Human pre-mRNA introns vary in size from under fifty to over a million nucleotides. We searched for essential factors involved in the splicing of human short introns by screening siRNAs against 154 human nuclear proteins. The splicing activity was assayed with a model HNRNPH1 pre-mRNA containing short 56-nucleotide intron. We identify a known alternative splicing regulator SPF45 (RBM17) as a constitutive splicing factor that is required to splice out this 56-nt intron. Whole-transcriptome sequencing of SPF45-deficient cells reveals that SPF45 is essential in the efficient splicing of many short introns. To initiate the spliceosome assembly on a short intron with the truncated poly-pyrimidine tract, the U2AF-homology motif (UHM) of SPF45 competes out that of U2AF65 (U2AF2) for binding to the UHM-ligand motif (ULM) of the U2 snRNP protein SF3b155 (SF3B1). We propose that splicing in a distinct subset of human short introns depends on SPF45 but not U2AF heterodimer.
here 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. Pre-mRNA splicing is dependent upon a set of signal RNA elements recognized by essential factors that is a ubiquitous and essential part of eukaryotic gene expression. However, our understanding about specific and distinct mechanisms for the precise recognition of degenerated splice site sequences within such extensively varied length of introns is fairly limited.

The canonical splicing mechanisms were 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. According to such optimal systems, the essential splicing sequences in pre-mRNA, namely the 5′ splice site, the branch-site sequence, and the poly-pyrimidine tract (PPT) followed by the 3′ splice site, are initially recognized by the U1 snRNP, SF1, and the U2AF heterodimer (U2AF65/U2AF35, U2AF2/U2AF1) as HGNC approved symbol), respectively. Following the assembly of this early spliceosomal E complex, SF1 is replaced by the U2 snRNP in the A complex, which commits the intron for splicing reaction (reviewed in ref. 6). The A complex is an asymmetric globular particle (~26 × 20 × 19.5 nm) that fully occupies 79–125 nucleotides (nt) of RNA, and recent high-resolution cryo-electron microscopy structures of the A complex have revealed molecular details of the overall architecture (reviewed in ref. 9). Interestingly, human ultrashort introns with much shorter lengths (43–65 nt) are nevertheless spliced. This raises the question of how such ultrashort introns can be recognized and committed to splicing by an ‘oversized’ A complex without steric hindrance.

We postulate that splicing of short introns depend on distinct specific factors, which utilize alternative ways for early spliceosomal assembly. Here, we have shown that this is the case in a subset of human short introns with the truncated PPT, which is recognized by a distinct constitutive splicing factor SPF45, but not by the authentic U2AF heterodimer.

Results

**SPF45 is a specific 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 (including many known RNA-binding proteins and splicing factors) for splicing activity of the HNRNPH1 pre-mRNA including 56-nt intron 7 (Fig. 1a; Supplementary 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 containing a 56-nt intron (Fig. 1a). The strongest splicing repression was markedly caused by knockdown of SPF45 (RBMI7 as HGNC approved symbol; Fig. 1b) that indeed effectively depleted SPF45 protein (Fig. 1b, left panel; Supplementary Table S1). We also confirmed that knockdown of the best seven factors including SPF45, showing significant splicing repression in this short 56-nt intron (Supplementary Table S1, PSI > 0.3), but no repression at all on pre-mRNA splicing of control 366-nt intron (Supplementary Fig. S1a). 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. The splicing efficiencies of these pre-mRNAs were also significantly repressed in SPF45-depleted HeLa cells (Fig. 1b).

Splicing inhibition was proportional to SPF45-knockdown efficiency induced by independent siRNAs (Supplementary Fig. S1b). These SPF45 siRNA-induced splicing defects were also observed in HEK293 cells, testifying to the robustness of our results (Supplementary Fig. S1c).

To test our hypothesis that SPF45 is indispensable 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 Supplementary Data 1 for the list of all 187 introns).

The analysis of these retained introns hinted at a potential mechanism for the role of SPF45. Remarkably, the length distribution of the retained introns in SPF45-depleted HEK293 cells is strongly biased towards shorter lengths compared to those in cells depleted of constitutive splicing factors, U2AF65 and SF3b155 (SF3B1 as HGNC approved symbols), which show a distribution comparable to the whole set of introns (Fig. 2b).

We validated these RNA-Seq-based profiles by RT-PCR. As assumed, splicing of pre-mRNAs with two control introns (366 and 329 nt) were not affected by SPF45 knockdown, while in contrast, three arbitrarily chosen pre-mRNAs with short introns (74, 75, and 76 nt) were repressed (Fig. 2c). These results demonstrate that SPF45 is required for the efficient splicing of a substantial population of pre-mRNAs with short introns.

**SPF45 is required for splicing on intron with truncated polypyrimidine tract (PPT).** Next we searched for a potential 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 are somewhat weaker than the average in RefGene (Supplementary Fig. S2a). Therefore, we first examined these cis-acting splicing signals using mini-gene splicing assays in SPF45-depleted cells. As expected, splicing of HNRNPH1 pre-mRNA (56-nt intron 7; use in our siRNA screening) was repressed by depletion of SPF45, whereas splicing of the control adenovirus 2 major late (AdML) pre-mRNA (231-nt intron 1; used as a standard splicing substrate previously) was unaffected (Fig. 3; top-row). The SPF45-dependent splicing of the HNRNPH1 pre-mRNA was not altered even after replacement of the 5′/3′ splice sites and the branch site individually, or all together, by those of the AdML pre-mRNA (Supplementary Fig. S2b). These results indicate that the requirement of SPF45 depend on neither the 5′/3′ splice sites nor the branch site.

We then examined whether the SPF45 dependency is attributed to the PPT. The PPT score (see Methods) is one of the criteria to evaluate effective PPTs: PPT scores are 19 for the PPT (13 nt) in HNRNPH1-intron 7 and 52 for the PPT (25 nt) in AdML-intron 1 (Fig. 3, second-row). Remarkably, the SPF45-dependent splicing of the HNRNPH1 pre-mRNA was not altered even after replacement of the 5′/3′ splice sites and the branch site individually, or all together, by those of the AdML pre-mRNA (Supplementary Fig. S2b). These results indicate that the requirement of SPF45 depend on neither the 5′/3′ splice sites nor the branch site.

We then examined whether the SPF45 dependency is attributed to the PPT. The PPT score (see Methods) is one of the criteria to evaluate effective PPTs: PPT scores are 19 for the PPT (13 nt) in HNRNPH1-intron 7 and 52 for the PPT (25 nt) in AdML-intron 1 (Fig. 3, second-row). Remarkably, the SPF45-dependent splicing of the HNRNPH1 pre-mRNA was not altered even after replacement of the 5′/3′ splice sites and the branch site individually, or all together, by those of the AdML pre-mRNA (Supplementary Fig. S2b). These results indicate that the requirement of SPF45 depend on neither the 5′/3′ splice sites nor the branch site.

We then examined whether the SPF45 dependency is attributed to the PPT. The PPT score (see Methods) is one of the criteria to evaluate effective PPTs: PPT scores are 19 for the PPT (13 nt) in HNRNPH1-intron 7 and 52 for the PPT (25 nt) in AdML-intron 1 (Fig. 3, second-row). Remarkably, the SPF45-dependent splicing of the HNRNPH1 pre-mRNA was not altered even after replacement of the 5′/3′ splice sites and the branch site individually, or all together, by those of the AdML pre-mRNA (Supplementary Fig. S2b). These results indicate that the requirement of SPF45 depend on neither the 5′/3′ splice sites nor the branch site.

We then examined whether the SPF45 dependency is attributed to the PPT. The PPT score (see Methods) is one of the criteria to evaluate effective PPTs: PPT scores are 19 for the PPT (13 nt) in HNRNPH1-intron 7 and 52 for the PPT (25 nt) in AdML-intron 1 (Fig. 3, second-row). Remarkably, the SPF45-dependent splicing of the HNRNPH1 pre-mRNA was not altered even after replacement of the 5′/3′ splice sites and the branch site individually, or all together, by those of the AdML pre-mRNA (Supplementary Fig. S2b). These results indicate that the requirement of SPF45 depend on neither the 5′/3′ splice sites nor the branch site.
The distance between the branch site and the 3’ splice site in ‘AdML PPT25’ since the same extended distance in SPF45-dependent pre-mRNAs did not lose the SPF45 dependency (Supplementary Fig. S3a, +12nt and AdML PPT13/+12nt). To determine whether SPF45 recognizes the strength or the length of a given 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 gain 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). Taken together, these results demonstrate that short PPT per se in the HNRNPH1-intron 7 is the determinant for the SPF45 dependency in splicing.

These observations were further recapitulated and validated in the distinct SPF45-dependent EML3 pre-mRNA, which contains a 71-nt intron (Supplementary Fig. S3b). 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. S4). Therefore, we can define introns shorter than ~100 nt as ‘short intron’ in this study. Together, it is the truncated PPTs in short introns that are crucial for the SPF45 function in splicing.

SPF45 replaces U2AF65 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...
A stable U2 snRNP-associated splicing complex is then formed by ATP-dependent displacement of SF1, where the p14 protein (a U2 snRNP-associated factor) contacts the branch site, and U2AF65 interacts with the U2 snRNP component SF3b155 (reviewed in ref. 6). Since the splicing activity of SPF45 depends on short PPTs, we hypothesized that insufficient U2AF65 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. S5) 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 SPF4512. 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 associated to all pre-mRNAs derived from these four mini genes irrespective of the length of the introns. The association of SPF45 to control AdML intron (231 nt) is consistent with previous proteomic detection of SPF45 in the human spliceosome13. SPF45 was also detected on MINX (131 nt) and PM5 (235 nt) introns14,15 (reviewed in ref.6). We next examined the binding of U2AF65 and SF3b155 to pre-mRNAs by UV crosslinking–immunoprecipitation (CLIP).

Whole-cell extracts from crosslinked cells were immunoprecipitated with anti-U2AF65 and anti-SF3b155 antibodies, and the precipitated RNAs were analyzed by RT-qPCR (Fig. 4b). SF3b155, as a component of the U2 snRNP, bound to all the pre-mRNAs, while significant U2AF65 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 U2AF65 and SF3b155 to the HNRNPH1 intron and control AdML intron by affinity pull-down assay. The biotinylated HNRNPH1 and AdML pre-mRNAs were transcribed in vitro, incubated with HEK293 cell nuclear extracts, and interacted proteins were examined by Western blotting (Fig. 5). In agreement with our formaldehyde crosslinking experiments, SPF45 associated with both AdML and HNRNPH1 pre-mRNAs (see Discussion). However, importantly, U2AF65 strongly bound only to AdML pre-mRNA but U2AF65 was allowed to associate HNRNPH1 pre-mRNA only if SPF45 was depleted from nuclear extracts. These results together support our proposed hypothesis that SPF45 replaces U2AF65

Fig. 4 SPF45 binds to all introns while U2AF65 cannot bind to 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 pre-mRNAs were quantified by RT-qPCR using specific primers (see the schematic mini gene below). Means ± SEM are given for three independent experiments and two-tailed Student t-test values were calculated (**P < 0.01, ***P < 0.0005, n.s. P > 0.05). b Cellular CLIP experiments shows strong U2AF65 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-U2AF65 and anti-SF3b155 antibodies and then immunoprecipitated RNAs were quantified by RT–qPCR using specific primers (see the schematic mini gene below). Means ± SEM are given for three independent experiments and two-tailed Student t-test values were calculated (**P < 0.01, ***P < 0.0005, n.s. P > 0.05). Source data of the above graphs are provided as a Source Data file.

The site are bound cooperatively by SF1, U2AF65, and U2AF35, respectively. A stable U2 snRNP-associated splicing complex is then formed by ATP-dependent displacement of SF1, where the p14 protein (a U2 snRNP-associated factor) contacts the branch site, and U2AF65 interacts with the U2 snRNP component SF3b155 (reviewed in ref. 6). Since the splicing activity of SPF45 depends on short PPTs, we hypothesized that insufficient U2AF65 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. S5) 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 SPF4512. 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 associated to all pre-mRNAs derived from these four mini genes irrespective of the length of the introns. The association of SPF45 to control AdML intron (231 nt) is
U2AF65 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 U2AF65. The graph shows quantification of the bands on Western blots (panel a as the representative blots). SPF45 and U2AF65 scanned data were normalized to SF3B155 (SPF45/SF3B155 and U2AF65/SF3B155) and plotted. Means ± SEM are given for four independent experiments and two-tailed Student t-test values were calculated (Left panel: p = 0.2982 for AdML vs HNRNPH1 intron in Control NE; p = 0.6973 for AdML vs HNRNPH1 intron in SPF45-KD NE; Right panel: p = 0.0142 for AdML vs HNRNPH1 intron in Control NE, p = 0.4647 for AdML vs HNRNPH1 intron in SPF45-KD NE). *P < 0.05, n.s. P > 0.05. Source data of all the above panels are provided as a Source Data file.

**Fig. 6 Depletion of U2AF65 rather increases splicing of pre-mRNAs with SPF45-dependent short introns.** a The co-depletion of U2AF65 and U2AF35 proteins from HeLa cells was observed by a Western blotting. Displayed blots are representative of three independent experiments. b After the siRNA transfection, HeLa cells were cultured for short time (4 h) to effectively observe splicing stimulation, and splicing efficiencies of the indicated four mini genes were analyzed by RT-PCR. Means ± SEM are given for three independent experiments and two-tailed Student t-test values were calculated (AdML intron: p = 0.0067 for Control vs U2AF65 siRNA; HNRNPH1 intron: p = 0.0038 for Control vs U2AF65 siRNA; EML3 intron: p = 0.0116 for Control vs U2AF65 siRNA; MUS81 intron: p = 0.0273 for Control vs U2AF65 siRNA. *P < 0.05, **P < 0.01. Source data of all the above panels are provided as a Source Data file.

in the assembly of U2 snRNP complexes as U2AF65 is poorly bound to truncated PPTs of short introns.

We noticed that endogenous U2AF65 knockdown barely repressed splicing of the SPF45-dependent short intron (Supplementary Table S1, No. 142; Supplementary Fig. S6a). Therefore, we checked splicing efficiencies of these four pre-mRNAs in U2AF65-knockdown HeLa cells (Fig. 6). This depletion of U2AF65 also caused effective co-depletion of U2AF35 (Fig. 6a) that is consistent with previous reports. In the control AdML pre-mRNA, spliced mRNA was reduced by the depletion of U2AF65, 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 U2AF65 (Fig. 6b).
In endogenous SPF45-dependent pre-mRNAs (Supplementary Fig. S6a), such marked activation was not observed that could be due to the almost saturated efficiency of splicing (see amounts of unspliced pre-mRNAs in Supplementary Fig. S6a). 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<sup>65</sup>/U2AF<sup>35</sup> is displaced by SF3b155–SPF45 via ULM–UHM binding.** The SPF45 protein contains a G-patch motif that may interact with nucleic acids and proteins<sup>18,19</sup>, and a C-terminal U2AF-homology motif (UHM) that binds the UHM-Ligand motifs (ULM) of its partner proteins. UHM–ULM interactions; e.g., U2AF<sup>65</sup>(ULM)–SF1(ULM), U2AF<sup>65</sup>(ULM)–SF3b155 (ULM), and U2AF<sup>35</sup>(ULM)–U2AF<sup>65</sup>(ULM), plays an essential role in the splicing reactions<sup>20–22</sup> (reviewed in ref. 23). Remarkably, in vitro binding analyses using the purified recombinant proteins showed that the UHM of SPF45 can bind to the ULMs of SF3b155, U2AF<sup>65</sup> and SF1; on the other hand, the UHM and G-patch motif of SPF45 cannot bind directly to RNA<sup>22</sup>. We therefore postulated that the SF3b155–U2AF<sup>65</sup>/U2AF<sup>35</sup> complex is remodeled to the SF3b155–SPF45 complex by switching of their ULM–UHM interactions and that SPF45 per se does not necessarily bind to the truncated PPT.

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 and its variants (Fig. 7a). The D319K mutant in the UHM (SPF45/ UHMmt) no longer binds any ULM, and G-patch motif-deleted mutant (SPF45/ΔG) loses potential interaction with nucleic acids and proteins<sup>22</sup> (Supplementary Fig. S7a). Although both SF3b155 and SF1 contain ULM and they can interact with SPF45 in vitro<sup>22</sup>, our GST pull-down assays with crude nuclear extracts, closer to physiological conditions, demonstrated that GST-SPF45 bound to SF3b155, but not to SF1 (Fig. 7b). As expected, the UHM of SPF45 is essential (Fig. 7b, GST-SPF45/UHMmt) while the G-patch of SPF45 is dispensable (GST-SPF45/ΔG), confirming the ULM–UHM interaction in the recruitment of SPF45 to the target short introns (Supplementary Fig. S8). The SPF45-UHM mutant (Flag-SPF45/UHMmt/siR) showed the remarkable impairment of the association to short introns, indicating that the recruitment of SPF45 to the truncated PPT is dependent on SF3b155 binding through ULM–UHM interaction. Using NMR analysis, we tested whether SPF45 can bind to truncated PPT RNA of the SPF45-dependent short introns in vitro. Our NMR titration experiments indicate that neither the G-patch motif nor the UHM domain of SPF45 shows significant affinity towards these two truncated PPT RNAs (Supplementary Fig. S9). Taken together, the recruitment of SPF45 to the target short introns may not depend on its direct binding to truncated PPT, but rather it requires the protein interaction with SF3b155 (Fig. 9).

The binding between U2AF<sup>65</sup>(ULM) and U2AF<sup>35</sup>(ULM) is very strong<sup>20</sup>. Remarkably, GST-SPF45 did not pull-down U2AF<sup>65</sup> and U2AF<sup>35</sup> in crude nuclear extracts (Fig. 7b). These data together suggest that the U2AF<sup>65</sup>(ULM) does not interact with the SPF45/ΔG in nuclear extracts, so that the SPF45 (UHM) and U2AF<sup>65</sup>(ULM) work together for a functional binding toward the SF3b155 (ULM) (Fig. 9). Therefore, we next investigated the competitive binding of U2AF<sup>65</sup> and SPF45 toward SF3b155 by titrating the dose of GST-SPF45 in the immunoprecipitation assays (Fig. 7c, d). Notably, GST-SPF45 interfered with the binding between SF3b155 and U2AF<sup>65</sup> in a dose-dependent manner; however, GST-SPF45/UHMmt did not
disturb this binding. These results indicate that the SPF45(UHM) competes with the U2AF65(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. S7a). We confirmed that the subcellular localization of these three mutant proteins did not change from that of endogenous SPF45 protein (Supplementary Fig. S7b). Protein expression levels of endogenous SPF45 and the G patch of SPF45 are essential for the SPF45-dependent splicing event on the shorter introns, is recognized by SPF45 but not by the authentic U2AF heterodimer; implicating that SPF45 is a distinct constitutive splicing factor in the spliceosomal complex A. This finding rationally answers the question of why SPF45, which was previously considered just to be an alternative splicing factor, is essential for cell survival and maintenance in vivo. Since the conditional knockout of SPF45 in mice causes extensive dysregulation of splicing, it is reasonable to assume that SPF45-dependent splicing of pre-mRNAs including short introns with truncated PPT could be a part of the targets of such dysregulation.

A mechanistic model of SPF45-dependent splicing. We found that both SPF45 and U2AF65 can bind on introns with SF3b155 irrespective of intron size presumably via interactions with the five ULMs in SF3b155, as previously shown in the simultaneous binding of U2AF65 and PUF60 to SF3b155. This observation is justified by the previous mass-spectrometry analysis using AdML, MINX, and PM5 pre-mRNAs with conventional introns; i.e., SPF45 is contained in E, A and B complexes as a U2 snRNP component SF3b155 to promote splicing of pre-mRNAs with short introns.

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 30). In the former model, the frequent lack of a canonical PPT in vertebrate short introns was noticed and an alternative mechanism that circumvents this problem were postulated. Here we provide one of answers to this puzzling question by demonstrating that a subset of human short introns, with significantly undersized pyrimidine tracts, is recognized by SPF45 but not by the authentic U2AF heterodimer; implicating that SPF45 is a distinct constitutive splicing factor in the spliceosomal complex A. This finding rationally answers the question of why SPF45, which was previously considered just to be an alternative splicing factor, is essential for cell survival and maintenance in vivo. Since the conditional knockout of SPF45 in mice causes extensive dysregulation of splicing, it is reasonable to assume that SPF45-dependent splicing of pre-mRNAs including short introns with truncated PPT could be a part of the targets of such dysregulation.
SPF45 was indeed identified as a member of the alternative splicing machinery, and cellular knockdown of SPF45 detected changes in alternative splicing and constitutive splicing. SPF45-mediated mechanisms in constitutive splicing and alternative splicing indicate that SPF45 is evolutionarily conserved from fruit fly to human.

In the SPF45-induced regulation of alternative splicing, there was no competition between SPF45 and U2AF heterodimer on the Sxl pre-mRNA. 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 heterodimer 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.

Another distinct subsets of human short introns. 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. Notably, Smu1/RED are human-specific splicing factors, whereas SPF45 is evolutionarily conserved from fruit fly. The distance between the 5' splice site and branch site needs 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. Since both U2AF and SPF45 can bind to long introns simultaneously via five ULMs in SF3b155, the alternative model that U2AF displaces SPF45 in long introns would be impossible.
distance, so that spliceosomes can overcome structural constraint associated with short introns. To explore the structural constraint on SPF45-dependent short intron, we definitely need to analyze the spliceosomal complex A formation by mass-spectrometry and cryo-electron microscopy. In short, Smu1/Red-dependent the spliceosomal complex A formation by mass-spectrometry subcloning the corresponding PCR-amplified product set into pcDNA3-Flag-SPF45 (Invitrogen –Thermo Fisher Scientific). The PCR products were performed using genomic DNA of HeLa cells and specific primer sets (Supplementary Table S2). For the pcDNA3-AdML, the PCR product was performed using the pBS-Ad2 plasmid and specific primer sets (Supplementary Table S2).

To construct expression plasmids, pcDNA3-HNRNP1, pcDNA3-EML3, and pcDNA3-MUS81 were constructed by subcloning PCR-amplified fragment into pcDNA3-Flag (Invitrogen –Thermo Fisher Scientific). The PCR products were analyzed using PCR-RFLP analysis to verify the amplified product set.

**Methods**

**Construction of expression plasmids.** The mini-gene expression plasmids, pcDNA3-HNRNP1, pcDNA3-AdML, and pcDNA3-AdML were constructed by subcloning PCR-amplified fragment into pcDNA3-Flag (Invitrogen –Thermo Fisher Scientific). The PCR products were analyzed using PCR-RFLP analysis to verify the amplified product set.

**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 Healthcare Life Sciences). The following commercially available antibodies were used to detect proteins: anti-tubulin (1:4000 dilution; HPA037487, Sigma–Aldrich), anti-SPF45 (1:3000 dilution; D221-3, MBL Life Science), anti-U2AF65 (1:5000 dilution; U4758, Sigma–Aldrich), anti-EML3 (1:5000 dilution; U4758, Sigma–Aldrich), and anti-Flag (1:3000 dilution; anti-DYKDDDK tag, M853L, MBL Life Science). The anti-hPRT3 antibody (1:1500 dilution) was described previously41. Immuno-reactive protein bands were detected using 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 HNRNP1 (Supplementary Table S2). The PCR products were separated on 5% PAGE and visualized by imaging analyzer (ImageQuant LAS 500, GE Healthcare Life Sciences). The unspliced premRNA and spliced mRNA were quantified using NIH Image J software, and PSI values were calculated according to the manufacturer’s protocol. The knockdown efficiencies of all the targeted 154 genes were analyzed by qPCR and ratios to the control knockdown were provided (Supplementary Table S1). The sequences of the utilized 308 primers are available upon request.

**siRNA knockdown and splicing assays.** HeLa cells (ATCC) and HEK293 cells (Invitrogen –Thermo Fisher Scientific) cultured in 35 mm dishes were transfected with 100 pmol siRNA using Lipofectamine RNAi Max (Invitrogen –Thermo Fisher Scientific) according to the 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, U2AF65 siRNA#1, and hPRT3 siRNA#1 were purchased (Nippon Gene; Supplementary Table S2 for the sequences).

To analyze endogenous splicing products derived from the HNRNP1, EML3, DUSP1, NFkbia, MUS81, RECL4, 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 (Supplementary Table S2). 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 U2AF65-depleted HeLa cells were transfected at 48 h and 68 h post-transfection, respectively, with 0.5 µg of mini-gene plasmid (Supplementary Table S2) using lipofectamine 2000 reagent (Invitrogen –Thermo Fisher Scientific). These cells were incubated for 24 h and 4 h (for Fig. 6) 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 (Supplementary Table S2). 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/UMHm/siR, or pcDNA3-Flag-SPF45/Ag/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; Supplementary Table S2) and all the PCR products were performed with Blend Taq polymerase (Toyobo Life Science).
irradiated with 254 nm UV light on ice. The collected cells were lysed and immunoprecipitated with anti-U2A8 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 (Supplementary Table S2).

**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 procedure48. Biotin-labeled HKRNP1 and AdML pre-mRNA 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 Streptavidin T1 magnetic beads (Invitrogen–Thermo Fisher Scientific) according to the manufacturer’s instruction. The immobilized pre-mRNA beads were incubated at 30 °C for 30 min in 30 µL reaction mixture containing 30% nuclear extract, RNase inhibitor (Takara Bio) and nuclease-free water. Then NET2 buffer [50 mM Tris (pH 7.5), 150 mM NaCl, 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).

**Gel pull-down assays.** GST–SPF45, GST–SPF45/UEHMmt, or GST–SPF45/AG were expressed in E. coli BL21 (DE3) CodonPlus (DE3) competent cells (Stratagene–Aglent) 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 (50 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 4B (GE Healthcare Life Sciences) 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).

**Immunofluorescence microscopic assays.** Immunofluorescence microscopic assays of ectopically expressed Flag-tagged SPF45 proteins were performed as described essentially previously48. HeLa cells (in 35 mm dishes) were transfected with 1 µg of plDNA3-Flag–SPF45-sir, pcDNA3-Flag–SPF45/UEHMmtsir, or pcDNA3-Flag–SPF45/AG-sir with 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 the following primary antibodies in 2% skimmed milk/PBS for 0.5 h: anti-SPF45 (1:200 dilution; Thermo Fisher Scientific). After washing three times with PBS, cells were incubated with Alexa Fluor 488 or Alexa Fluor 568 secondary antibody (Invitrogen–Thermo Fisher Scientific) and then washed five times with PBS. DNA in cells was counter-stained with 4’-6-diamidino-2-phenylindole (DAPI). The images were analyzed by fluorescence microscopy (Olympus).

**Preparation of recombinant proteins.** Recombinant SPF45–G–patch–UHM (234–401) was expressed from pET9d vectors with Hist–ProteinA TEV cleavable tag using E. coli BL21 (DE3) in minimal M9 medium supplemented with 3 HCl for [35S]–labeled protein. Protein expression was induced at OD600 around 0.8–1.0 with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG), followed by overnight expression at 18 °C. Cells were resuspended in 30 mM Tris/HCl (pH 8.0), 500 mM NaCl, 10 mM imidazole pre-inhibitors and lysed using french press. After centrifugation, the cleared lysate was purified with Ni-NTA resin column. The protein sample was further purified 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.

**NMR spectroscopy.** NMR experiments were recorded at 298 K on 500 MHz Bruker Avance NMR spectrometers equipped with cryogenic triple resonance gradient probes. NMR spectra were processed by TOPSPIN3.5 (Bruker), and analyzed using Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). Samples were measured at 100 µM protein concentration in the NMR buffer (20 mM sodium phosphate (pH 6.5), 150 mM NaCl, 3 mM DTT) with 10% D2O added as lock signal. The UHR NMR chemical shift assignment was transferred from the Biological Magnetic Resonance Biobank (BMRC) database (MRJ) to E-MRJ RNAs [EML: 5′-GACGUGUAAUUGGGAC-3′, HNRNPH1: 5′-CCUGUGUACUAUGAC-3′] used for the NMR titration was purchased (IBA Lifesciences).

**Data availability**

The data supporting the findings of this study are available from the corresponding authors upon reasonable request. The raw data from the RNA-Seq analysis of SPF45-knockdown HeLa cells have been deposited in the Sequence Read Archive (SRA) database under accession number GSE135128. Source data underlying the corresponding figures are provided with this paper.

**Code availability**

The Code for the analyses described in this study is available from the corresponding author upon request.

Received: 10 September 2020; Accepted: 6 July 2021; Published online: 13 August 2021

**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, 171–175 (1991).
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. Stamm, S., Smith, C., Lührmann, R.) (Wiley-Blackwell, 2012).
5. Lopes-Meija, I. & Taz, J. in *Alternative pre-mRNA splicing: Theory and protocols* (eds. Stamm, S., Smith, C., Lührmann, R.) (Wiley-Blackwell, 2012).
6. Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701–719 (2009).
7. Behazdnia, N. et al. Composition and three-dimensional EM structure of double affinity-purified, human pre spliceosomal 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 Acid Res.* 8, 5739–5731 (1980).
9. Zhang, L., Vieille, A., Espinoza, S. & Zhao, R. RNAs in the spliceosome: insight from cryoEM structures. *Wiley Interdiscip. Rev. RNA* 10, e1523 (2019).
10. Sasaki-Haraguchi, N., Shimada, M. K., Tanaguchi, 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 usually 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., Licklinder, L. J., Gygi, S. P. & Reed, R. Comprehensive proteomic analysis of the human spliceosome. *Nature* 449, 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. Alternative pre-mRNA splicing: Theory and protocols (eds. Stamm S., Smith C., Lührmann R.). (Wiley-Blackwell, 2012).
24. Hegele, A. et al. Dynamic protein-protein interaction wiring of the human spliceosome. Mol. Cell 45, 567–580 (2012).
25. Berget, S. M. Exon recognition in vertebrate splicing. J. Biol. Chem. 270, 2411–2414 (1995).
26. Tan, Q. et al. Extensive cryptic splicing upon loss of RBM17 and TDP43 in neurodegeneration models. Hum. Mol. Genet. 25, 5083–5093 (2016).
27. Corsini, I. et al. Dimerization and protein binding specificity of the U2AF motif of the spliceosome. Factor Prof. J. Biol. Chem. 284, 630–639 (2009).
28. Neubauer, G. et al. Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosomal complex. Nat. Genet. 20, 46–50 (1998).
29. Agafonov, D. E. et al. Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. Mol. Cell Proteomics 7, 2603–2611 (2008).
30. Wahl, M. C. & Lührmann, R. SnapShot: spliceosome dynamics I. Cell 161, 1474–e1 (2015).
31. 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 SFA4. Cell 109, 285–296 (2002).
32. 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).
33. Mackereth, C. D. et al. Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. Nature 475, 408–411 (2011).
34. Agrawal, A. A. et al. An extended U2A65-RNA binding domain recognizes the 3’ splice site signal. Nat. Commun. 7, 10950 (2016).
35. Banerjee, H., Rahn, A., Davis, W. & Singh, R. Sex lethal and U2 small nuclear ribonucleoprotein auxiliary factor (U2AF65) recognize polypryrimidine tracts using multiple modes of binding. RNA 9, 88–99 (2003).
36. Sartorius, U., Schmitz, I. & Krammer, P. H. Molecular mechanisms of death-receptor-mediated apoptosis. J. Immunol. 181, 1282–1289 (2008).
37. 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 U2A65. PLoS ONE 2, e538 (2007).
38. De Maio, A. A. et al. RBM17 interacts with U2SURF and CHERP to regulate expression and splicing of RNA-processing proteins. Cell Rep. 25, 726–736 (2018).
39. Kasowski, M. et al. Smau1 and RED are required for activation of spliceosomal B complexes assembled on short introns. Nat. Commun. 10, 3639 (2019).
40. Fukushima, 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).
41. Yoshimoto, R., Kataoka, N., Okawa, K. & Ohno, M. Isolation and characterization of post-splicing lariat-intron complexes. Nucleic Acids Res. 37, 891–902 (2009).
42. Yoshimoto, R. et al. Global analysis of pre-mRNA subcellular localization following splicing inhibition by spliceostatin A. RNA 23, 47–57 (2017).
43. Shen, H. et al. Reversal of multidrug resistance of gastric cancer cells by downregulation of TSG101 with TSG101iRNA. Cancer Biol. Ther. 3, 561–565 (2004).
44. 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).
45. Corvelo, A., Hallegger, M., Smith, C. W. J. & Eyras, E. Genome-wide characterization of the multi-protein spliceosome complex. Nucleic Acids Res. 50 (1998).
46. Kבחר, A. et al. The importance of the spliceosome in pre-mRNA splicing. Biochem. Biophys. Res. Commun. 496, 921–926 (2018).