

UC San Diego

UC San Diego Previously Published Works

Title

Large-scale evaluation of the ability of RNA-binding proteins to activate exon inclusion

Permalink

<https://escholarship.org/uc/item/9714311z>

Journal

Nature Biotechnology, 42(9)

ISSN

1087-0156

Authors

Schmok, Jonathan C

Jain, Manya

Street, Lena A

et al.

Publication Date

2024-09-01

DOI

10.1038/s41587-023-02014-0

Peer reviewed

Large-scale evaluation of the ability of RNA-binding proteins to activate exon inclusion

Received: 20 May 2023

Accepted: 29 September 2023

Published online: 2 January 2024

 Check for updates

Jonathan C. Schmok^{1,2,3,4}, Manya Jain^{1,2,3}, Lena A. Street⁵, Alex T. Tankka^{1,2,3}, Danielle Schafer^{1,2,3}, Hsuan-Lin Her^{1,2,3}, Sara Elmsaouri^{1,2,3}, Maya L. Gosztyla^{1,2,3}, Evan A. Boyle^{1,2,3}, Pratibha Jagannatha^{1,2,3}, En-Ching Luo^{1,2,3}, Ester J. Kwon⁴, Marko Jovanovic⁵ & Gene W. Yeo^{1,2,3}✉

RNA-binding proteins (RBPs) modulate alternative splicing outcomes to determine isoform expression and cellular survival. To identify RBPs that directly drive alternative exon inclusion, we developed tethered function luciferase-based splicing reporters that provide rapid, scalable and robust readouts of exon inclusion changes and used these to evaluate 718 human RBPs. We performed enhanced cross-linking immunoprecipitation, RNA sequencing and affinity purification–mass spectrometry to investigate a subset of candidates with no prior association with splicing. Integrative analysis of these assays indicates surprising roles for TRNAU1AP, SCAF8 and RTCA in the modulation of hundreds of endogenous splicing events. We also leveraged our tethering assays and top candidates to identify potent and compact exon inclusion activation domains for splicing modulation applications. Using these identified domains, we engineered programmable fusion proteins that outperform current artificial splicing factors at manipulating inclusion of reporter and endogenous exons. This tethering approach characterizes the ability of RBPs to induce exon inclusion and yields new molecular parts for programmable splicing control.

RNA-binding proteins (RBPs) mediate myriad layers of post-transcriptional gene regulation, including alternative pre-mRNA splicing (AS)¹. Despite the widespread importance of RBPs for cellular function, most of the more than 2,000 human proteins predicted or shown to bind RNA do not have an assigned molecular function^{1,2}. AS is a prevalent and critical RNA processing step, as up to 95% of human multi-exon genes exhibit multiple splice isoforms³. Aberrant splicing is also widespread in disease, especially cancer^{4–5}, driving proteomic imbalance and disruption of cellular homeostasis^{6,7}. Among the RBPs lacking functional annotation of their RNA-binding activity are RBPs

involved in AS. Systematic approaches to assign AS activity to RBPs are, thus, needed to bridge this knowledge gap.

Previous assays have employed luciferase and fluorescence-based reporter systems to identify and characterize RBPs that underscore AS. However, these have relied on global overexpression⁸ or knockdown^{9,10} of RBPs. Global perturbations of protein level are not able to separate effects caused by direct binding of RBPs from their indirect action through splicing regulatory networks. Furthermore, none of these previous studies has investigated how binding position relative to an alternatively spliced exon can modulate the effect of the RBP, even

¹Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA, USA. ²Sanford Stem Cell Institute Innovation Center and Stem Cell Program, University of California San Diego, La Jolla, CA, USA. ³Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA. ⁴Department of Bioengineering, University of California San Diego, La Jolla, CA, USA. ⁵Department of Biological Sciences, Columbia University, New York, NY, USA. ✉e-mail: geneyeo@ucsd.edu

though many splicing factors can exert different effects depending on the distance and orientation (upstream or downstream of the alternative exon) of their binding position^{11–14}. Reporter-based assays that recruit candidate proteins to a specific position, previously applied in studies of transcriptional effectors¹⁵ and modulators of RNA stability/translation¹⁶, are a promising avenue to address these limitations¹⁷.

Complementary to the important need to understand the mechanisms driving AS is the potential utility of tools for targeted modulation of splicing events. Engineered RBPs have been generated through fusion of exon activation domains to RNA-targeting PUF domains¹⁸ and RNA-targeting CRISPR systems^{19,20}. Such technologies are in their nascent stage, reliant on exon activation domains selected from historically well-known splicing factors. A molecular toolkit of potent and compact activation domains to be implemented in maturation of these technologies remains to be established.

In this study, we developed tethered function luciferase-based splicing reporter assays to investigate and quantify the capacity of any protein sequence to directly promote exon inclusion. We used this system to systematically assess proximity-dependent modulation of exon inclusion for 718 human RBPs at two separate tethering positions and to identify potent and compact exon inclusion activation domains. Altogether, our assays serve as both a biological discovery engine that reveals factors involved in splicing and a prototyping platform that can yield molecular parts for protein engineering applications.

Results

Development of tethered function splicing reporter assays

We constructed two dual-luciferase tethered AS minigene reporter systems based on the splicing event of *MAPT* (microtubule-associated protein tau) exon 10 (Fig. 1a and Extended Data Fig. 1a)²¹, which is predominantly excluded from the mature mRNA in HEK293T cells. The first reporter contains the MS2 hairpin 30 base pairs downstream of the 5' splice site (lucMAPT-30D), and the second contains the MS2 hairpin 30 base pairs upstream of the 3' splice site (lucMAPT-30U). The MS2 hairpin recruits MS2 coat protein (MCP) fused to RBP open reading frames (ORFs) to determine the effect on AS of the exon when RBPs are tethered to various positions on the RNA.

Both minigenes are flanked by a constitutively included Firefly luciferase ORF at the 5' end and a conditionally included Renilla luciferase ORF at the 3' end to permit inference of exon inclusion. Firefly luciferase is expressed independent of exon skipping, but inclusion of the tau exon harboring a stop codon terminates translation upstream of Renilla luciferase. We used changes in luminescence in experimental conditions to determine changes in the percent-spliced-in (ψ) of the AS exon when compared with a negative control (Fig. 1b). The AS exon is the penultimate exon, so we inserted the stop codon within 50 base pairs of the 5' splice site to minimize sensitivity of the long isoform to nonsense-mediated decay (NMD)²².

To validate our assay, we co-transfected the lucMAPT-30D reporter with fusion proteins composed of known regulators of exon inclusion and MCP. For a negative control (NC), we used a construct containing an array of three FLAG epitope tags fused to MCP (FLAG NC). We compared ψ value as measured by the reporter readout to an RNA-level validation (Fig. 1c,d). Compared with FLAG NC, MCP-fused proteins LUC7L2, SRSF5 and RBFOX1 increased exon inclusion as measured by both techniques in decreasing order of intensity. To verify that effector recruitment was mediated by the MS2–MCP system, we co-transfected lucMAPT-30D with an RBFOX1 plasmid lacking the MCP fusion. This did not activate the reporter (Extended Data Fig. 1b). As we designed our reporters to minimize sensitivity to NMD, we tested the response of the reporters to NMD perturbation by testing the reporter readout in response to shRNA-mediated knockdown of UPF1, the central effector of NMD²³, and SMG7, a non-essential NMD factor²⁴ (Extended Data Fig. 1c–e). We detected a minor (<10%) increase in long isoform abundance after NMD perturbation, indicating that the early

stop codon-containing long isoform is, to some degree, sensitive to NMD. For the purposes of our studies, where the NMD environment is consistent and candidates are recruited specifically to pre-mRNA by MS2-containing introns, we deemed it acceptable. Based on these validations, we moved forward with these reporters to screen our RBP–MCP library.

Tethering assays identify RBPs that induce exon inclusion

We evaluated 718 RBP ORFs fused to MCP for their ability to induce exon inclusion (Supplementary Table 1). Our laboratory previously developed the RBP–MCP library from subcloning of putative RBP ORFs¹⁶. We performed two arrayed co-transfection screens with candidate RBPs in HEK293T cells, one with lucMAPT-30D and one with lucMAPT-30U (Fig. 1e, left). We analyzed all ORFs in triplicate and compared with negative controls (FLAG NC) and positive controls (RBFOX1–MCP for lucMAPT-30D and SRSF5–MCP for lucMAPT-30U) on the same plate (Extended Data Fig. 1f). Because our analysis focused on ψ increases exclusively, we measured statistical significance when compared with the negative control by one-tailed independent two-sample *t*-test.

We moved forward with candidates that increased ψ significantly ($P < 0.05$; Supplementary Tables 2 and 3) and verified them with further rounds of screening (Fig. 1e, middle). First, we replicated the reporter results of all selected candidates and moved forward with those that again increased ψ significantly ($P < 0.05$; Supplementary Tables 4 and 5). We then verified that all positive hits induced exon inclusion of the reporter at the RNA level through agarose gel electrophoresis of amplified cDNA following the same transfection conditions (Extended Data Fig. 1g and Supplementary Tables 6 and 7). ψ was estimated by calculating the intensity ratio of the inclusion band to the skipping band in duplicate and comparing against control conditions distributed throughout the gel. We calculated *P* value by one-tailed independent two-sample *t*-test, and hits with Bonferroni-corrected $P < 0.05$ were kept. Finally, remaining hits that exclusively activated one of the two reporters were evaluated one more time with the opposite reporter in case they were missed by the initial screen (Supplementary Tables 8 and 9). After these rounds of screening, 26 hits were detected that exclusively activated lucMAPT-30D; 15 hits were detected that exclusively activated lucMAPT-30U; and 17 hits were detected that activated both reporters (Supplementary Table 10 and Fig. 1e, right).

We investigated the biology underlying the candidates detected from our screens. To verify that our assays robustly captured known regulators of AS, we performed Gene Ontology (GO) analysis on the full list of final hits. When compared with a background of the complete tethering library, GO analysis showed strong enrichment of RNA splicing-associated terms (Fig. 2a). As AS occurs in the nucleus, we investigated the subcellular localization of the candidates. We referenced the COMPARTMENTS subcellular localization database, which integrates evidence from text mining, high-throughput screens, literature and prediction methods, and extracted the nuclear localization confidence score for each candidate²⁵. All candidates, save two, have a nuclear confidence score of 4/5 or greater (Supplementary Table 10). The two candidates that scored lower than 4/5 were STAU1 and EIF4B. STAU1, which scored 2.68/5, has previously been linked to splicing regulation^{26,27}. EIF4B, which scored 3.82/5, initiates translation in the cytoplasm by binding RNA substrates and recruiting ribosomes. We hypothesize that this mechanism could drive a false positive when artificially driven to nuclear pre-mRNA in our tethering system, as the mechanism of spliceosome recruitment is similar. Nevertheless, a potentially nuclear role of EIF4B in splicing regulation merits future investigation. Altogether, the candidates determined by our screen are enriched for known regulators of mRNA splicing and are largely localized to the nucleus.

We also detected differences in the types of RBPs identified by each screen (Fig. 2b). Both RBFOX1 and RBFOX2 exclusively activated the reporter when tethered downstream (lucMAPT-30D), consistent

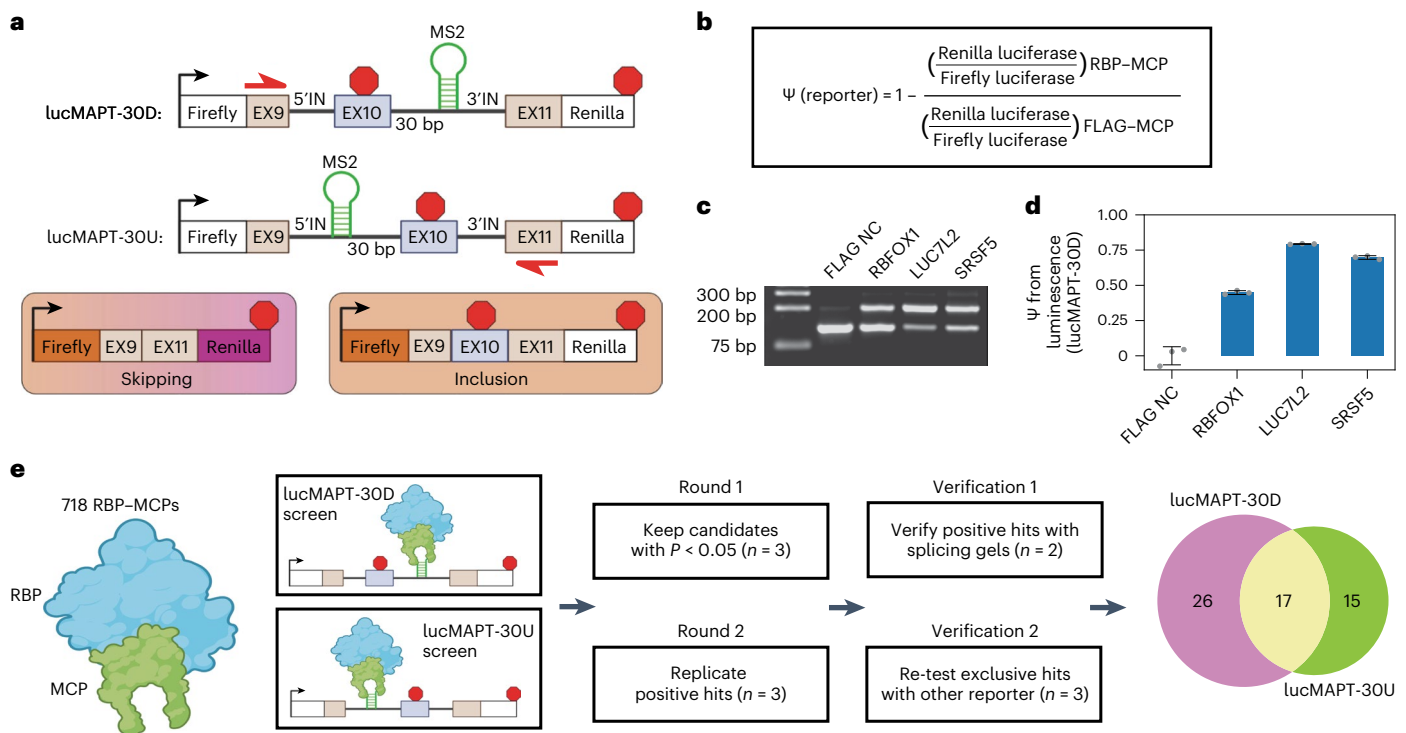


Fig. 1 Development of tethered function assays for detecting direct induction of exon inclusion. **a**, Schematic of luciferase reporters used in the assays and resulting isoforms after cellular mRNA processing. **b**, Analysis workflow for calculating percent-spliced-in from luminescence measurements. **c**, Splicing gels of lucMAPT-30D splicing in response to co-transfection with MCP-fused positive and negative controls. Bands are generated by agarose gel electrophoresis of RT-generated cDNA amplified by minigene specific primers (shown in **a**) that amplify skipping and inclusion isoforms. **d**, Bar graph of lucMAPT-30D reporter readout as calculated from the workflow in **b** with the

same conditions as **c** (mean \pm s.d., $n = 3$ replicate transfections). **e**, Experimental workflow of tethering assays. The effects of recruiting 718 MCP-fused RBPs are tested in both reporter contexts. P value was calculated by independent two-sample one-tailed t -test, comparing the co-transfection of the reporter and candidates to co-transfection of the reporter and FLAG NC performed concurrently. The displayed n refers to biological replicates of candidate transfections. For FLAG NC transfections, $n = 3$ biological replicates for the reporter experiments and $n = 6$ for the splicing gel experiment. Venn diagram of final hits after all rounds of screening and verification. bp, base pairs.

with the known effect of these proteins primarily causing exon inclusion when bound downstream of alternatively spliced exons^{11,28}. Three proteins associated with 3' splice site recognition exclusively activated the upstream tethering reporter (lucMAPT-30U): U2AF2 (the large subunit of the U2 auxiliary factor), SF1 and SNW1 (refs. 29,30). The RBPs tested from the Sm family (SNRPB, SNRPN, SNURF, SNRPG, SNRPE and SNRPA) exclusively and potentially activated the downstream tethering reporter, despite the Sm ring being found in spliceosomal subunits that form at either end of the splicing junction³¹. The SR family of splicing factors was primarily represented at the intersection of both screens (SRSF8, SRSF5, SRSF6, SRSF4, SRSF11 and SRSF10); however, SRSF7 exclusively activated the downstream tethering reporter, and SRSF12 exclusively activated the upstream tethering reporter.

As we were especially interested in candidates that have not previously been associated with AS regulation, we first determined candidates that were not annotated with splicing-associated GO terms and have not been specifically referenced in the literature as potential splicing factors and deemed them 'unexpected hits'. Most unexpected hits exclusively activated the upstream tethering reporter (UBAP2L, STAU2, EIF4B, CNOT3, MAZ, GTF2F1 and FIP1L1), which was uncommon for known splice modulatory factors. We detected three unexpected hits as exclusive activators of the downstream tethering reporter (TRNAUIAP, SCAF8 and RTCA) and one as an activator of both reporters (XPO1). Next, we searched for the unexpected hits on the spliceosome database (SpliceosomeDB) to determine if previous proteomics efforts have identified them as interactors with components of the spliceosome in humans³². This search yielded such evidence for SCAF8, CNOT3 and FIP1L1. SCAF8 has been detected in a supraspliceosome complex in vivo

assembled from HeLa cell extract³³ and after immunoprecipitation of CDC5L in HeLa cells³⁴. CNOT3 has been detected after immunoprecipitation of SRRM1 in HeLa extract³⁵. FIP1L1 has been detected after isolation of mixed spliceosome complexes assembled in vitro from the extracts of WERI-1 retinoblastoma cells³⁶ and HeLa cells³⁷. Finally, we also noted that XPO1 has a known, albeit indirect, role in mRNA splicing. XPO1 is a nuclear export receptor that shuttles the immature small nuclear RNAs (snRNAs) of the spliceosome to the cytoplasm for maturation³⁸. Despite the preliminary evidence linking a subset of the unexpected hits to mRNA splicing, the landscape of splicing events regulated by any of the unexpected hits has not currently been characterized in any biological system.

We binned hits into categories depending on whether they activated the downstream reporter only, activated the upstream reporter only or activated both reporters. Binned RBPs display effect size patterns associated with their categories (Fig. 2c–f). For the RBPs that activated both reporters, ψ for the two reporters is correlated. A population exists among the RBPs that activated both reporters with high strength, which includes the strongest overall hit, SRSF8. SRSF8 activated the highest ψ with the upstream tethering reporter and the second highest ψ for the downstream tethering reporter behind RNPS1. The downstream-only hits generally exhibited stronger activation than upstream-only hits. These categories of hits display trends in effect size; however, the variance within each category highlights the diversity of mechanisms by which RBPs influence AS by proximity.

We also tested our final collection of hits with orthogonal exon inclusion reporters. We screened our hits using lucMAPT reporters containing tethering sites 100 base pairs distal to the splice site instead of

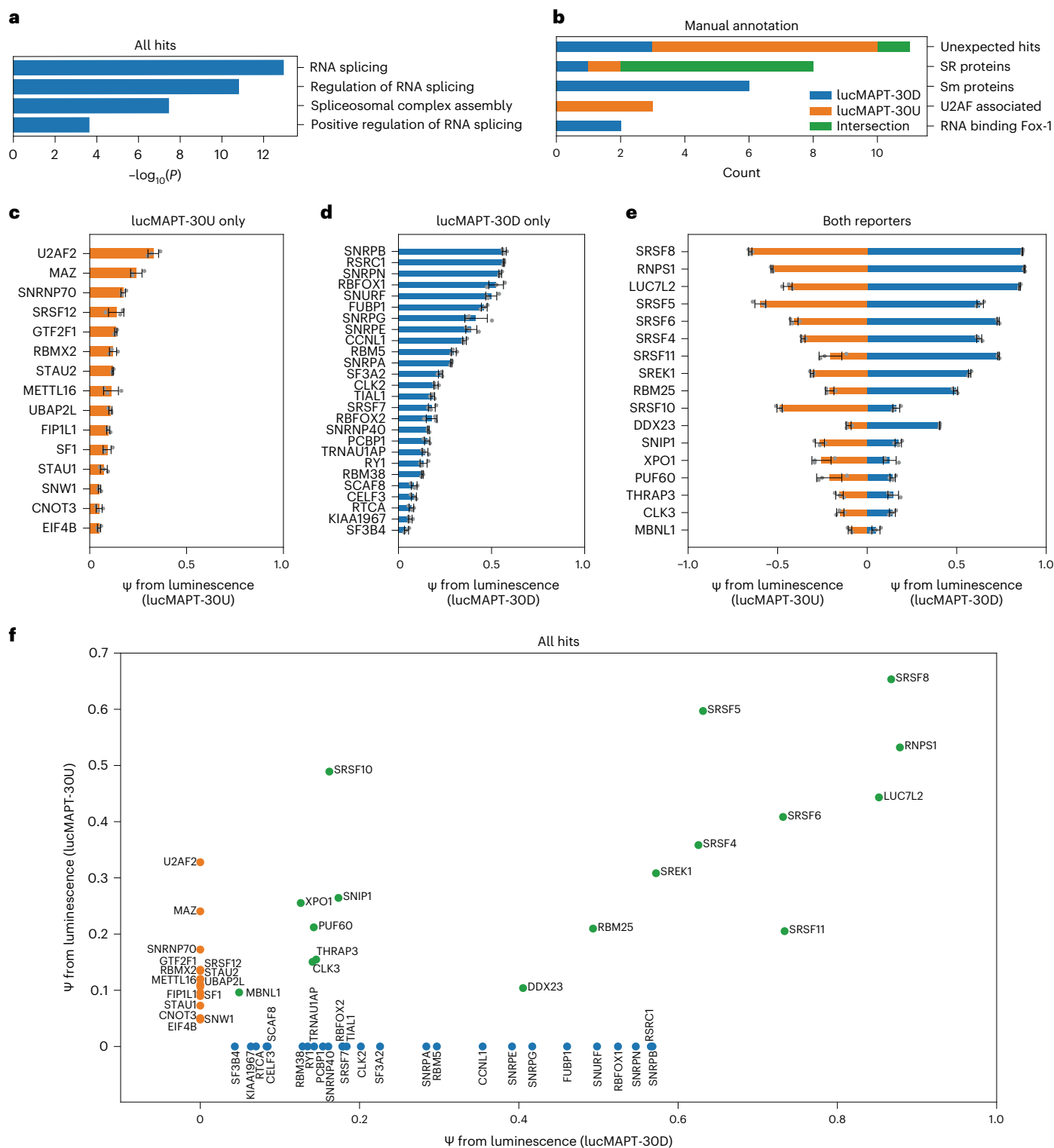


Fig. 2 | Tethering assays identify RBPs that induce exon inclusion. **a**, Bar graph displaying GO analysis of all hits emerging from the screen. The top four mostly significantly enriched biological processes are displayed. Background for analysis is the complete list of genes used in the screens. Unadjusted *P* value was calculated using Metascape⁵⁶ based on the accumulative hypergeometric distribution. **b**, Bar graph displaying manual annotation of protein families displaying reporter preferences in the screen and reporter preferences of unexpected hits. **c–e**, Bar graphs displaying reporter readout after screen and

validation for RBP–MCP fusions of all final hits passing the lucMAPT-30U screen only (**c**), the lucMAPT-30D screen only (**d**) and both screens (**e**) (mean \pm s.d., *n* = 3 replicate transfections). **f**, Scatter plot comparing activation of both reporters of all hits emerging from the screen. Hits that passed only for a single reporter are placed along the axis for the other reporter. Hits that passed the lucMAPT-30U screen only are displayed as orange markers; hits that passed the lucMAPT-30D screen only are displayed as blue markers; and hits that passed both screens are displayed as green markers.

30 base pairs (Extended Data Fig. 2a). Almost all hits exhibited reduced activity at the increased distance, but proximity dependence varied by RBP (Extended Data Fig. 2b–d and Supplementary Tables 11 and 12). Finally, we tested all hits with another exon inclusion reporter based around *MBNL1* exon 8 (lucMBNL1; Extended Data Fig. 2e). Although positive control SRSF5 successfully induced exon inclusion, the baseline inclusion rate was perturbed by a small subset of hits, implying some context dependence of proximity-dependent splicing activity of the tested RBPs (Supplementary Tables 13 and 14 and Extended Data Fig. 2f). Nevertheless, the lucMAPT screens provide one valid context, and we continued forward with their findings with the knowledge that we are capturing effects within it.

Initially, we also investigated a complementary approach to identify RBPs that induce exon skipping. We constructed a reporter using the same framework around *MAP3K7* exon 12, which is primarily included in HEK293T cells (Extended Data Fig. 3a). We validated the response of the *MAP3K7* reporter to HNRNPK and PCBP1, known activators of exon skipping, using the reporter readout and RNA-level validation when tethered 100 base pairs upstream of the AS exon (Extended Data Fig. 3b). Twenty-two of 44 RBPs induced exon skipping when tethered 30 base pairs downstream of the AS exon, and 154 of 194 induced exon skipping when tethered 100 base pairs upstream of the AS exon (Extended Data Fig. 3c,d and Supplementary Tables 15 and 16). The high proportion of hits suggests that recruitment of many proteins may simply act to sterically prevent spliceosome recognition; thus, we stopped the skipping screen here and constrained this study to focus on exon inclusion, a more specific molecular task.

Splicing events are modulated by unexpected hits

We followed up the screen with endogenous characterization of four hits from the screen, which, to this point, have no established role in AS regulation: STAU2, SCAF8, RTCA and TRNAUIAP. STAU2 is an important protein in neuronal mRNA localization³⁹ that shares 59.9% similarity with paralogue STAU1: a multi-functional RBP with implications for oncogenesis and neurodegeneration^{26,40}. SCAF8 was previously characterized for roles in selection of distal poly(A) sites and transcriptional elongation, and a selection of genes in the same family are known or predicted to be involved in AS, including SCAF1, SCAF4 and SCAF11 (ref. 41). Although SCAF8 was detected in two previous spliceosomal proteomics experiments, the significance of this finding has not been further investigated^{33,34}. RTCA has been previously characterized for its role in RNA metabolism by catalyzing the conversion of the 3' phosphate of RNA substrates to a 2',3'-cyclic phosphodiester⁴². TRNAUIAP is a poorly characterized protein predicted to play a role in selenocysteine (Sec) biosynthesis and incorporation into selenoproteins⁴³. The four unexpected candidates selected vary widely in structure and currently defined function. To assess whether these are bona fide splicing factors, we applied functional genomics approaches to investigate the activity of the unexpected candidates in cells.

We first interrogated endogenous RNA targets and transcriptome-wide binding sites of the unexpected candidates using enhanced cross-linking immunoprecipitation (CLIP) followed by sequencing (eCLIP)⁴⁴ in HEK293T cells. For TRNAUIAP, we performed eCLIP using an immunoprecipitation (IP)-grade specific antibody⁴⁵. For the other unexpected hits that did not have IP-grade antibodies available, we expressed V5-tagged ORFs and performed eCLIP with a validated V5 antibody. We successfully completed IP for all replicates (Extended Data Fig. 4a). We retrieved enriched windows using the Skipper pipeline⁴⁶ and found them to be reproducible across two independent replicates each for all eCLIP experiments (concordance odds ratio (OR) > 9× for all experiments; Extended Data Fig. 4b).

To determine the RNA region preferences of the candidate proteins, we examined the region annotation of all reproducible enriched windows from the eCLIP signals (Fig. 3a). STAU2 reproducible enriched windows were represented most frequently in intronic regions and

the 3' untranslated region (UTR) (also consistent with its known role in RNA localization). The reproducible enriched windows of SCAF8 were frequently near splice junctions, indicative of splicing regulation, with a relatively even distribution of regions otherwise. RTCA displayed widespread binding (>100,000 reproducible enriched binding windows), with a robust preference for coding sequence and 3' UTR (consistent with its role in 3' RNA processing) binding and a strong under-enrichment of intronic binding when compared with the other candidates. TRNAUIAP binding sites showed a stark preference for intronic binding, resembling the binding patterns of some well-described splicing factors, such as RBFOX2 and HNRNPC²⁸. From region binding alone, we saw patterns in SCAF8 and TRNAUIAP binding that are reflective of known splicing factor binding and patterns among the other candidates that indicate that, although the proteins may be able to modulate splicing, they play major roles in other RNA processing steps as well.

Next, we performed motif analysis on the reproducible enriched windows in the eCLIP signal for each of the unexpected hits (Fig. 3b). The top motif for RTCA is part of the known exonic splicing enhancer hexamer sequence 5'-GAAGAA-3' (ref. 47). The top motif for SCAF8 is a poly(G) run, associated with AS regulation^{48,49}. Overall, examination of the top motif contained within each of the eCLIP signals revealed that RTCA and SCAF8 bind to signals associated with splicing regulation.

To investigate whether these RBPs modulate AS of endogenous RNA, we performed shRNA-mediated knockdown followed by RNA sequencing (RNA-seq) analysis in HEK293T cells with shRNAs specific to these proteins. Knockdowns of all targets were successful, with knockdown of at least 50% as measured by transcripts per million (TPM) (Extended Data Fig. 4c). We examined the differential AS events after knockdown and detected differentially spliced events for all knockdowns (Fig. 3c). To simplify characterization, we performed further analysis on differentially spliced events of the skipped exon (SE) category. At least 30 differential SE events were driven by the knockdown of each of these candidates. For RTCA and TRNAUIAP, more than 500 differentially spliced events were detected. We determined the direction of splicing change for each differentially spliced SE event (Fig. 3d). As the initial screens were designed to detect RBPs with the potential to induce exon inclusion, we expected to observe splicing events with increased skipping upon knockdown. We observe this trend for TRNAUIAP, indicating that TRNAUIAP is endogenously driving exon inclusion, matching our prediction from the screens. The other candidates did not display the same trend. Nevertheless, they cannot be eliminated as direct drivers of exon inclusion at this stage, because final AS outcome also captures participation of the unexpected hits in upstream pathways and competitive effects with other splicing factors⁵⁰. The data here indicate that the candidates each play roles in AS regulation of some events, with TRNAUIAP and RTCA modulating many SE events.

To nominate AS exons that could be regulated by direct binding, we integrated findings from eCLIP and RNA-seq. We found that genes containing knockdown-sensitive exons are bound at a significantly higher rate than genes lacking knockdown-sensitive exons by SCAF8, RTCA and TRNAUIAP but not by STAU2 (Fig. 3e,f). Although the count of genes containing knockdown-sensitive SE events is low for STAU2 in comparison to the count of genes bound, the events in which there is overlap could be directly driven by binding; however, this appears to be a more specific than widespread phenomenon, at least in HEK293T cells. RTCA binds to most genes containing knockdown-sensitive SE events, indicating that the binding of RTCA directly drives many splicing changes. TRNAUIAP and SCAF8 both bind a substantial portion of genes with knockdown-sensitive SE events. Splicing modulation of these events may be directly driven by this binding. Some of the non-bound differential splicing events could be driven by their roles in pathways upstream of splicing outcome or could be bound at levels below the detection sensitivity of eCLIP.

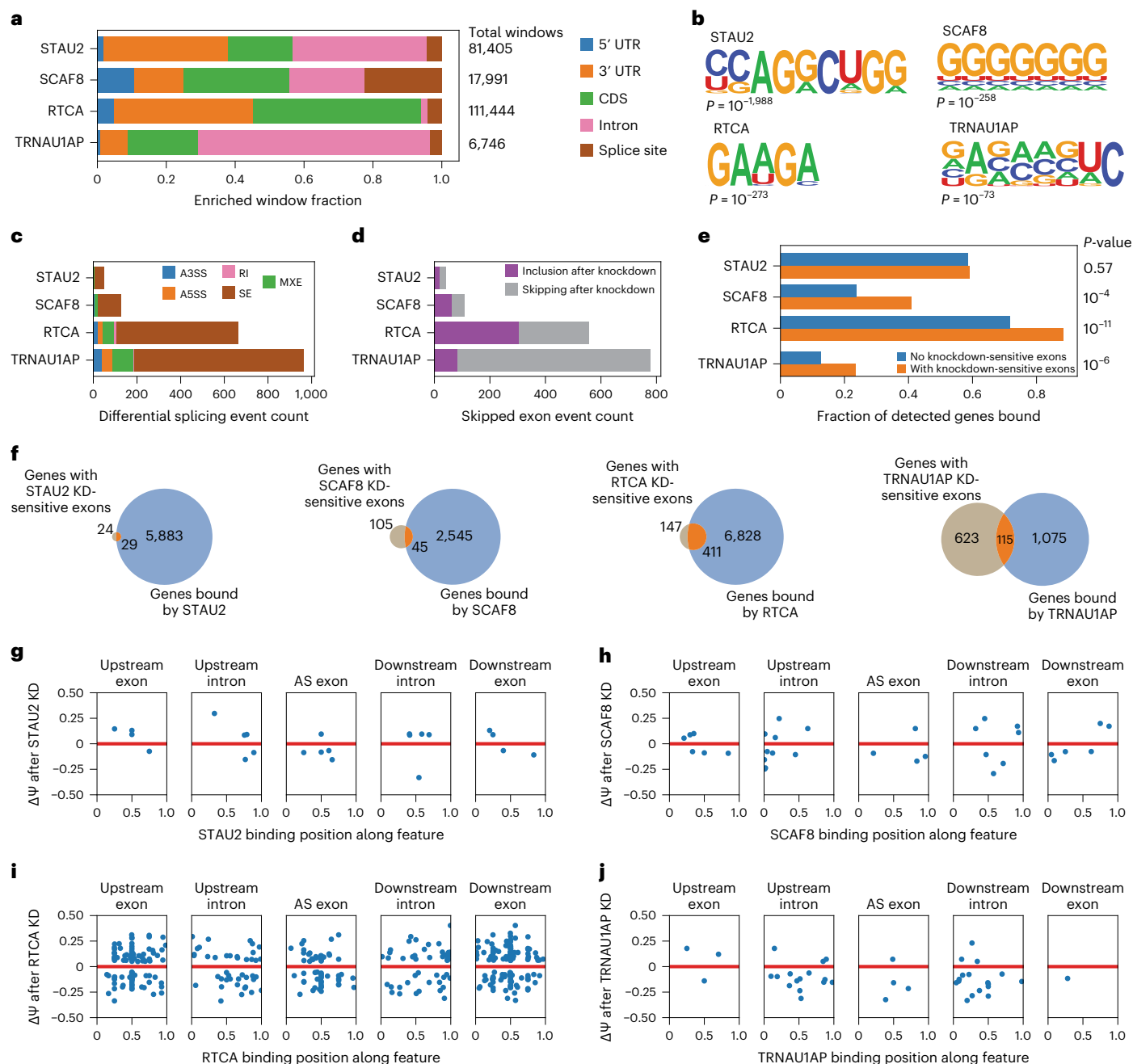


Fig. 3 | Integrated analysis of eCLIP and knockdown RNA-seq reveals splicing events modulated by unexpected hits from the tethering assay. a, Stacked bar graph displaying distribution of RBPs of unexpected hits identified by eCLIP and Skipper analysis ($n = 2$ for IP and sized-matched input). **b**, The most significantly enriched RNA binding motifs as identified by HOMER³⁷ analysis of eCLIP signal of unexpected hits along with corresponding P value, as calculated by the HOMER algorithm using default settings. **c**, Stacked bar graph of differentially spliced events after shRNA-mediated knockdown of unexpected hits as identified by rMATS analysis of RNA-seq data. Differentially spliced events are called as those with an inclusion level difference > 0.05 and a multiple hypothesis-adjusted $P < 0.05$ as calculated by the likelihood ratio test. **d**, Stacked bar graph of SE events after shRNA-mediated knockdown of unexpected hits as identified by rMATS analysis of RNA-seq data. Differentially spliced events are called as those with an inclusion level difference > 0.05 and a multiple hypothesis-adjusted $P < 0.05$ as calculated by likelihood ratio test. Events called as ‘skipping after

knockdown’ are differentially spliced events with $\text{InclLevelKD} - \text{InclLevelNT} < 0$. Events called as ‘inclusion after knockdown’ are differentially spliced events with $\text{InclLevelKD} - \text{InclLevelNT} > 0$. **e**, Bar graph showing the fraction of genes containing significantly enriched windows after eCLIP of unexpected hits, binned into genes containing corresponding knockdown-sensitive skipped exon events and those without. P value was calculated by one-tailed Fisher exact test. **f**, Venn diagrams displaying the number of genes containing differentially spliced skipped exon events after unexpected hit knockdown and the number of genes containing significantly enriched windows after the corresponding eCLIP. **g–j**, Scatter plots examining genes containing unexpected hit knockdown-sensitive skipped exon events and corresponding significantly enriched binding windows. Binding positions were stratified by feature relative to the skipped exon. When multiple binding windows were identified on the same feature relative to an exon, the median binding window is displayed. **g**, STAU2. **h**, SCAF8. **i**, RTCA. **j**, TRNAU1AP. KD, knockdown.

Altogether, RTCA, SCAF8 and TRNAUIAP appear to directly regulate many SE events through binding, whereas STAU2 appears to do this in a more limited capacity.

To investigate individual cases of our candidates directly driving AS modulation through position-dependent binding, we generated maps of knockdown-sensitive splicing events containing nearby binding signal. We found instances of candidate RBP binding to knockdown-sensitive exons as well as flanking introns and exons and plotted the center of the reproducible enriched binding windows across these features against the change in exon inclusion level after knockdown (Fig. 3g–j). At the few sites with STAU2 binding and STAU2 knockdown-sensitive splicing, no clear pattern emerges, indicating that direct STAU2-mediated splicing change is not a widespread and generalized phenomenon (Fig. 3g). Binding of SCAF8 is distributed throughout AS exons as well as the flanking introns and exons (Fig. 3h). SCAF8 frequently binds at the upstream 5' splice site of exons that are skipped after knockdown. RTCA binding is prevalent in AS exons, flanking introns and flanking exons, with most prevalent binding in the flanking exons (Fig. 3i). We detected knockdown-sensitive splicing changes in both directions with nearby RTCA binding. TRNAUIAP commonly binds the flanking introns of exons that are skipped after knockdown, with a cluster present at the downstream 5' splice site, implying that TRNAUIAP binds downstream of alternatively spliced exons and induces exon inclusion (Fig. 3j). This matches the position-dependent effect captured in the initial screen. To visualize specific instances of direct splicing regulation, we generated genome tracks of sample targets with knockdown-sensitive differential splicing and nearby eCLIP signal for TRNAUIAP, RTCA, SCAF8 and STAU2 (Extended Data Fig. 4d). In summary, we used integrated analysis of eCLIP and knockdown RNA-seq to identify instances of direct SE modulation by binding of STAU2, SCAF8, RTCA and TRNAUIAP with SCAF8, RTCA and TRNAUIAP displaying interesting position-dependent modulatory trends.

Splicing protein enrichment in pulldown of unexpected hits

Splicing occurs through assembly and action of complexes consisting of multiple proteins and RNAs, including core spliceosomal components and non-essential splicing factors. To examine if splicing-associated proteins interact with our candidates, we performed affinity purification–mass spectrometry (AP–MS) of V5-tagged TRNAUIAP, RTCA, SCAF8 and STAU2 expressed in HEK293T cells (Fig. 4 and Supplementary Table 17). We performed AP–MS in the absence of ribonuclease, allowing the detection of both proteins that interact directly with our candidates as well as proteins that our candidates associate with through nearby binding on RNA substrates. We aimed to include these RNA-mediated associations, because mutual binding to the snRNAs of the spliceosome or nearby splice sites on mRNA can indicate interactions during splicing. Replicates were highly correlated, and each bait protein was present among the top preys in corresponding samples (Extended Data Fig. 5). We also performed AP–MS with a known splicing-associated protein (CLK2), a tag-only control (FLAG-V5) and two RBPs from the screens that did not emerge as hits (PRKRA and GPATCH2).

We examined the enrichment of splicing-associated proteins (annotated with GO:0008380 RNA splicing, GO:0005681 Spliceosomal Complex or any of their child terms) in each of the AP–MS samples that were significantly enriched (z -score > 2) in at least one of the AP–MS samples (Fig. 4a). Setting aside the tag-only control, the baits separated into two clusters, one with high enrichment of splicing-associated proteins among the preys and the other with low enrichment. The low-enrichment cluster consists of the two non-activating controls and STAU2. Nevertheless, STAU2 is still enriched for interactions with a subset of splicing-associated proteins over the non-targeting controls, potentially due to it performing a limited, auxiliary role in splicing. The high-enrichment cluster consists of the known splicing-associated

protein CLK2 as well as TRNAUIAP, SCAF8 and RTCA, candidates that also displayed widespread direct modulation of AS of endogenous targets. Overall, the increased enrichment of splicing-associated proteins in the TRNAUIAP, SCAF8 and RTCA AP–MS samples provides supporting evidence for them performing widespread splicing regulation.

We also performed GO enrichment on the significantly enriched preys as detected by Spectronaut ($q < 0.05$ and \log_2 ratio IP/FLAG > 1) with each of the candidates as bait (Fig. 4b). The splicing-associated GO term 'regulation of mRNA splicing, via spliceosome' was among the most highly enriched in the significantly enriched preys pulled down by TRNAUIAP and SCAF8. No splicing-associated GO terms were enriched among the significantly enriched preys pulled down by RTCA. The splicing-associated GO term 'regulation of mRNA splicing, via spliceosome' was enriched in the preys pulled down by STAU2 but was not among the top terms. Following the initial evidence of splicing-associated protein enrichment after TRNAUIAP, SCAF8 and RTCA pulldown, we matched these experiments with ribonuclease-positive conditions as well as matching IgG controls in \pm ribonuclease conditions to distinguish between direct protein–protein interactions and RNA-mediated interactions (Fig. 4c,d)⁵¹. We applied a strict P value cutoff of 0.00000001 to visualize the most specific RBPs and splicing-associated proteins pulled down by each bait. The unfiltered output from follow-up experiments can be found in Supplementary Table 18. Overall, we used AP–MS to indicate that splicing-associated proteins are enriched after pulldown of TRNAUIAP, SCAF8 and RTCA and to identify the specific modes by which these proteins interact with RBPs and splicing-associated proteins.

AS modulation by TRNAUIAP

Owing to strong evidence across the eCLIP, knockdown RNA-seq and AP–MS data indicating the activity of TRNAUIAP as a splicing factor, we examined the protein in further detail. We first investigated the finding that most genes with TRNAUIAP knockdown-sensitive skipped exon events did not contain reproducible enriched binding windows from the eCLIP data. We considered the hypothesis that some of this effect could be explained by TRNAUIAP indirectly regulating splicing events through modulating the splicing of other splicing factors. This multi-layered control of splicing has been shown in the recently characterized splicing factor DAP3 (ref. 52) as well as in the SR family of splicing factors²⁰. To investigate this, we examined the top differentially expressed and differentially spliced genes with RNA splicing GO terms (splicing-associated genes) after TRNAUIAP knockdown.

The top differentially expressed splicing-associated gene was *PRPF39* (Fig. 5a), and the top two differentially spliced splicing-associated genes were *PRPF39* (at an unannotated poison exon) and *HNRNPA2B1* (at exon 2, responsible for isoform switching between HNRNPA2 and HNRNPB1) (Fig. 5b). In TRNAUIAP knockdown, presence of the *PRPF39* poison exon is virtually eliminated, and *PRPF39* TPM increases from 46.06 ± 3.62 to 117.34 ± 5.06 (mean \pm s.d.). TRNAUIAP binds in the intron downstream on this poison exon (Fig. 5c, left). We performed western blots to validate that the increase in *PRPF39* expression after TRNAUIAP knockdown is reflected at the protein level and detected a two-fold increase in HEK293T cells (Fig. 5d,e and Extended Data Fig. 6a). Due to the extent of poison exon elimination in the knockdown condition, TRNAUIAP appears to be the primary driver of poison exon-mediated expression control of *PRPF39* in HEK293T cells. As an initial investigation to test the hypothesis of *PRPF39* acting as a direct effector for certain TRNAUIAP knockdown-sensitive AS events, we analyzed *PRPF39* eCLIP signal in HepG2 cells generated by the ENCODE consortium⁴⁵. We found that *PRPF39* reproducible enriched binding windows are prevalent in a significantly higher percentage of introns flanking TRNAUIAP-sensitive exons than TRNAUIAP-insensitive exons, supporting the hypothesis (Fig. 5f). We also examined another TRNAUIAP-sensitive splicing factor exon, *HNRNPA2B1* exon 2, which also contains TRNAUIAP binding

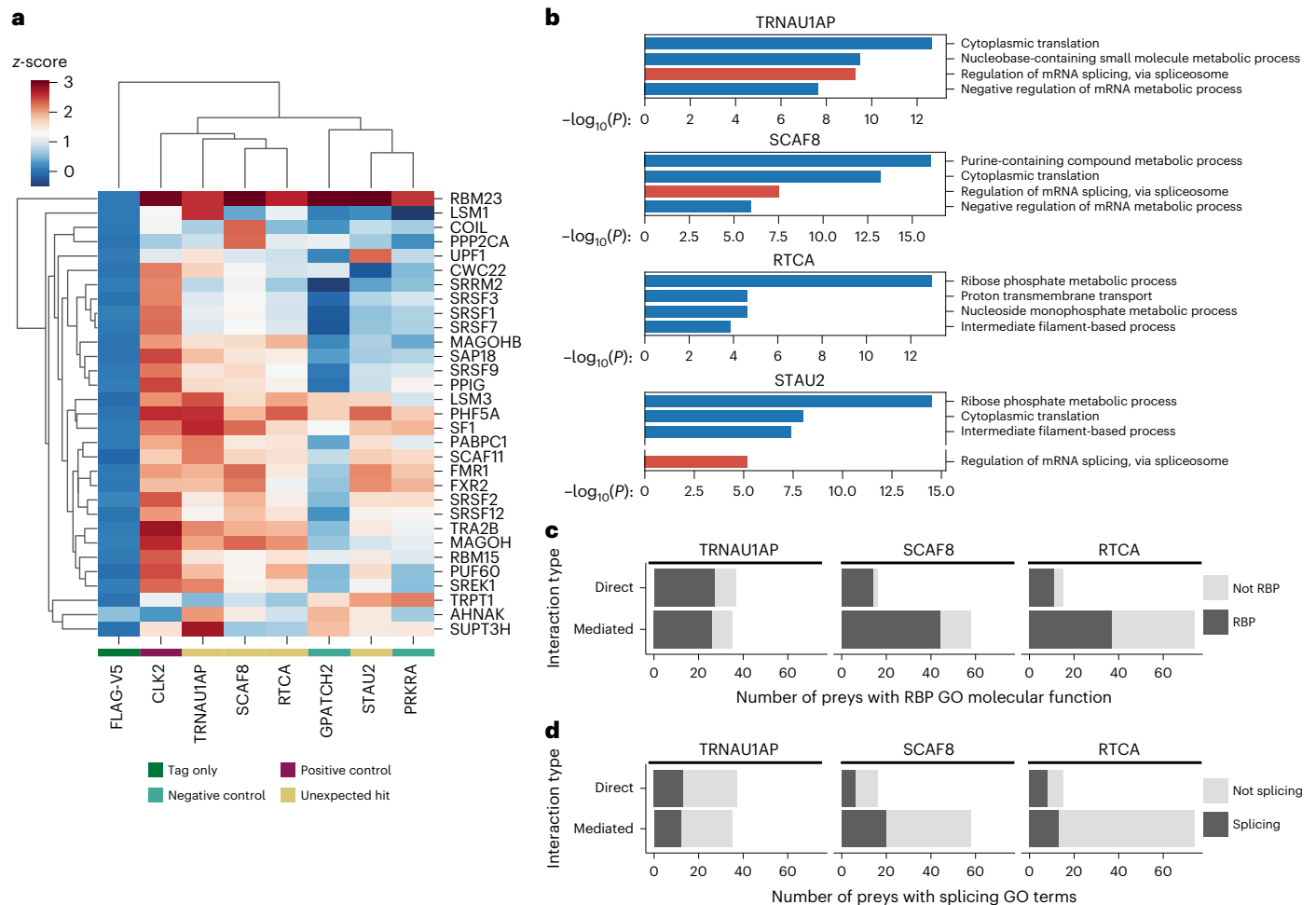


Fig. 4 | AP-MS identifies splicing-associated proteins after pulldown of candidate proteins. **a**, Hierarchically clustered heat map displaying z-scores from AP-MS of annotated splicing-associated preys that were detected with z-score > 2 as calculated by Spectronaut by mass spectrometry after affinity purification of any of the baits. Preys are displayed on the y axis, and baits are displayed on the x axis. **b**, Bar graphs displaying GO analysis of preys detected by mass spectrometry after affinity purification of the unexpected hits from the tethering assay. Significantly enriched preys for GO analysis are defined as having $q < 0.05$ and \log_2 ratio IP/FLAG > 1 (multiple hypothesis corrected, determined by Spectronaut algorithm). Splicing-associated GO terms are

highlighted in red. Gaps in the y axis are used to visualize the most highly enriched splicing-associated GO term when one was not present in the top four. Unadjusted P value of enrichment was calculated using Metascape⁵⁶ based on the accumulative hypergeometric distribution. **c,d**, Stacked bar graphs displaying the overall count of preys significantly detected in follow-up experiments over IgG controls (fold change > 0.5, unadjusted $P < 0.00000001$ as determined by the Spectromine algorithm), separated by interaction type (Methods). The count of preys that are annotated as RBPs (**c**) and contain RNA splicing GO terms (**d**) are displayed.

sites in the downstream intron and is virtually eliminated in TRNAU1AP knockdown (Fig. 5b,c, right). This implicates TRNAU1AP as the primary driver of isoform switching of *HNRNPA2B1* in HEK293T cells. Here, we showed that TRNAU1AP binds to the downstream intron of, and drives the inclusion of, exons in *PRPF39* and *HNRNPA2B1*, which likely drives further widespread splicing changes.

To identify the effector domain bestowing TRNAU1AP's ability to drive exon inclusion, we then performed a series of truncation experiments. We cloned truncations (Fig. 5g) into MCP fusions using the same backbone as the RBP library in the initial tethering screen. We co-transfected MCP-fused TRNAU1AP truncations with both splicing reporters, attempting to identify the region of the protein sufficient to drive the downstream-only effect captured in the screen (Fig. 5h). The C-terminal domain captured in truncations TRNAU1AP-4 and TRNAU1AP-5 appears to be responsible for most, but not all, of the exon inclusion driving activity of the full-length protein. This allowed us to build a domain model that matches the standard simplified model of an RBP, consisting of independent and separate effector and binding domains—in this case, an RNA-binding RRM-containing

domain at the N-terminus and an exon inclusion activating effector domain at the C-terminus.

To ensure that the exon-including capacity of TRNAU1AP and its C-terminal effector domain is not dependent on the MS2-MCP interaction, we cloned CRISPR artificial splicing factors by fusing TRNAU1AP-5 and full-length TRNAU1AP to catalytically dead Cas13d. We co-transfected these artificial splicing factors with a version of the luc-MAPT splicing reporter lacking MS2 stem loops, along with individual gRNA plasmids targeting the introns upstream and downstream of the alternatively spliced exons (Fig. 5i). Both full-length TRNAU1AP and TRNAU1AP-5 significantly drove exon inclusion as measured by the tethering-free reporter when co-transfected with gRNAs targeting downstream of the alternatively spliced exon but not with those targeting upstream (Fig. 5j,k and Extended Data Fig. 6b). These results are consistent with the downstream-only result from the tethering assays and show that the ability of TRNAU1AP and its C-terminal effector domain to induce exon inclusion is independent of the MS2-MCP interaction. In summary, we show that TRNAU1AP participates in splicing co-regulatory networks and drives exon inclusion through its C-terminal effector domain.

Employing identified domains in artificial splicing factors

Motivated by our results articulating that TRNAUIAP or its domain can be useful in artificial splicing factors, we returned to the original list of top RBPs that altered splicing of our reporter construct and tested various protein truncations of these with the aim of determining minimal splice-activating domains to repurpose for artificial splicing factors. LUC7L2 and SRSF8 were selected as strong hits that activated splicing both upstream and downstream of the alternative exon (Fig. 6a). SNRPB and FUBP1 were selected as strong hits that activated lucMAPT-30D only (Fig. 6b). U2AF2 and SRSF10 were selected as strong hits that primarily activated exon inclusion when tethered upstream (Fig. 6c). We designed and cloned truncations based on domain structure, assuming modularity of RBPs where effector and binding domains are separate and independent.

Selected truncations were fused to the MS2 coat protein using the same backbone and conditions as the RBP-MCP library (Fig. 6d-f). LUC7L2-4 recapitulated some of the activity of its full-length counterpart, however at substantially lower strength, implying important contributions from the other domains. SRSF8-2, the RS domain of the protein, captured much of the activity of SRSF8. FUBP1-3 captured much of the activity of full-length FUBP1, at a markedly reduced size. SNRPB-1 captured all the activity of SNRPB. Interestingly, SRSF10-2, the RS domain of SRSF10, displayed a different modulation pattern than the full-length protein, where a stronger effect was seen when tethered downstream of the alternatively spliced exon, more in line with all other tested SRSF proteins. U2AF2-2 was the most successful truncation of the proteins that activated only lucMAPT-30U.

We constructed CRISPR-based artificial splicing factors by fusing the truncations that most successfully activated the tethering reporter to catalytically dead Cas13d. These were tested with an MS2-free luciferase splicing reporter and compared with the recently reported RBFOX1N-dCasRx-C artificial splicing factor¹⁹ (Fig. 6g). As expected, RBFOX1N-dCasRx-C activated the reporter only when targeting sites downstream of the alternatively spliced exon, with a maximal ψ of 11.87% with g1. The SRSF8-2-based artificial splicing factor activated the reporter at all positions, with a maximal ψ of 31.34% with g2. The SNRPB-1-based artificial splicing factor activated the reporter only when targeting downstream of the alternatively spliced exon, as for RBFOX1N-dCasRx-C, but with a greater maximal ψ of 19.15% with g1. The U2AF2-2-based artificial splicing factor did not show activation only with upstream gRNAs as expected, although activation was maximized with upstream guide g5 at 18.60%. Altogether, the SNRPB-1 artificial splicing factor directly outperformed RBFOX1N-dCasRx-C; the SRSF8-2 artificial splicing factor provided a stronger tool with reduced position dependence; and the U2AF2 artificial splicing factor introduced a tool with upstream position association.

Activation of endogenous exon inclusion has remained challenging for the field, as the current solutions with antisense oligonucleotides (ASOs) are to block splicing repressor sites, which is not generalizable to exons that lack these. We employed a CRISPR artificial splicing factor based on our strongest activation domain, SRSF8-2, against an endogenous exon. We targeted exon 7 of *HNRNPD* in HEK293T cells, selected for its high expression for facile readout and endogenous inclusion rate of roughly 50% for perturbation detection. We compared our SRSF8-2 artificial splicing factor to the previous RBFOX1N-dCasRx-C artificial splicing factor by co-transfecting each with plasmids containing arrays of three gRNA sequences separated by repeats that are processed by Cas13d into independent guides. RBFOX1-dCasRx-C was not able to activate endogenous *HNRNPD* exon 7 inclusion with either of the gRNA arrays, whereas SRSF8-2 was able to with both arrays, especially the upstream array (Fig. 6h and Extended Data Fig. 6c,d). Exon 7 of *HNRNPD* appears to be most sensitive to inclusion, driving perturbation with effector domains guided to the upstream 3' splice site, which is incompatible with the downstream-only effect of RBFOX1-dCasRx-C but can be driven by SRSF8-2, exemplifying the importance of its generalizability. Furthermore, the stronger SRSF8-2 appeared to cross an activation threshold when guided to the downstream 5' splice site, whereas the weaker RBFOX1-dCasRx-C did not. In summary, our tethering assay and reporter system also allowed us to identify small and potent effector domains that we used to improve synthetic splicing modulatory proteins.

Discussion

We developed tethering assays and used these to assess the ability of 718 RBPs to induce exon inclusion after recruitment nearby an alternatively spliced cassette exon. Of the 718 RBPs evaluated, 58 reliably enhanced inclusion. Forty-seven of these 58 were annotated with splicing-associated GO terms, and 11 of these were previously unknown as performing any role in AS. We further applied our assays for technology development by using them to rapidly test exon inclusion activation domains identified from the top candidates for use in engineered splicing factors. By fusing these identified domains to catalytically dead Cas13d, we built CRISPR-based artificial splicing factors that are smaller, more potent and less restricted than current technologies. Our tethering assays served as fast, scalable and reliable platforms for both applications.

We employed eCLIP, AP-MS and shRNA knockdown followed by RNA-seq to endogenous TRNAUIAP, SCAF8, RTCA and STAU2, which, excitingly, provided evidence for regulation of splicing outcomes. We further implicated TRNAUIAP as a multi-layered regulator of splicing that also acts in splicing regulatory networks by modulating the splicing of other splicing factors. We performed AP-MS in ribonuclease-free conditions and detected splicing-associated

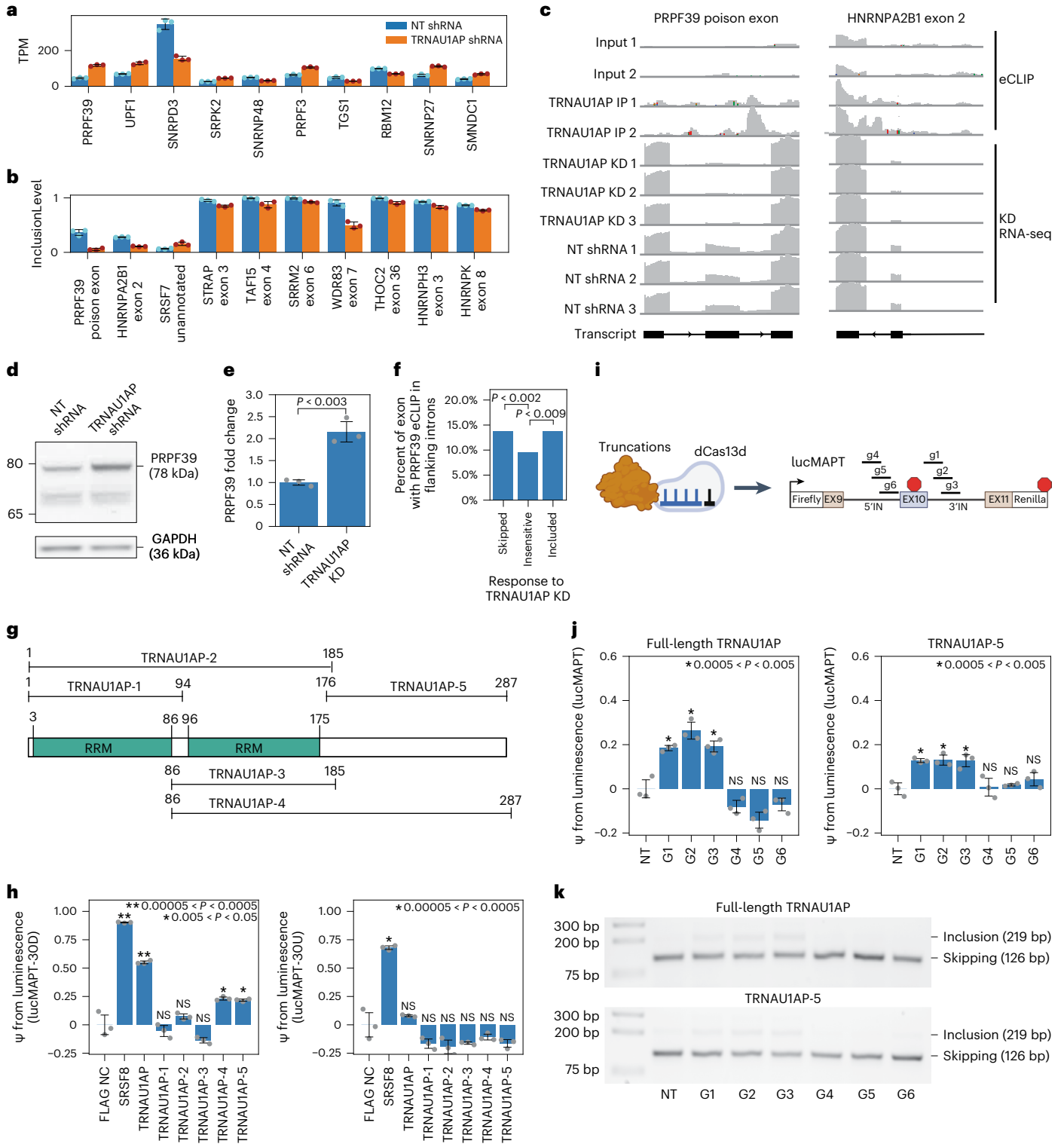
Fig. 5 | TRNAUIAP participates in splicing co-regulatory networks and activates exon inclusion through a C-terminal effector domain. a, Bar graph showing relative expression level of the top 10 differentially expressed splicing-associated genes as sorted by DeSeq2-determined adjusted *P* value after TRNAUIAP knockdown (mean \pm s.d., *n* = 3 replicate transductions). **b**, Bar graph showing relative exon inclusion level of the top 10 differentially spliced skipped exon events in splicing-associated genes as sorted by rMATS-determined adjusted *P* value after TRNAUIAP knockdown (mean \pm s.d., *n* = 3 replicate transductions). **c**, IGV browser tracks showing coverage of TRNAUIAP eCLIP signal relative to size-matched input and TRNAUIAP knockdown RNA-seq signal relative to non-targeting shRNA at a poison exon in PRPF39 and exon 2 of HNRNPA2B1. **d**, Representative western blot showing increased PRPF39 expression in HEK293T cells after TRNAUIAP knockdown. GAPDH is the loading control. **e**, Bar graph showing fold change of PRPF39 expression as quantified by western blot after TRNAUIAP knockdown (mean \pm s.d., *n* = 3 replicate transductions). *P* = 0.0024 by two-tailed independent two-sample *t*-test. **f**, Bar plot displaying percentage of exons containing PRPF39 reproducible enriched eCLIP windows in flanking introns

from ENCODE HepG2 data, separated by exon sensitivity to TRNAUIAP knockdown in HEK293T cells. *P* values are calculated using the two-sided chi-squared test. *P* = 0.0011 for PRPF39 binding to exons skipped after TRNAUIAP knockdown and 0.0088 for exons included after TRNAUIAP knockdown. **g**, Domain structure of TRNAUIAP with truncations used for effector domain identification. **h**, Bar graphs displaying reporter readout from both lucMAPT-30U and lucMAPT-30U co-transfected with MCP-fused truncations (mean \pm s.d., *n* = 3 replicate transductions). *P* value was calculated by one-tailed independent two-sample *t*-test. NS, not significant (*P* > 0.05). **i**, Schematic of truncation-dCas13d fusions used as for MS2-free tests. Schematic of MS2-free lucMAPT reporter used and associated guide RNAs. **j,k**, Reporter readouts from co-transfection of the MS2-free lucMAPT reporter, either full-length TRNAUIAP-dCas13d fusion or truncated TRNAUIAP-5-dCas13d fusion, and each guide RNA annotated in **i**. **j**, Bar graph showing PSI calculated from luminescence (mean \pm s.d., *n* = 3 replicate transductions). *P* value was calculated by one-tailed independent two-sample *t*-test. NS, not significant (*P* > 0.05). **k**, Splicing gels displaying lucMAPT AS. bp, base pairs; KD, knockdown.

proteins after pulldown of TRNAUIAP, RTCA and SCAF8, further supporting their role in splicing. Findings here are limited by the sensitivity and specificity of the assays chosen as well as potential tissue specificity of effects on splicing of the chosen proteins. Future work should investigate the role of these proteins on splice site selection in orthogonal models and employ further validation approaches, such as minigene assays of specific splicing events and co-IP western blots, to validate interaction partners.

Furthermore, the functional consequences of splicing modulation by TRNAUIAP, SCAF8, RTCA and STAU2 in health and disease remain to

be investigated. The splicing regulatory network formed by TRNAUIAP and PRPF39 deserves further investigation. TRNAUIAP and PRPF39 were recently identified as a co-dependency module that is selectively essential in cells carrying mutational signatures of DNA mismatch repair⁵³. The interaction of TRNAUIAP regulating PRPF39 expression through poison exon inclusion described here provides a mechanistic hypothesis for this finding. Furthermore, both genes are prognostic markers in a variety of cancer types⁵⁴. As our scope is limited to the introduction and initial characterization of these proteins in splicing regulation, we are excited for future investigations.



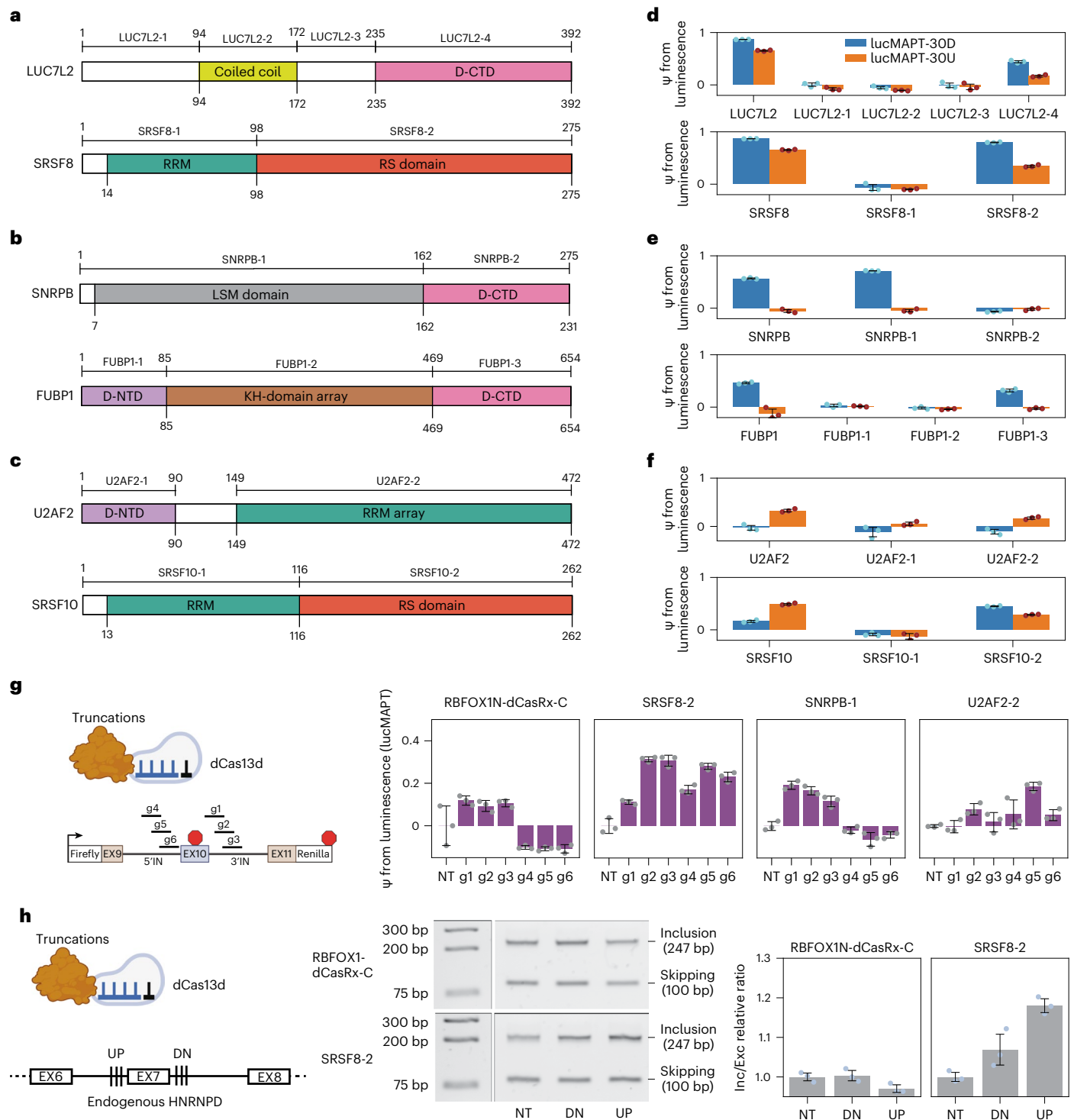


Fig. 6 | Truncation of the top RBP hits identify splice-enhancing domains that can be repurposed for artificial splicing factors. **a–c**, Domain structures of top hits used for truncation experiments; D-NTD and D-CTD represent N-terminal and C-terminal domains, respectively, containing MobiDB-lite consensus disorder prediction. All tested truncations are shown. Hits are separated into their position dependence from the initial screen: position-independent hits (**a**), hits that primarily activated the lucMAPT-30D reporter (**b**) and hits that primarily activated the lucMAPT-30U reporter (**c**). **d–f**, Bar graphs displaying reporter readout from both lucMAPT-30U and lucMAPT-30U of the full-length proteins next to their associated truncations (mean \pm s.d., $n = 3$ replicate transfections). Graphs are separated by position dependence of full-length protein from the initial screen: position-independent hits (**d**), hits that primarily activated the lucMAPT-30D reporter (**e**) and hits that primarily activated the lucMAPT-30U reporter (**f**). **g**, Left, top, schematic of truncation–dCas13d

fusion used as artificial splicing factors. Left, bottom, schematic of MS2-free lucMAPT reporter used for reporter-based assessment of artificial splicing factors. Right, bar graphs displaying reporter output from MS2-free lucMAPT reporter after co-transfection of reporter with truncation–dCas13d fusion and gRNA-containing plasmid (mean \pm s.d., $n = 3$ replicate transfections). **h**, Left, top, schematic of truncation–dCas13d fusion used as artificial splicing factor. Left, bottom, schematic of HNRNPd exon 7 used for endogenous splicing modulation, with the position of the two sets of three gRNAs that are co-transfected with the artificial splicing factors as gRNA arrays. Middle, agarose gel showing splicing of HNRNPd exon 7 of a sample replicate for both artificial splicing factors in co-transfection with both gRNA arrays and a non-targeting gRNA (NT). Right, bar graphs displaying quantification of inclusion/exclusion ratio normalized to the non-targeting gRNA (NT) from gels in Extended Data Fig. 6c,d (mean \pm s.d., $n = 3$ replicate transfections). bp, base pairs.

Our SNRPB-1 artificial splicing factor maintained the downstream targeting specificity of the prior RBFOX1N-dCasRx-C artificial splicing factor but with higher potency and a reduced size. We also identified exon activation domains with different specificity requirements. Our U2AF2 artificial splicing factor has maximum potency when targeted upstream of an AS exon, whereas our SRSF8-2 artificial splicing factor is the strongest thus far and maintains potency with proximity to the AS exon independent of orientation. This orientation independence proved important in our targeting of endogenous *HNRNP D* exon 7, where SRSF8-2 successfully activated exon inclusion and RBFOX1N-dCasRx-C did not.

A limitation of our assays is the potential of false negatives, and RBPs testing negative could still play a role enhancing exon inclusion in different contexts. Our work with lucMBNL1 exemplifies this by demonstrating a sequence context around an AS exon that responds only to a small subset of RBPs that induced exon inclusion in lucMAPT. Future studies that employ tethering approaches in a variety of minigene contexts could identify additional hits with different RNA sequence requirements. Loss of function due to the C-terminal MCP fusion might also explain false negatives in our screens. Nevertheless, these assays have provided the first of possibly many comprehensive investigations of proximity-dependent direct activators of exon inclusion. As the reporters were, to a small extent, sensitive to NMD, caution should be raised when using them in applications across different NMD environments or in applications that may detect changes in the processing of mature reporter mRNA. However, there is potential for NMD sensitivity to be engineered away in future versions of the reporter by relying on alternative exon-induced frameshift to halt translation in the final constitutive exon as opposed to introducing a stop codon in the alternative exon.

We anticipate utility in future studies from our methodology in large-scale discovery of RBPs that enhance exon inclusion by proximity, from our introduction and molecular characterization of previously uncharacterized AS proteins and from our development of small and potent molecular parts for engineered splicing modulation. Future studies could be used to examine the ~2,000 predicted human RBPs not included in our assays. Our engineered splicing domains can be used in future work for delivery through adeno-associated virus (AAV) with their reduced size over current technologies in models incompatible with transfection, and the increased potency can lower dose requirements and expand applicability of the technology. These minimal and potent splicing domains can also be recruited to RNA targets through other means than dCas13d, such as through PUF proteins¹⁸ or CRISPR-Cas-inspired RNA targeting systems (CIRTS)⁵⁵. Altogether, we are optimistic that future approaches will leverage the principles presented here to further explore the landscape of splicing regulation.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41587-023-02014-0>.

References

- Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15**, 829–845 (2014).
- Queiroz, R. M. L. et al. Comprehensive identification of RNA–protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nat. Biotechnol.* **37**, 169–178 (2019).
- Jiang, W. & Chen, L. Alternative splicing: human disease and quantitative analysis from high-throughput sequencing. *Comput. Struct. Biotechnol. J.* **19**, 183–195 (2021).
- Wheeler, E. C. et al. Integrative RNA-omics discovers *GNAS* alternative splicing as a phenotypic driver of splicing factor–mutant neoplasms. *Cancer Discov.* **12**, 836–855 (2022).
- Bradley, R. K. & Anczuków, O. RNA splicing dysregulation and the hallmarks of cancer. *Nat. Rev. Cancer* **23**, 135–155 (2023).
- Scotti, M. M. & Swanson, M. S. RNA mis-splicing in disease. *Nat. Rev. Genet.* **17**, 19–32 (2016).
- Rogalska, M. E., Vivori, C. & Valcárcel, J. Regulation of pre-mRNA splicing: roles in physiology and disease, and therapeutic prospects. *Nat. Rev. Genet.* **24**, 251–269 (2022).
- Zheng, S., Damoiseaux, R., Chen, L. & Black, D. L. A broadly applicable high-throughput screening strategy identifies new regulators of *Dlg4* (*Psd-95*) alternative splicing. *Genome Res.* **23**, 998–1007 (2013).
- Moore, M. J., Wang, Q., Kennedy, C. J. & Silver, P. A. An alternative splicing network links cell-cycle control to apoptosis. *Cell* **142**, 625–636 (2010).
- Tejedor, J. R., Papasaikas, P. & Valcárcel, J. Genome-wide identification of Fas/CD95 alternative splicing regulators reveals links with iron homeostasis. *Mol. Cell* **57**, 23–38 (2015).
- Sun, S., Zhang, Z., Fregoso, O. & Krainer, A. R. Mechanisms of activation and repression by the alternative splicing factors RBFOX1/2. *RNA* **18**, 274–283 (2012).
- Yeo, G. W. et al. An RNA code for the FOX2 splicing regulator revealed by mapping RNA–protein interactions in stem cells. *Nat. Struct. Mol. Biol.* **16**, 130–137 (2009).
- Lovci, M. T. et al. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat. Struct. Mol. Biol.* **20**, 1434–1442 (2013).
- Barash, Y. et al. Deciphering the splicing code. *Nature* **465**, 53–59 (2010).
- Tycko, J. et al. High-throughput discovery and characterization of human transcriptional effectors. *Cell* **183**, 2020–2035 (2020).
- Luo, E.-C. et al. Large-scale tethered function assays identify factors that regulate mRNA stability and translation. *Nat. Struct. Mol. Biol.* **27**, 989–1000 (2020).
- Bos, T. J., Nussbacher, J. K., Aigner, S. & Yeo, G. W. Tethered function assays as tools to elucidate the molecular roles of RNA-binding proteins. In *RNA Processing* (ed. Yeo, G. W.) 61–88 (Springer, 2016).
- Wang, Y., Cheong, C.-G., Tanaka Hall, T. M. & Wang, Z. Engineering splicing factors with designed specificities. *Nat. Methods* **6**, 825–830 (2009).
- Du, M., Jillette, N., Zhu, J. J., Li, S. & Cheng, A. W. CRISPR artificial splicing factors. *Nat. Commun.* **11**, 2973 (2020).
- Leclair, N. K. et al. Poison exon splicing regulates a coordinated network of SR protein expression during differentiation and tumorigenesis. *Mol. Cell* **80**, 648–665 (2020).
- Liu, F. & Gong, C.-X. Tau exon 10 alternative splicing and tauopathies. *Mol. Neurodegener.* **3**, 8 (2008).
- Popp, M. W. & Maquat, L. E. Leveraging rules of nonsense-mediated mRNA decay for genome engineering and personalized medicine. *Cell* **165**, 1319–1322 (2016).
- Chamieh, H., Ballut, L., Bonneau, F. & Le Hir, H. NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. *Nat. Struct. Mol. Biol.* **15**, 85–93 (2008).
- Boehm, V. et al. SMG5–SMG7 authorize nonsense-mediated mRNA decay by enabling SMG6 endonucleolytic activity. *Nat. Commun.* **12**, 3965 (2021).
- Binder, J. X. et al. COMPARTMENTS: unification and visualization of protein subcellular localization evidence. *Database* **2014**, bau012 (2014).
- Bondy-Chorney, E. et al. Staufen1 regulates multiple alternative splicing events either positively or negatively in DM1 indicating its role as a disease modifier. *PLoS Genet.* **12**, e1005827 (2016).
- Bondy-Chorney, E., Crawford Parks, T. E., Ravel-Chapuis, A., Jasmin, B. J. & Côté, J. Staufen1s role as a splicing factor and a disease modifier in myotonic dystrophy type I. *Rare Dis.* **4**, e1225644 (2016).

28. Van Nostrand, E. L. et al. A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719 (2020).
29. Ambrozková, M. et al. The fission yeast ortholog of the coregulator SKIP interacts with the small subunit of U2AF. *Biochem. Biophys. Res. Commun.* **284**, 1148–1154 (2001).
30. Selenko, P. et al. Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol. Cell* **11**, 965–976 (2003).
31. Matera, A. G. & Wang, Z. A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* **15**, 108–121 (2014).
32. Cvitkovic, I. & Jurica, M. S. Spliceosome Database: a tool for tracking components of the spliceosome. *Nucleic Acids Res.* **41**, D132–D141 (2013).
33. Chen, Y.-I. G. et al. Proteomic analysis of in vivo-assembled pre-mRNA splicing complexes expands the catalog of participating factors. *Nucleic Acids Res.* **35**, 3928–3944 (2007).
34. Ajuh, P. Functional analysis of the human CDC5L complex and identification of its components by mass spectrometry. *EMBO J.* **19**, 6569–6581 (2000).
35. McCracken, S. et al. Proteomic analysis of SRm160-containing complexes reveals a conserved association with cohesin. *J. Biol. Chem.* **280**, 42227–42236 (2005).
36. Sharma, S., Kohlstaedt, L. A., Damianov, A., Rio, D. C. & Black, D. L. Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nat. Struct. Mol. Biol.* **15**, 183–191 (2008).
37. Rappsilber, J., Ryder, U., Lamond, A. I. & Mann, M. Large-scale proteomic analysis of the human spliceosome. *Genome Res.* **12**, 1231–1245 (2002).
38. Azizian, N. G. & Li, Y. XPO1-dependent nuclear export as a target for cancer therapy. *J. Hematol. Oncol.* **13**, 61 (2020).
39. Heraud-Farlow, J. E. et al. Staufen2 regulates neuronal target RNAs. *Cell Rep.* **5**, 1511–1518 (2013).
40. Almasi, S. & Jasmin, B. J. The multifunctional RNA-binding protein Staufen1: an emerging regulator of oncogenesis through its various roles in key cellular events. *Cell. Mol. Life Sci.* **78**, 7145–7160 (2021).
41. Yuryev, A. et al. The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc. Natl Acad. Sci. USA* **93**, 6975–6980 (1996).
42. Tanaka, N. & Shuman, S. Structure–activity relationships in human RNA 3′-phosphate cyclase. *RNA* **15**, 1865–1874 (2009).
43. Hu, X. et al. Knockdown of *TrnaU1ap* inhibits the proliferation and migration of NIH3T3, JEG-3 and Bewo cells via the PI3K/Akt signaling pathway. *Biochem. Biophys. Res. Commun.* **503**, 521–527 (2018).
44. Van Nostrand, E. L. et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514 (2016).
45. Luo, Y. et al. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.* **48**, D882–D889 (2020).
46. Boyle, E. A. et al. Skipper analysis of eCLIP datasets enables sensitive detection of constrained translation factor binding sites. *Cell Genom.* **3**, 100317 (2023).
47. Fairbrother, W. G., Yeh, R.-F., Sharp, P. A. & Burge, C. B. Predictive identification of exonic splicing enhancers in human genes. *Science* **297**, 1007–1013 (2002).
48. Xiao, X. et al. Splice site strength-dependent activity and genetic buffering by poly-G runs. *Nat. Struct. Mol. Biol.* **16**, 1094–1100 (2009).
49. Georgakopoulos-Soares, I. et al. Alternative splicing modulation by G-quadruplexes. *Nat. Commun.* **13**, 2404 (2022).
50. Warf, M. B., Diegel, J. V., Von Hippel, P. H. & Berglund, J. A. The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing. *Proc. Natl Acad. Sci. USA* **106**, 9203–9208 (2009).
51. Street, L. et al. Large-scale map of RNA binding protein interactomes across the mRNA life-cycle. Preprint at *bioRxiv* <https://doi.org/10.1101/2023.06.08.544225> (2023).
52. Han, J. et al. Multilayered control of splicing regulatory networks by DAP3 leads to widespread alternative splicing changes in cancer. *Nat. Commun.* **13**, 1793 (2022).
53. Chen, X. et al. Context-defined cancer co-dependency mapping identifies a functional interplay between PRC2 and MLL-MEN1 complex in lymphoma. *Nat. Commun.* **14**, 4259 (2023).
54. Uhlén, M. et al. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
55. Rauch, S. et al. Programmable RNA-guided RNA effector proteins built from human parts. *Cell* **178**, 122–134 (2019).
56. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).
57. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024, corrected publication 2024

Methods

Generation of expression plasmids for MCP and dCas13d-fused RBPs and RBP truncations

Most ORF clones were obtained in pENTR vectors from the CCSB human ORFeome collection⁵⁸ (Dana-Farber Cancer Institute) or the DNASU Plasmid Repository (Arizona State University). For truncations, domain structures were determined using InterProScan⁵⁹ on the amino acid sequence of the full-length protein and informed truncation design. Truncations and ORFs that were ordered in standard expression vectors were amplified by PCR (Phusion polymerase, New England Biolabs (NEB)) with oligonucleotide primers containing attB recombination sites and recombined into pDONR221 using BP clonase II (Thermo Fisher Scientific). ORFs were then recombined into one of two custom pEF DEST51 destination vectors (Thermo Fisher Scientific). For MCP fusions, the destination vector is engineered to direct expression of the ORFs as fusion proteins with a V5 epitope tag and MCP appended C-terminally and under the control of the EF1- α promoter to create ORF-V5-MCP constructs. For dCas13d fusions, the MCP is simply replaced with dCas13d for the generation of ORF-V5-dCas13d constructs. Supplementary Table 19 contains sequences of both destination vectors. The identity of all cDNA clones was verified by Sanger sequencing. Plasmid libraries are available on Addgene (155390–156159). Supplementary Table 1 lists all ORFs and relevant information.

Cell lines

Lenti-X HEK293T cells were purchased from Takara Bio and were not further authenticated. Cells were routinely tested for mycoplasma contamination with a MycoAlert mycoplasma test kit (Lonza) and were found negative for mycoplasma.

Generation of constructs

lucMAPT reporter. Reporter was first constructed through a three-fragment Gibson Assembly using a homebrew enzyme mix (OpenWetWare). Fragments were generated by performing PCR on sub-fragments to generate complementary overhangs, followed by annealing, amplification and agarose gel extraction. The first fragment consists of Firefly luciferase, MAPT exon 9 and the 5'-most 500 base pairs of MAPT intron 9. The second fragment consists of the 3'-most 500 base pairs of MAPT intron 9, modified MAPT exon 10 and the 5'-most 500 base pairs of MAPT intron 10. The third fragment consists of the 3'-most 500 base pairs of MAPT intron 10, MAPT exon 11 and Renilla luciferase. Luciferase ORFs were cloned from plasmids used in our laboratory's previous work¹⁶. MAPT exons were ordered as synthetic oligonucleotides. MAPT intronic sequences were amplified from genomic DNA isolated from Lenti-X HEK293T cells. All PCR was performed using KAPA HiFi HotStart ReadyMix (Roche, 7958935001). The assembly strategy is summarized in Extended Data Fig. 1a.

lucMAPT-MS2 reporters. MAPT exon 10 and the flanking 100 intronic base pairs in either direction from the splice sites were removed from the construct and replaced with a cloning site containing BamHI and EcoRI cut sites through PCR, followed by two-fragment Gibson Assembly to generate a customizable backbone. Inserts containing MAPT exon 10, the flanking 100 base pairs and the MS2 stem-loop sequence in the desired position were cloned into this backbone through one-fragment Gibson Assembly into pcDNA3.1 (-) Mammalian Expression Vector (Thermo Fisher Scientific, V79520) to construct lucMAPT-MS2 reporters. Inserts containing other AS exons and flanking sequences were used to generate other reporters used. Sequences of reporters can be found in Supplementary Table 19.

Luciferase reporter screens

Reverse transfection. Ninety-six-well Solid Black Flat Bottom Polystyrene TC-treated Microplates (Corning, 3916) were coated with 75 μ l of

poly-D-lysine hydrobromide (Sigma-Aldrich, P6407-5MG), dissolved in water at 1 g L⁻¹ and further diluted 1:5 in 1 \times DPBS (Corning, 21-031-CV) overnight in a tissue culture incubator. Plates were rinsed two times with 1 \times DPBS and dried. A 1:1 mix of lucMAPT-MS2 reporter and an ORF-V5-MCP construct with a total of 100 ng of DNA were added to a mixture of Lipofectamine 3000 and P3000 reagents (Thermo Fisher Scientific, L3000001), diluted in Opti-MEM Reduced Serum Media (Gibco, 31985062) and incubated for 15 min. The mixture of DNA and transfection reagent was transferred to the PDL-coated 96-well plate. Then, 75 μ l of Lenti-X HEK293T cells was plated at a concentration of 266,666 cells per milliliter. Transfection was incubated for 48 h in a standard tissue culture incubator.

Dual-luciferase readout. Luminescence was generated using the Dual-Glo Luciferase Assay System (Promega, E2980). Cells were removed from the incubator to cool to room temperature for 30 min. Then, 75 μ l of Dual-Glo Luciferase Reagent was added directly to cells and thoroughly mixed using a Microplate Genie Plate Shaker (Scientific Industries). The reaction was briefly centrifuged and allowed to incubate at room temperature for 10 min. Luminescence was measured using a Spark Multimode Microplate Reader (Tecan) with a 500-ms signal interaction time at room temperature. The same process was repeated for Renilla luciferase luminescence using the Dual-Glo Stop & Glo Reagent.

Statistical analysis. Relative ψ values were calculated as described in Fig. 1b using the pandas library in Python version 3.10.11 (ref. 60). All plots generated from Python were generated using JupyterLab 4.04. Significance between candidate and negative control conditions was assessed by calculating P value through a one-tailed independent t -test using the `ttest_ind` function in `scipy`⁶¹.

RNA-level validation of luciferase screens

Transfection was performed as described for the luciferase reporter screens, using standard 96-well tissue culture plates (Costar, 3596). RNA was isolated from cells using the Direct-zol RNA Miniprep Kit (Zymo Research, R2052). cDNA was generated using the ProtoScript II First Strand cDNA Synthesis Kit (Promega, E6560L). cDNA was amplified using GoTaq Green Master Mix (Promega, M7122), and primers were designed for an amplicon stretching from MAPT exon 9 to the Renilla luciferase ORF. Amplicons were run through a 3% SeaKem Agarose Gel (Lonza, 5004) at 100 V for 25 min.

Statistical analysis. Relative band intensity was calculated using the Gel Analyzer feature in ImageJ version 1.53k software⁶². Significance between candidate and negative control conditions was assessed by calculating P value through a one-tailed independent t -test using the `ttest_ind` function in `scipy`⁶¹.

GO analysis

Metascape version 3.5 was used for GO analysis⁵⁶. Custom enrichment analysis for GO Biological Processes was performed using an appropriate set of background genes. `biomaRt` version 2.50.3 was used to identify genes matching specific GO terms from gene lists⁶³. We used `biomaRt` to generate a list of splicing associated genes by selecting genes annotated with `GO:0008380` RNA splicing, `GO:0005681` Spliceosomal Complex or any of their child terms.

Generation of samples overexpressing V5-tagged RBPs

HEK293T cells were plated in 10-cm plates at 10% confluency. Then, 28 ng of plasmid DNA encoding the V5-tagged RBPs was added to a mixture of Lipofectamine 3000 and P3000 reagents (Thermo Fisher Scientific, L3000001), diluted in Opti-MEM Reduced Serum Media (Gibco, 31985062) and incubated for 15 min. The mixture of DNA and transfection reagent was transferred to the plated cells. Cells were

collected 48 h later and washed with 10 ml of DPBS. Samples to be used for eCLIP were UV cross-linked (400 mJ cm^{-2} , 254 nm). Cells were resuspended in 1 ml of DPBS. Samples were centrifuged at 4°C and 18,000g for 1 min. Supernatant was removed, and cells were flash frozen in dry ice before storage at -80°C until experimentation.

eCLIP library preparation and sequencing

eCLIP was performed as per Yeo laboratory standard operating procedures⁴⁴. Antibodies used are listed in Supplementary Table 20. For V5-tagged eCLIPs, overexpression samples were generated as described herein. Samples for endogenous eCLIP were generated using the same procedure without transfection. Two replicates were generated for each experiment. Pellets were lysed, and lysates were subjected to sonication and RNase I to fragment RNA. Ninety-eight percent of each lysate was immunoprecipitated using either V5 (Bethyl, A190-120A) or TRNAUIAP-specific (GeneTex, GTX121631) antibodies, and the remainder was stored for preparation of a SMInput library. Ten micrograms of antibody was used per sample. Pulled-down RNA fragments were dephosphorylated and 3'-end ligated to an RNA adaptor. Immunoprecipitates and SMInputs were run on an SDS-polyacrylamide gel and transferred to a nitrocellulose membrane. Membrane regions from the RBP size to that size plus 75 kDa were excised, and RNA was released with proteinase K. SMInput samples were then dephosphorylated and 3'-end ligated to an RNA adaptor. All samples were reverse transcribed with SuperScript III Reverse Transcriptase (Life Technologies). cDNAs were ligated to a DNA adaptor at the 5' end. cDNA was quantified by qPCR and amplified to 100–500 fmol of library using Q5 PCR Master Mix (NEB). Sequencing was performed using the NovaSeq 3000 platform, with a targeted number of single-ended reads of 40 million per sample.

Computational analysis of eCLIP data

Computational analysis of eCLIP data was performed using the default settings of Skipper resources available on GitHub (<https://github.com/YeoLab/skipper>). Reads were mapped to human genome assembly GRCh38 (ref. 64). For V5-tagged eCLIPs, reproducible enriched windows were first found after transfection and eCLIP of a V5-FLAG negative control plasmid and added to the blacklist file to reduce spurious enrichment from V5 binding to RNA.

shRNA lentiviral production, transduction and sequencing

To generate lentiviral particles for RBP knockdown, we seeded 500,000 HEK293T cells per well in six-well plates. After 24 h, cells in each well were transfected with 500 ng of sequence-verified shRNA plasmid (pLKO.1; Supplementary Table 21) and packaging plasmids (50 ng of pMD2.G; Addgene, 12259; 500 ng of psPAX2; Addgene, 12260—both gifts from Didier Trono, École polytechnique fédérale de Lausanne) using Lipofectamine 3000 (Thermo Fisher Scientific). Transfection media was replaced with 2.5 ml of fresh media after 6 h. Virus-containing medium was collected 48 h later, replaced with 2.5 ml of fresh media and collected again a further 24 h later. Virus-containing media were pooled and stored at -80°C until transduction.

For lentiviral transduction, 500,000 HEK293T cells were seeded per well in each well of a six-well tissue culture plate. After 24 h, media were replaced with 2 ml of virus-containing media supplemented with 16 μg of polybrene. We replaced the virus-containing media with fresh media 24 h later. Twenty-four hours after this, media were replaced with fresh media containing 3 $\mu\text{g ml}^{-1}$ puromycin. Cells were either given fresh puromycin-containing media or passaged every 48 h and expanded to 10-cm plates. Cells were pelleted and flash frozen once all replicates for a given construct had reached 70% confluency or higher.

Total mRNA was extracted from samples using the Direct-zol RNA Miniprep Kit (Zymo Research). RNA quality was verified using TapeStation 3000 (Agilent Technologies). Library preparation was performed using the Stranded mRNA Prep Ligation Kit (Illumina).

Sequencing was performed using the NovaSeq 3000 platform, with a targeted number of paired ended reads of 60 million per sample. Read counts and uniquely mapped reads were verified after STAR version 2.6.7a alignment.

Differential expression analysis

Differentially expressed genes were detected from RNA-seq data using DeSeq2 (ref. 65). We only considered genes expressed with TPM > 10 in the control sample.

Differential splicing analysis

Differential AS events were detected using rMATS 4.0.2 (ref. 66). Splicing events were identified as significantly differentially spliced if the absolute value of inclusion-level difference was detected as greater than 5% and with a false discovery rate (FDR) of less than 5%. We only considered differential splicing events with a sum of ≥ 150 reads across all conditions.

Integrated analysis of eCLIP and shRNA knockdown followed by RNA-seq data

The fraction of knockdown-sensitive or knockdown-insensitive genes containing binding sites from eCLIP was calculated using the number of genes expressed with TPM ≥ 10 from the eCLIP size-matched input as the denominator.

Binding position relative to knockdown-sensitive exons is visualized as the midpoint of the significantly enriched window. For events where multiple significantly enriched windows were present in a single feature, the midpoint of the median window is displayed.

Western blots

Cells were lysed in lysis buffer (see eCLIP protocol) on ice for 15 min and sonicated for 5 min. Lysates were centrifuged at 15,000g for 10 min at 4°C to pellet debris and transferred to a clean tube. Total protein concentration was quantified using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, 23225). For gel electrophoresis, 20 μg was loaded per well onto 4–12% Bis-Tris gels and subsequently transferred to PVDF membranes. Membranes were blocked in 5% milk in TBST solution for 60 min at room temperature. Primary antibodies for UPF1 (Cell Signaling Technology, D15G6, 1:1,000), PRPF39 (Invitrogen, PA5-21627, 1:1,000) and GAPDH (Millipore, MAB374, 1:10,000) were diluted in 5% milk in TBST and probed overnight at 4°C . Secondary antibodies (goat anti-rabbit IgG, HRP-linked, Cell Signaling Technology, 7074, and 800CW, goat anti-mouse IgG, Licor, 926-32210) were diluted at 1:2,000 in 5% milk in TBST and probed for 120 min at room temperature.

AP-MS

HEK293T cells overexpressing V5-tagged RBPs were generated as described herein. Cells were lysed and affinity purified using 10 μg per sample of a V5-specific antibody. In brief, the cell lysates with antibody were incubated with magnetic beads overnight in the cold room. Then, 5 μl of 10 mg ml^{-1} RNase A was added to ribonuclease-positive conditions at this step. Supernatants were removed, and beads were washed four times with NP-40 buffer, twice in Buffer 2 (50 mM Tris (pH 7.5), 150 mM NaCl, 10 mM MgCl_2 , 0.05% NP-40 and 5% glycerol) and twice in Buffer 3 (50 mM Tris (pH 7.5), 150 mM NaCl, 10 mM MgCl_2 and 5% glycerol). After the last wash, the wash buffer was aspirated completely, and the beads were resuspended in 80 μl of trypsin buffer (2 M urea, 50 mM Tris (pH 7.5), 5 $\mu\text{g ml}^{-1}$ trypsin) to digest the bound proteins at 37°C for 1 h with agitation. The beads were centrifuged at 100g for 30 s, and the partially digested proteins (the supernatant) were collected. The beads were then washed twice with 60 μl of urea buffer (2 M urea, 50 mM Tris (pH 7.5)). The supernatant of both washes was collected and combined with the partially digested proteins (final volume, 200 μl). After brief centrifugation, the combined partially digested proteins were cleared from residual beads. Then, 80 μl of these partially digested proteins was

used; disulfide bonds were reduced with 5 mM dithiothreitol (DTT); and cysteines were subsequently alkylated with 10 mM iodoacetamide. Samples were further digested by adding 0.5 µg of sequencing-grade modified trypsin (Promega) at 25 °C. After 16 h of digestion, samples were acidified with 1% formic acid (final concentration). Tryptic peptides were desalted on C18 StageTips according to ref. 67 and evaporated to dryness in a vacuum concentrator and reconstituted in 15 µl of 3% acetonitrile/2% formic acid for liquid chromatography with tandem mass spectrometry (LC–MS/MS).

LC–MS/MS analysis was performed on a Q Exactive HF. Five microliters of total peptides was analyzed on a Waters M-Class UPLC using a 25-cm Thermo Fisher Scientific EASY-Spray column (2 µm, 100 Å, 75 µm × 25 cm) coupled to a benchtop Thermo Fisher Scientific Orbitrap Q Exactive HF mass spectrometer. Peptides were separated at a flow rate of 400 nl min⁻¹ with a 100-min gradient, including sample loading and column equilibration times. Data were acquired in data-independent (DIA) mode for initial experiments and data-dependent (DDA) mode for follow-up experiments. DIA MS1 spectra were measured with a resolution of 120,000, an automatic gain control (AGC) target of 5 × 10⁶ and a mass range from 350 *m/z* to 1,650 *m/z*; 34 isolation windows of 38 *m/z* were measured at a resolution of 30,000, an AGC target of 3 × 10⁶, normalized collision energies of 22.5, 25 and 27.5 and a fixed first mass of 200 *m/z*. DDA MS1 spectra were measured with a resolution of 120,000, an AGC target of 3 × 10⁶ and a mass range from 300 *m/z* to 1,800 *m/z*; MS2 spectra were measured at a resolution of 15,000, an AGC target of 1 × 10⁵, a TopN of 12, an isolation window of 1.6 *m/z* and a mass range from 200 *m/z* to 2,000 *m/z*.

Proteomics raw data were analyzed by Spectronaut version 16.0 (ref. 68) (Biognosys) using a UniProt database (*Homo sapiens*, UP000005640), and MS/MS searches were performed under Biognosys factory settings. UniProt GO term annotations (downloaded on 14 January 2022) were used for the differential enrichment analysis conducted by the Spectronaut software. Spectromine version 4.2.230428.52329 was used to analyze proteomics data in follow-up experiments using the same UniProt databases and default parameters. Preys identified in both the RNase treatment and non-treatment IPs for a particular bait were called ‘direct interactors’, and preys identified in only RNase non-treatment were called ‘RNA-mediated interactors’.

Modulation of splicing with dCas13d fusions

Transfection was performed as described for the luciferase reporter screens. The plasmid DNA transfected consisted of 10 ng of lucMAPT Reporter DNA, 45 ng of gRNA plasmid and 45 ng of dCas13d–RBP fusion. Dual-luciferase readout was collected as described for the luciferase reporter screens. gRNA sequences were designed using the cas13d-design tool^{69,70}. Transfection for modulation of endogenous targets was performed in 24-well plates with 250 ng of gRNA plasmid DNA and 250 ng of dCas13d–RBP fusion.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

RNA-seq and eCLIP-seq data of this study are available at the National Center for Biotechnology Information’s Gene Expression Omnibus (accession code [GSE232599](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE232599))⁷¹. Source data are provided with this paper.

References

- Rual, J.-F. et al. Human ORFeome version 1.1: a platform for reverse proteomics. *Genome Res.* **14**, 2128–2135 (2004).
- Blum, M. et al. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **49**, D344–D354 (2021).

- The pandas development team. *pandas-dev/pandas*. <https://doi.org/10.5281/ZENODO.3509134> (2023).
- Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
- Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
- Durinck, S. et al. *biomaRt*. <https://doi.org/10.18129/B9.BIOC.BIOMART> (2017).
- Frankish, A. et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- Shen, S. et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-seq data. *Proc. Natl Acad. Sci. USA* **111**, E5593–E5601 (2014).
- Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification enrichment pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).
- Bruderer, R. et al. Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol. Cell. Proteomics* **14**, 1400–1410 (2015).
- Wessels, H.-H. et al. Massively parallel Cas13 screens reveal principles for guide RNA design. *Nat. Biotechnol.* **38**, 722–727 (2020).
- Guo, X. et al. Transcriptome-wide Cas13 guide RNA design for model organisms and viral RNA pathogens. *Cell Genom.* **1**, 100001 (2021).
- Schmok, J. C. et al. Systematic identification of RNA-binding proteins and tethered domains that activate exon splicing inclusion. *Gene Expression Omnibus*. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE232599> (2023).

Acknowledgements

We thank members and alumni of the Yeo laboratory, in particular F. Tan, A. Smargon, T. Yu, P. Le, J. Xiang, N. Ahmed, J. Mueller, K. Brannan, N. Al-Azzam, K. Rothamel, S. Aigner and S. Blue, for advice and support. J.C.S. was awarded a Natural Sciences and Engineering Research Council of Canada Postgraduate Scholarships–Doctoral (PGS D-532649-2019). A.T.T. was supported by the Cancer Systems Biology Training Program (U54 CA209891) and the Cancer Biology, Informatics, and Omics Training Program (T32CA067754). A National Science Foundation (NSF) Graduate Research Fellowship (grant no. DGE-2038238), a Myotonic Dystrophy Foundation Doctoral Research Fellowship and an Association for Women in Science Scholarship were awarded to M.L.G. E.A.B. was supported by the Helen Hay Whitney Foundation. An ARCS Scholarship was awarded to P.J. M. Jovanovic is funded by the National Institutes of Health (NIH) (R35GM128802, R01AG071869 and R01HG012216), the NSF (award no. 2224211) and Columbia startup funding. G.W.Y. is supported by NIH R01 HG004659, U24 HG009889 and an Allen Distinguished Investigator Award, a Paul G. Allen Frontiers Group advised grant of the Paul G. Allen Foundation. Figures were created, in part, using BioRender. This work includes data generated at the UC San Diego IGM Genomics Center using an Illumina NovaSeq 6000 that was purchased with funding from an NIH Scientific Interest Groups grant (S10 OD026929).

Author contributions

J.C.S. designed the reporter assays and was primarily responsible for designing and executing experiments, data analysis and writing the manuscript, under the supervision of G.W.Y. M. Jain carried out several of the experiments, under the supervision of J.C.S. L.A.S. carried out

all mass spectrometry measurements described in the manuscript as well as analyzed and interpreted data, under the supervision of M. Jovanovic. A.T.T., D.S., S.E. and M.L.G. contributed to experimental execution and design. H.-L.H. contributed to data analysis. E.A.B., P.J. and E.-C.L. contributed to overall study conception and design. E.J.K. consulted throughout the project and contributed use of vital equipment. All authors interpreted data and revised the paper.

Competing interests

G.W.Y. is a co-founder, member of the board of directors, scientific advisory board member, equity holder and paid consultant for Locanabio and Eclipse BioInnovations. G.W.Y. is a visiting professor at the National University of Singapore. G.W.Y.'s interests have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. The authors declare no other competing financial interests.

Additional information

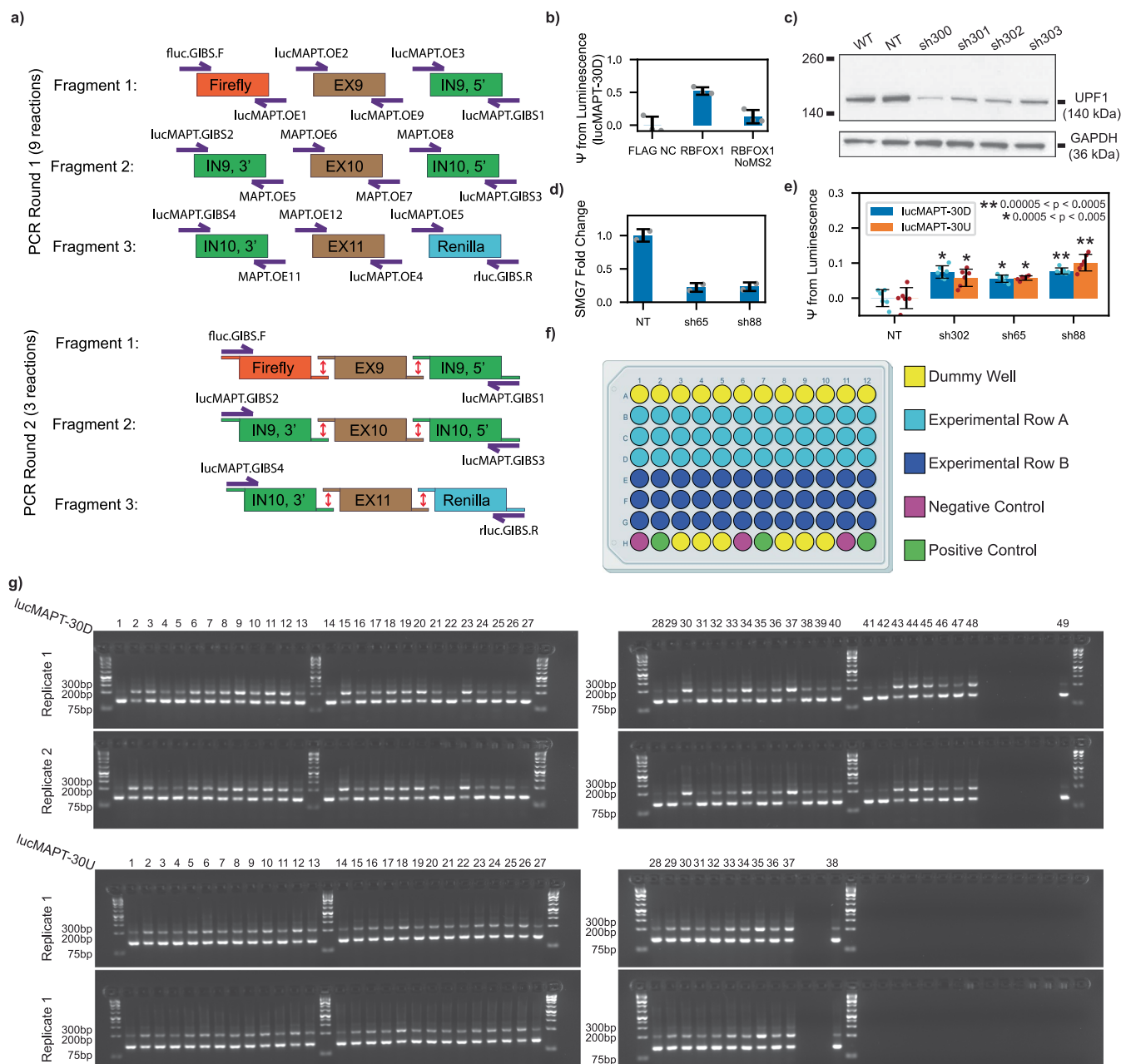
Extended data is available for this paper at <https://doi.org/10.1038/s41587-023-02014-0>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41587-023-02014-0>.

Correspondence and requests for materials should be addressed to Gene W. Yeo.

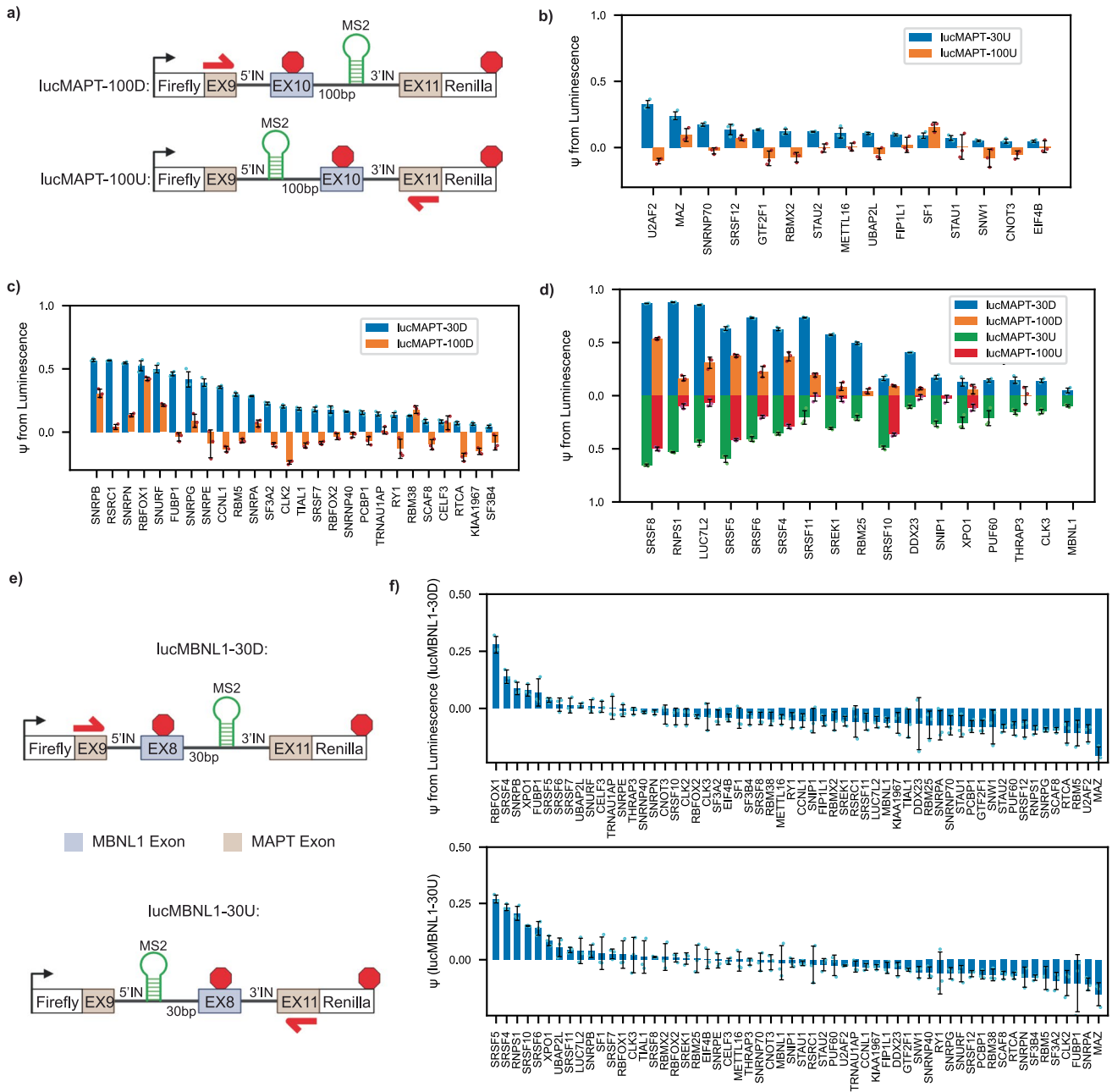
Peer review information *Nature Biotechnology* thanks Jeremy Sanford and Sika Zheng for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Reporter construction strategy, tethering validation, reporter layout, splicing gels. a Schematic of strategy used for assembling luciferase based minigene splicing reporters. **b** Bar graph of lucMAPT-30D reporter readout following co-transfection with FLAG NC, RBFOX1-MCP fusion (RBFOX1), and RBFOX1 lacking an MCP fusion (RBFOX1 NoMS2) (mean \pm s.d., n = 3 replicate transfections). **c** Western blots for validation of UPF1 shRNA constructs qualitatively showing decreased UPF1 protein levels for each of four UPF1 shRNA constructs tested in HEK293T cells. **d** qPCR for validation of SMG7 shRNA constructs showing decreased SMG7 expression levels as quantified using the delta-delta Ct method in RNA extracted from MDAMB231 and MCF10A cells

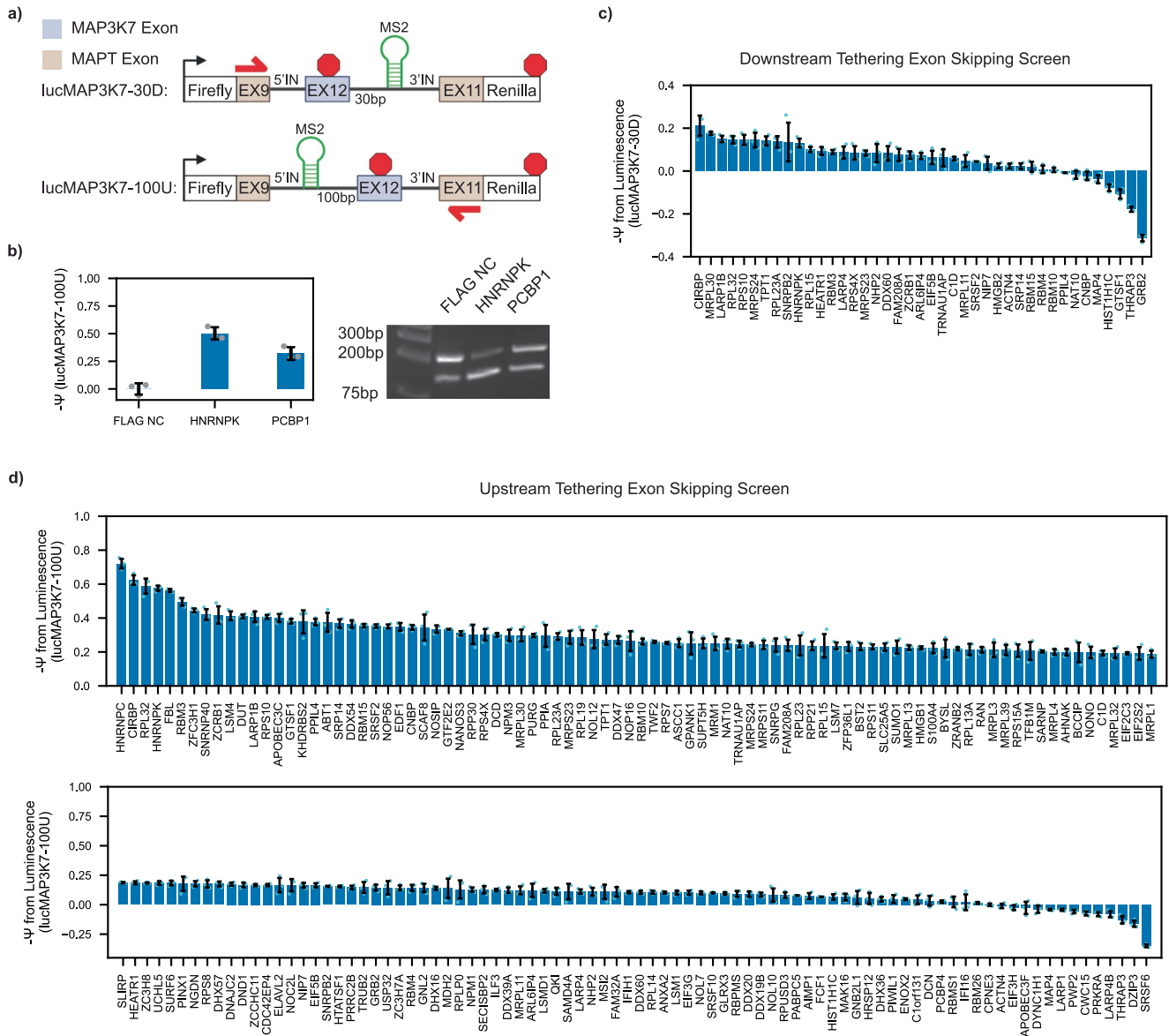
stably expressing the constructs (n = 2 biological replicates (1 replicate/line), n = 2 technical replicates). **e** Bar graph of reporter readouts in HEK293T cells stably expressing a non-targeting shRNA (NT), a UPF1-targeting shRNA (sh302), and two SMG7-targeting shRNAs (sh65 and sh88), co-transfected with reporter plasmids and FLAG NC (mean \pm s.d., n = 6 replicate transfections). P-value is calculated by two-tailed independent two-sample t-test. **f** Layout of 96-well transfections used throughout the screens. **g** Agarose gels of RNA-level validation of hits from the splicing screen. All hits were tested for lucMAPT-30D (top) and lucMAPT-30U (bottom). Numbers along the top correspond to lane number in Supplementary Table 6-7. n = 2 replicate transfections.



Extended Data Fig. 2 | Survey of screen hits with complementary reporters.

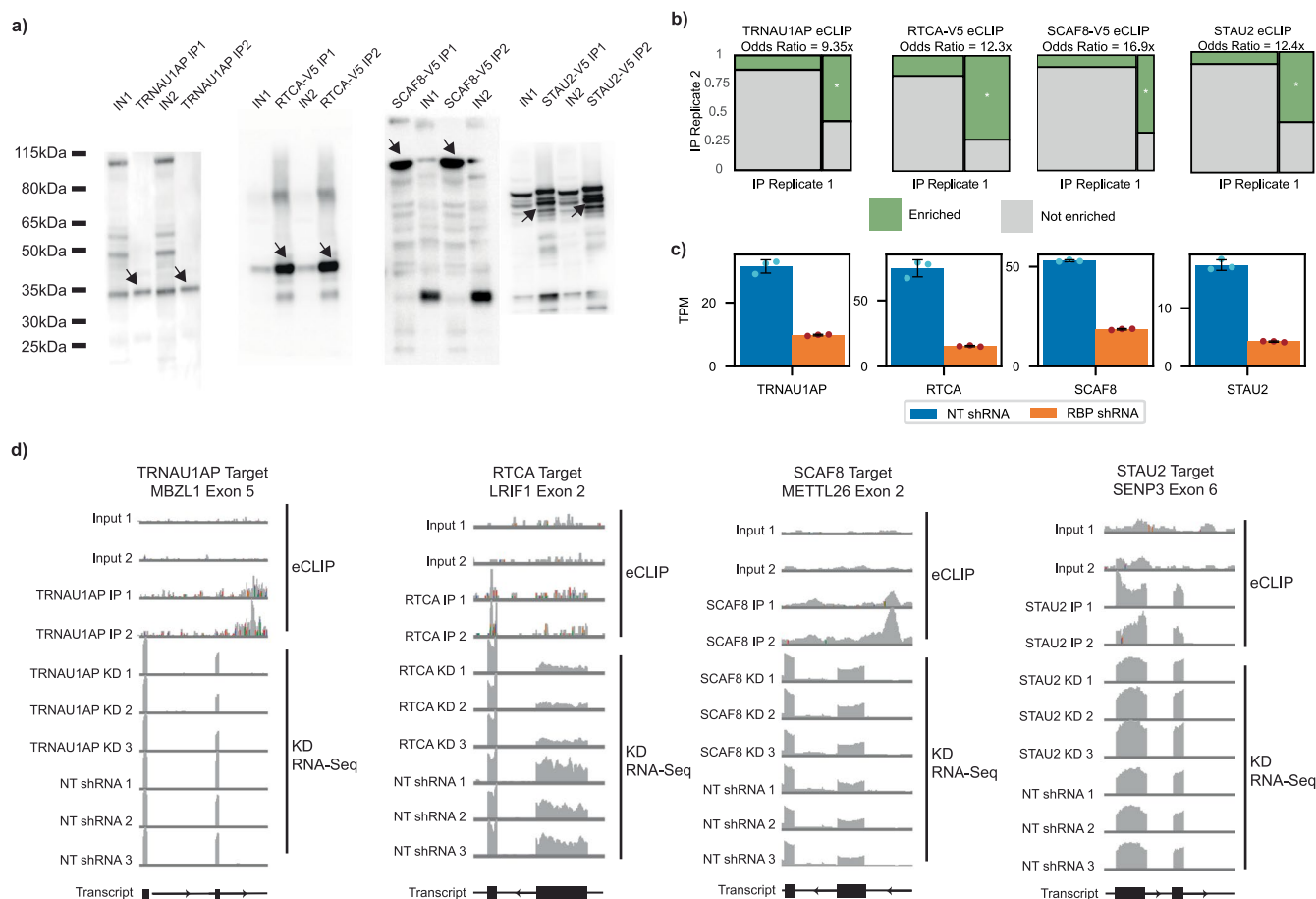
a Schematic of luciferase reporters for tethering 100 base pairs away from the splice site. **b** Clustered bar graph of upstream tethering only hits from the screen comparing results from the original screen (lucMAPT-30U) to results from co-transfection of the RBP-MCP fusions and lucMAPT-100U (mean \pm s.d., $n = 3$ replicate transfections). **c** Clustered bar graph of downstream tethering only hits from the screen comparing results from the original screen (lucMAPT-30D) to results from co-transfection of the RBP-MCP fusions and lucMAPT-100D (mean \pm s.d., $n = 3$ replicate transfections). **d** Clustered bar graph of hits that activated

both reporters from the screen comparing results from the original screens (lucMAPT-30D, lucMAPT-30U) to results from co-transfection of the RBP-MCP fusions and the long-distance reporters (lucMAPT-100D and lucMAPT-100U) (mean \pm s.d., $n = 3$ replicate transfections). Results where hits displayed a mean ψ from luminescence < 0 are omitted for clarity. **e** Schematic of lucMBNL1 reporters used as orthogonal exon inclusion reporters. **f** Bar graphs of reporter readout from co-transfection of all hits from the original screens with lucMBNL1-30D and lucMBNL1-30U (mean \pm s.d., $n = 3$ replicate transfections).



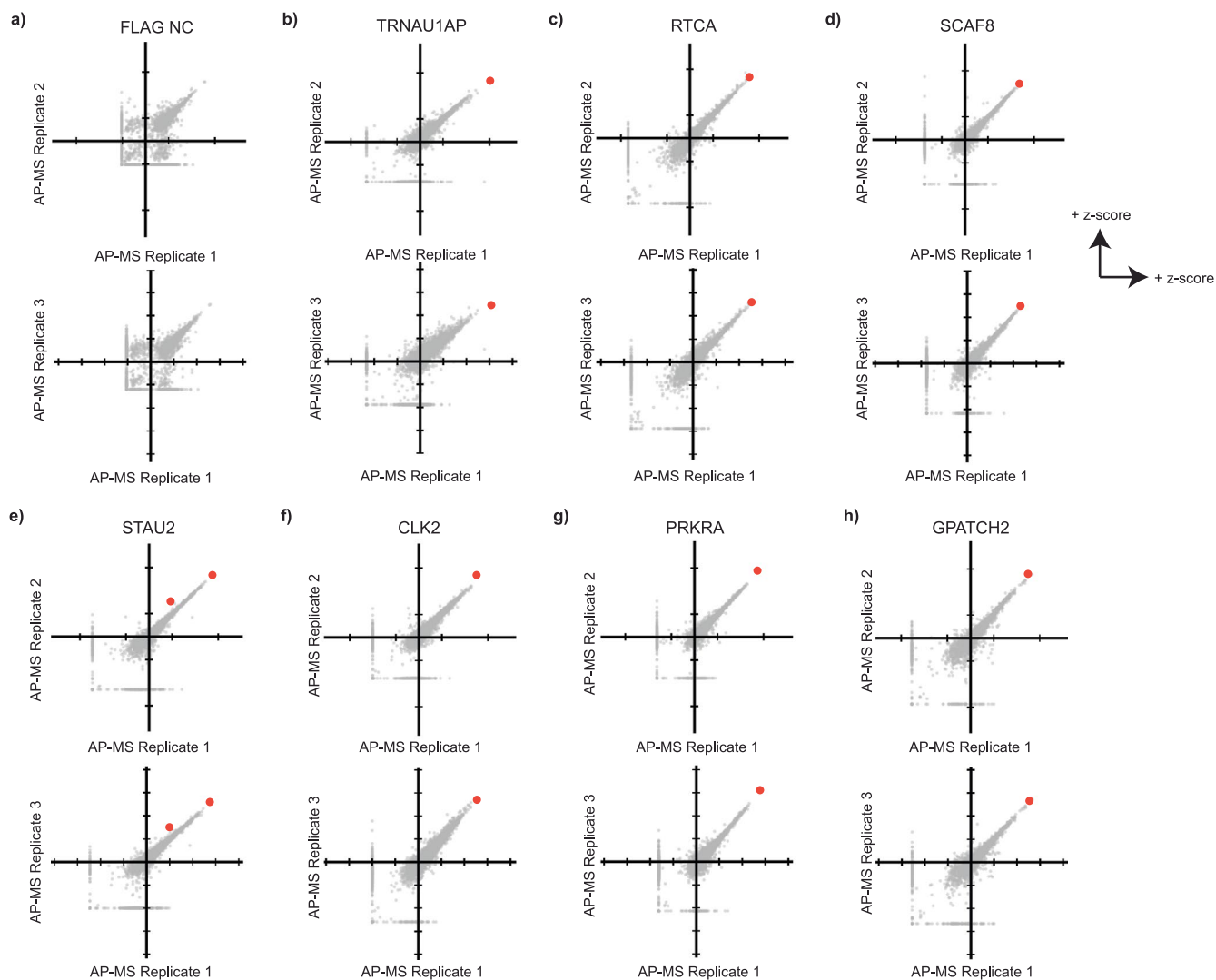
Extended Data Fig. 3 | Exon skipping screen. a Schematic of luciferase reporters for skipping readout. **b** lucMAP3K7-100U splicing in response to co-transfection with MCP-fused positive and negative controls. (left) Bar graph of lucMAP3K7 reporter readout (mean \pm s.d., $n = 3$ replicate transfections). (right) Agarose gel electrophoresis of RT-generated cDNA amplified by minigene specific primers

(shown in panel a) that amplify skipping and inclusion isoforms. **c** Bar graph of lucMAP3K7-30D reporter readout when co-transfected with RBP-MCP fusions from the library. (mean \pm s.d., $n = 3$ replicate transfections). **d** Bar graph of lucMAP3K7-100U reporter readout when co-transfected with RBP-MCP fusions from the library (mean \pm s.d., $n = 3$ replicate transfections).

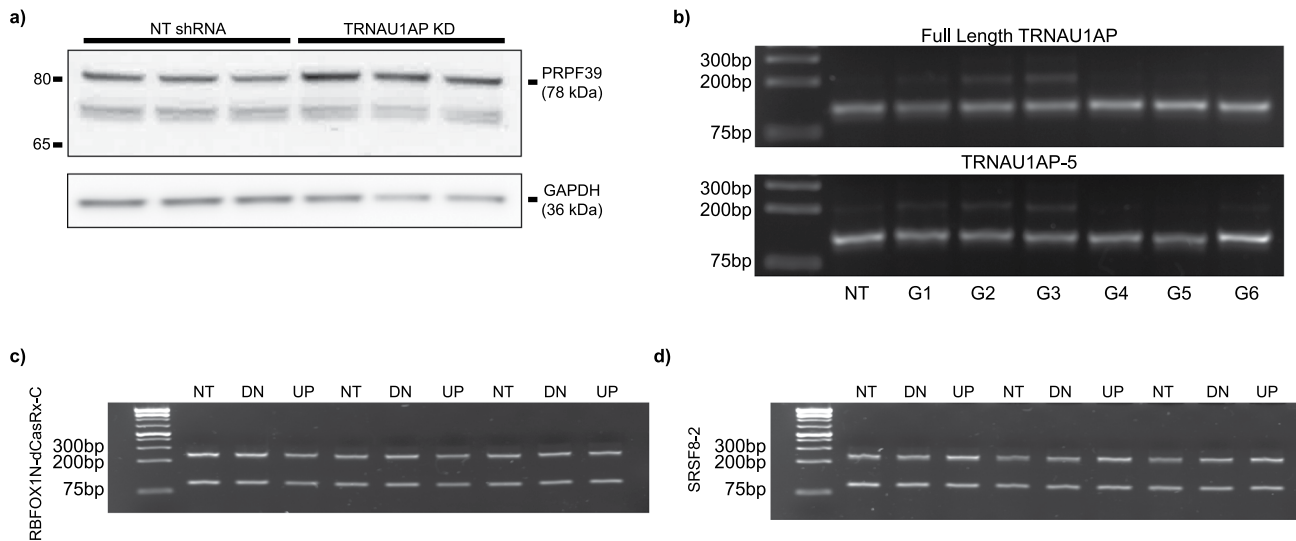


Extended Data Fig. 4 | Quality control of eCLIP and shRNA knockdown followed by RNA-seq. **a** Western blots of cold gels from eCLIP protocol for TRNAU1AP, SCAF8, STAU2 and RTCA. Size-matched input and immunoprecipitation conditions are compared. *n* = 2 independent samples, with size-matched input and IP conditions extracted from both. **b** Mosaic plots from Skipper showing concordance between eCLIP replicates. Odds ratios and significance from Fisher's exact test. **c** TPM of unexpected hits following shRNA

knockdown as measured from aligned RNA-seq data. (mean \pm s.d., *n* = 3 replicate knockdowns). **d** IGV browser tracks showing coverage of RBP eCLIP signal relative to sized-matched input and the RBP KD RNA-Seq signal relative to non-targeting shRNA. From left to right: comparison of TRNAU1AP eCLIP and KD RNA-Seq signal near MBZL1 Exon 5, comparison of RTCA eCLIP and KD RNA-Seq signal near LRIF Exon 2, comparison of SCAF8 eCLIP and KD RNA-Seq signal near METTL26 Exon 2, comparison of STAU2 eCLIP and KD RNA-Seq signal near SENP3 Exon 6.



Extended Data Fig. 5 | Quality control of AP-MS. a-h Scatter plots showing concordance between AP-MS replicates. Each point represents a detected protein and its z-score in two replicates per plot. Red points represent the detection of the bait protein among the preys. Multiple red points indicate multiple major isoforms detected with average Z-score > 1.



Extended Data Fig. 6 | Full western blots and splicing gels for TRNAU1AP follow-up experiments and modulation of endogenous HNRNP Exon 7. **a** Western blot replicates used for quantification showing increased PRPF39 expression in HEK293T cells following TRNAU1AP knockdown. GAPDH is the loading control. $n = 3$ independent transductions. **b** Additional replicate displaying lucMAPT alternative splicing from co-transfection of the MS2-free lucMAPT reporter, either full-length TRNAU1AP-dCas13d fusion or truncated

TRNAU1AP-5-dCas13d fusion, and each reporter targeting guide RNA annotated in Fig. 5i. $n = 2$ independent transfections **c-d** Agarose gels of amplified cDNA collected from HEK293T cells co-transfected with artificial splicing factors (RBFOX1-dCasRx-C, SRSF8-2) and gRNA arrays (NT = non-targeting gRNA, DN = downstream 3-gRNA array, UP = upstream 3-gRNA array). $n = 3$ independent transfections.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

RNA-seq and eCLIP-seq data of this study are available at NCBI-GEO (accession code GSE232599). We referenced the UniPort database (Homo sapiens, UP000005640) and GO Term annotations for analysis of proteomics raw data. We referenced the Ensembl BioMart database for determining splicing-associated proteins using GO Term annotation. We referenced the COMPARTMENTS database for subcellular localization scores of candidates. We referenced the SpliceosomeDB database to identify previous experiments that captured our candidates in the spliceosome.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	<input type="text" value="n/a"/>
Population characteristics	<input type="text" value="n/a"/>
Recruitment	<input type="text" value="n/a"/>
Ethics oversight	<input type="text" value="n/a"/>

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size for each experiment is indicated in the figure legend for each experiment. The sample size was chosen based on previous experience for each experiment. No statistical methods were used to predetermine sample size. Experiments were performed in triplicate as per previously published high-throughput tethered function assays (Luo, EC., Nathanson, J.L., Tan, F.E. et al. Large-scale tethered function assays identify factors that regulate mRNA stability and translation. Nat Struct Mol Biol 27, 989–1000 (2020). https://doi.org/10.1038/s41594-020-0477-6). RNA-level validation with agarose gels was performed in duplicate due to the experimental burden required and the availability of a visual readout.
Data exclusions	No data were excluded.
Replication	Independent replication was performed for candidates from the screens as described in the text. All reported findings from the screens were successfully replicated.
Randomization	No randomization was employed. All comparisons are made across experiments in HEK293T cells under controlled experimental conditions, therefore random allocation of samples is not relevant.
Blinding	Blinding is not relevant to this study since group allocation does not occur.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

Primary Antibodies:

Rabbit V5 Epitope Tag - Bethyl A190-10A, Lot 7, No Clone Listed (used for IP)
 Rabbit TRNAU1AP - GeneTex GTX121631, Lot 40863, No Clone Listed (used for IP)
 Rabbit PRPF39 - Thermo Fisher PA5-21627, Lot yg3989185b, No Clone Listed (used for Western Blot at 1:1000 dilution)
 Rabbit UPF1 - Cell Signaling Technology D15G6, Lot 3, No Clone Listed (used for Western Blot at 1:1000 dilution)
 Mouse GAPDH - Millipore MAB374, Lot 3855179, Clone 6C5 (used for Western Blot at 1:10000 dilution)

Secondary Antibodies:

Goat Anti-Rabbit IgG, HRP Linked - Cell Signaling 7074, Lot 29, No Clone Listed (used for Western Blot at 1:2000 dilution)
 800CW Goat Anti-Mouse IgG - Licor 926-32210, Lot d2012535, No Clone Listed (used for Western Blot at 1:2000 dilution)

Validation

Extended Data Figure 4 provided with the manuscript shows validation of antibodies used for eCLIP. All antibodies used for Western Blots were validated by the manufacturer. Reactivity for human samples is noted for each antibody on the manufacturer's website. Validation for other antibodies from the manufacturers' websites are noted below.

PA5-21627: PA5-21627 targets PRPF39 in IHC (P) and WB applications and shows reactivity with Human samples. Validated from manufacturer's website and included citations. Validation included blotting for PRPF39 in PRPF39-transfected HEK293T cells. <https://www.thermofisher.com/antibody/product/PRPF39-Antibody-Polyclonal/PA5-21627>

D15G6: UPF1 (D15G6) Rabbit mAb recognizes endogenous levels of total UPF1 protein. Species Reactivity: Human, Mouse, Rat, Monkey. Validated from manufacturer's website and included citations. Validation included blotting for UPF1 in various cell lines. <https://www.cellsignal.com/products/primary-antibodies/upf1-d15g6-rabbit-mab/12040>

MAB374: species reactivity: human, feline, pig, mouse, rabbit, fish, canine, rat. Validated from manufacturer's website and included citations. Validation included blotting for GAPDH in various cell lines. https://www.emdmillipore.com/US/en/product/Anti-Glyceraldehyde-3-Phosphate-Dehydrogenase-Antibody-clone-6C5,MM_NF-MAB374

7074: Species Reactivity: Rabbit. Validated from manufacturer's website and included citations. Validated with a variety of CST primary antibodies for western immunoblotting. <https://www.cellsignal.com/products/secondary-antibodies/anti-rabbit-igg-hrp-linked-antibody/7074>

926-32210: Species Reactivity: Mouse. Validated from manufacturer's website and included citations. Validation includes applications using ApoTrack primary mAb cocktail in Staurosporine treated and non-treated HeLa cells. <https://www.licor.com/bio/reagents/irdye-800cw-goat-anti-mouse-igg-secondary-antibody>

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)

Lenti-X HEK293T cells were purchased from Takara Bio.

Authentication

Cell lines were not further authenticated.

Mycoplasma contamination

The cell lines used in this study have been routinely and frequently tested negative for mycoplasma contamination.

Commonly misidentified lines
(See [ICLAC](#) register)

No cell lines used are listed in the database of commonly misidentified cell lines.