ARTICLE

Check for updates

psiCLIP reveals dynamic RNA binding by DEAH-box helicases before and after exon ligation

Lisa M. Strittmatter [1,6], Charlotte Capitanchik[2,6], Andrew J. Newman[1], Martina Hallegger[2,3], Christine M. Norman[1], Sebastian M. Fica [1,7], Chris Oubridge[1], Nicholas M. Luscombe [2,4,5,7], Jernej Ule [2,3,7✉] & Kiyoshi Nagai [1]

RNA helicases remodel the spliceosome to enable pre-mRNA splicing, but their binding and mechanism of action remain poorly understood. To define helicase-RNA contacts in specific spliceosomal states, we develop purified spliceosome iCLIP (psiCLIP), which reveals dynamic helicase-RNA contacts during splicing catalysis. The helicase Prp16 binds along the entire available single-stranded RNA region between the branchpoint and 3′-splice site, while Prp22 binds diffusely downstream of the branchpoint before exon ligation, but then switches to more narrow binding in the downstream exon after exon ligation, arguing against a mechanism of processive translocation. Depletion of the exon-ligation factor Prp18 destabilizes Prp22 binding to the pre-mRNA, suggesting that proofreading by Prp22 may sense the stability of the spliceosome during exon ligation. Thus, psiCLIP complements structural studies by providing key insights into the binding and proofreading activity of spliceosomal RNA helicases.

---

[1] MRC Laboratory of Molecular Biology, Cambridge, UK. [2] The Francis Crick Institute, London, UK. [3] Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, Queen Square, London, UK. [4] UCL Genetics Institute, Department of Genetics, Environment and Evolution, University College London, London, UK. [5] Okinawa Institute of Science & Technology Graduate University, Okinawa, Japan. [6] These authors contributed equally: Lisa M. Strittmatter, Charlotte Capitanchik. [7] These authors jointly supervised this work: Sebastian M. Fica, Nicholas M. Luscombe, Jernej Ule. ✉email: jernej.ule@crick.ac.uk

Splicing is an essential step in pre-mRNA processing in eukaryotes, during which non-coding introns are removed and coding exons are ligated. Splicing is catalysed by a large, dynamic molecular machine called the spliceosome, which comprises about a hundred proteins in yeast. During the first splicing step—branching—the spliceosome catalyses attack by the branch point adenosine (brA) at the 5′-splice site (5′-SS), producing a lariat-intermediate structure[1]. During the second step of splicing—exon ligation—the cleaved 5′-SS attacks the 3′-splice site (3′-SS) to ligate the exons and form the mature RNA (mRNA). To faithfully remove introns, the spliceosome must accurately recognize the 5′-SS, the brA, and the 3′-SS (Fig. 1A). Use of the wrong splice sites could ultimately lead to errors in protein translation and erroneous splicing has been linked to cancer in humans, leading to the development of therapeutics designed to modulate splicing[2]. Eight ATP-dependent helicases—comprising three DEAD-box helicases (Sub2, Prp5, and Prp28), four DEAH-box helicases (Prp2, Prp16, Prp22, and Prp43), and a Ski2-like helicase (Brr2)—ensure splicing fidelity by actively promoting correct spliceosome assembly and splice site usage[3]. All eight helicases are essential in yeast and conserved in humans[4,5].

The DEAD-box helicases Sub2, Prp5, and Prp28 are involved in early assembly of the spliceosome on introns, promoting correct splice site recognition[5]. The Ski2-like helicase Brr2 is involved in spliceosome activation prior to the first catalytic step of splicing. The four DEAH-box helicases Prp2, Prp16, Prp22, and Prp43 remodel the spliceosome through the first (branching) and second (exon ligation) catalytic steps, and coordinate spliceosome disassembly once the reaction has completed[5]. In vitro experiments and structural studies of spliceosomal DEAH-box helicases showed that they translocate on RNA in the 3′ to 5′ direction to unwind duplexes[6,7]. Prp16 repositions the substrate after the branching reaction by removing the branch helix, formed between the branch point sequence and U2 snRNA, from the active site of the C-complex spliceosome[8–10]. Following this remodelling event, Prp16 dissociates from the C complex and causes dissociation of the branching factors that stabilize the branch helix[11]. Prp16 action generates binding sites for the exon-ligation factors Slu7 and Prp18 (refs. [10,12,13]), resulting in formation of the C* complex, which is competent for exon ligation. In the C* complex, Prp16 is replaced by Prp22, which proofreads exon ligation[14]. Together with the exon-ligation factors Slu7 and Prp18, Prp22 promotes exon ligation for most pre-mRNA substrates, acting in an ATP-independent manner[15]. Exon ligation results in the formation of the spliceosomal P complex that retains both the excised intron and the mature mRNA, which is held by spliceosomal proteins and the U5 snRNA[11]. The mature mRNA is then released from the spliceosome through the ATP-dependent activity of Prp22 (refs. [15–17]), which also results in the release of the exon-ligation factors Prp18 and Slu7. Prp22 was shown to unwind duplex RNA in vitro in an NTP-dependent manner[6,17]. After exon ligation, Prp22 crosslinks to the second exon and was proposed to release the mRNA by translocating along the second exon in a 3′ to 5′ direction, thus disrupting the U5 snRNA–mRNA duplex[18]. Nonetheless, it remained unclear whether Prp22 translocates directly through the duplex or acts at a distance.

Despite advances in structural and compositional analysis of the spliceosome, much is still unknown about helicase-mediated transitions between spliceosome intermediates[11,19]. Although previous research indicates the stages at which helicases are required, it is not known exactly where on the snRNA and/or pre-mRNA they bind, how they promote conformational changes in the spliceosome, nor the mechanism by which they ensure splicing fidelity. Both steps of splicing are reversible, offering the

potential to prevent suboptimal pre-mRNA transcripts from completing splicing[20]. Indeed, DEAD/H-box helicases were proposed to promote the backwards reaction for suboptimal splice sites that engage the spliceosome active site[3,21]. Two helicase-mediated proofreading mechanisms were suggested to increase splicing fidelity. The kinetic model posits that the ATPase activity of DEAH-box helicases acts as a timer. ATP hydrolysis promotes either the productive pathway towards exon ligation, for optimal splice sites, or the discard pathway, for suboptimal splice sites, depending on the rate of catalysis, which is inferred to be faster than ATP hydrolysis for optimal substrates[3]. In the discard pathway, spliceosomes assembled on suboptimal splice sites are dissociated by the ATPase Prp43, whose activity safeguards against the accumulation of splicing errors[22]. The thermo-dynamic model proposes that splicing fidelity is enhanced through modulating the stability of competing spliceosome conformations, with suboptimal substrates being rejected because they promote conformations that compete with catalysis[3]. Indeed, DEAH-box helicases were proposed to shift the equilibrium between different catalytic conformations of the spliceosome[23]. Although these two models are not mutually exclusive, their relative role in proofreading is unknown.

Most helicases bind to the spliceosome only transiently, making them difficult to study. Electron cryo-microscopy (cryo-EM) structures have shown helicases to be at the spliceosome periphery, where they are detected at much lower local resolutions than other components due to their flexibility. Given their peripheral location in these structures, it remains unclear by what mechanisms helicases promote remodelling of the RNA-based catalytic core of the spliceosome.

Additional insights have been gained by site-specific cross-linking of helicases to pre-mRNA substrates in splicing extracts upon incorporation of single photo-reactive residues into the RNA[18,24,25]. However, the identified binding positions were limited to the nucleotides chosen for site-specific modification and were not linked to a specific conformation of the spliceosome. Single-molecule fluorescence resonance energy transfer (smFRET) studies have advanced our understanding of the mechanism of DEAH-box helicases, suggesting they remodel the spliceosome active site from a distance[9]. However, this approach does not provide precise positional information about RNA–protein contacts and is technically cumbersome, as the positioning of FRET probes must be carefully calibrated for each substrate. Perhaps for this reason, most studies focused on one specific pre-mRNA transcript, making it difficult to distinguish transcript-specific effects from general principles.

Transcriptome-wide studies, such as spliceosome iCLIP using an antibody against the small nuclear ribonucleoprotein-associated protein B (SmB)[26], profiling using Prp19-TAP affinity purification[27], and affinity purification of post-catalytic spliceosomes after mRNA release[28], have uncovered new positional principles for assembly of specific spliceosomal complexes on pre-mRNAs. However, the contribution of individual splicing factors is difficult to discern in such profiles. In addition, iCLIP of individual splicing factors has revealed their position-dependent capacity to control alternative splicing decisions in mammals[29], but has not led to direct insights into step-specific mechanisms for the core spliceosome. A key issue with these methods is that spliceosomes are not stalled at a specific step, therefore the resulting profiles represent an ensemble of spliceosomal snapshots across many splicing steps, with a likely bias towards those that are rate limiting. This prevents us from assigning specific binding sites to individual steps. To understand the molecular mechanisms that underlie splicing fidelity and splice site choice, a method is needed that is capable of identifying RNA contacts formed at defined steps of splicing. Such a method would be
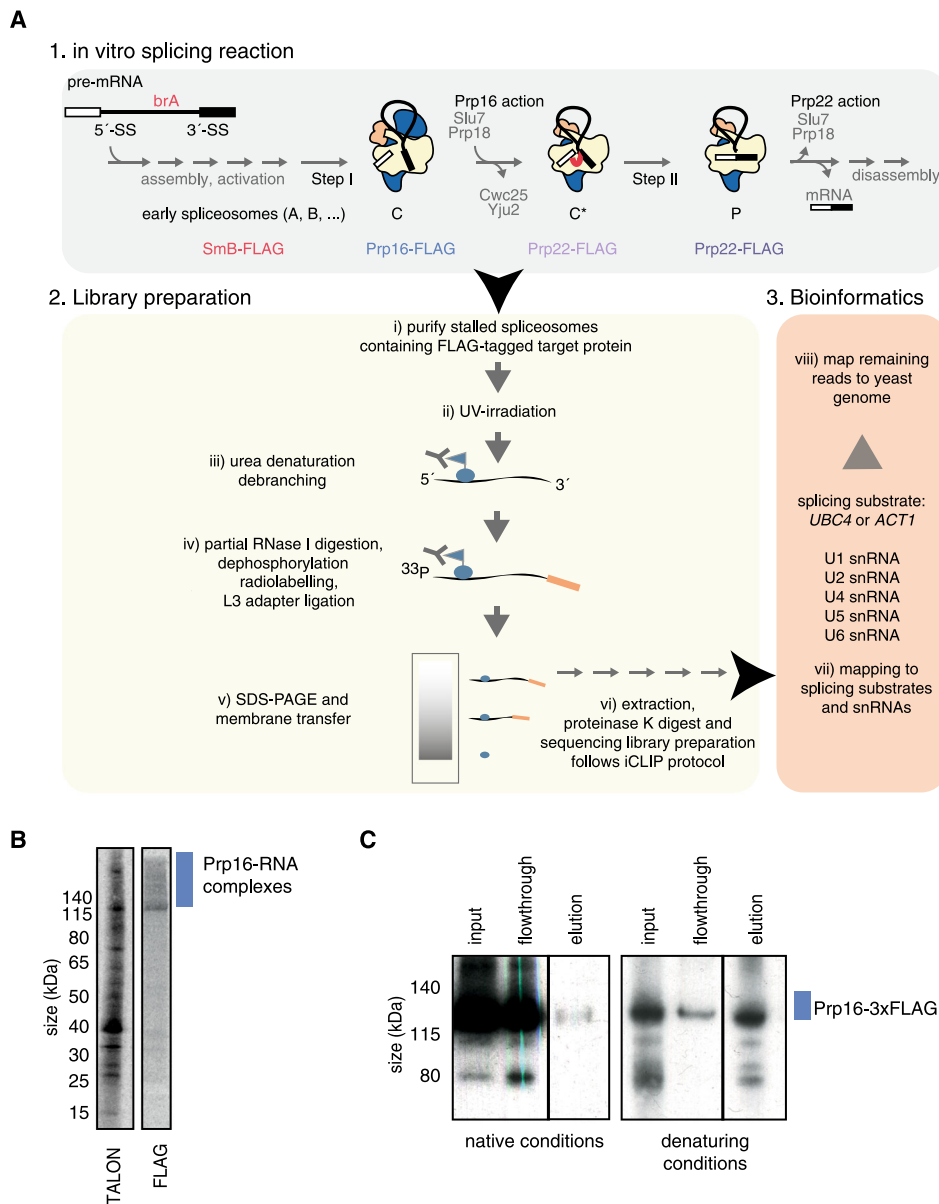
Fig. 1 The psiCLIP method reveals the complete RNA-binding profile of an RBP in a specific spliceosomal complex. A Spliceosomes are assembled from yeast cell extract on an in vitro transcribed pre-mRNA and stalled through different strategies that block the transition towards later stages. Spliceosomes are depicted in schematics, where U2 snRNP is coloured in yellow, U5 snRNP in blue, and U6 snRNP in orange. The RNA-binding protein of interest is fused to a C-terminal FLAG tag. After purification of the stalled spliceosome population, spliceosomes are irradiated with a defined UV dose at 254 nm, and denatured to release the target protein from the complex and allow FLAG capture of the RNA–protein complex. For spliceosomes stalled at steps containing lariat structures, an additional debranching step was introduced. Controlled RNase I digestion fragments the RNA into suitable pieces for Illumina NGS. The 5′ end is radio-labelled, while a DNA adapter is attached to the 3′ end, carrying a sequence for primer annealing for reverse transcription. Finally, the crosslinked RNA–protein complexes are resolved on SDS–PAGE and extracted for RNA library preparation. See "Methods" for details. **B** Autoradiographs of RNA after the psiCLIP procedure following capture of Prp16 with either TALON beads for His-tag proteins (left) or anti-FLAG beads for 3× FLAG tag proteins (right). Signal that arises from Prp16 crosslinked to RNA is expected to be larger than the size of Prp16 at 122 kDa. **C** anti-FLAG western blot detecting Prp16-3× FLAG shows that denaturing conditions are required to capture the tagged protein. Purified spliceosomal complex C containing endogenous Prp16-3× FLAG was captured using anti-FLAG beads after either treatment with 6 M urea (right) or under native conditions (left). In **B** and **C**, a representative autoradiograph for wild-type Prp16 is shown; a similar result was obtained with Prp16-G378A.

particularly useful in resolving the principles that guide the binding of spliceosomal helicases to RNA substrates, in identifying potential transcript-specific effects, and for understanding how the ATPase activities of helicases are linked to their RNA-binding behaviour and thereby to their proofreading activities.

Here, we present purified spliceosome iCLIP (psiCLIP), a method to determine the complete range of contacts between specific splicing substrates and helicases in native spliceosomal complexes stalled at specific stages of the splicing cycle. UV crosslinking of an enriched complex makes it possible to capture the weak and transient binding of DEAH-box helicases. We used a system of in vitro-assembled spliceosomes in budding yeast (*Saccharomyces cerevisiae*), which allows targeted isolation of helicases at defined points in the splicing pathway and enables functional characterization of helicases using specific mutants. We used psiCLIP to analyse two DEAH-box helicases, Prp16 and

Prp22, which act before and after exon ligation, respectively (C, C*, and P complexes), in the context of multiple pre-mRNA transcripts. The high sensitivity and resolution of our psiCLIP data provide insights into helicase dynamics and lay the foundation for future mechanistic studies to dissect helicase functions in the spliceosome.

Results

Optimized psiCLIP detects RNA binding with positional specificity. In addition to conventional iCLIP, which uses crosslinked cells or tissues as starting material[30], iCLIP has also been adapted to study the binding of purified U2AF2 to pre-mRNAs in vitro[31]. Here we establish psiCLIP, which adapts iCLIP for in vitro studies of spliceosomes stalled at defined stages of the splicing cycle. Native spliceosomes from yeast cell extracts are assembled in vitro on a defined pre-mRNA substrate and stalled, using substrate modifications or dominant-negative, recombinant helicases. Enriched step-specific spliceosomes are then purified in a similar way to those used for structural analysis[32] (Fig. 1A), using affinity tags on the specific pre-mRNA substrate (Supplementary Fig. 1A). As in conventional iCLIP, spliceosomes are irradiated with 254 nm ultraviolet light to crosslink proteins to the pre-mRNA substrate and snRNAs. Next, the RNA-binding protein (RBP) of interest is purified under stringent, denaturing conditions, along with any crosslinked RNA, followed by cDNA library preparation and high-throughput sequencing. The computationally processed sequencing reads provide a profile of RBP–RNA contacts that occur within the spliceosome enriched for a specific state (see Methods).

In this study, we used psiCLIP primarily to investigate the binding profiles of the DEAH-box helicases Prp16 and Prp22 that act in the spliceosomal C, C*, and P complexes (Fig. 1A). Therefore, we optimized purification, crosslinking, and RNA fragmentation for these helicases. Purification using a 3× FLAG tag was more specific than His-tag purification via TALON beads, which has been a common choice for crosslinking and immunoprecipitation[33], therefore, we proceeded with 3× FLAG for all experiments in this study (Fig. 1B). We found that stringent 6 M urea denaturing conditions were required to capture 3× FLAG-tagged Prp16 bound to a range of RNA fragment sizes (Fig. 1C). To optimize crosslinking and RNA fragmentation, we titrated UV and RNase I doses in parallel (Supplementary Fig. 1B), which showed that the signal was saturated at a UV dose of $312 \times 100 \mu J/cm^2$. We chose the UV dose so as to give efficient crosslinking, while minimizing the potential for multiple proximal protein–RNA crosslinks, which could bias reverse transcription truncations to the 3′ end of the binding sites. We also titrated the amount of RNase for treatment of purified spliceosomes so as to minimize the biases that can result from RNA sequence and structure preferences of the RNase[34,35].

To demonstrate the positional specificity of psiCLIP, we first assayed binding of the integral protein SmB, which is known to bind the Sm site in U1, U2, U4, and U5 snRNAs, and is thus present in spliceosomes at all stages of splicing. The Sm binding sites have been characterized by genetic studies, which identified a U-rich consensus sequence as essential for binding[36], while the Sm proteins have been modelled in crystal and cryo-EM structures[37–40]. However, the exact nucleotide-binding positions of the yeast Sm proteins, including their unstructured tail regions, are unknown for most snRNPs, providing an initial test of our method.

Spliceosomes were assembled on a substrate with a 3′-SS mutation from the canonical 3′-UAG| to 3′-UAC| (where | marks the splice site), which impairs exon ligation. A mixture of pre-catalytic and catalytic (complex C) spliceosomes were purified by glycerol gradient and subsequent substrate affinity purification using MS2 stem–loops on the pre-mRNA substrate. In agreement with a previous study[26], psiCLIP performed under stringent 6 M urea denaturing conditions identified SmB crosslinks primarily on snRNAs, with little signal on the pre-mRNA substrate (Fig. 2A and Supplementary Fig. 2A). We observed crosslinks within, and adjacent to, the expected Sm sites, with little signal in the negative controls (no UV/tagged protein; Fig. 2B, C and Supplementary Fig. 2B–G). The three-dimensional structure for the pre-catalytic spliceosomal complex pre-B[40] shows that the SmB protein is in close spatial proximity to the U1 snRNA nucleotides detected using psiCLIP (Fig. 2D). The 5′ end of U1 snRNA extends into the direction of the SmB protein, thus explaining why psiCLIP detected crosslinks mainly at the beginning of the Sm site motif. In addition, the crosslinks upstream of the Sm site identified by psiCLIP (G535 and A536) interact with the long tail of the SmB protein in the cryo-EM structure. Only part of this SmB tail could be modelled as an alpha-helix into the U1 snRNP in the cryo-EM structure of the yeast pre-B complex, due to its high flexibility[3,6]. The psiCLIP data suggest that SmB has a similar configuration in the other snRNPs (Supplementary Fig. 2B–G). The SmB example shows that psiCLIP detects RNA–RBP interactions with high positional specificity and can also provide information on more transient interactions between flexible protein regions and RNA, which are challenging to detect by structural methods.

psiCLIP reveals broad binding of Prp16 downstream of brA. Next, we set out to study the RNA contacts made by spliceosomal helicases. The DEAH-box helicase Prp16 is essential for the transition from complex C, the complex after branching, to complex C*, which catalyses exon ligation[13]. Prp16 binds transiently to the spliceosome to reposition the splicing substrate, removing the brA from the active site, creating space that will later allow the 3′-SS to dock in the active site[10,41]. This Prp16-dependent spliceosome remodelling dissociates branching factors Cwc25 and Yju2, and promotes stable binding of the exon-ligation factors Slu7 and Prp18 (refs. [9,10,12,13,42,43]).

To further optimize psiCLIP, we introduced a debranching step for spliceosomes bound to the lariat RNA structure produced after branching, in order to reduce cDNA truncations at branch points, which could otherwise be erroneously interpreted as protein–RNA crosslinks[26,28] (Supplementary Fig. 3). Despite debranching, we still observed some cDNA truncations at the brA, which likely result from truncation at partially digested three-way lariat junctions (Supplementary Fig. 3A). Therefore, we removed reads mapping to a five nucleotide region around the brA (−2 to +2 around brA) to avoid confounding the analysis. Moreover, we established a way to visualize data across multiple experiments on a single plot for comparative purposes. The psiCLIP data provide nucleotide-resolution data on crosslinking that can be visualized with histograms, with the size of each bar corresponding to the number of crosslink events at each nucleotide (Fig. 3A). While nucleotide-resolution histograms convey precise positional information, as seen in the case of SmB data (Fig. 2), they cannot be used to present data across multiple experimental conditions on a single plot since overlapping bars would not be visible. Therefore, we present the data from helicase psiCLIP experiments as Gaussian smoothed curves with a window size of ten nucleotides, which reflects well the crosslinking trajectory (Fig. 3B), and is in agreement with the nine nucleotides of RNA bound to the two RecA domains in structures of DEAH-box helicases[7].
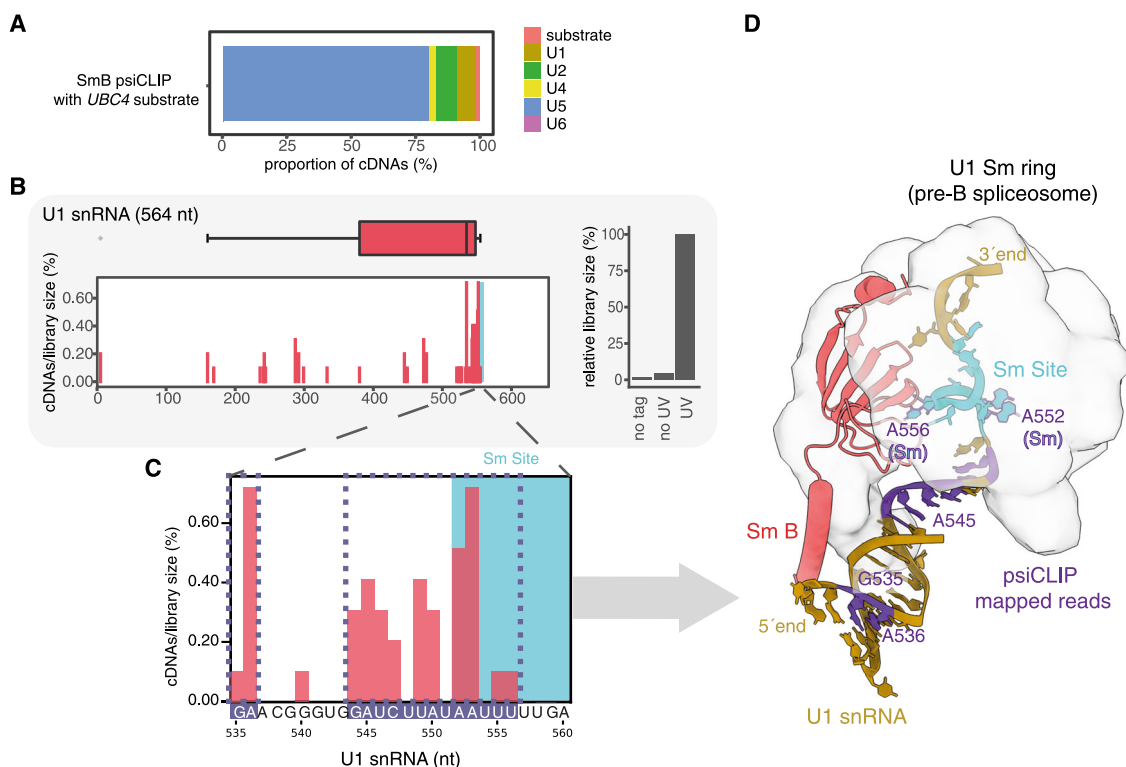
Fig. 2 Validation of the positional specificity of psiCLIP with SmB. A Proportion of cDNAs mapping to substrate and snRNAs for one SmB psiCLIP experiment on the short *UBC4* substrate. **B** Crosslinking to U1 snRNA. Crosslink events (−1 position of cDNA start) are represented as a histogram and box plot aligned to the indicated snRNA. On the histogram, the *y*-axis indicates the proportion of cDNAs out of all cDNAs mapped to snRNAs and pre-mRNA substrate. The box plot represents crosslink positions from one sample along the transcript, weighted by normalized cDNA count. The line across the box represents the median; the lower and upper bounds correspond to the first and third quartiles. The whiskers end at the largest and smallest value no further than 1.5 times the inter-quartile range. Outliers outside of this range are plotted as dots. The Sm site is highlighted in blue throughout. The bar chart represents the proportion of cDNAs in samples, normalized to the UV condition, which is shown as 100%. No tag: UV-irradiated untagged SmB, no UV: SmB-3× FLAG without UV irradiation, and UV: UV-irradiated SmB-3× FLAG. **C** Zoom in to show crosslinking specifically at the Sm sequence element on U1 snRNA, which is highlighted with a blue background. The U1 snRNA region highlighted in purple is shown in **D** mapped onto the cryo-EM structure. **D** Structure of the U1 snRNP Sm ring in the spliceosomal complex pre-B (ref. [40]; PDB: 5ZWN) with the SmB protein in salmon and the nucleotides of U1 snRNA in gold, with those showing the highest psiCLIP signal in purple or with a purple outline, corresponding to the purple boxes in **C**.

We performed Prp16 psiCLIP using the purified C-complex spliceosome, which was stalled and enriched using three different approaches: mutation of the 3′-SS from the canonical 3′-UAG| to 3′-UAC| (where | marks the splice site), addition of recombinant, ATPase-deficient Prp16-G378A mutant[44], or a combination of 3′-SS and Prp16 mutations. To assess substrate-specific effects, we compared *UBC4* and *ACT1*—two pre-mRNAs that have been used extensively in previous spliceosome functional and structural studies. Under similar conditions that were used to pull down Prp16 after UV crosslinking, Prp16 binds the lariat intermediate but not the pre-mRNA, demonstrating that our method captures reads derived from spliceosomes after branching (Supplementary Fig. 4A).

psiCLIP of Prp16 shows crosslinks predominantly on the pre-mRNA substrate, with little signal on snRNAs (Fig. 3C). The snRNAs that are not present in C complex, such as U1 and U4 snRNA, are under-represented compared with U2 and U5 snRNA, which are part of C complex (Fig. 3D). Our three stalling methods were expected to enrich for the same spliceosomal state—the C complex stalled right after branching. Indeed, we found that Prp16 makes similar RNA contacts regardless of the stalling method, demonstrating the high reproducibility of the psiCLIP method (Supplementary Figs. 3E, F and 4C, E). Nonetheless, we found that poorer crosslinking

efficiency for Prp16, compared to Prp22, for example, resulted in increased stochastic variation in the crosslinking profiles between replicates (Supplementary Fig. 4B–G).

On both splicing substrates, the binding pattern of Prp16 is widely spread, producing prominent crosslinks up to 40 nucleotides downstream of the brA and extending well beyond the predicted 9 nucleotides for the occluded site of a DEAD/H-box helicase[7]. On *ACT1* Prp16 shows a prominent peak ~30 nucleotides downstream of the brA, whereas on *UBC4* the main peak occurs 20 nucleotides downstream of the brA, though binding spans the entire region between the brA and the second exon for both substrates (Fig. 3E, F—sum of all replicates, and Supplementary Fig. 4C, E—three replicates shown separately). Although the position of the main crosslinking peak is significantly different between the two splicing substrates ($p < 0.001$, alpha = 0.05, Student's unpaired two sided *t* test, Supplementary Fig. 4H, I), comparisons between substrates need to be interpreted with caution, as crosslinking efficiency may be affected by the underlying RNA sequence. Our findings include positions previously identified using photo-reactive nucleotides[9,25] and confirm inferences from the C-complex cryo-EM structure, in which at least 18 nucleotides of RNA were predicted to span the distance between the brA and the RNA entry site of Prp16. This prediction was drawn from the six
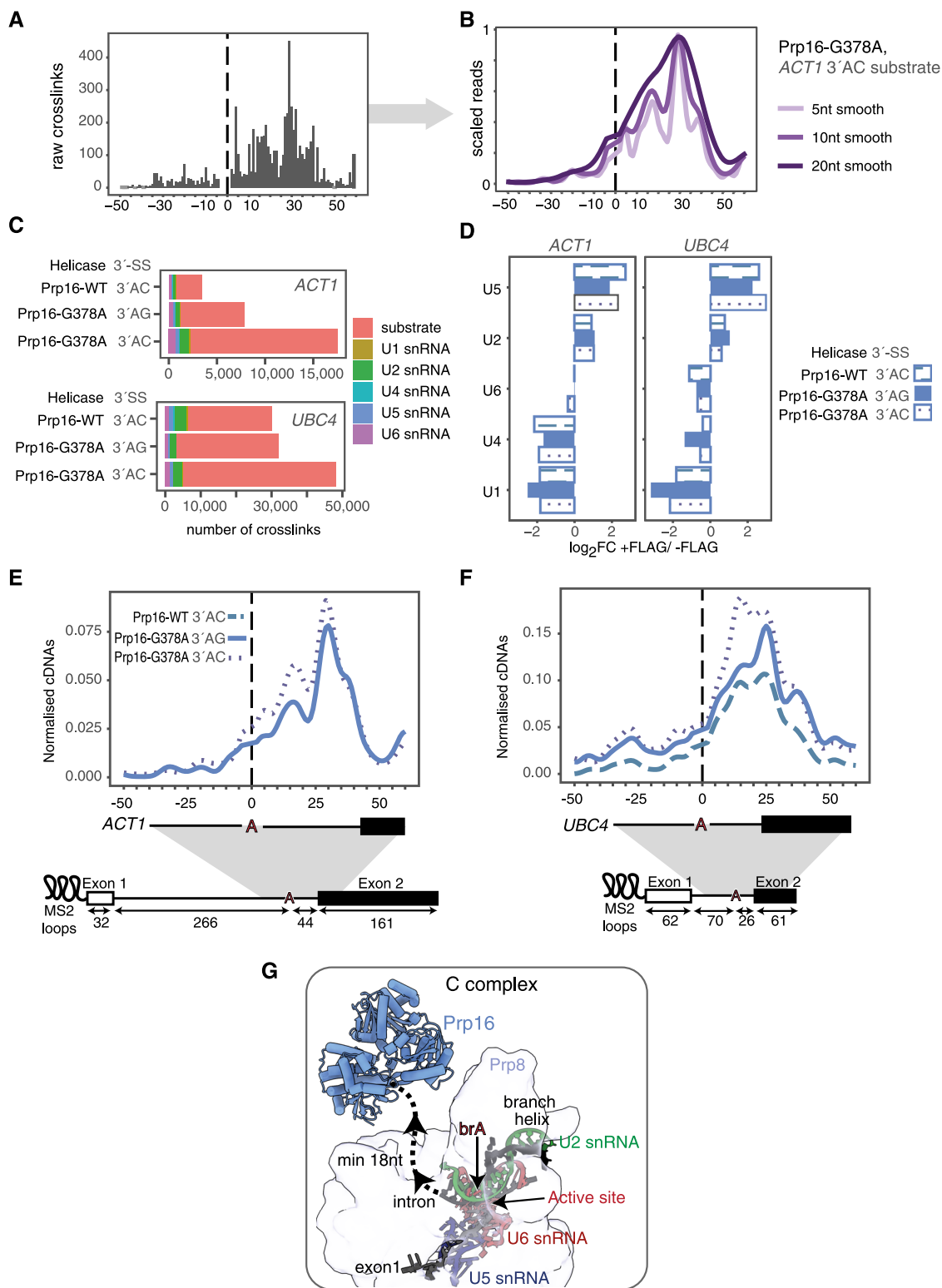
nucleotides closest to the brA, which could be modelled into ordered density, with the additional 12 added based on the approximate length that would be needed to reach the RNA entry site of Prp16 (ref. [8]; Fig. 3G). Importantly, Prp16 binding did not appear strictly constrained by the 3′-SS, as binding could be detected in the second exon for *UBC4*. Overall psiCLIP provides comprehensive positional information for Prp16 binding, showing an extensive binding profile in C complex, downstream of the brA.

Prp16 binding depends on the brA to 3′-SS distance. To test the constraints on Prp16 binding, we designed transcripts based on *ACT1* with increasing lengths between the brA and 3′-SS. Randomized 20 and 40 nucleotide sequences with the same GC content as the *ACT1* intron were inserted after the brA, creating transcripts that spliced with similar efficiency to the original transcript in a wild-type Prp16 background (Fig. 4A–C). psiCLIP was performed with C-complex spliceosomes assembled on these substrates and stalled using the ATPase-deficient Prp16 mutant

Fig. 3 Prp16 psiCLIP data show substrate-specific binding in a region downstream of the brA. A, **B** An example of Prp16 data shown as raw crosslinks and after Gaussian smoothing using different window sizes. Raw crosslink signal is shown as a grey histogram. **C** Total number of crosslinks summed across three replicates for each condition (UV crosslinked, 3× FLAG-tagged samples). Colour denotes the RNA species to which the crosslinks map. **D** Enrichment of snRNAs in FLAG-tagged Prp16 vs. untagged samples, represented as log$_2$ fold change. Note that for this analysis, crosslinks were first normalized to the total number of snRNA crosslinks in each sample. **E**, **F** Mapping of Prp16 psiCLIP data onto *ACT1* and *UBC4* splicing substrates. Three replicates were summed. The smoothed lines show the truncation events normalized to endogenous yeast RNA, with the control signal subtracted from the UV signal. Lines were Gaussian smoothed with a window size of ten nucleotides. Positions along the transcript are shown relative to the brA. Crosslinking signals derived from the three different stalling methods show the same pattern. The main peak that reflects the position of the Prp16 helicase is slightly further downstream on *ACT1* than on *UBC4*. Note that wild-type Prp16 binding on 3′-AC *ACT1* substrate is excluded here due to small library sizes and low reproducibility (Supplementary Fig. 4F). **G** Cryo-EM structure of Prp16 in C complex shows the helicase at a distance from the brA (ref. [8]; PDB: 5LJ3). Six nucleotides were built into the density after the brA (brA +6 is the last ordered nucleotide). The predicted path of the intron and second exon is shown as a dotted line, implying a minimum distance of 18 nucleotides between the brA and the entry site of Prp16.

(Prp16-G378A). As the distance between brA and 3′-SS was increased, the binding region of Prp16 expanded to fill the full intronic region between the brA and 3′-SS (Fig. 4E–G). In all cases, a clear drop of signal was observed at the intron–exon boundary (Supplementary Fig. 5A–C), even though the normalized crosslinking signal generally decreases on the extended transcripts (Supplementary Fig. 5E). We observed crosslinking ~30 nucleotides upstream of the brA, but this was disregarded as it was inconsistent between replicates and it was also detected in the control samples (Fig. 4G, H). Thus, our results indicate that Prp16 binds downstream of the brA on the entire length of accessible single-stranded RNA, and does not strictly discriminate between introns and exons.

Prp16-mediated remodelling of C complex removes the brA from the active site to allow 3′-SS docking for exon ligation. Prp16 action thus results in remodelling of the branch helix, which forms between U2 snRNA and the substrate region upstream of brA (Fig. 3G). The branch helix may be disrupted during Prp16 action and such disruption may allow Prp16 to promote use of alternative 3′-SS[9]. A long-standing puzzle has been how Prp16, transiently located at the spliceosomal periphery, can remodel the branch helix, which is buried within the spliceosome core. Previous work suggests Prp16 can act indirectly without translocating through the branch helix[8,9]. To investigate whether Prp16 could disrupt RNA helices during remodelling of the spliceosome, we introduced a 40 nucleotide stem–loop into the *ACT1* intron between the brA and the 3′-SS (Fig. 4D, H, I and Supplementary Fig. 5D). This substrate was spliced in a wild-type Prp16 background, indicating that the Prp16 action is not appreciably hindered by the inserted sequence (Fig. 4D). We then performed psiCLIP on C complexes assembled on this substrate and stalled with the Prp16-G378A mutant. We observed Prp16 crosslinking mainly on the predicted single-stranded region between the stem–loop and the 3′-SS (Fig. 4H, I). The 5′ arm of the stem–loop showed very little crosslinking and the 3′ arm showed reduced crosslinking, suggesting that the stem–loop sequence indeed formed a secondary structure that hindered Prp16 binding. Since RBPs preferentially crosslink to single-strand RNA[45], the small amount of crosslinking observed towards the 3′ end of the 3′ arm (Fig. 4I) suggests that the Prp16 helicase mutant, which retains residual ATPase activity, may have partially unwound the stem–loop. Nonetheless, the impaired Prp16 binding we observe is consistent with a model in which Prp16 remodels the branch helix without translocating fully through the brA[9]. Although future experiments will be necessary to determine whether wild-type Prp16 could fully disrupt such stem–loop sequences, our findings for the mutant Prp16 helicase demonstrate that psiCLIP can be used to test hypotheses about the activity of DEAH-box helicases in the context of intact native spliceosomes, thus providing complementary functional assays to smFRET and cryo-EM studies.

psiCLIP of Prp22 detects repositioning of the helicase after exon ligation. During remodelling of C complex to the C* complex, Prp16 dissociates from its binding site on Prp8, allowing Prp22 to replace Prp16 on Prp8 and to stabilize the C* complex for exon ligation[10–12] (Figs. 1A, 3G, and 5A). Exon ligation forms the spliceosomal P complex that still contains both the excised intron and the mature mRNA. RNA density observed in the Prp22 helicase core in the cryo-EM structure of the C* complex was assigned to the second mRNA exon, with ~16 nucleotides needed to span the distance between the last nucleotide of the first exon and the centre of the Prp22 RNA-binding pocket[10]. However, biochemical experiments suggested Prp22 binds on the intron close to the 3′-SS before exon ligation[24]. In the structure of the P-complex spliceosome Prp22 was positioned between nucleotides 14–21 on the second exon of the *UBC4* transcript[12], consistent with biochemical experiments[18]. Thus the precise position of Prp22 on the RNA substrate before and after exon ligation remained unclear.

We performed Prp22 psiCLIP using multiple splicing substrates after enriching for spliceosomes stalled either before (C* complex) or after exon ligation (P complex). We enriched for the C* complex by assembling spliceosomes on *UBC4* and *ACT1* transcripts, in which the 3′-SS G(−1) ribonucleotide was replaced with a 2′-deoxynucleotide (dG), which blocks exon ligation[46]. psiCLIP was then performed using both the wild-type and the dominant-negative Prp22-K512A ATPase-deficient mutant. Again, there was little signal on snRNAs, with underrepresentation of U1 and U4 snRNAs, which are absent in C* complex (Fig. 5C, D). Prp22 binding covered a broad region downstream of the brA on both *UBC4* and *ACT1* transcripts. Crosslinks were detected throughout the intron between the brA and the 3′-SS, including a site at eight nucleotides before the 3′-SS, which was previously observed in splicing extracts for a substrate that could not undergo exon ligation[24]. Crosslinks were also detected in the second exon, with a major peak at 10–25 nucleotides downstream of the 3′-SS. The psiCLIP profile for Prp22 covers similar positions as those shown previously to crosslink to Prp8 and Prp22 using site-specific crosslinking in spliceosomes at the C*-complex stage[47,48], further validating our method. Thus Prp22 binds broadly on both the intron and second exon of the lariat intermediate before exon ligation. Importantly, both the wild-type and mutant Prp22 show a similar psiCLIP profile, confirming that RNA binding is ATP-independent, as suggested previously from cryo-EM studies, where RNA density was observed bound to an open, ATP-free, conformation of the helicase[10,49] (Fig. 5E, F and Supplementary Fig. 6A).

To determine if exon ligation changes Prp22 binding, we performed Prp22 psiCLIP in P-complex spliceosomes, which was enriched by adding recombinant Prp22-K512A dominant-negative mutant protein to block release of the mRNA. P-complex psiCLIP reads were enriched for mRNA junctions, with 10–15% of mapped
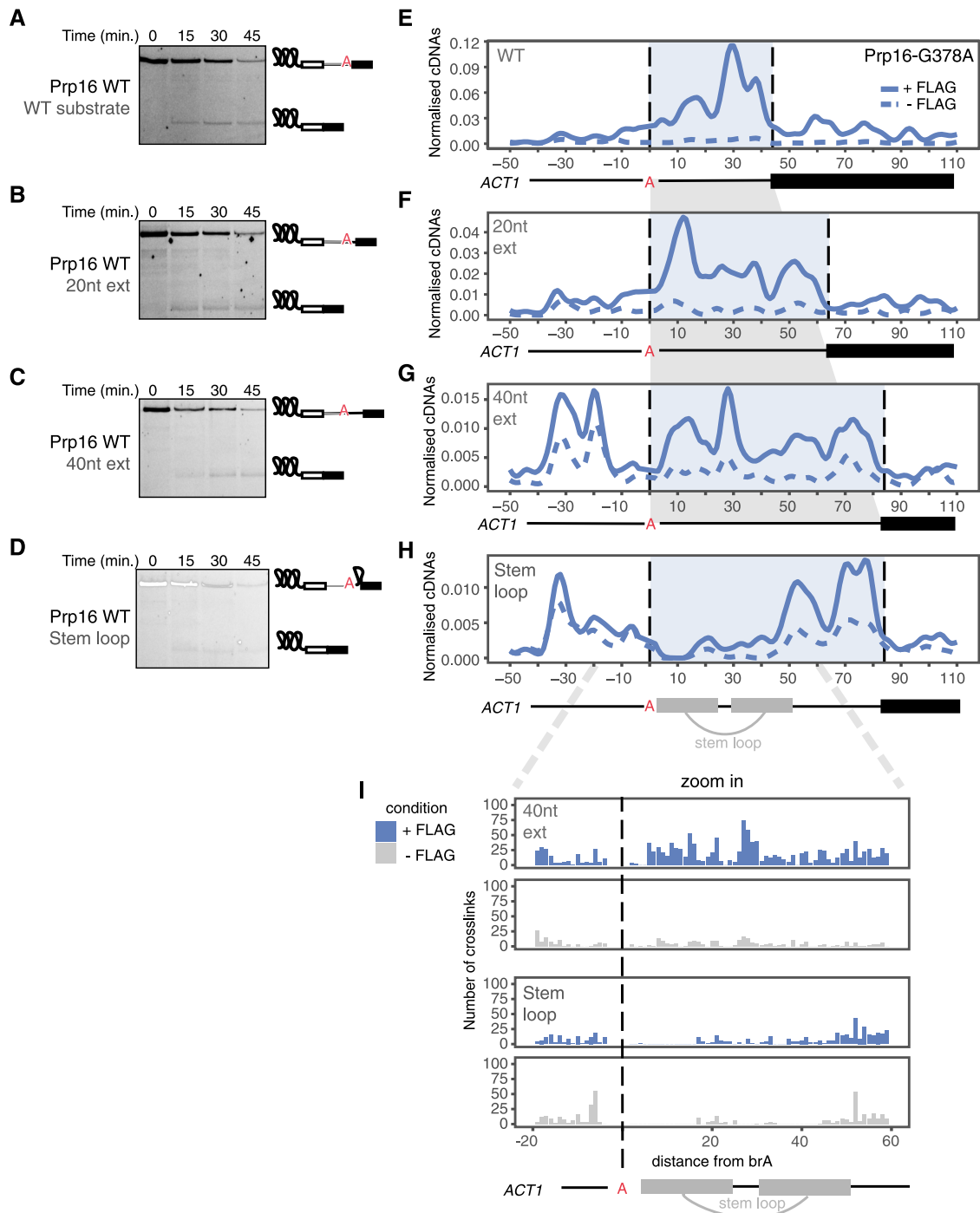
Fig. 4 Prp16 binds the entire available single-stranded RNA between brA and the 3′-SS in *ACT1* extended substrates. A–D All versions of the extended *ACT1* substrate are capable of being spliced. Splicing assays were performed in a Prp16-WT background. Only splicing products containing the second exon are visible, as the 3′ end of the transcript is fluorescently labelled. Gels are representative of two biological replicates. **E–H** Mapping of the psiCLIP data generated using Prp16-G378A mutant protein and various forms of the *ACT1* splicing substrate, including artificial extensions between the brA and the 3′-SS. The splicing substrates are aligned onto the brA, and the distance between the brA and the 3′-SS is highlighted in light blue. The smoothed lines show the truncation events normalized to endogenous yeast RNA, with the untagged signal subtracted from the tagged signal. Lines were Gaussian smoothed with a window size of ten nucleotides. Positions along the transcript are shown relative to the brA. The crosslinking signal spreads across the whole length of the sequence between the brA and 3′-SS. **I** Zoom in on the region around the brA shows the almost complete loss of crosslinking downstream of brA in the stem–loop experiment.

reads containing the canonical splice junction, compared with a maximum of 0.09% junction reads for the C* complex (Supplementary Fig. 6B). We found sharp binding peaks in the second exons of both splicing substrates at +10 and +20 nucleotides downstream of the exon–exon junction in *UBC4* and

ACT1, respectively, with binding extending as far as +40 nucleotides (Fig. 5G, H and Supplementary Fig. 6A). The binding region determined by psiCLIP is in perfect agreement with the binding site implied for *UBC4* by the complex P cryo-EM structure, and contains positions +10, +17, +22, and +35 on the
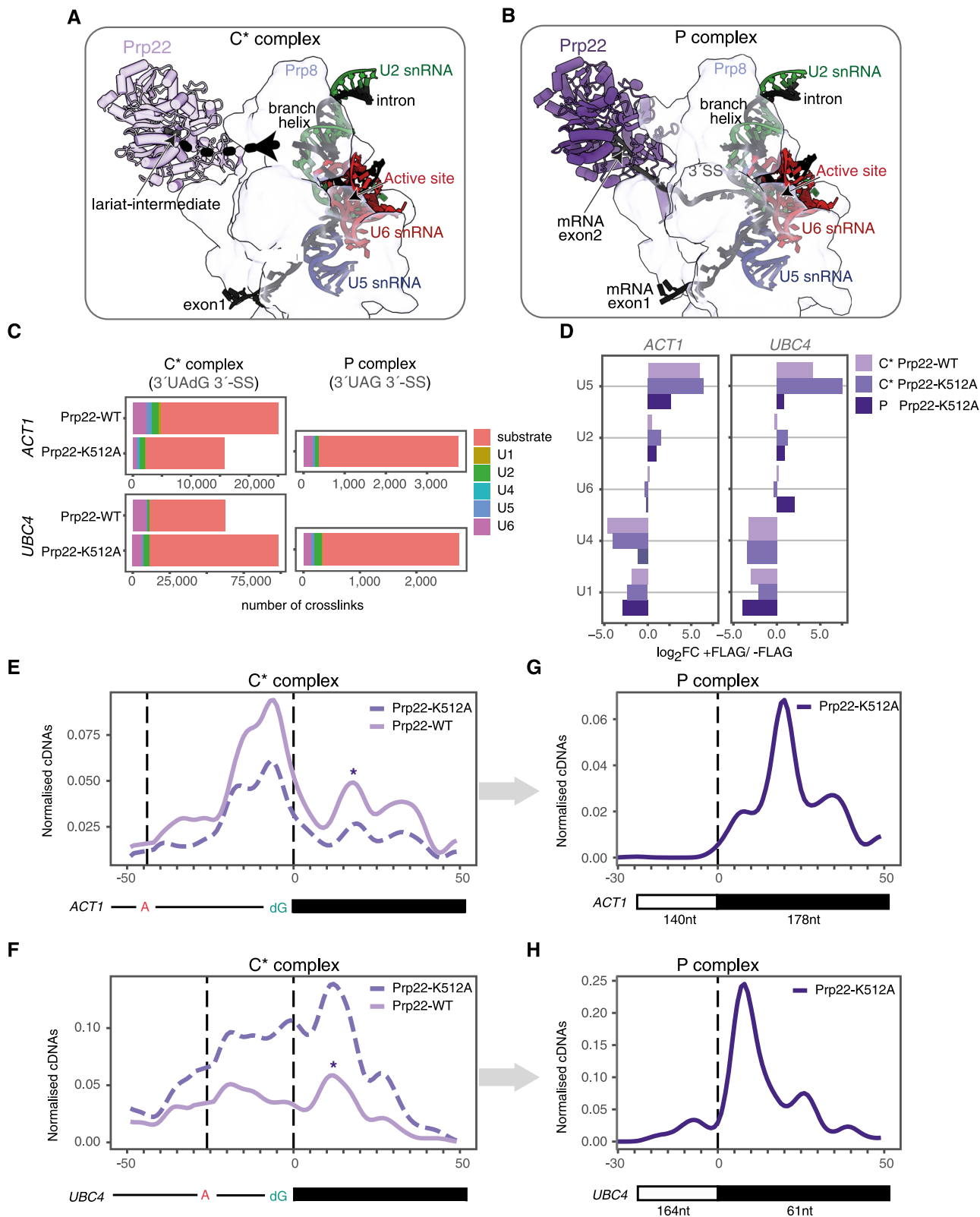
second exon of *ACT1*, which were previously identified by site-specific crosslinking[18]. Importantly, the major psiCLIP peak observed for Prp22 in complex P on the second exon is also observed in complex C* (Fig. 5E–H), indicating that Prp22 binds at similar positions within the second exon both before and after exon ligation. In contrast, psiCLIP suggests that Prp22 no longer binds the intron after exon ligation, in the P complex.

psiCLIP of Prp22 is sensitive to the presence of auxiliary factors. Exon ligation of *ACT1* and *UBC4* requires the exon-ligation factors Slu7 and Prp18 (ref. [50]). It has been suggested that Slu7 binds the spliceosome first, followed by Prp18 and then Prp22, although Slu7 and Prp18 may bind together as a heterodimer[43,50]. In the C* and P complexes, Slu7 and Prp18 bind the Prp8 RH domain to stabilize the exon-ligation

Fig. 5 Prp22 psiCLIP indicates a shift from broad binding in C* complex to narrow binding in P complex. A, B Cryo-EM structures of C* complex and P complex show Prp22 helicase in light and dark purple, respectively (refs. [8,12]; PDB: 5MQ0; PDB: 6EXN). Differential colouring for Prp22 is used simply for representation purposes; the helicase binds the same site in both complexes. The dotted line represents a possible path for unbuilt RNA regions. **C** Total number of crosslinks summed across two UV crosslinked, FLAG-tagged replicates for each condition. Colour denotes the RNA species to which the crosslinks map. **D** Enrichment of snRNAs in FLAG-tagged Prp16 vs. untagged samples, represented as log$_2$ fold change. Note that for this analysis, crosslinks were first normalized to the total number of snRNA crosslinks in each sample. **E–H** Mapping of Prp22 psiCLIP data onto *ACT1* and *UBC4* transcripts. The smoothed lines show the truncation events normalized to endogenous yeast RNA, with the control signal subtracted from the UV signal. Crosslinks are shown after Gaussian smoothing with a ten nucleotide window. Positions on the substrate are given relative to the 3′-SS. P-complex crosslinks are shown with the intron mapping reads removed. **E** Prp22-WT and Prp22-K512A psiCLIP on *ACT1* substrate in C* complex. **F** Prp22-WT and Prp22-K512A psiCLIP on *UBC4* substrate in C* complex. In **E** and **F**, the proposed P-complex-like signal, potentially diagnostic of 3′-SS docking, is indicated with an asterisk (*). **G** Prp22-K512A psiCLIP on *ACT1* substrate in P complex. **H** Prp22-K512A psiCLIP on *UBC4* substrate in P complex.

conformation[10,12] (Fig. 6A) and a loop from Prp18 binds near the 3′-SS to stabilize its docking in the active site[12]. Prp18 was also suggested to promote Prp22 binding to the spliceosome[50], yet the structures of the C* and P complexes did not identify direct contacts between Prp18 and Prp22. We therefore sought to investigate whether Prp18 affects Prp22 binding patterns by performing psiCLIP upon depletion of Prp18 from splicing extracts. We also compared wild-type Prp22 and ATPase-deficient Prp22-K512A to assess how the Prp22 binding pattern depends on its ATPase activity.

As expected, the Prp18-depleted extract accumulates lariat intermediate with both the wild-type and mutant Prp22, indicating a defect in exon ligation, which could be rescued by adding recombinant Prp18 (Supplementary Fig. 7A). Next, psiCLIP was performed using *ACT1* transcripts containing either the canonical 3′-SS or the suboptimal dG modified 3′-SS (3′-UAdG), which in the absence of Prp18 would both be expected to enrich for spliceosomes stalled at the C* stage. Indeed, purified complexes contained lariat intermediate and thus appeared stalled after branching (Supplementary Fig. 7B, C). We observed a striking difference in the substrate binding of wild-type and mutant Prp22 helicases in Prp18-deficient spliceosomes. Wild-type Prp22 displayed very little psiCLIP signal in the Prp18-deficient spliceosomes (Fig. 6B–D and Supplementary Fig. 7D–F), whereas the ATPase-deficient Prp22 helicase showed similar binding in the presence or absence of Prp18 (Fig. 6B–D). In contrast to wild-type Prp22, the ATPase-deficient mutant Prp22-K512A allows more exon ligation at the canonical splice site, as indicated by the presence of more spliced mRNA reads in the psiCLIP samples (Supplementary Fig. 7G). Nonetheless, the majority of the psiCLIP signal derives from lariat intermediates, suggesting that the depletion of Prp18 destabilizes binding of wild-type Prp22 before exon ligation, in agreement with previous biochemical data[50]. Overall, our psiCLIP data indicate that ATP hydrolysis destabilizes Prp22 binding when the optimal composition of the spliceosome is disrupted.

Discussion

Splicing fidelity is safeguarded by eight ATP-dependent helicases, but their precise binding sites and dynamic interactions with snRNAs and pre-mRNAs remain poorly understood. Detailed knowledge of helicase binding profiles is required to understand how they promote conformational changes in the spliceosome and ensure splicing fidelity. We developed psiCLIP to define the RNA contacts of two of the most elusive DEAH-box helicases: Prp16 and Prp22. We establish new data normalization and visualization approaches to present insights from comparative experiments upon mutation of these helicases or depletion of auxiliary factors in various defined spliceosomal states. Our study thus presents a strategy that can complement functional insights from structural studies by combining biochemical isolation of

defined conformations of large RNA–protein complexes with high-resolution interactomics (psiCLIP) and bioinformatics.

psiCLIP provided mechanistic insights into helicase activity by revealing the complete binding profile of the helicases Prp16 and Prp22 on their native substrates, in defined spliceosome complexes. DEAH-box helicases like Prp16 and Prp22 were generally proposed to bind away from their targets and hydrolyse ATP to translocate through RNA structures in the 3′ to 5′ direction[51]. However, cryo-EM structures of the C, C*, and P complexes showed that Prp16 and Prp22 bind at the periphery of the spliceosome at least 15–20 nucleotides away from the RNA structures that they remodel: the branch helix and the mRNA (or lariat intermediate), respectively. Thus both Prp16 and Prp22 appeared unable to remodel these sites by directly translocating through them. In agreement, we found extensive binding of both Prp16 and Prp22 on the substrate downstream of the sites that they remodel, supporting a mechanism of action at a distance rather than processive translocation through the remodelled RNA sites.

Single-molecule FRET studies suggested that Prp16 translocates on the lariat intermediate, while remaining anchored on the C complex (Fig. 3G), acting as a molecular winch that pulls the substrate out of the active site and disrupts the branch helix to allow remodelling[9]. Our psiCLIP data provides further support for this winching model. Instead of binding at a fixed distance from the brA, we found that Prp16 binds over the full available single-stranded RNA region between the brA and 3′-SS, even when this distance is extended with unstructured RNA. Nonetheless, most of the Prp16 psiCLIP signal occurred within ~45 nucleotides downstream of the brA, which correlates with the most common distance between brA and 3′-SS for unstructured yeast introns[52]. This defined region of accessible RNA is likely determined by the binding of Prp16 to spliceosomal components, such as Prp8 (Fig. 3G), as well as by RNA secondary structures formed between the brA and the 3′-SS. Additional single-molecule and chase experiments are necessary to determine if the broad binding we observed reflects movement during translocation or other forms of dynamic contacts, such as binding, release, and re-binding. However, since broad contacts were detected for both the wild-type helicase and the ATPase mutant, which is likely impaired in translocation[44] (Fig. 3E, F and 4E–G), we propose that the broad binding we observe reflects, at least in part, multiple rounds of binding and dissociation by one or more helicase molecules.

Our Prp16 psiCLIP data shows that Prp16 binding can extend into unstructured RNA inserted after the branch helix. By contrast, an ATPase-deficient Prp16 mutant could not appreciably bind into a structured stem–loop region, suggesting that it cannot disrupt strong RNA helices, even though the structured stem–loop allowed splicing, and thus C to C* remodelling. Our data thus provide support for a mechanism of action at a distance, in which Prp16 winches towards the branch helix but remodels the spliceosomal active site without translocating through the branch helix.
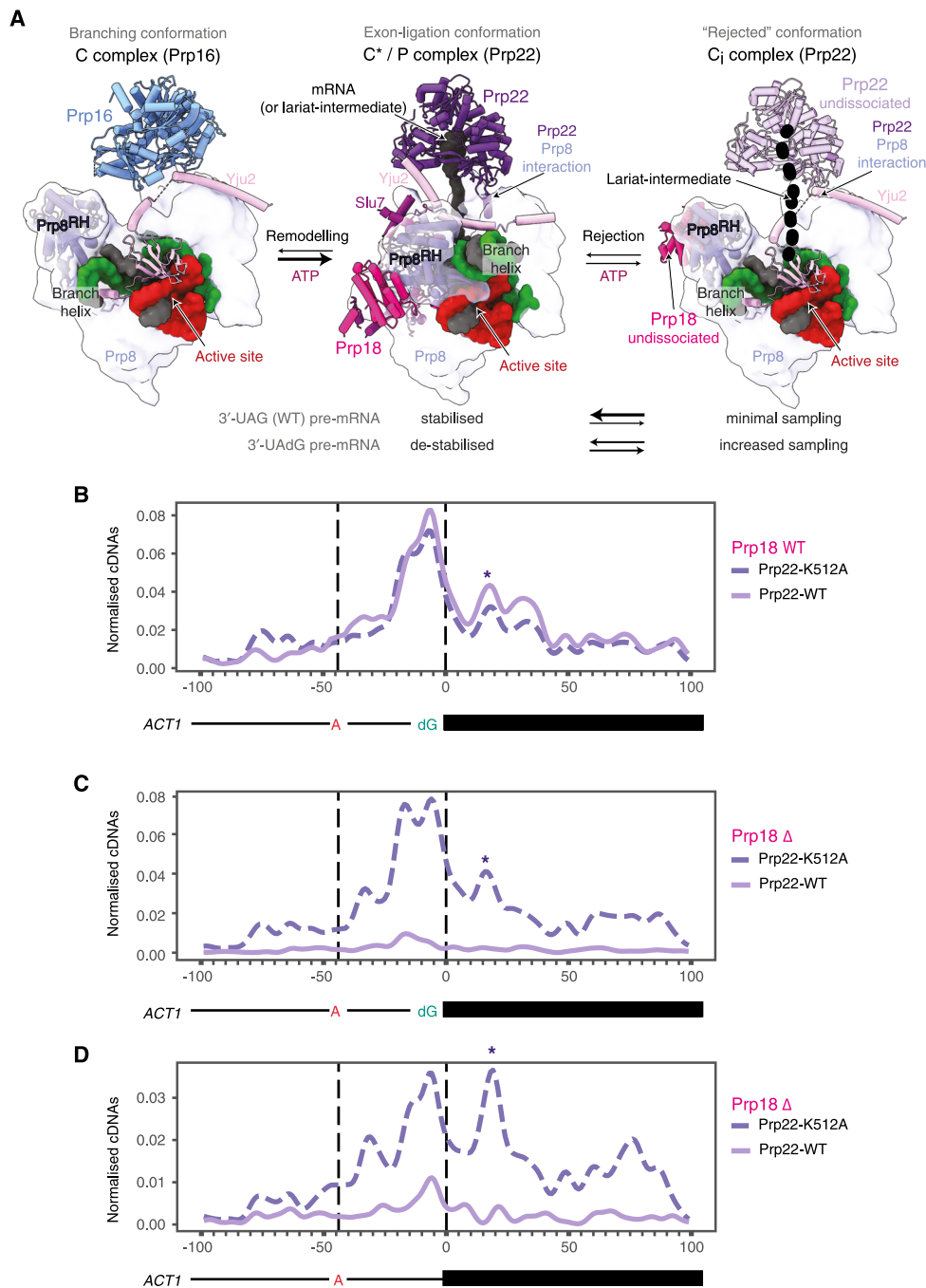
Fig. 6 Prp18 depletion destabilizes wild-type Prp22 binding on the pre-mRNA substrate, but not the binding of the ATPase-deficient Prp22 mutant.
A Binding of Prp22 in the exon-ligation conformation in C* (ref. [10]; PDB: 5MQ0) and P (ref. [12]; PDB: 6EXN) complexes is stabilized by Prp18 and Slu7 (middle panel). For comparison, binding of Prp16 is shown in the C complex (left panel). During rejection by Prp22, the spliceosome may revert to a C-like conformation (C$_i$, ref. [43]; PDB: 7B9V), in which both Prp18 and Prp22 may remain loosely bound (right panel). **B** Prp22 binding to spliceosomes assembled on the suboptimal 3'-UAdG substrate in mock-depleted extracts. **C** psiCLIP of Prp22 in spliceosomes assembled on the 3'-UAdG substrate following depletion of Prp18 (Prp18Δ). **D** psiCLIP of Prp22 in spliceosomes assembled on the wild-type ACT1 substrate following depletion of Prp18. Both wild-type and ATPase mutant Prp22 are shown for all experiments. The proposed P-complex-like signal, potentially diagnostic of 3'-SS docking, is indicated with an asterisk (*). In all panels, crosslinks are shown normalized to endogenous yeast RNA, with reads from untagged Prp22 subtracted from tagged Prp22 reads, and after applying Gaussian smoothing with a ten nucleotide window.

Similarly to Prp16, psiCLIP revealed broad binding of Prp22 downstream of the brA before exon ligation, independently of ATPase activity. The high reproducibility of binding profiles observed at specific substrate positions in several replicates and across stalling conditions (Supplementary Fig. 4B–G) likely represents an average of interactions in defined complexes, rather than resulting from heterogeneous binding or processive translocation events across multiple complexes. Thus, we interpret the broad binding as evidence that the helicases can bind at multiple defined regions along the substrate. In the C* complex, we detected Prp22 binding on the intron before the 3'-SS and on the second exon downstream of the 3'-SS. The strong signal observed for the
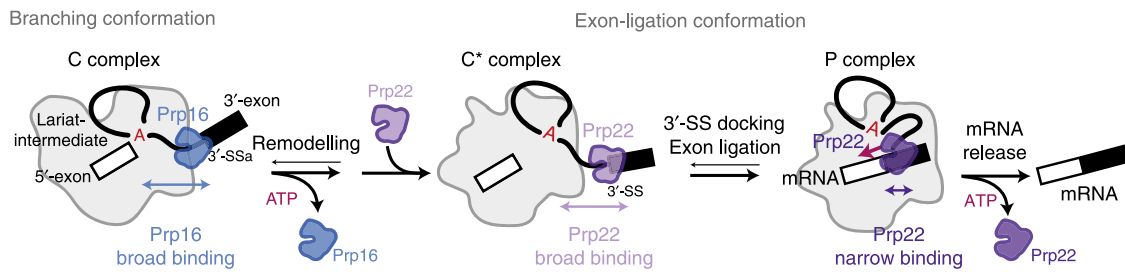
Fig. 7 Model for substrate binding by spliceosomal helicases during the catalytic stage of pre-mRNA splicing. After branching, Prp16 binds the lariat intermediate broadly downstream of the brA, though the main binding peaks are observed between the brA and the 3′-SS. Prp16 dissociates during remodelling to the C* conformation, in which it is replaced by Prp22 before docking of the 3′-SS. In complex C*, before exon ligation, Prp22 binds the lariat intermediate broadly on the 3′-exon and the intron. Such broad binding may facilitate 3′-SS docking to allow exon ligation. After exon ligation, in complex P, Prp22 engages the mRNA in a narrow binding window on the 3′-exon, from which it pulls on the RNA substrate to release the mRNA and dissociates following mRNA release.

wild-type Prp22 protein, as well as the broad profile seen for the ATPase-deficient K512A mutant, which is unable to translocate, suggest multiple rounds of stochastic binding and dissociation, rather than a single ATP-dependent translocation and dissociation event. Notably, the psiCLIP-binding profile is maintained upon the inhibition of Prp22 ATPase activity, consistent with an ATP-independent role for Prp22 in exon ligation[15].

The broad accessibility for Prp22 binding in the intron on the 3′-UAdG substrate indicates that the 3′-SS is not stably docked in the active site in a significant population of C* spliceosomes assembled on this substrate. Indeed, the 3′-UAdG substrate does not produce mRNA and no stably docked 3′-SS was detected in the cryo-EM structure of the C* complex[10]. The extended intron binding profile in the C* complex suggests that in the absence of a stably docked 3′-SS Prp22 can translocate along the lariat-intermediate upstream of the 3′-SS, or bind and rebind in this region, potentially to reject these spliceosomes, thus ensuring that exon ligation occurs only at the correctly docked 3′-SS. Supporting this idea, recent cryo-EM studies indicate that a population of spliceosomes assembled on the 3′-UAdG substrate revert to the branching conformation, potentially as a result of rejection of the C* complex by Prp22 (refs. [10,43]).

psiCLIP also detected substrate rearrangements resulting from substrate docking and catalysis. In complex P, Prp22 binding shifted to a narrower region 10–25 nucleotides downstream of the 3′-SS, and we detected the same major peak in the 3′-exon in both complex C* and complex P (Fig. 5G, H). This exonic binding position is likely associated with a stably docked 3′-SS and consequently this 3′-exon peak may be diagnostic for complexes in the exon-ligation conformation. Given the 3′ to 5′ direction of helicase movement on RNA, the observed shift in the Prp22-binding profile from major peaks in the intron in C* to major peaks in the 3′-exon in P complex is inconsistent with a simple mechanism of processive 3′ to 5′ translocation during Prp22 action. Instead, we propose that Prp22 may initially engage the substrate on the 3′-exon before exon ligation. After exon ligation, Prp22 would thread the substrate from the 3′-exon to release the mRNA from the spliceosome, potentially through repeated binding and dissociation (Fig. 7 and Supplementary Fig. 8A). Before exon ligation, Prp22 may promote, and proofread, 3′-SS sampling and docking by dynamic engagement with the lariat intermediate, as evidenced by binding in the intron upstream of the 3′-SS (Figs. 5E, F and 7). Consistent with this idea, Prp22 was shown to promote sampling of alternative 3′SS (ref. [9]).

The psiCLIP-binding profile may thus reflect the proofreading activity of Prp22 and could be influenced by factors that stabilize the P complex. Indeed, we found that Prp22 binding to the

substrate depended on the exon-ligation factor Prp18, which promotes 3′-SS docking. Prp18 indirectly stabilizes Prp22 binding by locking Prp8 in the exon-ligation conformation in the C* and P complexes (Fig. 6A). As expected, the Prp22 ATPase-deficient K512A mutant remained bound to both the canonical and sub-optimal 3′-UAdG substrate upon Prp18 depletion, consistent with the ATPase-deficient mutant being intrinsically less prone to dissociation, even when the exon-ligation conformation is destabilized in the absence of Prp18 (Supplementary Fig. 6G and Fig. 6A). Strikingly, binding by the wild-type Prp22 was significantly reduced in the absence of Prp18. In the absence of Prp18, Prp22 may dissociate from the spliceosome and may not rebind. Depletion of Prp18 may also indirectly increase the ATPase activity of Prp22, promoting dissociation and reducing stable binding to the substrate. Indeed, structural studies have shown that the absence of Prp18 destabilizes the C* and P conformation[43], and would likely weaken Prp22 binding onto Prp8, which may be necessary for stable association with the substrate. Similar to Prp22's proofreading activity for correct 3′-SS selection[14], wild-type Prp22 may not bind stably in the absence of Prp18 and could therefore prevent potentially erroneous exon ligation by incompletely assembled spliceosomes. Consistent with this idea, in the presence of Prp22-K512A we detected exon–exon spanning reads corresponding to mRNA (Suplementary Fig. 7G). Thus, when Prp22 ATPase activity is compromised splicing can proceed in the absence of Prp18, suggesting that the ATPase activity of Prp22 may reject spliceosomes lacking Prp18.

Our data imply that productive Prp22 association with the substrate is dependent more broadly on the stability of the exon-ligation conformation (Supplementary Fig. 8). In this proofreading model, Prp22 monitors the thermodynamic stability of the spliceosome by discriminating against complexes that cannot complete exon ligation before Prp22 action, whether due to a suboptimal, non-canonical, 3′-SS or a suboptimal spliceosome, resulting for example from poor binding of auxiliary factors. For such suboptimal spliceosomes the exon-ligation conformation would be destabilized. Before exon ligation can occur, Prp22 bound initially in the 3′-exon may either translocate from this position or dissociate and rebind upstream of the 3′-SS to reject lariat intermediates that cannot stably reach the exon-ligation conformation (Fig. 5E, F and Supplementary Fig. 8B). Dissociation of Prp22 would reject these complexes and may cause the spliceosome to revert from the C* complex to an intermediate C-like conformation (C_i). This C_i conformation may not allow stable re-binding by Prp22 (Fig. 6A), thus preventing progress to the P complex and blocking exon ligation by suboptimal spliceosomes. Indeed, the C_i conformation, in which Prp18 remains loosely bound, was

observed by cryo-EM for spliceosomes assembled on the sub-optimal 3′-UAdG substrate[43]. Depletion of Prp18 may further destabilize the C* and P complexes, and push spliceosomes towards such a C-like conformation. Thus, psiCLIP complements the interpretation of structural studies and provides mechanistic insights into the proofreading activities of spliceosomal ATPases.

Taken together, our study raises new questions about the activity of Prp22 and Prp16 ATPases and, more generally, provides valuable information about active and passive helicase activities. Further biochemical work will be required to fully elucidate the mechanisms behind the phenomena uncovered by psiCLIP. While psiCLIP is well suited for comparisons between conditions, the absolute crosslinking levels in psiCLIP are affected by the variable crosslinking efficiencies of nucleotides and amino acids, which can lead to biases at the sequence level, as in all techniques utilizing UV crosslinking. It is known that uridine is the most favourable crosslinker, followed by guanosine, cytosine, and adenosine, respectively[53]. Thus, comparisons of psiCLIP data between transcripts and/or proteins should take into account such sequence-specific biases. Looking to the future, we envision that psiCLIP can be extended to study helicases and other RBPs in additional biochemically defined RNP complexes, such as the RBP–rRNA interactions during ribosome biogenesis. For example, time-resolved psiCLIP could be used to track the binding profile of a helicase in motion during RNP remodelling. Thus, psiCLIP provides a versatile method to investigate fundamental mechanisms of RNP dynamics and remodelling.

Methods

Yeast strains. All strains used in this study were derived from BCY123 (*MATa, can1, ade2, trp1, Ura3-52, his3, leu2-3, 112, pep4::his+, prb1::leu2+,bar1::HisG+, lys2::pGAL1/10-GAL4+*; Supplementary Table 1). In brief, sequences coding for protein tags and the respective resistance cassette were amplified from plasmids coding for the protein tag and resistance cassette with ~60–90 nucleotides homology to the end of the target gene and the beginning of its 3′ UTR (Supplementary Tables 1 and 3). The linear PCR product was transformed into BCY123 and the cells were plated on selective media containing either 100 μg/mL ClonNAT, 300 μg/mL hygromycin, or 250 μg/mL G418. Positive clones were verified by Sanger sequencing and expression of the tagged protein was analysed by western blot using monoclonal ANTI-FLAG® M2-Peroxidase antibody (Sigma Aldrich) with 1:3000 dilution, c-Myc antibody (9E10) HRP (Santa Cruz Biotechnology) with 1:1000 dilution, or HA-probe antibody (F-7) HRP (Santa Cruz Biotechnology) with 1:1000 dilution to detect the respective tag. Strains containing several C-terminal tags underwent the cycle several times.

For recombinant protein expression, plasmids based on pRS424 and pRS426 vectors carrying *TRP1* and *URA3* markers, respectively, were transformed into BCY123 and grown on YM4 selective media (Supplementary Table 1). Both Prp16 and Prp22 constructs were encoded on pRS424 under the GAL-GAPDH hybrid promoter.

Preparation of splicing substrates. In vitro transcriptions were performed from pUC-based vectors with the desired sequence following the T7 promoter (Supplementary Table 2) and containing a 3× MS2 stem–loop sequence either at the 3′ or 5′ end[54]. Artificial extensions of the *ACT1* substrate were generated using oligonucleotides with a random sequence that maintained the same GC content as that of the original sequence between brA and 3′-SS. These were cloned into a restriction site following the brA position (Supplementary Table 3). Single clones were selected with either a single 20 nucleotide insertion (5′-CGA TTT TAT TTA TTT GAT CT-3′), a 40 nucleotide tandem insertion (5′-CGA AAA TCA AGA TAA ATA ATC GAA AAT CAA GAT AAA TAA T-3′), or a 40 nucleotide head-to-head insertion to generate an RNA stem–loop (5′-CGA TTA TTT ATC TTG ATT TTC GAA AAT CAA GAT AAA TAA T-3′). Long RNA pieces were generated by run-off transcription. Some splicing substrates (Supplementary Table 2), including substrates containing a 2′-deoxynucleotide, were generated by ligation from a long 5′ piece ending before the 3′-SS and short 3′ oligonucleotides. For the respective *UBC4* substrates, oligonucleotides carrying a 3′-Cy5 fluorophore (purchased from Sigma) were added by ligation (Supplementary Table 2). For *ACT1*, after ligation of an oligonucleotide and an RNA piece generated by run-off transcription, 3′ end labelling was performed by enzymatic ligation of pCp-Cy5 (Supplementary Table 2). To generate precise 3′ or 5′ ends for ligation, transcripts included a hepatitis delta virus or hammerhead ribozyme sequence at the respective end. To join RNA pieces by splint-mediated ligation, the 5′-ends were first phosphorylated with T4 PNK (New England Biolabs). RNAs were annealed to bridge DNA oligonucleotides complementary to the junction by slowly decreasing the temperature from 80 to 25 °C before ligation with T4 DNA ligase (Supplementary Table 3)[46]. Ligated transcripts were gel purified after each ligation step.

Expression and purification of recombinant Prp16 and Prp22. BCY123 cells were co-transfected with expression vectors pRS426 and pRS424, the latter coding for the recombinant protein (Supplementary Table 1). Positive transformants were grown in 24 L YM4 selective media supplemented with 1% raffinose at 30 °C. Protein expression was induced with 2% final concentration of galactose at OD600 1.0. After 12–16 h further growth at 30 °C, cells were harvested, resuspended in one volume 2× lysis buffer (2 M NaCl, 100 mM Tris-Cl pH 9.0, 2 mM imidazole, 20 mM β-mercaptoethanol (β-ME), 0.2% IGEPAL CA-630, 4 mM CaCl₂, 2 mM magnesium acetate, and EDTA-free protease inhibitor cocktail (Roche)), and frozen in liquid nitrogen in droplet form. Cells were disrupted in a 6870 Freezer/Mill (SPEX SamplePrep). After thawing, the pH of the extract was raised to 8.5 by addition of 1 M Tris base. Cell debris were removed by ultracentrifugation at 195,000 × g for 90 min. The supernatant was incubated with 2 mL Calmodulin-sepharose beads (home made) for 12–16 h at 4 °C. Beads were washed with 5 × 50 mL CAL wash buffer (500 mM NaCl, 20 mM Tris-Cl pH 8.0, 2 mM CaCl₂, 1 mM magnesium acetate, 1 mM imidazole, and 10 mM β-ME) and the proteins eluted in 10 × 5 mL CAL elution buffer (500 mM NaCl, 20 mM Tris-Cl pH 8.0, 2 mM EGTA, 1 mM magnesium acetate, 1 mM imidazole, and 10 mM β-ME). Protein-containing fractions were pooled and dialyzed against Ni-NTA binding buffer (1 M NaCl, 20 mM Tris-Cl pH 8.5, 5 mM imidazole, and 10 mM β-ME) for 4 h at 4 °C. After 14 h binding to 4 mL Ni-NTA agarose beads (Qiagen), the beads were first washed with 15 mL Ni-NTA binding buffer and then with 15 mL Ni-NTA wash buffer (1 M NaCl, 20 mM Tris-Cl pH 8.5, 15 mM imidazole, and 10 mM β-ME). The protein was eluted in about six 4 mL fractions with Ni-NTA elution buffer (1 M NaCl, 20 mM Tris-Cl pH 8.5, 250 mM imidazole, and 10 mM β-ME). Helicases were dialyzed against the respective helicase storage buffer for 4 h at 4 °C and stored at −80 °C until further use (20 mM HEPES pH 7.9, 0.2 mM EDTA, 0.5 mM DTT, 20% glycerol with 250 mM KCl for Prp16 and 300 mM KCl for Prp22).

Expression and purification of recombinant Prp18. The coding sequence for Prp18 was cloned into pET14. The protein was expressed as N-terminal His6-thrombin-Prp18 in *Escherichia coli* BL21 (DE3)RIL cells. The protein was first purified on Ni-NTA-agarose beads and peak fractions dialyzed against buffer containing 10 mM potassium phosphate and 50 mM KCl pH 7.4. The protein was further purified on a hydroxyapatite column, eluted with 0.5–3.0% ammonium sulfate, and dialyzed into a suitable buffer for splicing reactions (20 mM HEPES pH 7.9, 0.2 mM EDTA, 0.5 mM DTT, 20% glycerol, and 250 mM KCl).

Splicing extract preparation and in vitro splicing. Yeast splicing extract was prepared using the liquid nitrogen method[55]. In vitro splicing reactions were performed with 2.5 nM splicing substrate in 40% splicing extract, 2 mM ATP, 2.5 mM MgCl₂, and 60 mM potassium phosphate pH 6.5 (ref. [56]). The resulting RNA species were phenol extracted and analysed on 5% (for *ACT1*) or 10% (for *UBC4*) denaturing polyacrylamide gels. Quantification of gel bands was done using ImageJ (ref. [57]).

Protein depletion from splicing extract. Prp18-3× HA was depleted from splicing extract by increasing the KCl concentration to 750 mM and incubating twice with 1/10 volume anti-HA magnetic beads at 4 °C for several hours, followed by dialysis against 20 mM HEPES pH 7.9, 0.2 mM EDTA, 0.5 mM DTT, 20% glycerol, and 200 mM KCl for 16 h. To ensure complete depletion of Prp18 while preventing co-depletion of Slu7, western blots against Prp18-3× HA and Slu7-9xcmyc were performed. To rescue the splicing defect caused by Prp18 depletion, rPrp18 was added to the splicing reaction at a final concentration of 165 ng/mL (ref. [58]).

Spliceosome assembly and purification. Complexes were assembled in a 1.5 mL splicing reaction on the indicated pre-mRNA substrate, which was pre-bound to a 1.25 fold excess of MS2-MBP fusion protein[54]. For reactions containing recombinantly expressed helicase mutants, the splicing extract was firstly incubated with 15 ng/mL final concentration of dominant-negative mutant protein (Prp16-G378A or Prp22-K512A). Reactions were incubated at 23 °C for 30 min (*UBC4* transcripts) or 60 min (*ACT1* transcripts). For C and C* complex purification, the reaction mixture was adjusted to 2 mM glucose and incubation prolonged for 5 min. For the C* complex, beads were incubated with 2 mM ATP/2 mM MgCl₂ for 30 min at room temperature before the wash steps[10]. For P-complex purification, to remove spliceosomes before the P-complex stage, reactions were incubated with 5 μM of a DNA oligonucleotide complementary to the 3′-exon for an additional 20 min (ref. [12]; Supplementary Table 3).

For all spliceosome preparations, the reaction mixture was centrifuged through a 40% glycerol cushion in buffer A (20 mM HEPES, pH 7.9, 75 mM KCl, and 0.25 mM EDTA). The cushion was collected and applied to amylose resin. After 15 h of incubation at 4 °C, the resin was washed three times with 1 mL buffer A and eluted for 20 min in 200 μL buffer A containing 5% glycerol, 0.01% NP-40, and 12 mM maltose.

Spliceosome crosslinking, immunoprecipitation, and cDNA library preparation. For each helicase experiment, at least two replicates were produced from independently assembled and purified spliceosomes, although not necessarily from independent extract preparations. In addition, control samples containing no tag on the protein and non-irradiated samples were prepared. Crosslinking and immunoprecipitation were adapted from the original iCLIP protocol[34] with the following modifications. The 200 μL spliceosome eluate was irradiated by UV light using a Stratalinker 2400 at 254 nm. Spliceosomes with the target protein SmB were irradiated with $312 \times 100 \, \mu J/cm^2$ and spliceosomes with the target protein Prp16 or Prp22 with $312 \times 100 \, \mu J/cm^2$ (assembled on *UBC4*) or $625 \times 100 \, \mu J/cm^2$ (assembled on *ACT1*). A total of 72 mg urea was added and dissolved by shaking to gain a 6 M solution to denature the complex. The protein of interest was captured by magnetic anti-FLAG beads. The beads were incubated in 20% S100, 8 mM EDTA, and 40 μL debranching buffer (20 mM HEPES 7.9, 20% glycerol, 100 mM KCl, and 0.5 mM DTT) for 15 min to debranch intron lariats[59]. A total of 0.5–5 units (depending on complex and splicing transcript) of RNase I was added over 3 min at 37 °C for fragmentation. Dephosphorylation, adapter ligation, radioactive labelling, isolation of the RNA–protein complex, and cDNA library preparation were performed as in iCLIP. cDNA libraries were sequenced at the Illumina HiSeq 2500 platform with cycle lengths between 100 and 250 nucleotides in single-read mode using HiSeq control software v2.2.68. An overview of all psiCLIP samples with respective adapter and RT primer sequences is given in Supplementary data 1.

Data processing. A custom psiCLIP software pipeline was developed in Snakemake[60] and is available from github.com/luslab/psiclip. The code is easily extended to other experimental scenarios and substrates. Reads were demultiplexed using iCount demultiplex and trimmed for quality using Trim Galore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Quality of sequencing data was assessed with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Trimmed reads were mapped to a custom transcriptome index consisting of *S. cerevisiae* U1, U2, U4, U5, and U6 snRNAs alongside the pre-mRNA substrate sequence, using STAR aligner[61]. Key parameters were: `−alignEndsType EndToEnd` to prevent soft-clipping of reads which would obscure the truncation site, `−outFilterMismatchNmax 2` to allow a maximum of two mismatches and `−seedSearchStartLmax 16`, which means the read is split into more seeds resulting in a potentially more sensitive search given our short reads. Subsequently, Sambamba was used to retrieve reads mapping only in the forward orientation to the custom transcriptome[62], as psiCLIP is stranded. Finally, the cDNA start position −1 is taken as the crosslink site and bed files are generated, using Bedtools and R[63]. Reads that did not map to the custom transcriptome index were then mapped to the yeast genome (SacCer3) and similarly processed to generate crosslinks. We used the proportion of yeast genome mapping RNA in the library as internal normalization, much like a spike-in[64]. Further plotting and analysis is performed in R using dplyr (https://cran.r-project.org/web/packages/dplyr/index.html), ggplot2 (ref. [65]), and smoother (https://CRAN.R-project.org/package=smoother) packages. Data for Prp22 and Prp16 are presented as Gaussian smoothed curves with a window size of ten nucleotides. In the main figures, the sum of all replicates is shown, where available, with replicates shown separately in the Supplementary Material.

Data availability
psiCLIP data are available to download from ArrayExpress at E-MTAB-8895. Structural biology data that was used for comparison was taken from the following publicly available sources: https://doi.org/10.2210/pdb5ZWN/pdb, https://doi.org/10.2210/pdb5LJ3/pdb, https://doi.org/10.2210/pdb5MQ0/pdb, https://doi.org/10.2210/pdb6EXN/pdb, and 7B9V. The data supporting the findings of this study are available from the corresponding author upon reasonable request. Source data are provided with this paper.

Code availability
Scripts for data pre-processing and downstream processing are available from github.com/luslab/psiclip (ref. [66]), https://doi.org/10.5281/zenodo.4439637.

References
1. Will, C. L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**, a003707–a003707 (2011).
2. Lee, S. C.-W. & Abdel-Wahab, O. Therapeutic targeting of splicing in cancer. *Nat. Med.* **22**, 976–986 (2016).
3. Semlow, D. R. & Staley, J. P. Staying on message: ensuring fidelity in pre-mRNA splicing. *Trends Biochem. Sci.* **37**, 263–273 (2012).
4. Giaever, G. et al. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
5. Cordin, O. & Beggs, J. D. RNA helicases in splicing. *RNA Biol.* **10**, 83–95 (2012).
6. Tanaka, N. & Schwer, B. Characterization of the NTPase, RNA-binding, and RNA helicase activities of the DEAH-Box splicing factor Prp22 †. *Biochemistry* **44**, 9795–9803 (2005).
7. He, Y., Staley, J. P., Andersen, G. R. & Nielsen, K. H. Structure of the DEAH/RHA ATPase Prp43p bound to RNA implicates a pair of hairpins and motif Va in translocation along RNA. *RNA* **23**, 1110–1124 (2017).
8. Galej, W. P. et al. Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537**, 197–201 (2016).
9. Semlow, D. R., Blanco, M. R., Walter, N. G. & Staley, J. P. Spliceosomal DEAH-Box ATPases remodel pre-mRNA to activate alternative splice sites. *Cell* **164**, 985–998 (2016).
10. Fica, S. M. et al. Structure of a spliceosome remodelled for exon ligation. *Nature* **542**, 377–380 (2017).
11. Wilkinson, M. E., Charenton, C. & Nagai, K. RNA splicing by the spliceosome. *Annu. Rev. Biochem.* **89**, 1–30 (2019).
12. Wilkinson, M. E. et al. Postcatalytic spliceosome structure reveals mechanism of 3′-splice site selection. *Science* **358**, 1283–1288 (2017).
13. Ohrt, T. et al. Molecular dissection of step 2 catalysis of yeast pre-mRNA splicing investigated in a purified system. *RNA* **19**, 902–915 (2013).
14. Mayas, R. M., Maita, H. & Staley, J. P. Exon ligation is proofread by the DExD/H-box ATPase Prp22p. *Nat. Struct. Mol. Biol.* **13**, 482–490 (2006).
15. Schwer, B. & Gross, C. H. Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *EMBO J.* **17**, 2086–2094 (1998).
16. Company, M., Arenas, J. & Abelson, J. Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. *Nature* **349**, 487–493 (1991).
17. Wagner, J. D. O. The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes. *EMBO J.* **17**, 2926–2937 (1998).
18. Schwer, B. A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. *Mol. Cell* **30**, 743–754 (2008).
19. Plaschka, C., Newman, A. J. & Nagai, K. Structural basis of nuclear pre-mRNA splicing: lessons from yeast. *Cold Spring Harb. Perspect. Biol.* **11**, a032391 (2019).
20. Tseng, C. K. & Cheng, S. C. Both catalytic steps of nuclear pre-mRNA splicing are reversible. *Science* **320**, 1782–1784 (2008).
21. Koodathingal, P. & Staley, J. P. Splicing fidelity: DEAD/H-box ATPases as molecular clocks. *RNA Biol.* **10**, 1073–1079 (2013).
22. Mayas, R. M., Maita, H., Semlow, D. R. & Staley, J. P. Spliceosome discards intermediates via the DEAH box ATPase Prp43p. *Proc. Natl Acad. Sci. USA* **107**, 10020–10025 (2010).
23. Query, C. C. & Konarska, M. M. Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants. *Mol. Cell* **14**, 343–354 (2004).
24. McPheeters, D. S., Schwer, B. & Muhlenkamp, P. Interaction of the yeast DExH-box RNA helicase prp22p with the 3′ splice site during the second step of nuclear pre-mRNA splicing. *Nucleic Acids Res.* **28**, 1313–1321 (2000).
25. McPheeters, D. S. & Muhlenkamp, P. Spatial organization of protein-RNA interactions in the branch site-3′ splice site region during pre-mRNA splicing in yeast. *Mol. Cell Biol.* **23**, 4174–4186 (2003).
26. Briese, M. et al. A systems view of spliceosomal assembly and branchpoints with iCLIP. *Nat. Struct. Mol. Biol.* **26**, 930–940 (2019).
27. Burke, J. E. et al. Spliceosome profiling visualizes operations of a dynamic RNP at nucleotide resolution. *Cell* **173**, 1014–1030.e17 (2018).
28. Chen, W. et al. Transcriptome-wide interrogation of the functional intronome by spliceosome profiling. *Cell* **173**, 1031–1044.e13 (2018).
29. Ule, J. & Blencowe, B. J. Alternative splicing regulatory networks: functions, mechanisms, and evolution. *Mol. Cell* **76**, 329–345 (2019).
30. Lee, F. C. Y. & Ule, J. Advances in CLIP technologies for studies of protein-RNA interactions. *Mol. Cell* **69**, 354–369 (2018).
31. Sutandy, F. X. R. et al. In vitro iCLIP-based modeling uncovers how the splicing factor U2AF2 relies on regulation by cofactors. *Genome Res.* **28**, 699–713 (2018).
32. Wilkinson, M. E., Lin, P.-C., Plaschka, C. & Nagai, K. Cryo-EM studies of pre-mRNA splicing: from sample preparation to model visualization. *Annu. Rev. Biophys.* **47**, 175–199 (2018).
33. Hahn, D., Kudla, G., Tollervey, D. & Beggs, J. D. Brr2p-mediated conformational rearrangements in the spliceosome during activation and substrate repositioning. *Genes Dev.* **26**, 2408–2421 (2012).
34. Huppertz, I. et al. iCLIP: protein–RNA interactions at nucleotide resolution. *Methods* **65**, 274–287 (2014).
35. Haberman, N. et al. Insights into the design and interpretation of iCLIP experiments. *Genome Biol.* **18**, 7 (2017).
36. Guthrie, C. & Patterson, B. Spliceosomal SNRNAS. *Annu. Rev. Genet.* **22**, 387–419 (1988).
37. Price, S. R., Evans, P. R. & Nagai, K. Crystal structure of the spliceosomal U2B''-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* **394**, 645–650 (1998).

38. Leung, A. K. W., Nagai, K. & Li, J. Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. *Nature* **473**, 536–539 (2011).

39. Kondo, Y., Oubridge, C., Roon, A.-M. M. van & Nagai, K. Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5′ splice site recognition. *eLife* **4**, e04986 (2015).

40. Bai, R., Wan, R., Yan, C., Lei, J. & Shi, Y. Structures of the fully assembled *Saccharomyces cerevisiae* spliceosome before activation. *Science* **360**, 1423–1429 (2018).

41. Schwer, B. & Guthrie, C. A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. *EMBO J.* **11**, 5033–5039 (1992).

42. Tseng, C.-K., Chung, C.-S., Chen, H.-C. & Cheng, S.-C. A central role of Cwc25 in spliceosome dynamics during the catalytic phase of pre-mRNA splicing. *RNA* **23**, 546–556 (2017).

43. Wilkinson, M. E., Fica, S. M., Galej, W. P. & Nagai, K. Structural basis for conformational equilibrium of the catalytic spliceosome. Preprint at *bioRxiv* https://doi.org/10.1101/2020.09.22.306647 (2020).

44. Schneider, S., Hotz, H.-R. & Schwer, B. Characterization of dominant-negative mutants of the DEAH-box splicing factors Prp22 and Prp16. *J. Biol. Chem.* **277**, 15452–15458 (2002).

45. Rogelj, B. et al. Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. *Sci. Rep.* **2**, 603 (2012).

46. Moore, M. J. & Sharp, P. A. Site-specific modification of pre-mRNA: the 2′-hydroxyl groups at the splice sites. *Science* **256**, 992–997 (1992).

47. Teigelkamp, S., Newman, A. J. & Beggs, J. D. Extensive interactions of PRP8 protein with the 5′ and 3′ splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. *EMBO J.* **14**, 2602–2612 (1995).

48. Chung, C.-S. et al. Dynamic protein-RNA interactions in mediating splicing catalysis. *Nucleic Acids Res.* **47**, 899–910 (2019).

49. Fica, S. M. & Nagai, K. Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine. *Nat. Struct. Mol. Biol.* **24**, 791–799 (2017).

50. James, S.-A., Turner, W. & Schwer, B. How Slu7 and Prp18 cooperate in the second step of yeast pre-mRNA splicing. *RNA* **8**, 1068–1077 (2002).

51. Pyle, A. M. Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu. Rev. Biophys.* **37**, 317–336 (2008).

52. Meyer, M., Plass, M., Pérez-Valle, J., Eyras, E. & Vilardell, J. Deciphering 3′ss selection in the yeast genome reveals an RNA thermosensor that mediates alternative splicing. *Mol. Cell* **43**, 1033–1039 (2011).

53. Sugimoto, Y. et al. Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol.* **13**, R67 (2012).

54. Zhou, Z., Sim, J., Griffith, J. & Reed, R. Purification and electron microscopic visualization of functional human spliceosomes. *Proc. Natl Acad. Sci. USA* **99**, 12203–12207 (2002).

55. Umen, J. G. & Guthrie, C. Prp16p, Slu7p, and Prp8p interact with the 3′ splice site in two distinct stages during the second catalytic step of pre-mRNA splicing. *RNA* **1**, 584–597 (1995).

56. Lin, R. J., Newman, A. J., Cheng, S. C. & Abelson, J. Yeast mRNA splicing in vitro. *J. Biol. Chem.* **260**, 14780–14792 (1985).

57. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).

58. Horowitz, D. S. & Krainer, A. R. A human protein required for the second step of pre-mRNA splicing is functionally related to a yeast splicing factor. *Genes Dev.* **11**, 139–151 (1997).

59. Ruskin, B. & Green, M. An RNA processing activity that debranches RNA lariats. *Science* **229**, 135–140 (1985).

60. Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522 (2012).

61. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

62. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).

63. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

64. Chen, K. et al. The overlooked fact: fundamental need for spike-in control for virtually all genome-wide analyses. *Mol. Cell Biol.* **36**, 662–667 (2016).

65. Wickham, H. ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* **3**, 180–185 (2011).

66. Capitanchik, C. psiCLIP reveals dynamic RNA binding by DEAH-box helicases before and after exon ligation. https://github.com/luslab/psiclip, https://doi.org/10.5281/zenodo.4439637 (2020).

Author contributions

L.M.S. established the method psiCLIP, with advice from M.H. and J.U. L.M.S. and C.C. designed psiCLIP experiments for SmB, Prp16, and Prp22. L.M.S. generated yeast strains, purified the helicase mutant proteins, and performed the psiCLIP experiments. L.M.S. and A.J.N. prepared the splicing substrates and performed the biochemical experiments. A.J.N. cloned, expressed, and purified Prp18. S.M.F., A.J.N., and C.O. provided experimental advice and contributed to data interpretation. C.C. performed all computational analysis. C.N. supported the generation of yeast strains and optimized debranching steps. The manuscript was written by L.M.S. and C.C., with input from all authors. K.N., J.U., and N.M.L. initiated and coordinated the project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-21745-9.

Correspondence and requests for materials should be addressed to J.U.

Peer review information *Nature Communications* thanks Jean Beggs, David Tollervey, and other, anonymous, reviewers for their contributions to the peer review of this work. Peer review reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information

psiCLIP reveals dynamic RNA binding by DEAH-box helicases before and after exon ligation

Lisa M. Strittmatter[1]*, Charlotte Capitanchik[2]*, Andrew J. Newman[1], Martina Hallegger[2,3], Christine M. Norman[1], Sebastian M. Fica[1,†], Chris Oubridge[1], Nicholas M. Luscombe[2,4,5,†], Jernej Ule[2,3,§,†], and Kiyoshi Nagai[1]

* These authors contributed equally. † Joint corresponding authors
§ Lead contact
1. MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK
2. The Francis Crick Institute, London NW1 1AT, UK
3. Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, Queen Square, London WC1N 3BG, UK
4. UCL Genetics Institute, Department of Genetics, Environment and Evolution, University College London, London WC1E 6BT, UK
5. Okinawa Institute of Science & Technology Graduate University, Okinawa 904-0495, Japan
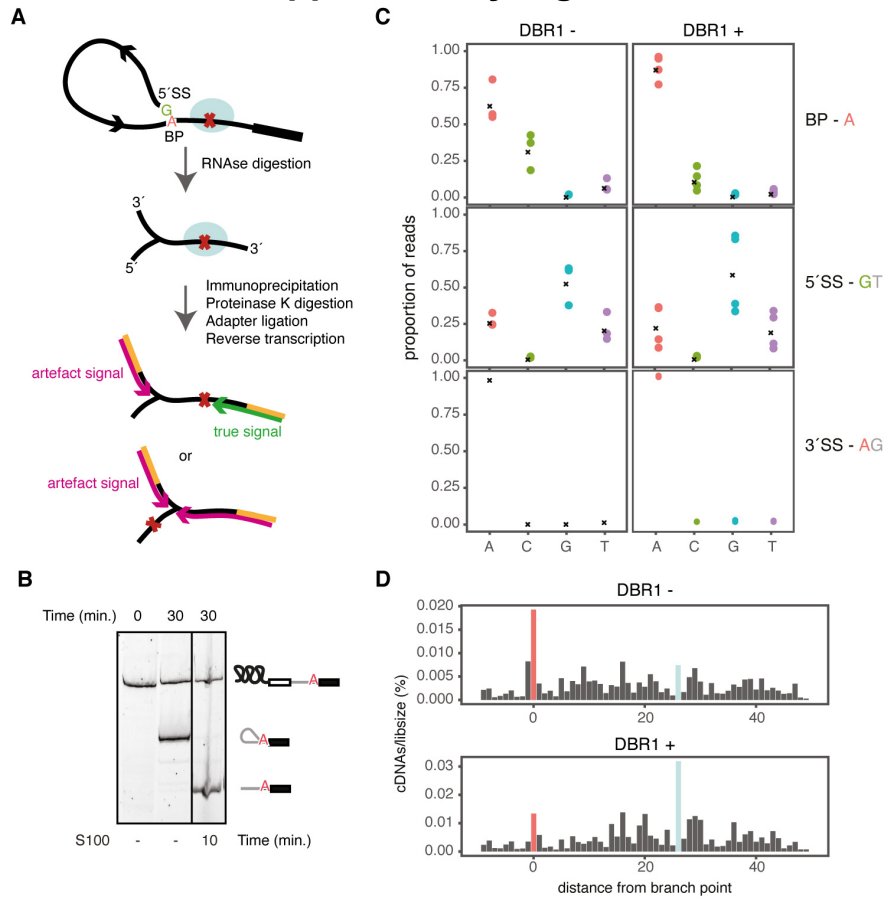
Supplementary Figures

A

B
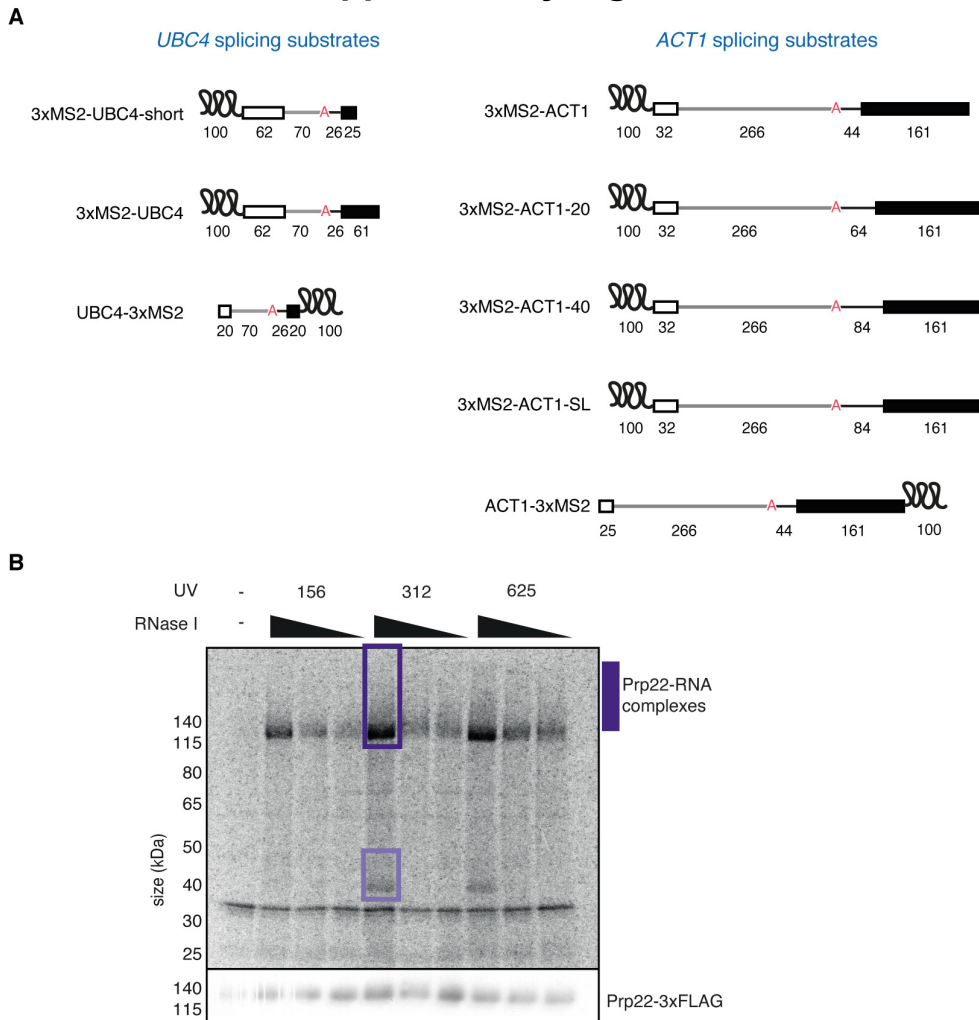
Supplementary Fig. 1 Splicing substrates used in psiCLIP. A All transcripts were generated to include 3 MS2 stem loops for spliceosome purification. The transcripts are derived from the budding yeast *UBC4* (left) and *ACT1* genes (right). Distances between the splice sites are given in nucleotides. **B** Autoradiograph for parallel fine-tuning of UV-irradiation and RNase I digest (top panel). Purified complex C* containing endogenous Prp22-3xFLAG was irradiated with the indicated doses of UV (in 100 x µJ/cm2), and fragmented by the addition of 5, 0.5 or 0.05 units of RNase I following the psiCLIP procedure as described in the methods section. Bottom panel shows subsequent Western probing with anti-FLAG antibody to visualise Prp22 protein levels. The dark purple box indicates the membrane area selected for cDNA library preparation from the optimised condition for Prp22 crosslinking. The light purple box shows a co-crosslinked protein that was released with the given RNase I concentration. A high RNase concentration was optimal for liberating a multi-protein crosslink (light purple box).
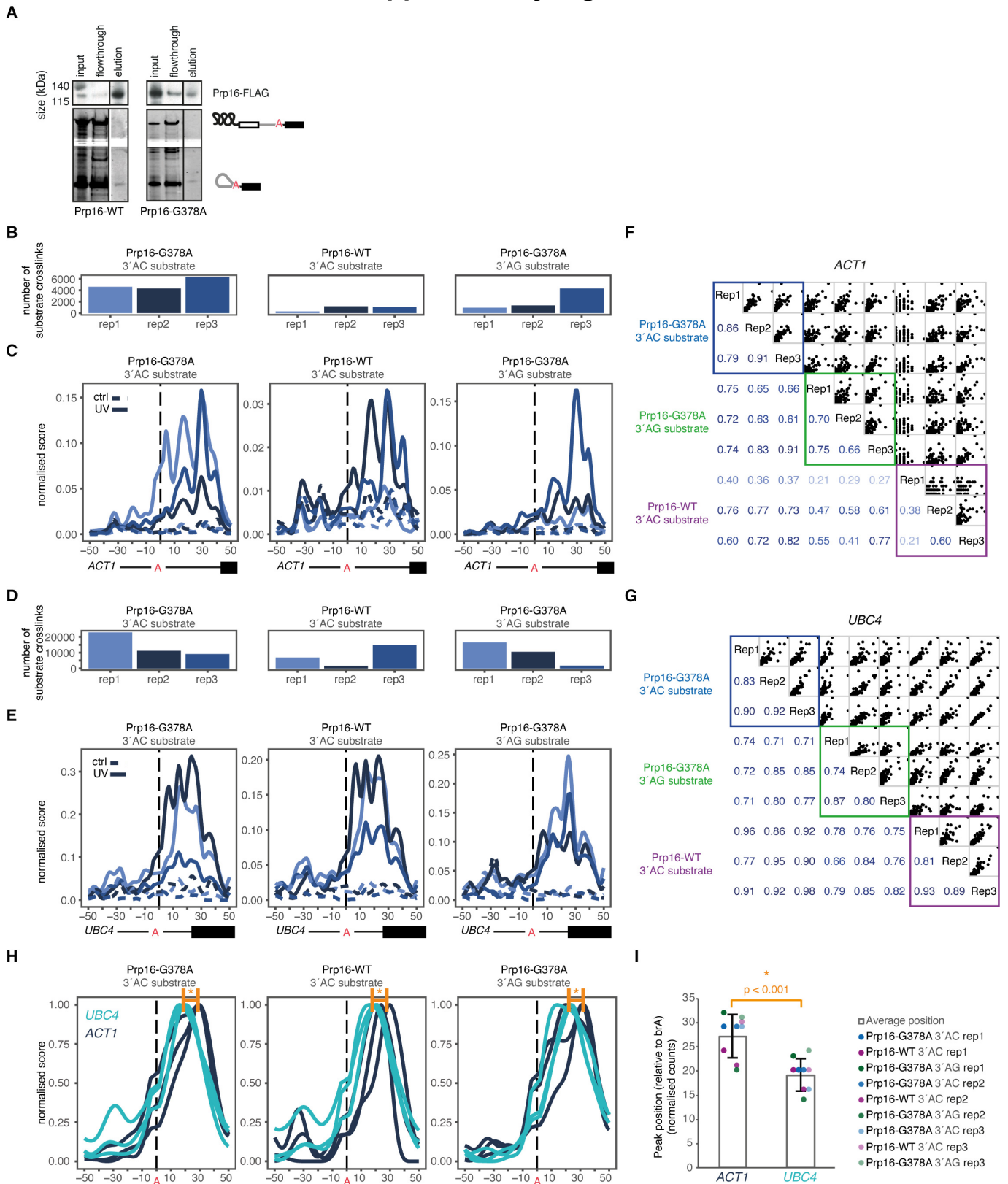
Supplementary Figures
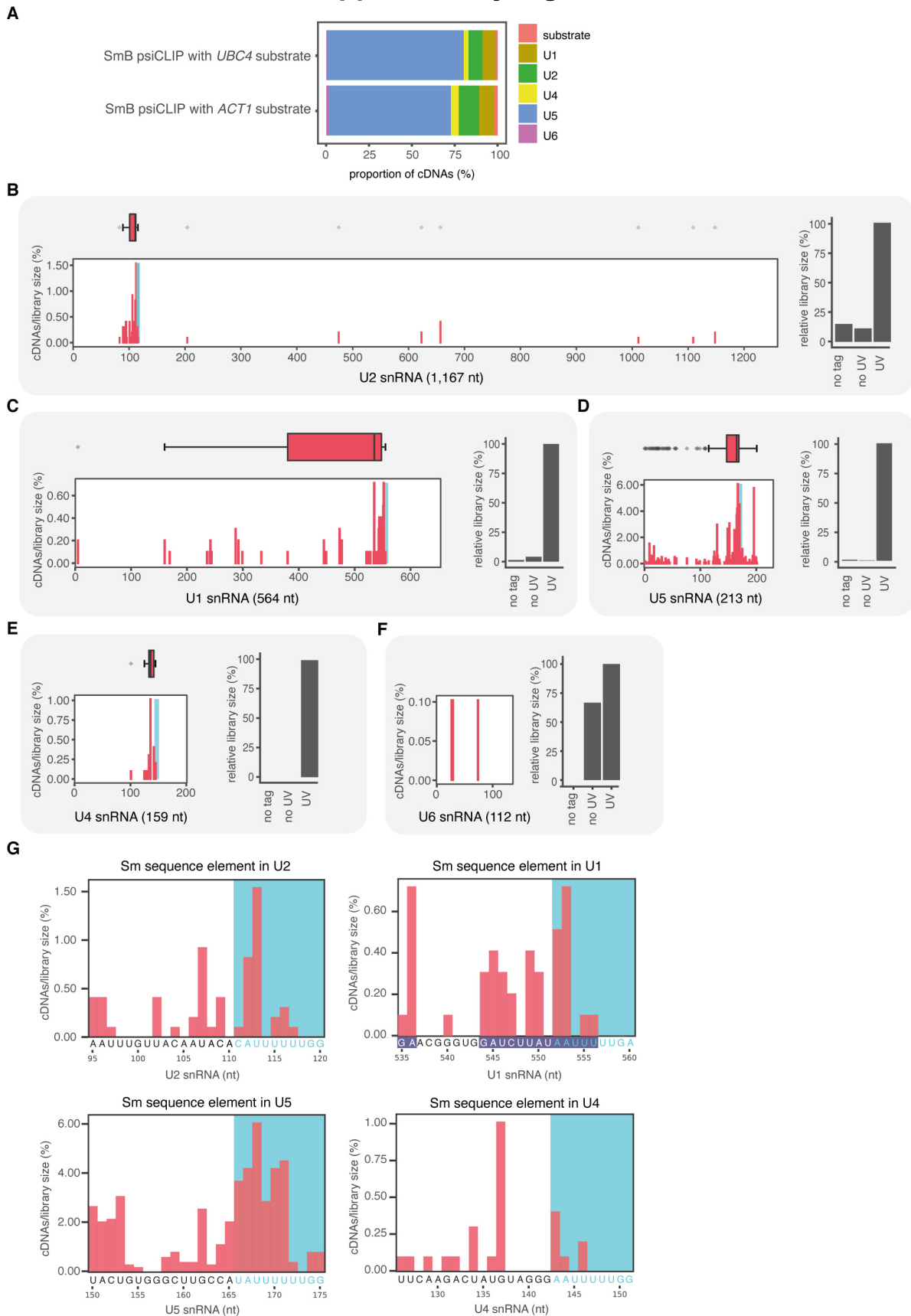
Supplementary Fig. 2 Validation of the positional specificity of psiCLIP with SmB. A Proportion of cDNAs mapping to substrate and snRNAs for two SmB psiCLIP experiments. **B-F** Crosslinking to snRNAs from experiment that used the short *UBC4* substrate. Crosslink events (-1 position of cDNA start) are represented as histograms and box plots aligned to the indicated snRNA. On the histograms, the y-axis indicates the proportion of cDNAs out of all cDNAs mapped to snRNAs and pre-mRNA substrate. The box plots represent crosslink positions from one sample along the transcript, weighted by normalised cDNA count. The line across the box represents the median, the lower and upper bounds correspond to the first and third quartiles. The whiskers end at the largest and smallest value no further than 1.5 times the inter-quartile range. Outliers outside of this range are plotted as dots. The bar charts represent the proportion of cDNAs in samples, normalised to the UV condition, which is shown as 100%. No tag: UV irradiated untagged SmB, No UV: SmB-3xFLAG without UV irradiation, UV: UV irradiated SmB-3xFLAG. As expected, very few crosslinks are found on the U6 snRNA, which is bound by the LSm ring that lacks the SmB protein; due to very low crosslinking on U6 snRNA the box plot is not shown. **G** Zoom in to show crosslinking at the Sm sequence elements, which are highlighted in blue letters and blue background. The U1 snRNA region highlighted in purple is shown in Fig 2D, mapped onto the cryo-EM structure.

2

Supplementary Fig. 3 DBR1 treatment resolves intron-lariats. A A schematic showing how intron-lariats can cause artefactual signal. If a protein (in blue) crosslinks near to the branch point, upon RNase digestion part of the lariat may from a Y-shaped structure. In this situation, the free 3′ end of the lariat is available for adaptor ligation (in orange). Thus the final library can have many reads that begin at the 5′-SS and branch-point and which originate from the lariat-intermediate and not from a protein-RNA crosslink. **B** Treatment of lariats with S100 cell lysate fraction leads to debranching. Intron-lariat intermediates were generated by *in vitro* splicing reactions of 3xMS2-UBC4-AC (lane 2). Treatment with S100 for 10 min in debranching buffer leads to full debranching. Only splicing products containing the second exon are visible as the 3′ end of the transcript is fluorescently labelled. This experiment was repeated three times (n=3) with different incubation times for debranching of 10min, 30min and 60min; a similar effect was observed for all experiments. **C** Reverse transcription of the 5′-SS guanosine and branch-point adenosine can result in mutation. In the final reads, the branch-point nucleotide is often read as a cytosine (top) and the 5′-SS guanosine can be read as an adenosine or thymine (middle). For comparison, the 3′-SS adenosine is rarely mutated (bottom). DBR1 treatment reduces the incidence of mutations (right). Three samples of DBR1- Prp16 psiCLIP and four samples of DBR1+ Prp16 psiCLIP are shown as individual data points. **D** Prp16 crosslinks in a DBR1- and DBR1+ sample, shown normalised to library size. The ratio between the artefact signal (branch-point A in red) and true signal (a representative crosslink shown in blue) shifts dramatically upon DBR1 treatment.
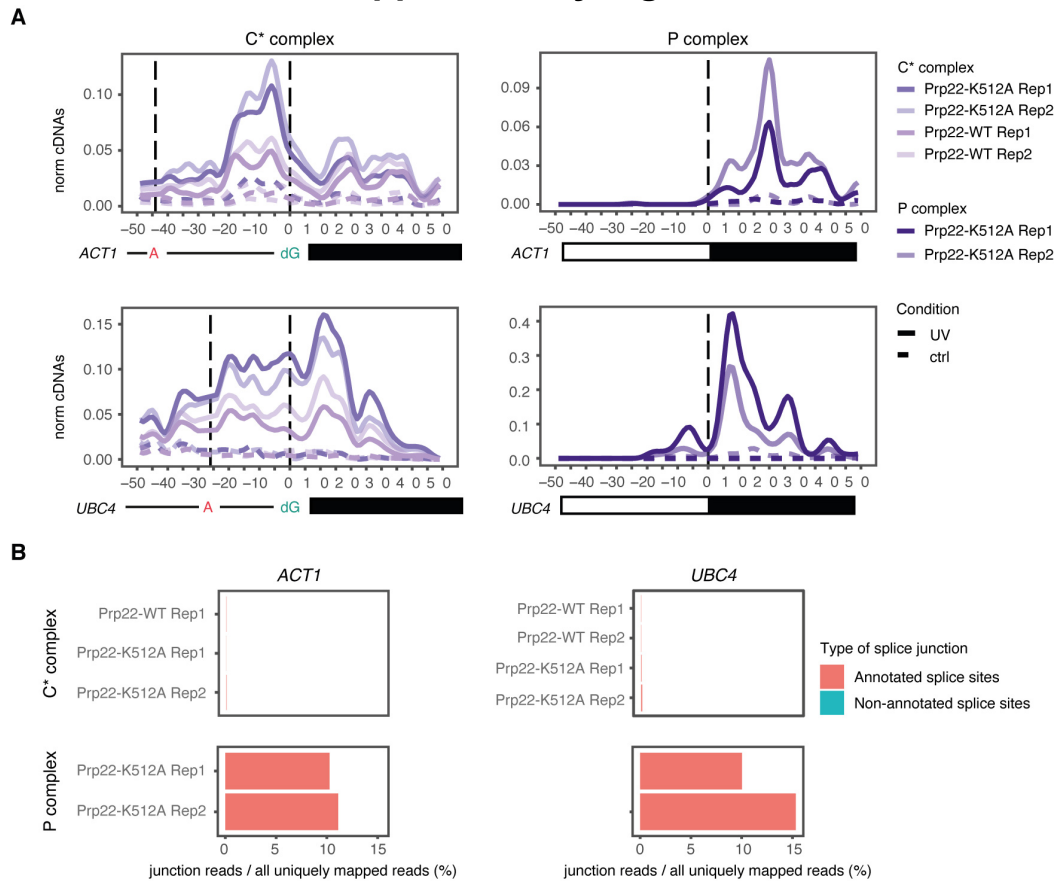
Supplementary Figures

Supplementary Fig. 4 Prp16 psiCLIP data show substrate specific binding in a region downstream of the brA. A Western blot (top) probing the FLAG tag shows Prp16 levels at various steps of the psiCLIP procedure for spliceosomes assembled on *UBC4-3′UAC* and purified via the MS2-tag on the splicing substrate. In this experiment, the sample was not crosslinked to allow protein and RNA to run independently on the gel. Before membrane transfer the gel was scanned for the fluorescently labelled splicing substrate (bottom) and shows that Prp16 binds to lariat-intermediate but not pre-mRNA. Western blots were performed once for each sample. **B** Total number of crosslinks on the *ACT1* pre-mRNA substrate for three replicates of Prp16 psiCLIP. "3′AC" and "3′AG" denote mutant and canonical 3′-SS respectively. **C** Mapping of Prp16 psiCLIP data onto *ACT1*. Positions along the transcript are shown relative to the brA. The smoothed lines show the truncation events normalised to the number of cDNAs mapping to the yeast genome. Lines were Gaussian smoothed with a window size of 10 nucleotides. **D** Same as **A** but for three replicates of *UBC4* Prp16 psiCLIP. **E** Same as **B** but for three replicates of *UBC4* Prp16 psiCLIP. **F-G** Correlation analysis between replicates of Prp16 psiCLIP. The data were normalised to the maximum signal across the substrate so that reproducibility of positional enrichment, rather than crosslinking intensity, could be compared. Each dot represents the normalised crosslinking signal at each position along the substrate. The numbers are Pearson correlation coefficients. Rep1, replicate 1; rep2, replicate 2; rep3, replicate 3. **H-I** Summary of reads normalised to the maximum peak in each sample. Note significant difference between average *ACT1* and *UBC4* peaks, as determined by a two-sided unpaired t-test ($\bar{x}1 - \bar{x}2=8$ nucleotides, 95% CI [4, 12], $\alpha=0.05$, d.f.=16, tα=4.35, p<0.001). Peak positions for individual replicates are shown in **I**.

4

Supplementary Figures
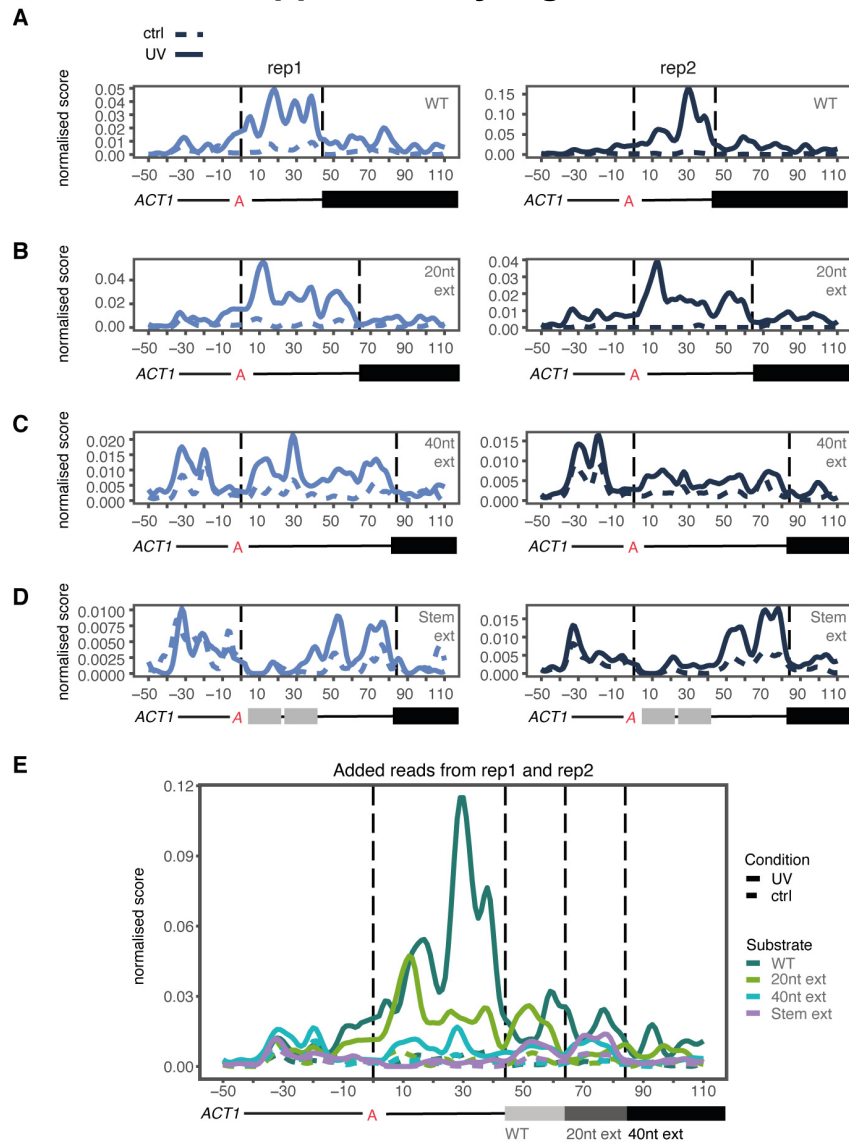
Supplementary Fig. 5 Prp16 binding spans the entire available distance between brA and the 3′-SS. Two replicates for each condition are shown. Mapping of Prp16 mutant psiCLIP data onto the respective *ACT1* splicing substrate. cDNAs are normalised to total cDNAs mapping to the yeast genome. Crosslinks are shown after Gaussian smoothing with a 10 nucleotide window. Positions on the substrate are given relative to brA. **A** Prp16-G378A on the wild-type *ACT1* substrate. **B** Prp16-G378A on *ACT1* substrate with a 20 nucleotide extension between brA and the 3′-SS. **C** Prp16-G378A on *ACT1* substrate with a 40 nucleotide extension between brA and the 3′-SS. **D** Prp16-G378A on *ACT1* substrate with a stem loop in the 40 nucleotide extension between brA and the 3′-SS. The stem-loop region is indicated in grey underneath the graph. **E** The two replicates are summed for each condition, as shown in the main figure, and are shown here on the same scale for comparison. Prp16-G378A binding is reduced on the extended substrates. A sharp decrease in signal from intron to exon for all substrates indicates that the extension of signal observed over the intron results from genuine binding events.
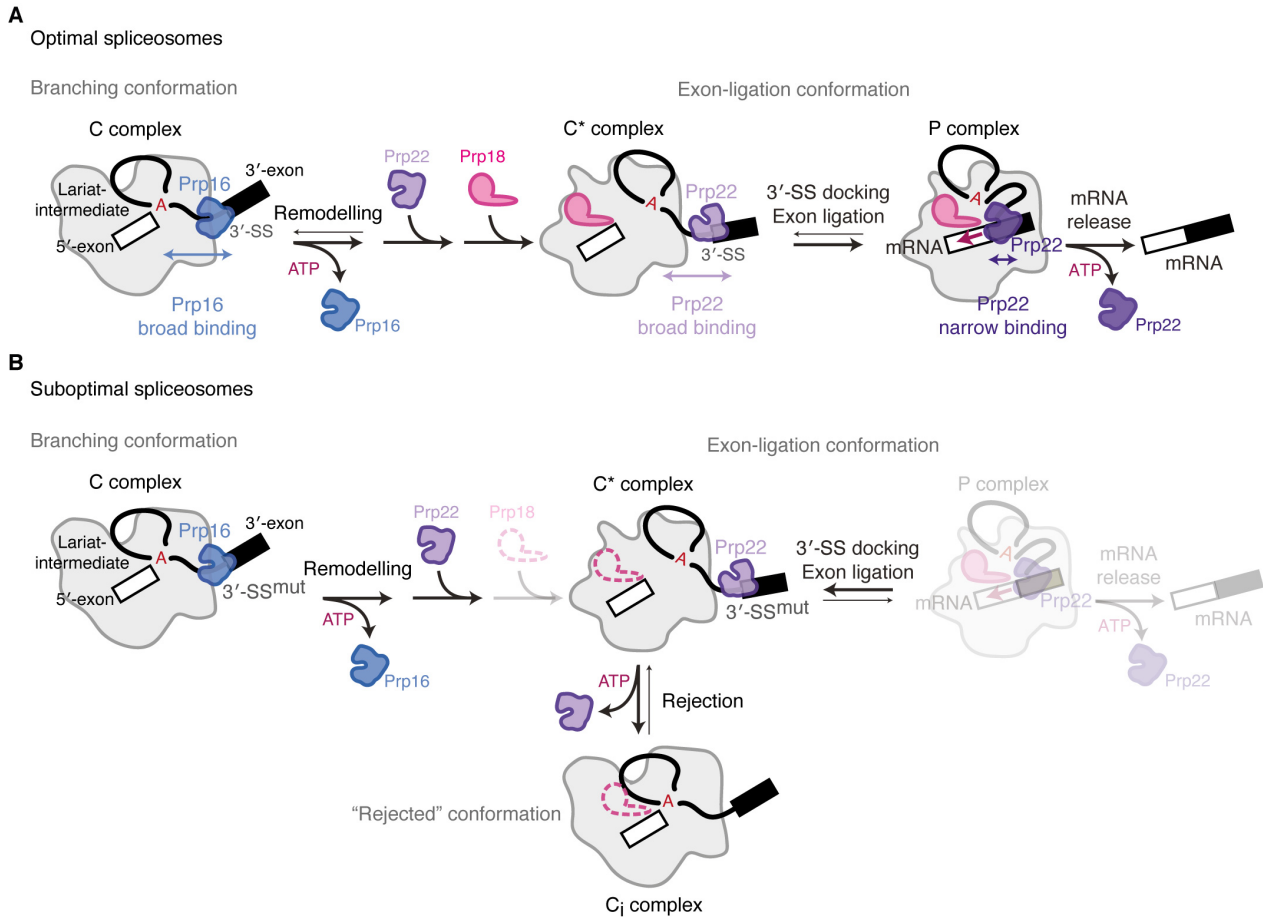
Supplementary Figures

A

B

Supplementary Fig. 6 The data pattern for psiCLIP of Prp22 in C* and P complex is consistent between replicates. A Two replicates for each condition are shown. Mapping of Prp22 (WT and mutant) psiCLIP data onto the respective splicing substrate. cDNAs are normalised to total cDNAs mapping to the yeast genome. Crosslinks are shown after Gaussian smoothing with a 10 nucleotide window. Positions on the substrate are given relative to the 3′-SS. P complex crosslinks are shown with the intron removed. Solid lines are tagged experimental conditions and dashed lines are untagged controls. Rep1, replicate 1; rep2, replicate 2. **B** The proportion of spliced reads in each library, showing that C* complex libraries have very few spliced reads, whereas P complex libraries are enriched for spliced reads. The colour denotes whether the splice junction occurs at an annotated, canonical splice site, or a non-annotated, cryptic site. The lack of reads resulting from non-annotated sites indicates that the spliceosome is using the canonical splice sites, as expected in the P complex data.

Supplementary Figures
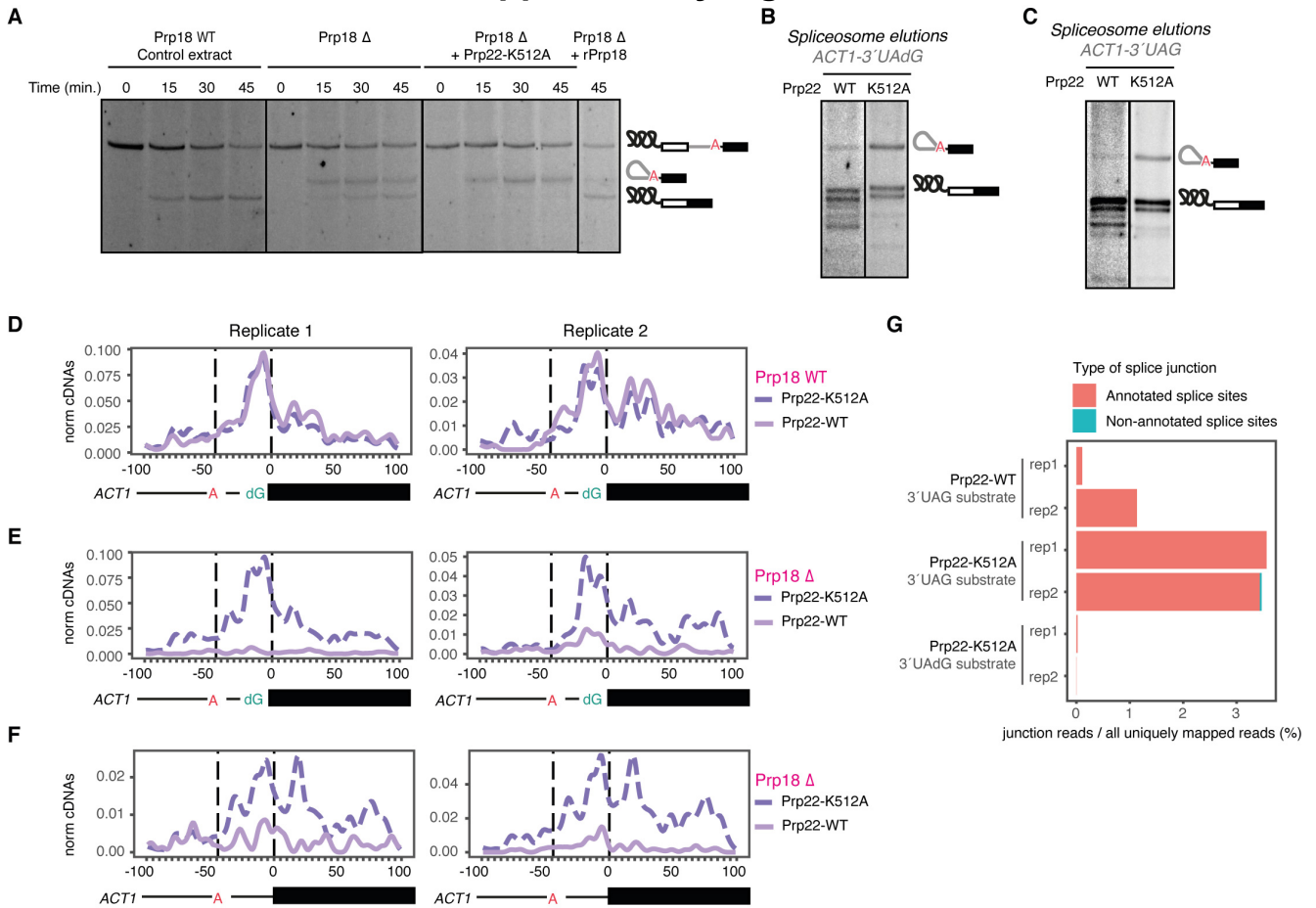
Supplementary Fig. 7 The psiCLIP profile for Prp22 binding following immunodepletion of Prp18 is consistent between replicates.
A Denaturing gel after splicing assays with 3xMS2-*UBC4* demonstrates that there is a step two splicing defect following depletion of Prp18 both in the background of Prp22-WT protein or with the addition of the dominant negative Prp22-K215A. The second step splicing defect can be rescued after 45 minutes by *in vitro* addition of recombinant Prp18 and is shown with the Prp22 WT background. Prp18Δ denotes extracts depleted of Prp18. **B** Substrate composition of eluted spliceosomes assembled on the 3′-UAdG substrate following depletion of Prp18. Denaturing gel of the spliceosomes eluted from amylose column showing the input for UV-crosslinking and immunoprecipitation. The amount of spliceosomes of the first step are reduced for Prp22 WT background compared to Prp22-K512A. Gels in **A-C** are representative of two independent depletion experiments (n=2). **C** Substrate composition of eluted spliceosomes assembled on the canonical 3′-UAG substrate following depletion of Prp18. Denaturing gel of the spliceosomes eluted from amylose column showing the input for UV-crosslinking and immunoprecipitation. The amount of spliceosomes of the first step are reduced for Prp22 WT background compared to Prp22-K512A. **D-F** psiCLIP of Prp22-WT and Prp22-K512A in spliceosomes assembled on wild-type and 3′-UAdG *ACT1* substrates in wild-type extracts or in extracts depleted of Prp18. Crosslinks are shown as normalised to the total yeast genome mapping reads and with the untagged ctrl subtracted from the tagged UV condition, after Gaussian smoothing with a 10 nucleotide window. **G** The proportion of spliced reads in each Prp18 depletion library, showing that the wild-type substrate can be spliced to make mRNA and this occurs to a higher extent with the Prp22-K512A mutant. The colour denotes whether the splice junction occurs at an annotated, canonical splice site, or a non-annotated, cryptic site. Only a small proportion of reads occurs at non-annotated sites, indicating that the majority of splicing observed occurred at the canonical splice junctions. Rep1, replicate 1; rep2, replicate 2.

Supplementary Figures

Supplementary Fig. 8 Model for substrate binding by spliceosomal helicases during the catalytic stage of pre-mRNA splicing in yeast. A Binding to optimal spliceosomes. After branching, Prp16 binds the lariat-intermediate broadly downstream of the brA, though the main binding peaks are observed between the brA and the 3′-SS. Prp16 dissociates during remodelling to the C* conformation, in which it is replaced by Prp22 before docking of the 3′-SS. Prp18 stabilises the exon-ligation conformation. In complex C*, before exon ligation, Prp22 binds the lariat-intermediate broadly on the 3′-exon and the intron. Such broad binding may facilitate 3′-SS docking to allow exon ligation. After exon ligation, in complex P, Prp22 engages the mRNA in a narrow binding window on the 3′-exon, from which it pulls on the RNA substrate to release the mRNA and dissociates following mRNA release. **B** Proofreading of suboptimal spliceosomes by Prp22. When the 3′-SS is non-canonical (3′-SS^{mut}), or in the absence of exon-ligation factors like Prp18, the exon-ligation conformation is destabilised. Prp22 binds along the 3′ exon and upstream of the 3′-SS on the intron, thus disfavouring stable 3′-SS docking, which prevents spliceosomes from catalysing exon ligation. Instead, Prp22 ATPase activity rejects these suboptimal C* complexes. As a result, Prp22 may completely or partially dissociate, thus further destabilising the exon-ligation conformation. Rejected spliceosomes may revert back to a C-like conformation (C$_i$), thus further minimising the potential for docking of the suboptimal 3′-SS.

Supplementary Tables

Supplementary Table 1. Yeast strains used in this study.

Strain	Resistance casette	Casette plasmids	Integration
SmB-3XFLAG-His8	TEF:nat:TEF	pAG25-3xFLAG-His8	genomic
Prp16-3XFLAG-His8	TEF:nat:TEF	pAG25-3xFLAG-His8	genomic
Prp22-3XFLAG-His8	TEF:nat:TEF	pAG-3xFLAG-His8	genomic
Prp22-3XFLAG-His8 Prp18-3XHA Slu7-9Xcmyc	TEF:nat:ADH1 TEF:KanMX:TEF TEF:hph:scCYC1	pBP83 pYM14 pYM20	genomic
Prp18-3XHA Slu7-9Xcmyc	TEF:KanMX:TEF TEF:hph:scCYC1	pYM14 pYM20	genomic
Prp16-G378A-CBP-3XFLAG-His6	URA, TRP		plasmid: pRS424, pRS426
CBP-Prp22-K512A-3XFLAG-His6	URA, TRP		plasmid: pRS424, pRS426

The first column indicates the modification compared to the parent strain BCY123. For strains with a genomic integration (indicated in the third), the resistance gene that was used for integration of the C-terminal tag is indicated with its promoter and terminator sequence, as well as the cassette plasmid used. Plasmids for over-expression of recombinant proteins are given in the fourth column together with the respective selection marker.

Supplementary Tables

Supplementary Table 2. pre-mRNA substrates used in this study.

Substrate	Exon 1	Intron distance (5′-SS to brA)	Intron distance (brA to 3′-SS)	Exon 2	Synthesis method	Label (3′ end)
3XMS2-UBC4-3′-UAc short	100 MS2 + 62	70	26	25	transcription + ligation	Cy5 (pre-labeled oligo)
3XMS2-UBC4	100 MS2 + 62	70	26	61	transcription + ligation	Cy5 (pre-labeled oligo)
3XMS2-UBC4-3′-UAC	100 MS2 + 62	70	26	61	transcription + ligation	Cy5 (pre-labeled oligo)
3XMS2-UBC4-3′-UAdG	100 MS2 + 62	70	26	61	transcription + ligation	Cy5 (pre-labeled oligo)
UBC4-3XMS2	20	70	26	20 + 100 MS2	transcription	Fluorescein (Cy2)
3XMS2-ACT1	100 MS2 + 32	266	44	161	transcription	pCp-Cy5 (enzymatic ligation)
3XMS2-ACT1-3′-UAC	100 MS2 + 32	266	44	161	transcription	pCp-Cy5 (enzymatic ligation)
3XMS2-UBC4-3′-UAdG	100 MS2 + 32	266	44	161	transcription + ligation	pCp-Cy5 (enzymatic ligation)
ACT1-3XMS2	25	266	44	161 + 100 MS2	transcription	pCp-Cy5 (enzymatic ligation)
3XMS2-ACT1-20	100 MS2 + 32	266	64	161	transcription	pCp-Cy5 (enzymatic ligation)
3XMS2-ACT1-40	100 MS2 + 32	266	84	161	transcription	pCp-Cy5 (enzymatic ligation)
3XMS2-ACT1-SL	100 MS2 + 32	266	84	161	transcription	pCp-Cy5 (enzymatic ligation)

The first column indicates the name of the transcript, including any modifications. The length of different transcript segments is indicated as number of nucleotides. Transcription + ligation indicates enzymatic ligation of a nucleotide pre-labelled with Cy5 during chemical synthesis.

Supplementary Tables

Supplementary Table 3. Oligonucleotides used in this study.

Number	Name	Sequence	Type
1	SmB_FLAG-His_a	GTTTAATAATGAAGCGCCCCCTCAAACAAGGAAGTTTCAGCCCCC ACCAGGTTTTAAAAGAAAAGACTACAAAGACCATGACGGTG	DNA
2	SmB_FLAG-His_b	CACATGCGTACACAAAAAAGTATACGGAAACTATATTAGACTACA CTACATCAACCTTAGCGGCCGCATAGGCCACTAGTGG	DNA
3	Prp16_FLAG-His_a	GCAAAATATACTGAACGGCAAAGAAAATTCAATGAAACCTTTCAAA AGAAGGAAGCCTTTTTTTGACTACAAAGACCATGACGGTG	DNA
4	Prp16_FLAG-His_b	GCATGCATATAACTATATAATAACATATATGAATATTTTGCCTATTAGC ACGCTCTTCCCATAAAGCGGCCGCATAGGCCACTAGTGG	DNA
5	Prp22_FLAG-His_a	GGATCAAAATTCATGGAGACTAAGCTCAATAAGGCAGTCAAGGGA AAGGGCATTAGGTATCAAGAGGGACTACAAAGACCATGACGGTG	DNA
6	Prp22-FLAG-His_b	GTTGTTAAAAAATTAAATATAGGTCTATAAAACTCGATAATTATAATG CATAAAAAGCTAACAATGGCGGCCGCATAGGCCACTAGTGG	DNA
7	Prp22_FLAG_a	GTTAAAAAATTAAATATAGGTCTATAAAACTCGATAATTATAATGCATA AAAAGCTAACAATGTTAATCGATGAATTCGAGCTCGATTACAAC	DNA
8	Prp22_FLAG_b	CAAAATTCATGGAGACTAAGCTCAATAAGGCAGTCAAGGGAAAGG GCATTAGGTATCAAGAGGCGTACGCTGCAGGTCGACGACGAT	DNA
9	Prp18_HA_a	CCAGTATTAAAAGATTAATAACTTTTGAAGAATGGTATACCAGCAAC CACGATAGCTTAGCCCGTACGCTGCAGGTCGAC	DNA
10	Prp18_HA_b	CTTCTTATTTTGGCCGCATGATATCGTGCCACGCGATAACGAAAAC AATAGTTCAACAATTAATCGATGAATTCGAGCTCG	DNA
11	Slu7_cmyc_a	GGAAAGTAGATGGTACAAAGCAATCTGAGGAACAACGGAACCACT TAAAAGATTTATATGGTCGTACGCTGCAGGTCGAC	DNA
12	Slu7_cmyc_b	CTAAATTCTTTACCTCTCGTGCTCCGTGACTAGGATATTGGAAATT CAAATGATATAACTTAATCGATGAATTCGAGCTCG	DNA
13	Extension1a	CGATTTTATTTATTTGATCT	DNA
14	Extension1b	CGAGATCAAATAAATAAAAT	DNA
15	Extension2a	CGATTATTTATCTTGATTTT	DNA
16	Extension2b	CGAAAATCAAGATAAATAAT	DNA
17	dG-instert_ACT1	AUAUUAUAUGUUUA(dG)ACGUUGCUGCUUUGG	RNA with dG
18	Bridge_1_ACT1	GCAACGTCTAAACATATAATATGACAACAAAAGAATGAAGC	DNA
19	Bridge_2_ACT1	CATACCAGAACCGTTATCAATAACCAAAGCAGC AACGTCTAAACATATA	DNA
20	dG-insert_UBC4	UAUUGAACUA(dG)ACAUCCACCUACUUCAUGUUCACGC	RNA with dG
21	AC-insert_UBC4	UAUUGAACUACACAUCCACCUACUUCAUGUUCACGC-Cy5*	RNA *with or without Cy5
22	AG-insert_UBC4	UAUUGAACUAGACAUCCACCUACUUCAUGUUCACGC	RNA
23	UBC4_extension	GGUCCACUCGGCGAUGAUCUAUAUCACUGGCAA GCA-Cy5	RNA with Cy5
24	Bridge_UBC4	AGATCATCGCCGAGTGGACCGCGTGAACATGAAGTAGGTGGATG TCTAGTTCAATAGCATGTTGAT TTTGTTAGTAAATA	DNA
25	RNase-H_UBC4	ATGAAGTAGGTGGAT	DNA
26	RNase-H_ACT1	AACCAAAGCAGCAAC	DNA

Sequences are given in the 5′ to 3′ direction. The last column indicates any additional modifications, such as Cy5-labelling. Oligonucleotides 1-12 were used for genetic modification of yeast strains. Oligonucleotides 13-16 were inserted to generate artificial *ACT1* transcripts. Oligonucleotides 17-24 were used to generate pre-mRNA substrates. Oligonucleotides 25-26 were used to induce RNase H digestion during P-complex purification.