splicing of the indicated three representative short introns were analyzed by RT–PCR (right 3 panels). Means ± standard error (SE) are given for three independent experiments and two-sided t test values were calculated (**P < 0.005, ***P < 0.0005).

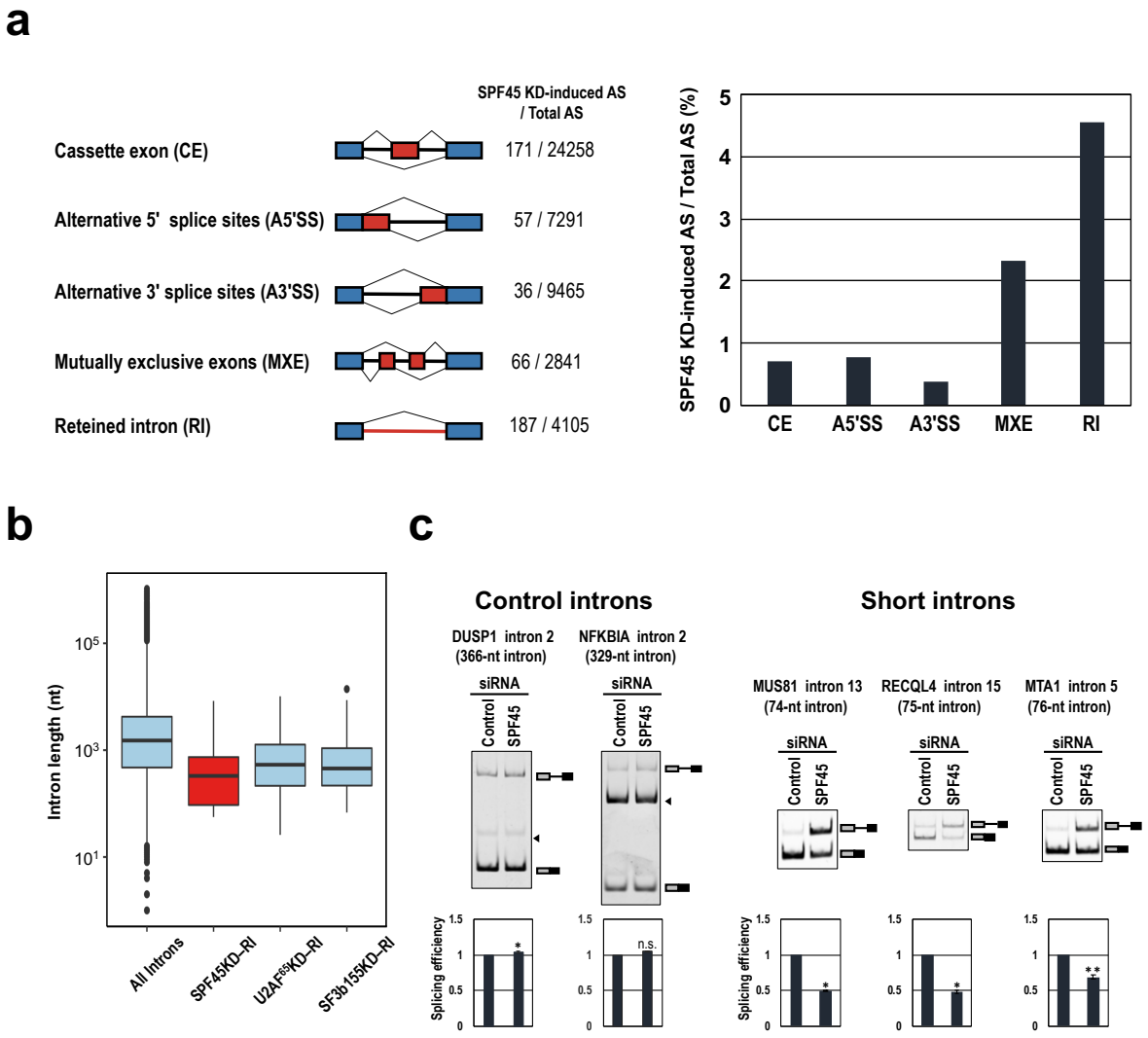

Fig. 2 | SPF45 is generally required for splicing of pre-mRNAs including short introns. a, RNA-Seq data exhibits deferential splicing patterns between control siRNA- and SPF45 siRNA-treated HEK293 cells. The numbers of significant splicing changes and total splicing events are indicated and the ratios are shown on the right. **b,** Box plots are comparing the intron-length distributions of all introns in human RefGene with those of the retained introns in SPF45-knockdown HEK293 cells. The retained introns in U2AF^{65}- and SF3b155-knockdown HeLa cells obtained from the RNA-Seq data in GEO database are shown for comparison (significant for all pairs, $P < 0.05$). **c,** SPF45-knockdown selectively repressed splicing of pre-mRNAs with short introns. After the siRNA transfection in HEK293 cells, endogenous splicing of the indicated two control introns and three short introns were analyzed by RT–PCR. Arrowheads indicate non-specific PCR products. Means ± SE are given for three independent experiments (*$P < 0.05$, **$P < 0.005$, n.s.=not statistically significant $P > 0.05$).

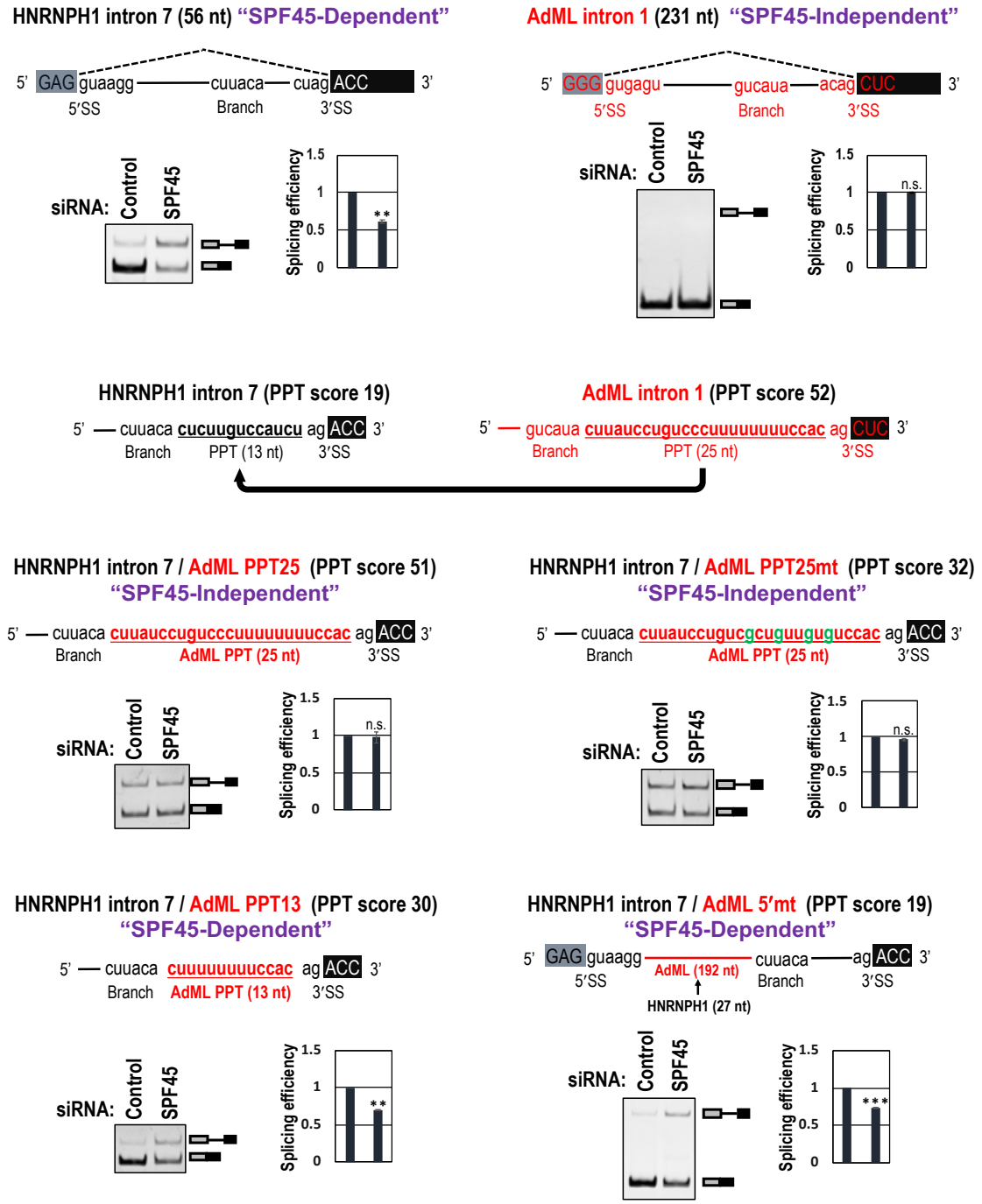

Fig. 3 | The polypyrimidine tracts (PPTs) determine the SPF45-dependency of splicing. The PPT sequences have effects on SPF45-dependent splicing. Original HNRNH1 and AdML pre-mRNAs and chimeric HNRNH1 pre-mRNAs are schematically shown (red color indicates AdML derived sequences). These pre-mRNAs were expressed from mini-genes in HeLa cells and their splicing was assayed by RT–PCR. PAGE images and quantifications of RT–PCR are shown. Means ± SE are given for three independent experiments (**P < 0.005, ***P < 0.0005, n.s.=not statistically significant P > 0.05).

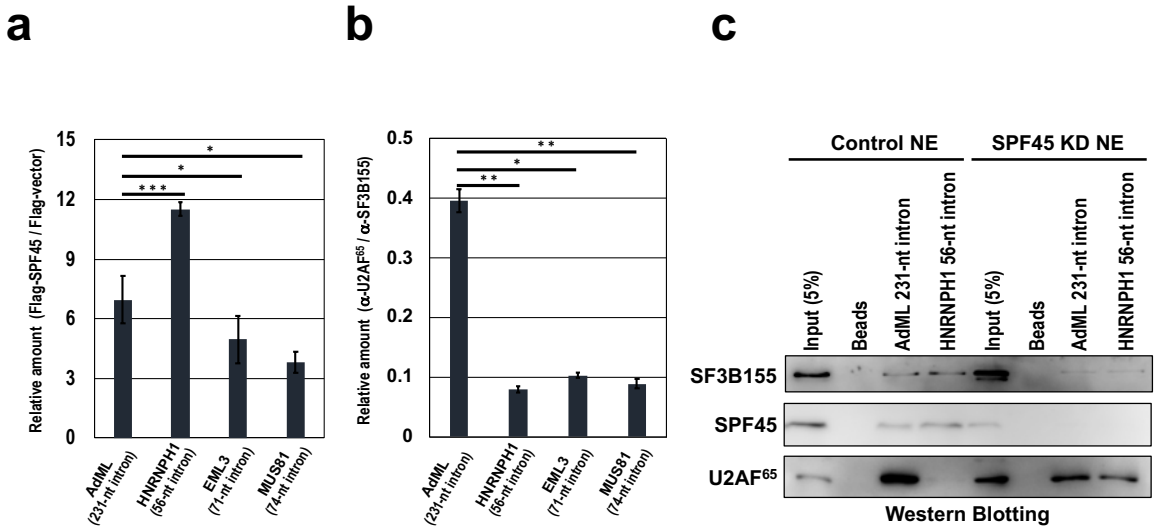

Fig. 4 | Binding of SPF45 competes out U2AF^{65} on truncated PPTs in short introns. **a,** Cellular formaldehyde crosslinking and immunoprecipitation experiments shows SPF45 binding to all the indicated introns. Mini-genes containing these four introns were individually co-transfected into HEK293 cells with a plasmid expressing Flag-SPF45 protein. The Flag-SPF45 was immunoprecipitated after formaldehyde crosslinking and then co-precipitated RNAs were quantified by RT–qPCR using specific primers. Means ± SE are given for three independent experiments (*P < 0.05, ***P < 0.0005). **b,** Cellular CLIP experiments shows strong U2AF^{65} binding to control AdML pre-mRNA but not much to the three indicated short introns. Mini-genes containing these four introns were individually co-transfected into HEK293 cells and irradiated with UV light. The lysates were immunoprecipitated with anti-U2AF^{65} and anti-SF3b155 antibodies and then immunoprecipitated RNAs were quantified by RT–qPCR using specific primers. Means ± SE are given for three independent experiments (*P < 0.05, **P < 0.005). **c,** Affinity pull-down experiments of biotinylated RNA indicates U2AF^{65} binding to the short intron only if SPF45 was depleted. Biotinylated pre-mRNAs including short HNRNPH1 intron (56 nt) and control AdML intron (231 nt) were incubated with nuclear extracts from either control siRNA- or SPF45 siRNA-treated HEK293 cells. The biotinylated RNA-bound proteins were pulled down with streptavidin-coated beads and analyzed by Western blotting with antibodies against SF3b155, SPF45 and U2AF^{65}.

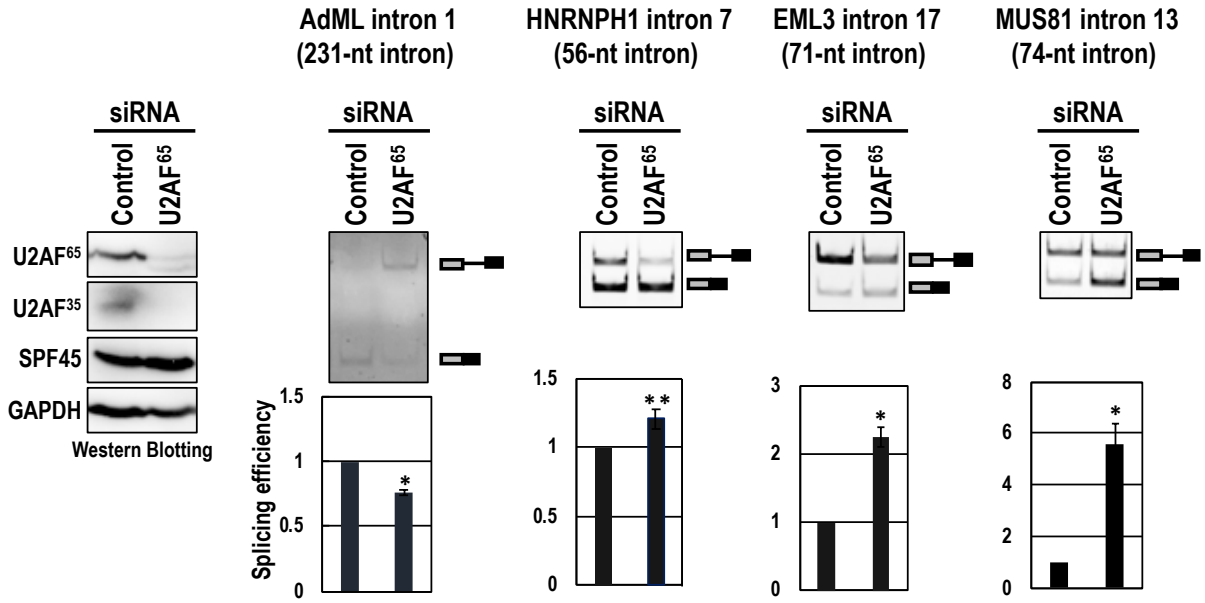

Fig. 5 | Depletion of U2AF^{65} rather increases splicing of pre-mRNAs with SPF45-dependent short introns. The co-depletion of U2AF^{65} and U2AF^{35} proteins from HeLa cells was observed by a Western blotting (left panel). After the siRNA transfection, splicing efficiencies of the indicated four mini-genes were analyzed by RT–PCR (right 4 panels). Means ± SE are given for three independent experiments (*P < 0.05, **P < 0.005).

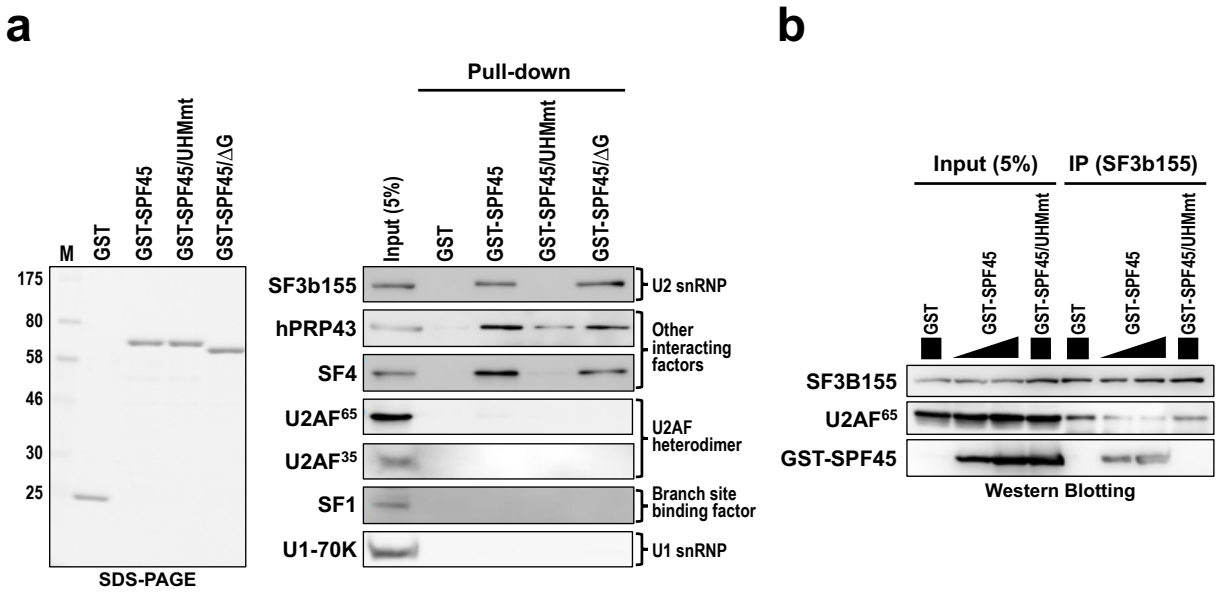

Fig. 6 | ULM–UHM binding between SF3b155 and SPF45 promotes splicing of pre-mRNA with short introns. **a,** A GST pull-down assay in RNase A-treated HeLa nuclear extract shows the UHM-dependent binding of GST-SPF45 to SF3b155 but not to U2AF heterodimer. SDS-PAGE analysis of the indicated purified recombinant proteins (left panel). Proteins that associated with these GST-fusion proteins were detected by Western blotting using the indicated antibodies (right panel). **b,** Immunoprecipitation of SF3b155 with the indicated GST-fusion proteins shows the competitive binding between $U2AF^{65}$ and SPF45 to SF3b155 *via* UHM–ULM interactions. The same reaction mixtures in panel **a** were immunoprecipitated with anti-SF3b155 antibody and the associated proteins were analyzed by Western blotting with antibodies against SF3b155, $U2AF^{65}$ and SPF45.

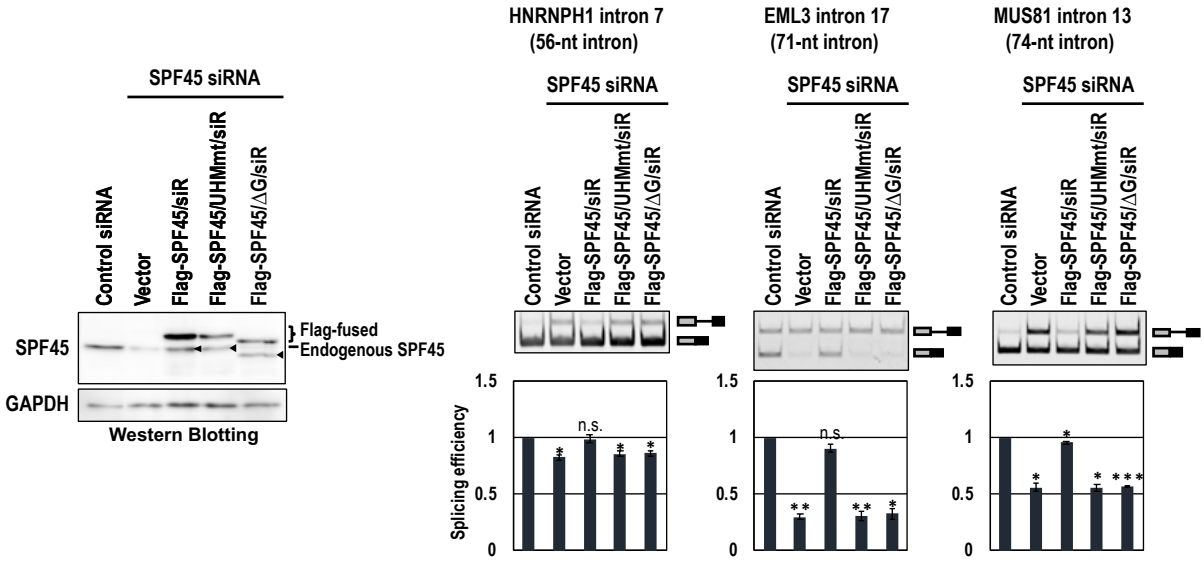

Fig. 7 | The expression of siRNA-resistant SPF45 proteins rescue splicing of short introns in SPF45-depleted cells. The expressed Flag-fused siRNA-resistant (siR) proteins and endogenous SPF45 in HeLa cells were checked by Western blotting (left panel). Arrowheads indicate degraded siRNA-resistant (siR) proteins, but not endogenous SPF45 (see 'vector' lane for the depletion efficiency of endogenous SPF45). After the co-transfection of the indicated siRNA-resistant plasmids and three mini-gene plasmids, splicing efficiencies of the indicated three mini-genes were analyzed by RT–PCR (right 3 panels). Means ± SE are given for three independent experiments (*P < 0.05, **P < 0.005, n.s. P > 0.05).

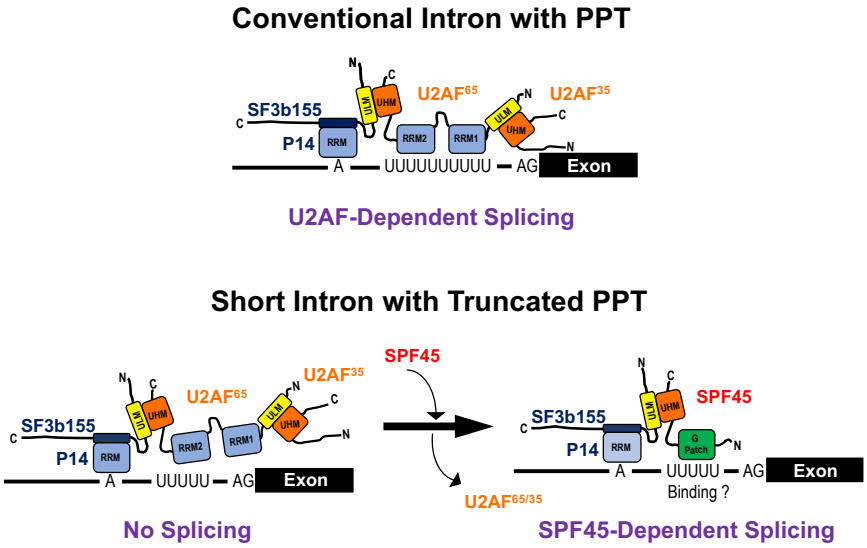

Fig. 8 | The model of U2AF-dependent splicing on a conventional-sized intron with a regular PPT and SPF45-dependent splicing on a short intron with a truncated PPT. The associated splicing factors with the domain structures and the target sequences of pre-mRNAs are represented schematically. PPT is indicated with a stretch of 'U'. On short intron with truncated PPT, U2AF-heterodimer is replaced by SPF45 and interacts with SF3b155 (U2 snRNP component) *via* UHM–ULM binding to promote splicing.

The general splicing factor SPF45 (RBM17) replaces the U2AF heterodimer in a distinct subset of human short introns

Kazuhiro Fukumura[1]*, Rei Yoshimoto[1], Tetsuro Hirose[2], Kunio Inoue[3] and Akila Mayeda[1]*

[1]Division of Gene Expression Mechanism, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Aichi 470-1192, Japan
[2]Institute for Genetic Medicine, Hokkaido University, Sapporo, Hokkaido 060-0815, Japan
[3]Department of Biology, Graduate School of Science, Kobe University, Kobe, Hyogo 657-8501, Japan

*e-mails: fukumura@fujita-hu.ac.jp (K.F.), mayeda@fujita-hu.ac.jp (A.M.)

Contents

Table S1
Table S3
Fig. S1
Fig. S2
Fig. S3
Fig. S4
Fig. S5
Fig. S6

Table S2
(Uploaded in a separate Excel file)

Table S1 List of siRNA library targeting human nuclear proteins

No.	Protein name / Gene symbol	No.	Protein name / Gene symbol	No.	Protein name / Gene symbol	No.	Protein name / Gene symbol
1	Acinus (ACIN1)	41	GRSF1	81	Nucleolin (NCL)	121	SERBP1
2	AHDC1	42	GRY-RBP(SYNCRIP)	82	Peptidylprolyl isomerase E (PPIE)	122	Set1 (SETD1A)
3	AKAP149 (AKAP1)	43	hnRNP A/B (HNRNPAB)	83	POLDIP3	123	SF2/ASF (SRSF1)
4	AKAP8L	44	HnRNP D-like (HNRNPDLP1)	84	PPARGC1 (PPARGC1A)	124	SFRS10 (TRA2B)
5	Aly (ALYREF)	45	hnRNP RALY (RALY)	85	PSF (SFPQ)	125	**SPF45 (RBM17)**
6	ASR2B (SRRT)	46	hnRNPA0 (HNRNPA0)	86	PSP1 (PSPC1)	126	SR140 (U2SURP)
7	Ataxin1 (ATXN1)	47	hnRNP A1 (HNRNPA1)	87	PSP2 (RBM14)	127	SRM300 (SRRM2)
8	Barentsz (CASC3)	48	hnRNPA2/B1 (HNRNPA2B1)	88	PTB (PTBP1)	128	SRp20 (SRSF3)
9	CArG-binding factor A (NFYB)	49	hnRNP A3 (HNRNPA3)	89	PUF60	129	SRp30c (SRSF9)
10	CBP20 (NCBP2)	50	hnRNPC1/C2 (HNRNPC)	90	Puralpha (PURA)	130	SRp38 (SRSF10)
11	CBX7	51	hnRNP D (HNRNPD)	91	p54nrb (NONO)	131	SRp40 (SRSF5)
12	CELF5	52	hnRNP E1 (PCBP1)	92	RAVER1	132	SRp54 (SRSF11)
13	CELF6	53	hnRNP F (HNRNPF)	93	RBM10	133	SRp75 (SRSF4)
14	Cflm68 (CPSF6)	54	hnRNP H1 (HNRNPH1)	94	RBM12	134	SRrp86 (SREK1)
15	CIRBP	55	hnRNP H2 (HNRNPH2)	95	RBM15	135	STRBP
16	CLK1	56	hnRNP H3 (HNRNPH3)	96	RBM19	136	TAF15
17	CLK4	57	hnRNP K (HNRNPK)	97	RBM22	137	TAP (NXF1)
18	CPSF5 (NUDT21)	58	hnRNP L (HNRNPL)	98	RBM28	138	TAT-SF1 (HTATSF1)
19	CPSF7	59	hnRNP M (HNRNPM)	99	RBM3	139	TEP1
20	CUGBP (CELF1)	60	hnRNP R (HNRNPR)	100	RBM30(RBM4B)	140	TIA1
21	DAZAP1	61	hnRNP UL1 (HNRNPUL1)	101	RBM42	141	$U2AF^{35}$ (U2AF1)
22	DGCR8	62	HP1a (CBX5)	102	Rbm4a(RBM4)	142	$U2AF^{65}$ (U2AF2)
23	ECM2	63	IMP1 (HM13)	103	RBM5	143	UBAP2L
24	ELAVlike (ELAVL1)	64	IMP3	104	RBM6	144	vigilin (HDLBP)
25	ES349 (FUBP1)	65	Kin 17 (KIN)	105	RBM7	145	vparp (PARP4)
26	EWS (EWSR1)	66	KSRP (KHSRP)	106	RBM8A	146	WF9/2C7 (ZRSR2)
27	F9/11A1 (DDX18)	67	La autoantigen (SSB)	107	RBMX	147	WF9/5A7 (RBM39)
28	F9/17A4 (NUP153)	68	Matrin 3 (MATR3)	108	RBMX2	148	wig-1 (ZMAT3)
29	F9/29D5 (SRM160)	69	MBD1	109	RBPMS	149	YB-1 (YBX1)
30	F9/3B6 (ZNF346)	70	MBD2	110	RDBP (NELFE)	150	ZC3H6
31	FAM113A (PCED1A)	71	MBNL (MBNL1)	111	RIP-1 (KRR1)	151	ZFR
32	FAM98A	72	MECP2	112	RNPS1	152	ZNF335
33	FIGN	73	MEX3A	113	Ro autoantigen 60 kDa (RO60)	153	ZNF74
34	FLJ10005 (SLTM)	74	MEX3B	114	RPP25	154	9G8 (SRSF7)
35	FLJ10968 (IM3)	75	MEX3D	115	RUNX3		
36	FLJ20273 (RBM47)	76	MINT (SPEN)	116	SAF-A (HNRNPU)		
37	FUBP3	77	MSP58 (MCRS1)	117	SAFB		
38	FUS	78	MSY2 (YBX2)	118	Sam68 (KHDRBS1)		
39	FXR1	79	NOL8	119	SART3		
40	FXR2	80	Nopp34 (NIFK)	120	SC35 (SRSF2)		

Table S3 List of synthetic oligonucleotides used in the experiments

siRNAs (sense sequences) for cellular knockdown (dT: deoxyribonucleotide)		
SPF45-siRNA#1	5'-GAACAAGACAGACCGAGAUdTdT-3'	Knockdown of SPF45
SPF45-siRNA#2	5'-GACCCUAUGUUUCCUAAUGdTdT	Knockdown of SPF45
U2AF^{65}-siRNA	5'-GCACGGUGGACUGAUUCGUdTdT-3'	Knockdown of U2AF^{65}
PRP43-siRNA	5'-AAACAGAAAUGCAGGAUAAdTdT-3'	Knockdown of PRP43
SF4-siRNA	[SMARTpool ON-TARGETplus Human SUGP1 siRNA (horizon)]	Knockdown of SF4
Primer DNAs for plasmid constructions		
HNRNPH1-E7F-EcoRI	5'-GGAATTCCGCTATGGAGGCTATGATGA-3'	Construction of pcDNA3-HNRNPH1
HNRNPH1-E8R-XhoI	5'-CCCTCGAGGGCAGTAGCTCTGTAAGGTAAT-3'	
AdMLF-BamHI	5'-CGGGATCCCGACTCTCTTCCGCATCGCTG-3'	Construction of pcDNA3-AdML
AdMLR-EcoRI	5'-GGAATTCCCGTTCGGAGGCCGACGGGTT-3'	
EML3-E17F-EcoRI	5'-GGAATTCCGTGGTTGGTTTTGGACACAG-3'	Construction of pcDNA3-EML3
EML3-E17R-XhoI	5'-CCCTCGAGGGCATACAGCGGCCAAAGCGGC-3'	
MUS81-E13F-EcoRI	5'-GGAATTCCCCCTGGGAACCCTGAATCAG-3	Construction of pcDNA3-MUS81
MUS81-E14R-XhoI	5'-CCGCTCGAGGGTGCTGTATCGATCCACCA-3	
HNRNPH1-5'SSAdML-S	5'-GGAAGGGGTGAGTTAAGAATTGAATTTCTC-3'	Construction of pcDNA3-HNRNPH1/5'SSAdML
HNRNPH1-5'SSAdML-AS	5'-CTTAACTCACCCCTTCCAAATCTATCTGAC-3'	
HNRNPH1-BranchAdML-S	5'-GAAGGAGTCATACACTCTTGTCCATCTAGA-3'	Construction of pcDNA3-HNRNPH1/BranchAdML
HNRNPH1-BranchAdML-AS	5'-AAGAGTGTATGACTCCTTCAACTGAGAAAT-3'	
HNRNPH1-3'SSAdML-S	5'-TCCATACAGCTCTCAATTACTGTTTTTCAG-3'	Construction of pcDNA3-HNRNPH1/3'SSAdML
HNRNPH1-3'SSAdML-AS	5'-ATTGAGAGCTGTATGGACAAGAGTGTAAGC-3'	
HNRNPH1-AdMLPPT25-S	5'-TTATCCTGTCCCTTTTTTTTCCACAGACCTCAATTACTGTTTTTC-3'	Construction of pcDNA3-HNRNPH1/AdML-PPT25
HNRNPH1-AdMLPPT25-AS	5'-GAAAAAAAAGGGACAGGATAAGTGTAAGCATCCTTCAACTGAG-3'	
HNRNPH1-AdMLPPT25mt-S	5'-TTATCCTGTCGCTGTTGTGTCCACAGACCTCAATTACTGTTTTTC-3'	Construction of pcDNA3-HNRNPH1/AdML-PPT25mt
HNRNPH1-AdMLPPT25mt-AS	5'-GACACAACAGCGACAGGATAAGTGTAAGCATCCTTCAACTGAG-3'	
HNRNPH1-AdMLPPT13-S	5'-CTTTTTTTTCCACAGCTCTCAATTACTGTTTTTCAG-3'	Construction of pcDNA3-HNRNPH1/AdML-PPT13
HNRNPH1-AdMLPPT13-AS	5'-GAGCTGTGGAAAAAAAAGGGCTGGATCAGTGGGGGT-3'	
HNRNPH1-5'AdML-S1	5'-AGATAGATTTGGAAGAGGTAAGGACTCCCTCTCAAAAGCGG-3'	Construction of pcDNA3-HNRNPH1/5'AdML
HNRNPH1-5'AdML-AS1	5'-GATGGACAAGAGTGTAAGCAATCATCAAGGAAACCCTGG-3'	
HNRNPH1-5'AdML-S2	5'-CCAGGGTTTCCTTGATGATTGCTTACACTCTTGTCCATC-3'	
HNRNPH1-5'AdML-AS2	5'-CCGCTTTTGAGAGGGAGTCCTTACCTCTTCCAAATCTATCT-3'	
EML3-AdMLPPT25-S	5'-TTATCCTGTCCCTTTTTTTTCCACAGATGGGTTGTACCTGGCCAT-3'	Construction of pcDNA3-EML3/AdML-PPT25
EML3-AdMLPPT25-AS	5'-GAAAAAAAAGGGACAGGATAAGACTCAGAGAGGGAAGGGCCA-3'	
EML3-AdMLPPT13-S	5'-CTTTTTTTTCCACAGACCGGTTGTACCTGGCCATTG-3'	Construction of pcDNA3-EML3/AdML-PPT13
EML3-AdMLPPT13-AS	5'-GGTCTGTGGAAAAAAAAGCTCAGAGAGGGAAGGGCC-3'	
EML3-5'AdML-S1	5'-GTACAGCCCAGGTGGGAACTCCCTCTCAAAAGCGG-3'	Construction of pcDNA3-EML3/5'AdML
EML3-5'AdML-AS1	5'-GCAAATACAGTCACTCAGAGATCATCAAGGAAACCCTGG-3'	
EML3-5'AdML-S2	5'-CCAGGGTTTCCTTGATGATCTCTGAGTGACTGTATTTGC-3'	
EML3-5'AdML-AS2	5'-CCGCTTTTGAGAGGGAGTTCCCACCTGGGCTGTAC-3'	
SPF45-F-EcoRI	5'-GGAATTCTGATGTCCCTGTACGATGACCT-3'	Construction of pcDNA3-Flag/SPF45
SPF45-R-XhoI	5'-CCCTCGAGGGTCAAACTTGTTCTGCCAAATCC-3'	
SPF45-UHMmt-F	5'-GTGCGGGAGAGGTGAAGGAAGACTTGGAAG-3'	Construction of pcDNA3-Flag-SPF45/UHMmt
SPF45-UHMmt-R	5'-CACCTCTCCCGCACCAACCA-3'	
SPF45-ΔG-F	5'-CTTCCTCGCTGGCGACGCCACAGAGAAAGA-3'	Construction of pcDNA3-Flag-SPF45/ΔG
SPF45-ΔG-R	5'-TGGCGTCGCCAGCGAGGAAGGAGTTGCTAG-3'	
SPF45-siR-F	5'-CCAATGTTCCCAAACGATTATGAGAAAGTA-3'	Construction of pcDNA3-Flag-SPF45/siR, pcDNA3-Flag-SPF45/UHMmt/siR & pcDNA3-Flag-SPF45/ΔG/siR
SPF45-siR-R	5'-TTTGGGAACATTGGATCATATTCGTCAGCT-3'	

Primer DNAs for analysis		
HNRNPH1-E7F-EcoRI	See above	Splicing assay of endogenous HNRNPH1 intron 7
HNRNPH1-E8R-XhoI	See above	
RFC4-E9F-EcoRI	See above	Splicing assay of endogenous HNRNPH1 intron 7
RFC4-E10R-XhoI	See above	
EML3-E17F-EcoRI	See above	Splicing assay of endogenous EML3 intron 17
EML3-E17R-XhoI	See above	
DUSP1-F	5'-TGCAGTACCCCACTCTACGA-3'	Splicing assay of endogenous DUSP1 intron 2
DUSP1-R	5'-GAGACGTTGATCAAGGCAGTG-3'	
NFKBIA-F	5'- TCCTCAACTTCCAGAACAACC-3'	Splicing assay of endogenous NFKBIA intron 2
NFKBIA-R	5'-TCAGCAATTTCTGGCTGGT-3'	
MUS81-E13F-EcoRI	5'-GGAATTCCCCCTGGGAACCCTGAATCAG-3'	Splicing assay of endogenous MUS81 intron 13
MUS81-E14R-XhoI	5'-CCGCTCGAGGGTGCTGTATCGATCCACCA-3'	
RECQL4-E15F-EcoRI	5'-GGAATTCCAAGACCTGCGAGAGCTGCG-3'	Splicing assay of endogenous RECQL4 intron 15
RECQL4-E16R-XhoI	5'-CCGCTCGAGGCAGTTCAGACGGCAATGGG-3'	
MTA1-E5F-EcoRI	5'-GGAATTCCGAAATGGAGAACCCGGAAATG-3'	Splicing assay of endogenous MTA1 intron 5
MTA1-E6R-XhoI	5'-CCGCTCGAGGCTCCAGGTAGGACTTGAG-3'	
AdML-E1F	5'-CGTTCGTCCTCACTCTCTTCCGC-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	See above	
T7	See above	Splicing assay of EML3 mini-gene pre-mRNA
EML3-E17R-XhoI	See above	
T7	See above	Splicing assay of NUS81 mini-gene pre-mRNA
MUS81-E14R-XhoI	See above	
AdML-I1F	5'-TGATGGCTATGGATTTGGGTC-3'	qPCR detection of AdML pre-mRNA in CLIP assay
AdML-I1R	5'-GGACAAGAGTGTAAGCATCCT-3'	
HNRNPH1-E7F-EcoRI	See above	qPCR detection of HNRNPH1 pre-mRNA in CLIP assay
HNRNPH1-I7R	5'-GGTCTAGATGGACAAGAGTGT-3'	
EML3-E17F-EcoRI	See above	qPCR detection of EML3 pre-mRNA in CLIP assay
EML3-I17-R	5'-GCATGTGAGTCCAGGGTT-3'	
MUS81-E13F-EcoRI	See above	qPCR for detection of EML3 pre-mRNA in CLIP assay
MUS81-I13R	5'-CAGGCCATGTCTGAGAAGCT-3'	
SP6	5'-ATTTAGGTGACACTAT-3'	Reverse transcription in CLIP assay

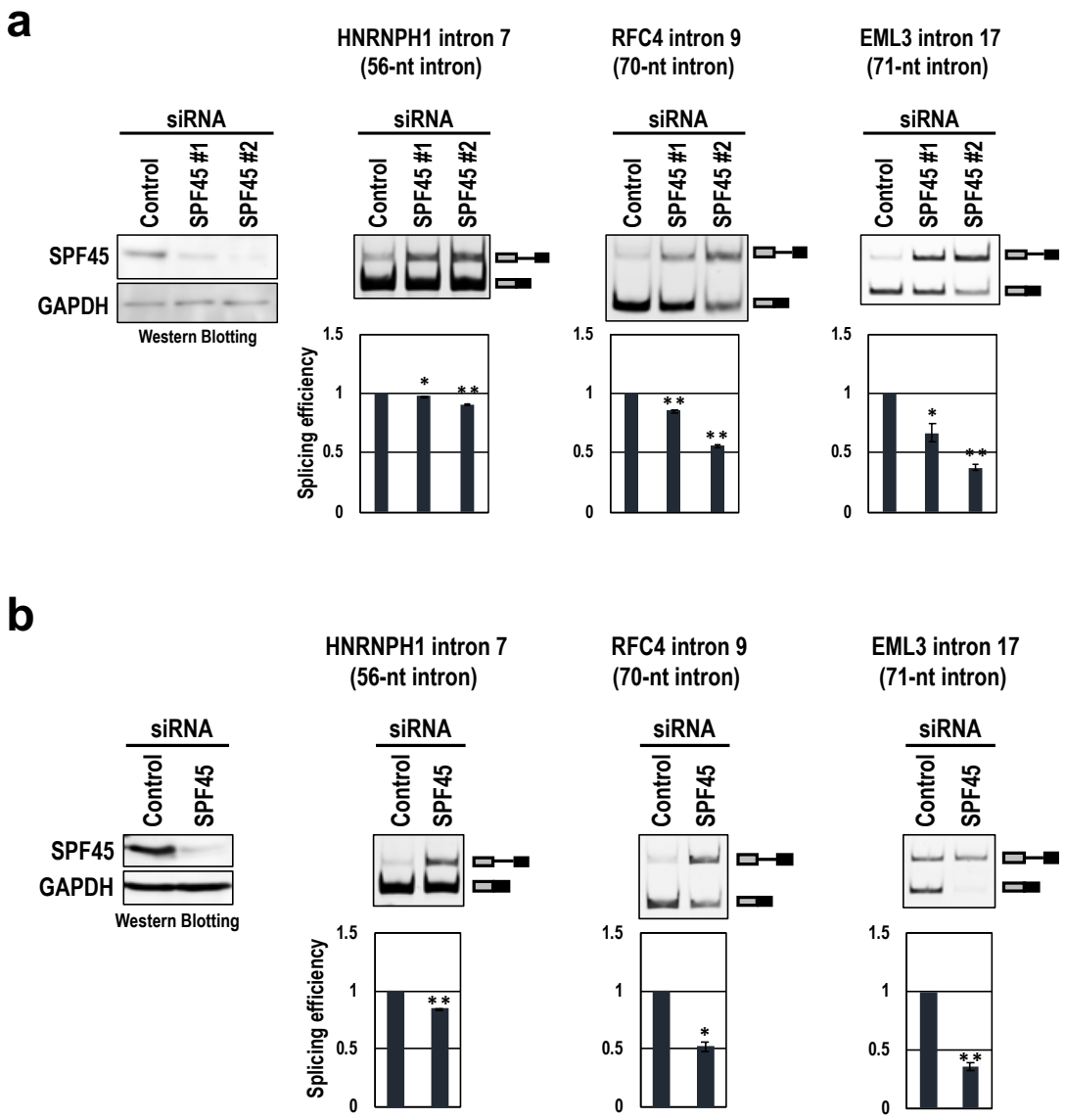

Fig. S1 | Validation of two different siRNAs targeting SPF45 and splicing assay in SPF45-depleted HEK293 cells.

a, HeLa cells were transfected with control siRNA (Ctl) and two different siRNAs (#1, #2) targeting SPF45. At 72 h post-transfection, the SPF45 protein depletion was checked by a Western blotting with indicated antibodies (left panel). *In cellulo* splicing assays of the indicated three short introns using RT–PCR (see also Figs. 1, 2c). Means ± SE are given for three independent experiments (*P < 0.05, **P < 0.005).

b, The same *in cellulo* splicing assays were performed using SPF45-knockdown HEK293 cells. Means ± SE are given for three independent experiments (*P < 0.05, **P < 0.005).

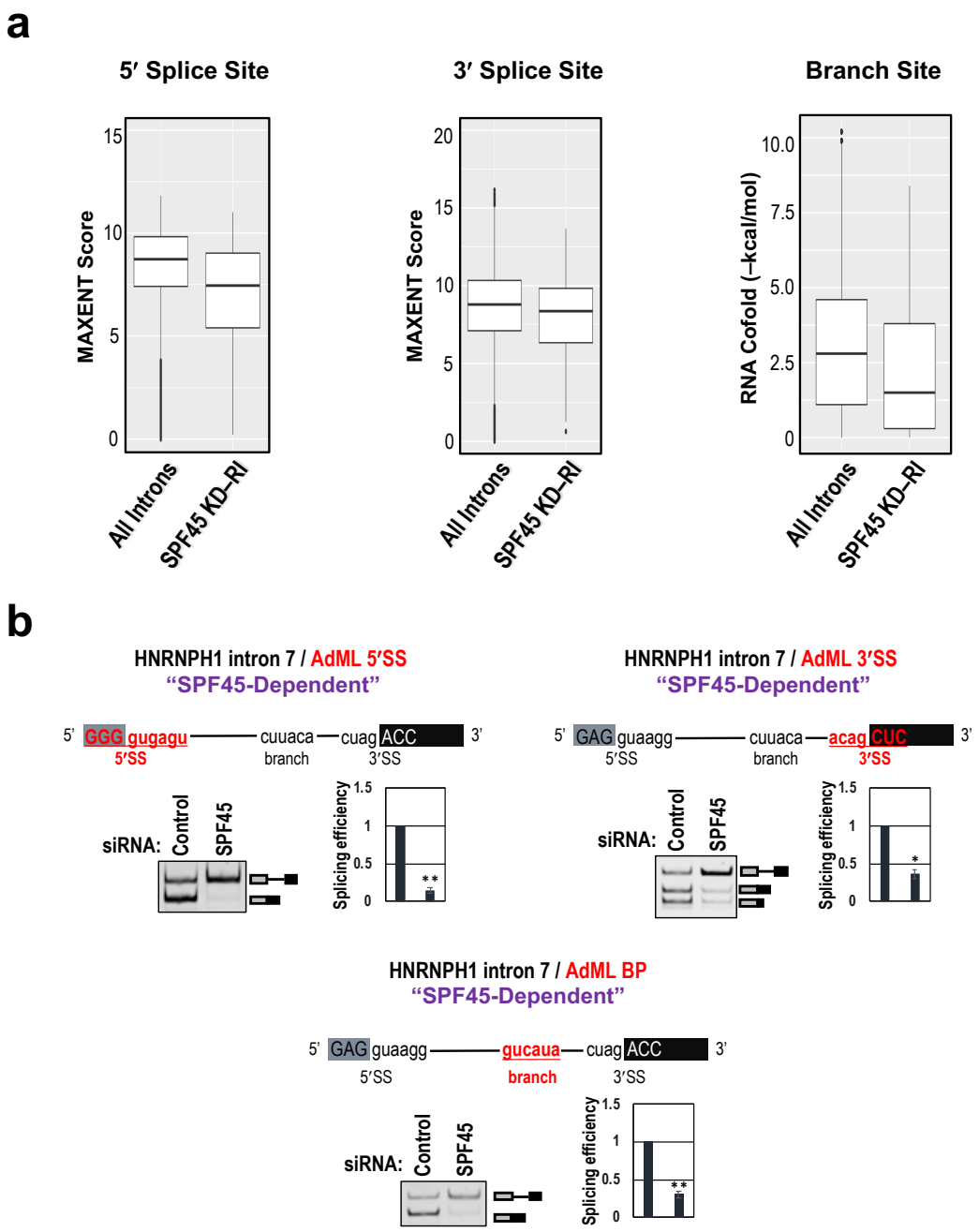

Fig. S2 | Characterization of flanking splicing signal sequences of retained-introns in SPF45-knockdown HEK293 cells.

a, The box plots compare the strengths of the 5′/3′ splice sites and the branch site of all introns (in human RefGene) with those of the retained introns in SPF45-knockdown HEK293 cells.

b, The splice sites and branch site sequences have no effect on SPF45-dependent splicing. Chimeric HNRNH1-intron 7 pre-mRNAs are schematically shown (red color indicates AdML derived sequences). These pre-mRNAs were expressed from mini-genes in HeLa cells and their splicing was assayed by RT–PCR. PAGE images and quantifications of RT–PCR are shown. Means ± SE are given for three independent experiments (*P < 0.05, **P < 0.005).

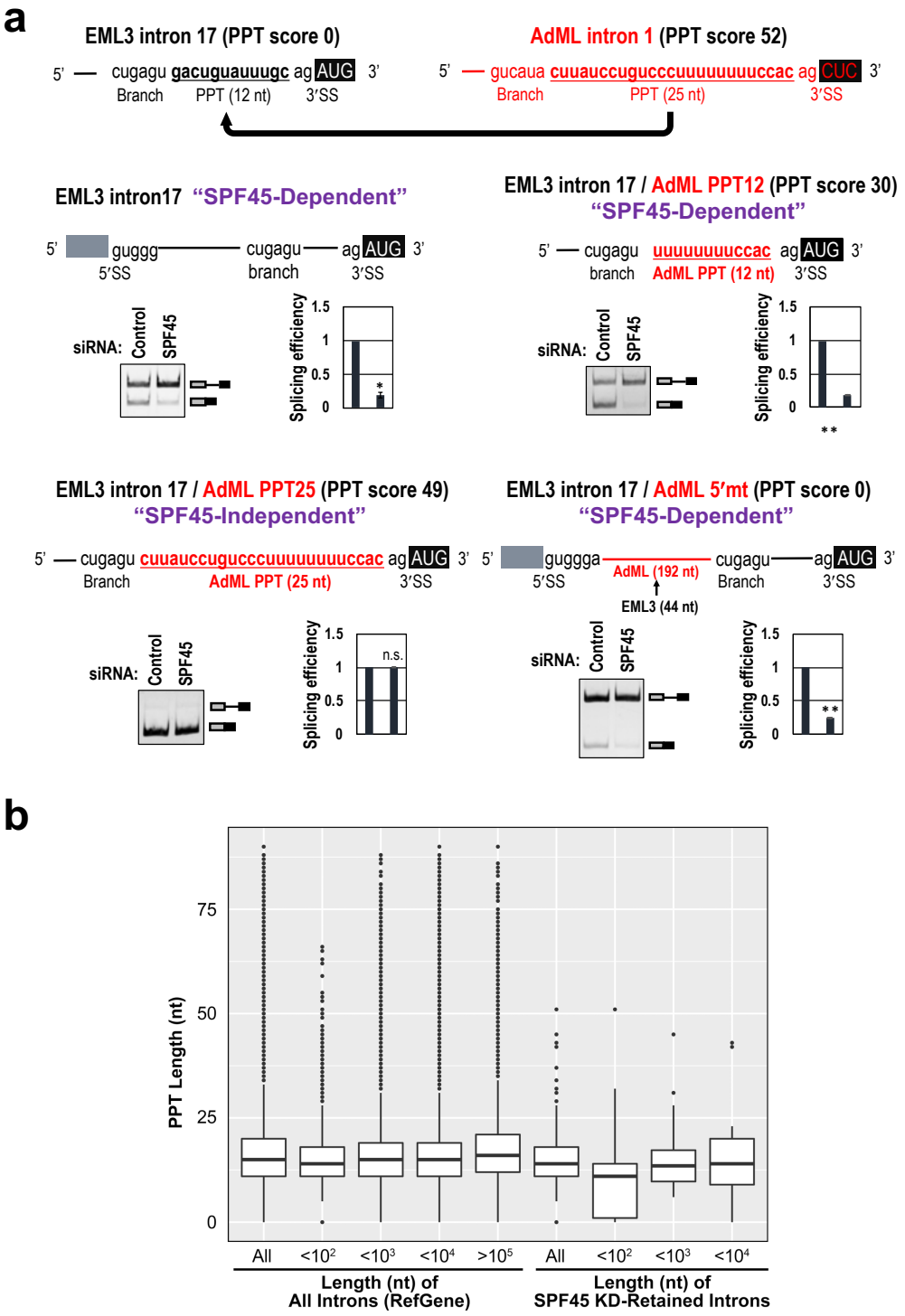

Fig. S3 | Truncated polypyrimidine tract (PPT) is determinant for SPF45-dependent splicing on short introns.

a, *In cellulo* splicing assays of chimeric EML3 intron 17 pre-mRNAs carrying PPT and its variants derived from AdML pre-mRNA (red color). See Fig. 3 for the schematic representation of pre-mRNAs, RT–PCR panels, and graphs. Means ± SE are given for three independent experiments (*P < 0.05, **P < 0.005, n.s. P > 0.05).
b, The box plots compare the PPT-length distributions of all introns (in human RefGene) with those of the retained introns in SPF45-knockdown HEK293 cells. The determinant and mechanism for SPF45-dependent splicing on the longer introns (>10^2 nt) are under investigation.

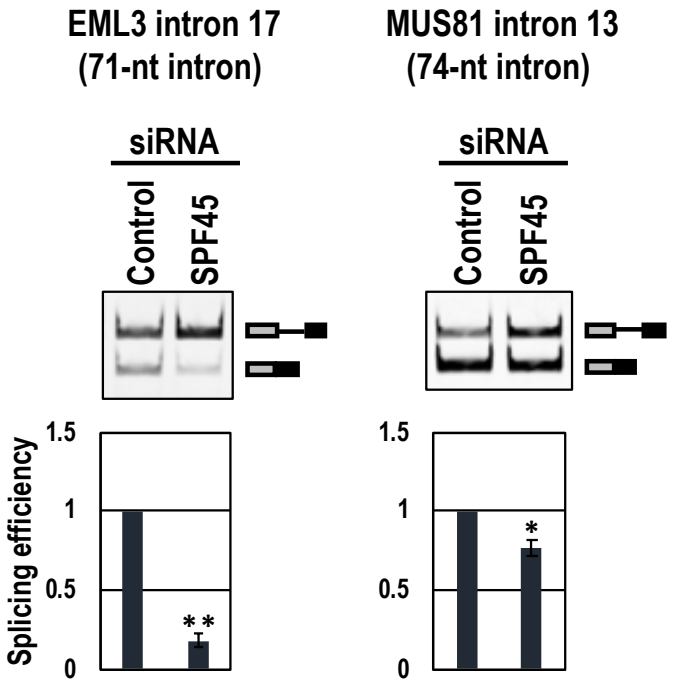

Fig. S4 | The SPF45-dependency of ectopically expressed two pre-mRNAs from transfected mini-genes including short introns.
In cellulo splicing assays of EML3-intron 17 and MUS81-intron 13 pre-mRNAs were performed using control siRNA- and SPF45 siRNA-treated HeLa cells (see Fig. 1 for the depletion efficiency of SPF45). See Fig. 3 for splicing assays with HNRNPH1-intron 7 and AdML-intron 1 pre-mRNAs). Means ± SE are given for three independent experiments (*P < 0.05, **P < 0.005).

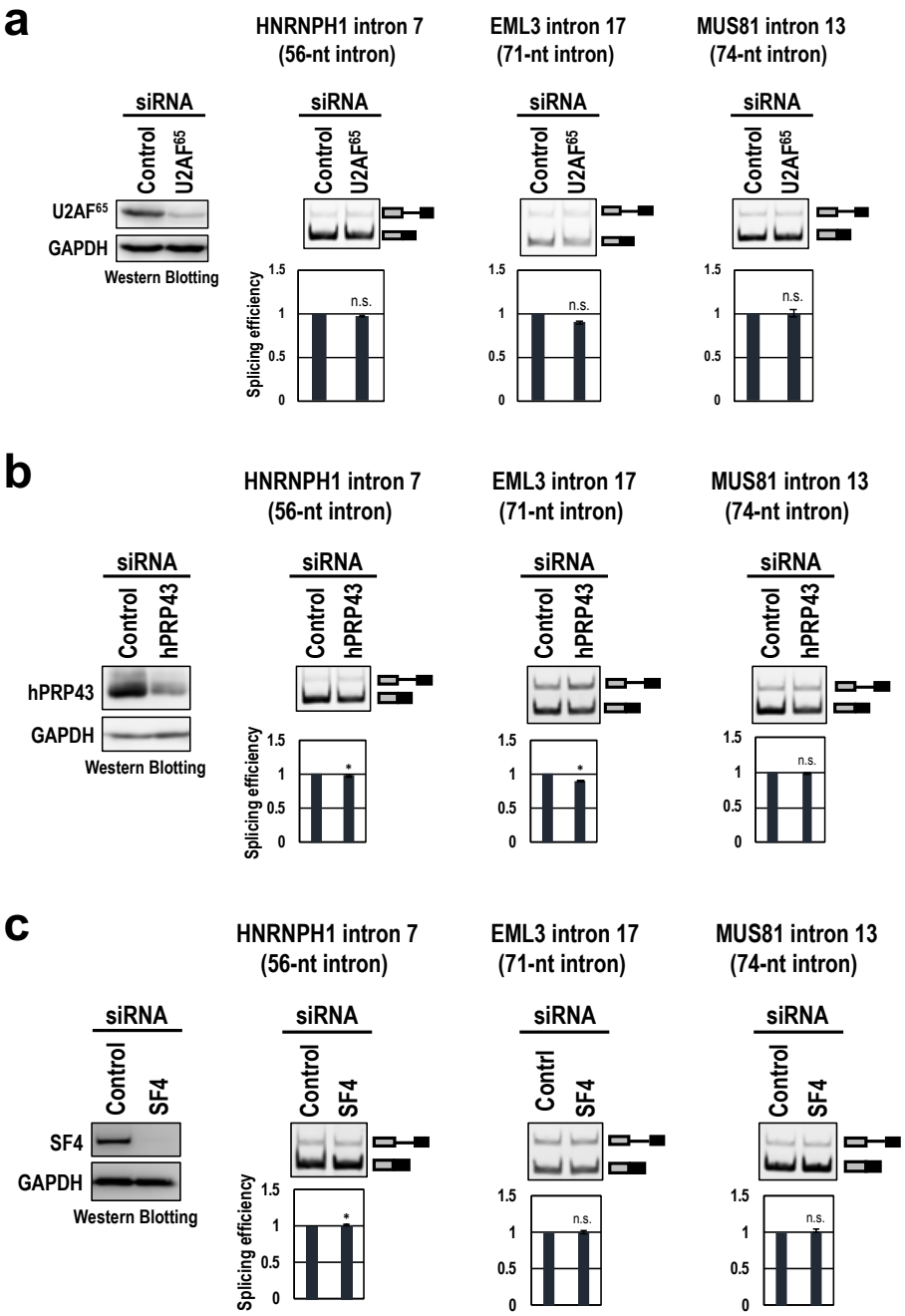

Fig. S5 | Unaffected splicing efficiencies of pre-mRNAs including SPF45-dependent short introns in U2AF^{65}-, hPRP43- and SF4-knockdown HeLa cells.

a, HeLa cells were transfected with control siRNA (Ctl) and siRNA targeting U2AF^{65}. At 72 h post-transfection, the U2AF^{65} protein depletion was checked by a Western blotting with indicated antibodies (left panel). *In cellulo* splicing assays of the indicated three short introns using RT–PCR (see also Fig. S1). Means ± SE are given for three independent experiments (*P < 0.05, n.s. P > 0.05).

b, HeLa cells were transfected with control siRNA (Ctl) and siRNA targeting hPRP43. See above for Western blotting and *in cellulo* splicing assays.

c, HeLa cells were transfected with control siRNA (Ctl) and siRNA targeting SF4. See above for Western blotting and *in cellulo* splicing assays.

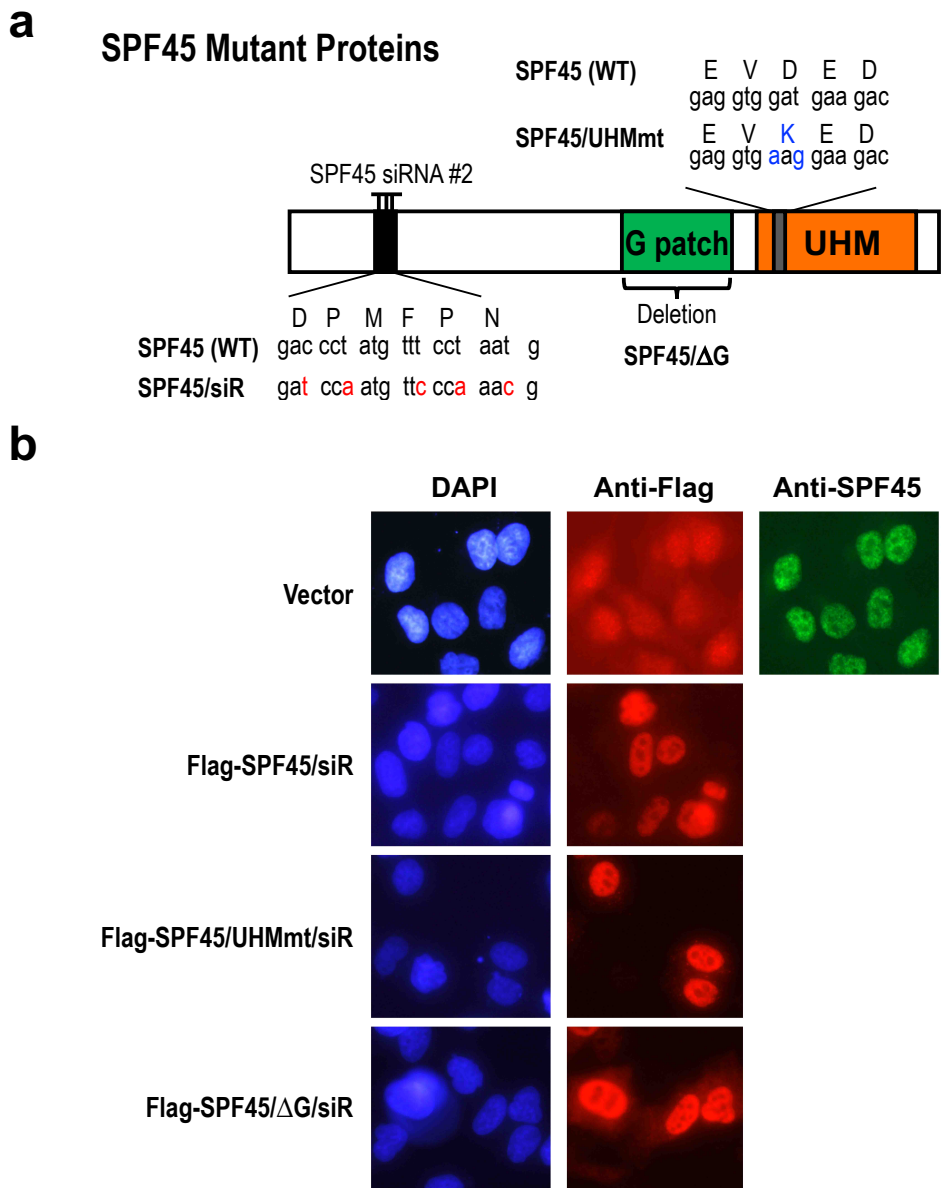

Fig. S6 | Subcellular localization of SPF45/siR, SPF45/UHMmt/siR, and SPF45/ΔG/siR proteins.
a, Schematic structures of SPF45 and the induced mutations in the indicated expression plasmids. The D→K mutation in SPF45/UHMmt, whose protein cannot binds ULM, is highlighted in blue. The target of SPF45-siRNA#2 is depicted and introduced five silent siRNA-resistant mutations were highlighted in red.
b, HeLa cells were transfected with Flag-SPF45/siR, Flag-SPF45/UHMmt/siR, and Flag-SPF45/ΔG/siR expression plasmids and their subcellular localizations were examined by immunofluorescence microscopy using indicated antibodies. Cells were also stained with DAPI to detect nuclei.