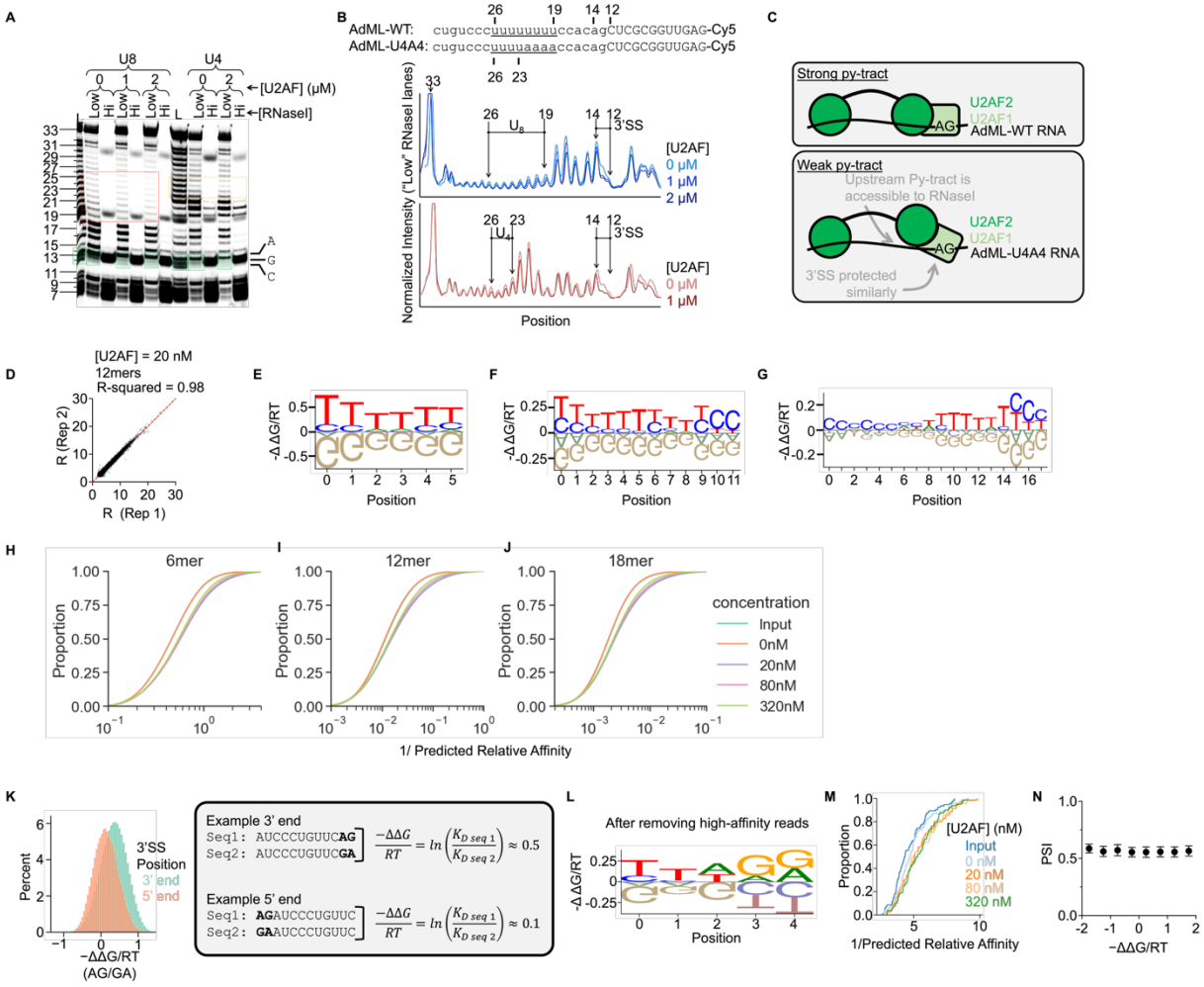
SUPPLEMENTARY FIGURES



Extended Data Fig. 1. Details on RBNS and Probound Assays

- **A)** Measuring U2AF footprintg on AdML-WT and AdML-U4A4 RNA subjected to low (0.0003 Units) and high (0.003 Units) RNaseI treatment. Digested RNAs were resolved on a 20% polyacrylamide, 7M urea denaturing gel.
- **B)** Anayzing band intensity (normalized to position 33) for AdML-WT and AdML-U4A4 RNAs mixed with increasing concentration of U2AF.
- **C)** Interpretation of footprinting result. While we observe distinct differences in py-tract protection for the two sequences, U2AF1 appears to protect the 3'SS equally well on both sequences.
- **D)** RBNS was performed with two replicates of 0, 20, 80, 320 nM U2AF. Comparing enrichment (R) for significantly enriched kmers (Z=3) for U2AF = 20 nM indicates strong agreement between two measurements.
- **E)** U2AF binding model from probound for 6mers.
- **F)** U2AF binding model from probound for 12mers.
- **G)** U2AF binding model from probound for 18mers.
- **H)** Cumulative distribution of predicted binding affinity from the 6mer Probound model for enriched RNA in different replicates of the RBNS assay (input, 20 nM, 80 nM, 320 nM). The RNAs enriched for 20 nM U2AF have the highest binding affinity. Reads from the the initial library (input) have the lowest binding affinity.

I) Same as (E) but for 12mer probound model.

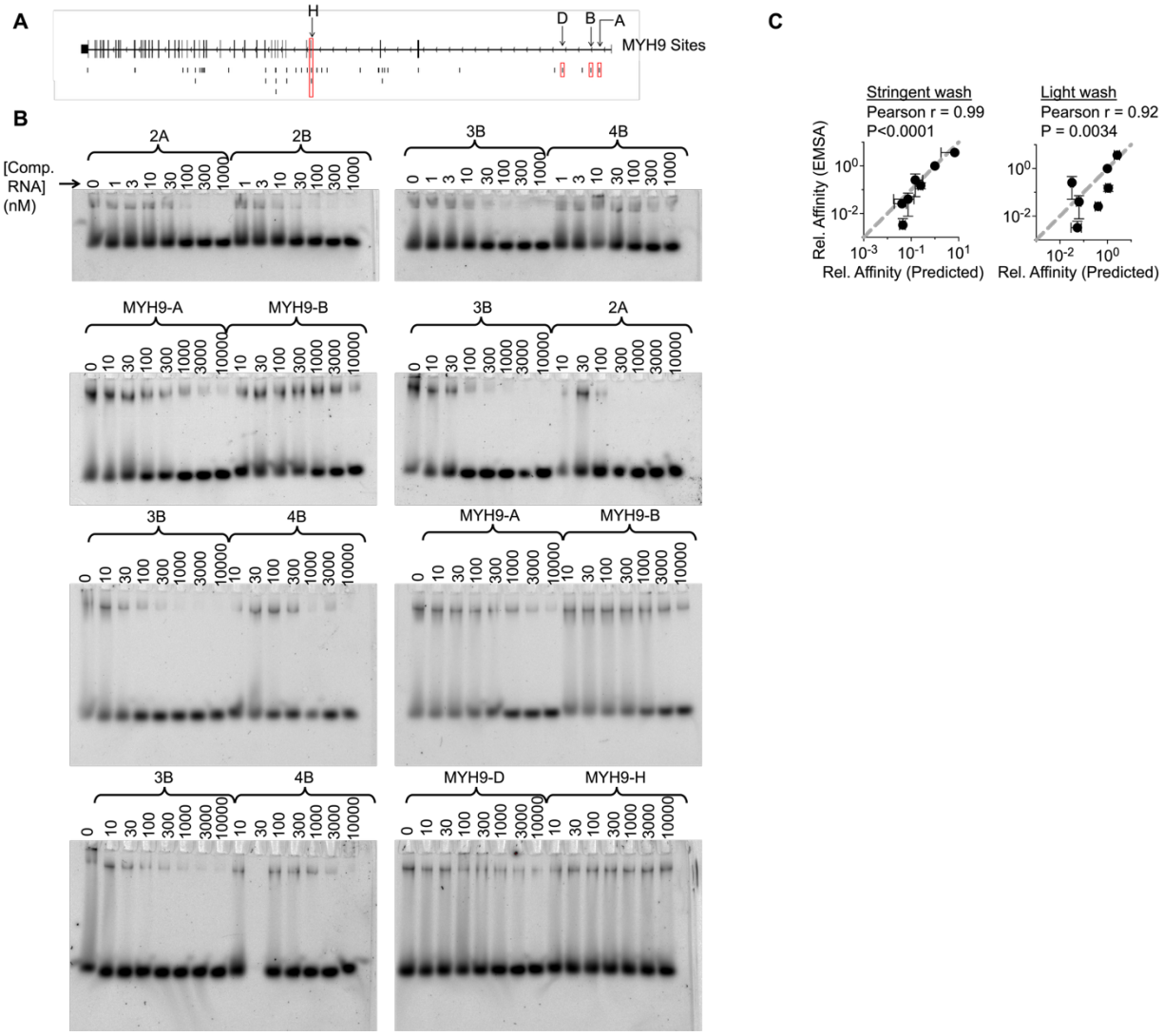
J) Same as (E) but for 18mer probound model.

K) Comparing predicted binding affinity of a 3' AG compared to a 3' GA (expressed in terms of $\frac{-\Delta\Delta G}{RT} = ln\frac{K_{D\,AG}}{K_{D\,GA}}$) for all 10mers (green distribution). As a control, we also performed the same analysis comparing predicted affinities for 5'AG relative to 5'GA.

L) Producing a secondary binding model for U2AF using a filtered set of sequencing reads. We originally identified three models. This motif is the result of a probound run on a set of raw sequencing reads that does not contain high affinity sequences corresponding to the other two models.

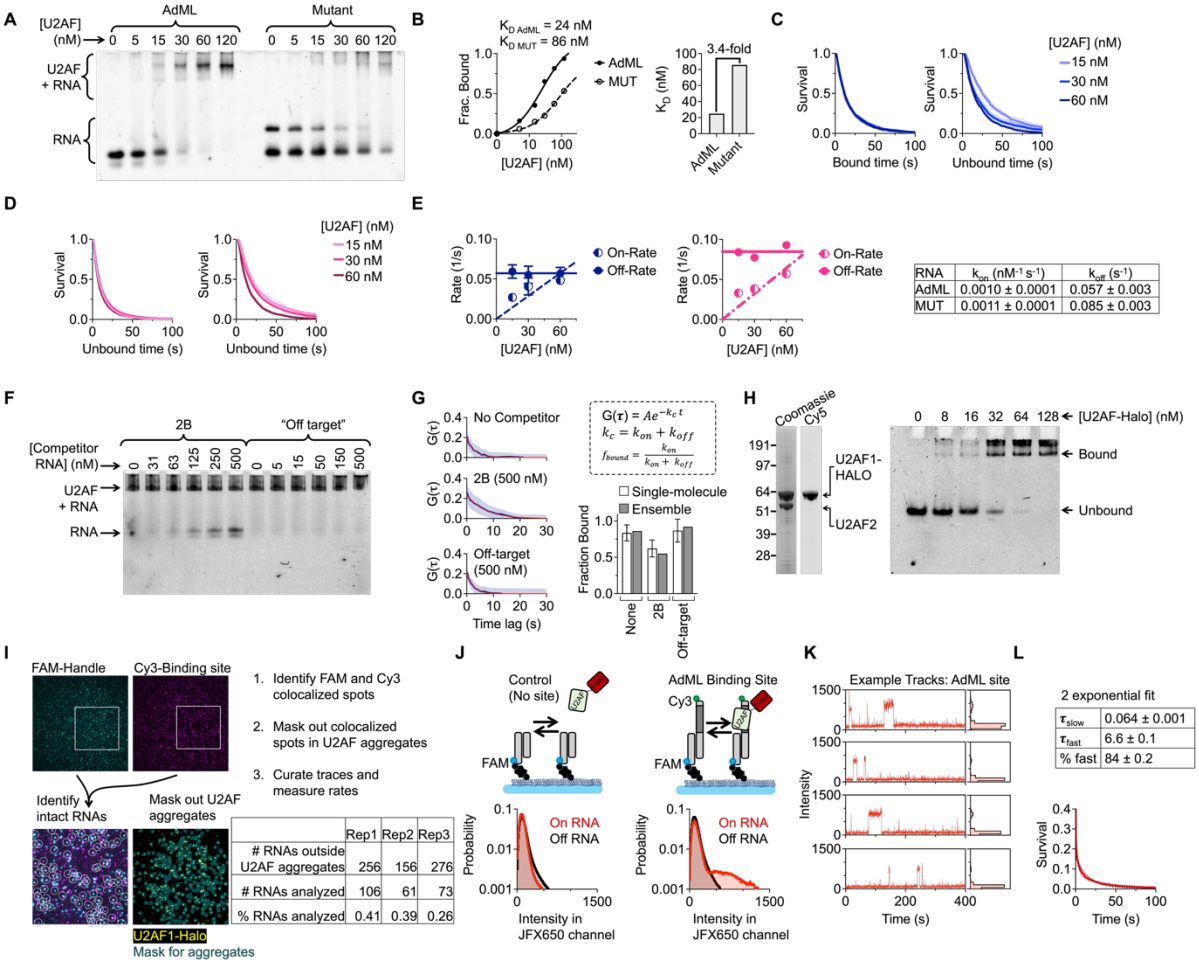
M) Measuring the distribution of predicted binding affinities from this secondary probound model in the input, 0 nM, 20 nM, 80 nM, and 320 nm datasets. We observe clear enrichment of high affinity sequences in the 20 nM, 80 nM, and 320 nM datasets.

N) The relationship between relative free energy ($-\Delta\Delta G$, from the secondary probound model) between two alternative 3'SS pairs and PSI.



Extended Data Fig. 2. Details on Probound validation by competitive binding assay.

A) We tested U2AF binding to various binding sites identified in PAR CLIP in the MYH9 pre-mRNA.
B) Competitive binding EMSAs used to validate probound predictions.
C) Correlation between predicted relative binding affinity and measured relative binding affinity for Probound models used in figure 1 and figure 3. In figure 3, we used a less stringent wash to preserve complex formation (see methods).



Extended Data Fig. 3. Details on in vitro binding assays.

A) Measuring U2AF binding affinity to AdML and Mutant 3'SSs by EMSA.

B) Quantification of EMSA measurement reveals a 3.4-fold difference in binding affinity (n=1 replicate).

C) Dwell time distributions of U2AF in the bound state and unbound state for 3 concentrations of U2AF binding to the AdML 3'SS (n=3 replicates for each concentration). Dwell time distributions were fit to a single exponential distribution to determine the off-rate (from bound time survival plot) and on-rate (from unbound time survival plot) for U2AF binding the AdML 3'SS.

D) Same as (C) but for the mutant 3'SS.

E) Measured on-rates and off-rates for U2AF binding to the AdML sequence (blue) and the mutant sequence (pink) from the single-exponential fits in C & D. The table contains the reported rates from these plots.

F) Competing U2AF off Cy3-AdML RNA using 2B (UUUUUUUUUCCC) or off-target RNA (CCCCCAGCUCAG).

G) Repeating the competition assay at the single-molecule level where fluctuations in the Cy3 channel were measured by autocorrelation and quantified with a single exponential distribution ($k_c = k_{off} + k_{on,\ apparent}$). Using k_{off} = 0.057 ± 0.003 s^{-1} (see Fig S3E), we determined $k_{on,\ apparent}$ and fraction bound. The bar plot compares fraction bound determined in the single-molecule assay with the gel shift in figure S3F.

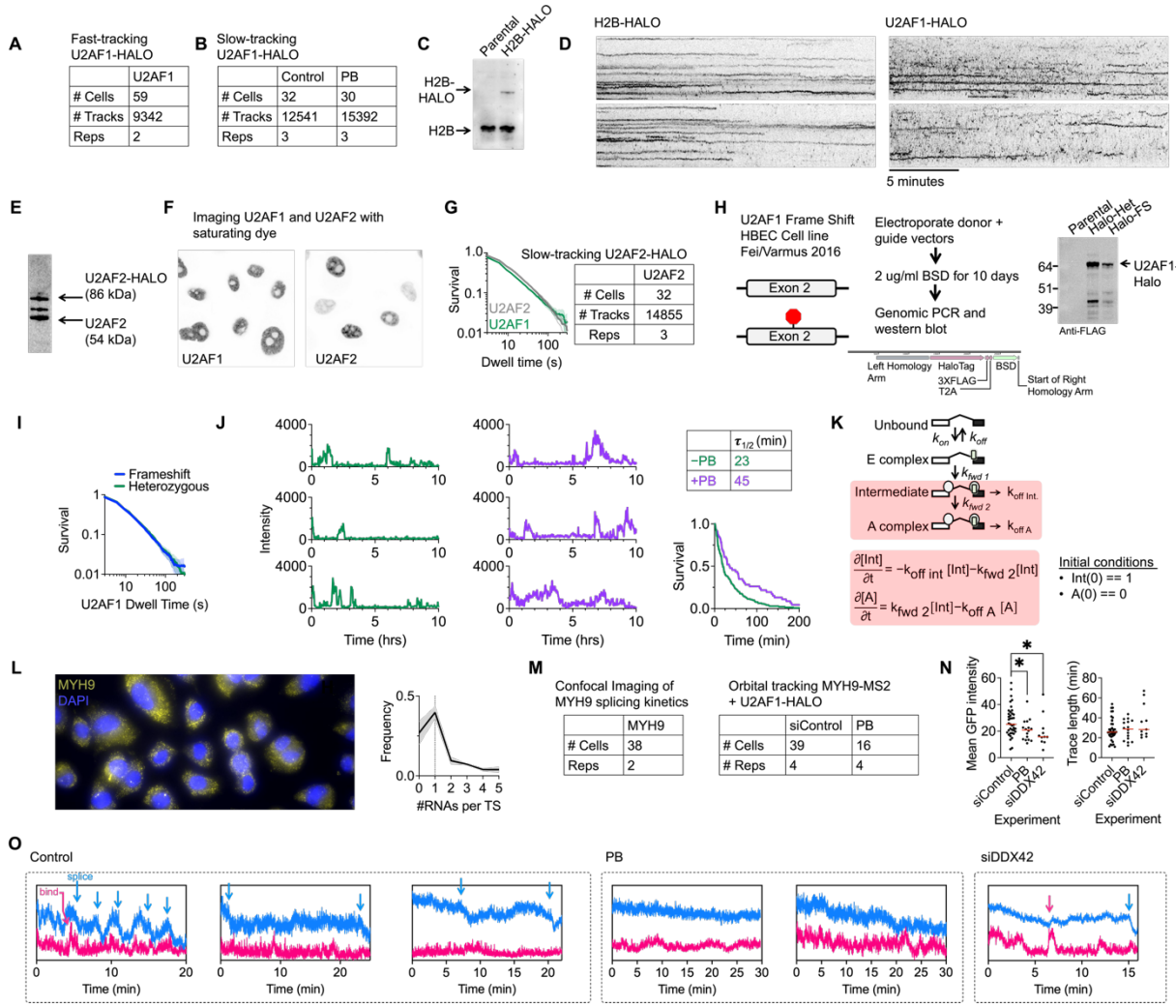
H) Purification and functional validation of U2AF2 in complex with U2AF1-HALO purified from insect cells. U2AF1-HALO was labeled with equal concentration of JFX650 before FLAG purification to remove free dye.

I) Workflow for colocalization assay (see methods). We observe binding fluctuations on about 35% of RNA outside U2AF aggregates.

J) Monitoring spot intensity in JFX650 channel on RNA and off RNA. Only when we include a binding site do we observe increased U2AF-HALO-JFX650 occupancy on RNA compared to background locations.

K) Example time traces of 1 nM U2AF-HALO binding to the AdML RNA. Histograms represent occupancy. At this concentration of U2AF, occupancy is very low.

L) Dwell time distribution of U2AF-HALO measured by colocalization. Data is fit to a double exponential where the fast component represents non-specific binding.



Extended Data Fig. 4. Live cell single-molecule tracking of U2AF in spliceosome assembly.

A) Experimental details for fast tracking dataset (100 Hz acquisition rate, 10 seconds) for U2AF1-HALO.

B) Experimental details for slow tracking dataset (0.33 Hz acquisition rate, 20 minutes) for U2AF1-HALO (with and without PB treatment).

C) Western blot of H2B-HALO HBEC cell line.

D) Kymographs of H2B-HALO and U2AF1-HALO in slow tracking conditions.

E) Western blot of U2AF2-HALO HBEC cell line.

F) Confocal imaging of U2AF1-HALO and U2AF2-HALO at saturating dye concentration. Generally the staining pattern appears smooth. We observe small U2AF2 speckles.

G) Left: Survival plot comparing U2AF1-HALO and U2AF2-HALO dwell times in slow SMT assay. Dwell times appear similar for both cell lines. Right: Experimental details for U2AF2-HALO slow SMT dataset.

H) HALO-Tagging U2AF1 in an HBEC cell line where a frameshift was introduced in U2AF1 Exon 2. We observe Halo Tag integration both by genomic PCR and FLAG western blot.

I) Comparing U2AF1 dwell time distribution in heterozygous knock-in cell line vs frameshift knock-in cell line.

J) Time traces from overnight confocal imaging of splicing kinetics in polyclonal gene trap cell line where hundreds of introns contain MS2 stem loops. Imaging was performed in the presence and absence of the splicing inhibitor pladienolide B.

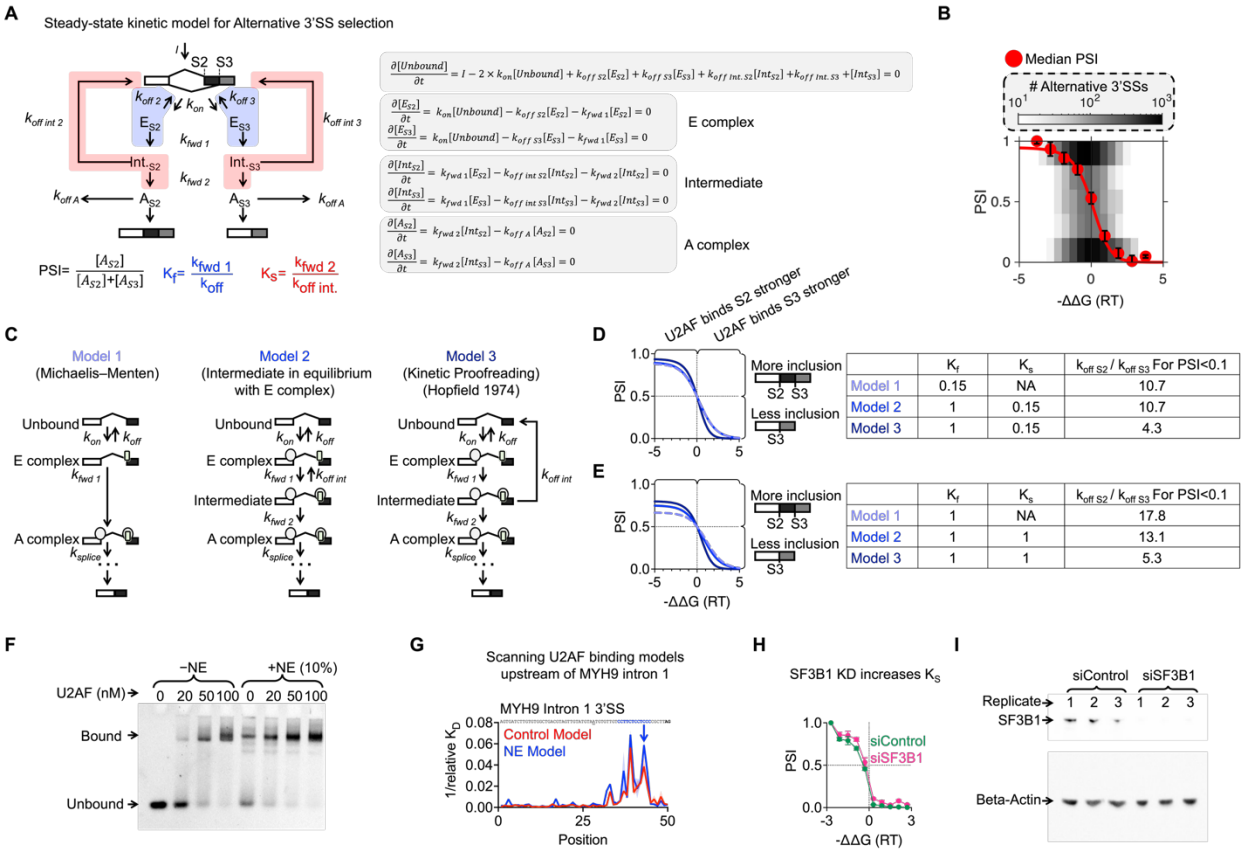
K) Binding model used to interpret slow SMT data. Equation 1 comes from master equation with initial conditions. We assume that PB affects only $k_{fwd\ 2}$, the rate of forward progression from the intermediate state to the A complex. Therefore, both datasets were fit simultaneously so that $k_{off\ Int.}$ and $k_{off\ A}$ were shared.

L) Single-molecule FISH measurements using probes that target the 3'UTR of MYH9. Histogram of number of RNAs per TS reveals MYH9 is actively transcribed in many cells.

M) Experimental details for overnight imaging of MYH9 transcription and splicing kinetics and for orbital tracking measurements for cells treated with and without pladienolide B (PB).

N) Comparison of mean GFP intensity of transcription site and trace duration across all three orbital tracking experiments.

O) Raw time traces from orbital tracking measurements of MYH9-MS2 (blue) and U2AF1-HALO (pink). This assays requires U2AF1-HALO to be underlabeled with 5-10 pM JF646 dye. For this reason, while splicing appears in most tracks (blue arrows), we did not observe clear U2AF binding in every cell. See methods for details.



	K_f	K_s	$k_{off\,S2}$ / $k_{off\,S3}$ For PSI<0.1
Model 1	0.15	NA	10.7
Model 2	1	0.15	10.7
Model 3	1	0.15	4.3

	K_f	K_s	$k_{off\,S2}$ / $k_{off\,S3}$ For PSI<0.1
Model 1	1	NA	17.8
Model 2	1	1	13.1
Model 3	1	1	5.3

Extended Data Fig. 5. Alternative splicing model

A) Schematic and master equations for alternative splicing model (equation 2).

B) Plotting median PSI from alternative splicing analysis as in figure 3E. PSI values were binned by −ΔΔG, the free energy of U2AF binding the S2 site relative to the S3 site. Red circles represent median PSI of each bin. Error bars represent the average standard deviation of PSI for each alternative 3'SS (n = 2 replicates). The red curve represents a fit to the alternative splice model (equation 2) (K_f = 1.0 ± 0.5, K_s = 0.15 (not fit)). To fully describe the variability of this data, we have also included a binned scatter plot showing the underlying variability. The color of each bin corresponds to the number of 3'SSs with a particular −ΔΔG and PSI combination as indicated by scale bar.

C) Three models of splice site selection. Model 1 represents a standard Michelaelis-Menten reaction scheme. Model 2 represents a 4 state model where Unbound, E complex and Intermediate states are in equilibrium and Model 3 represents the Kinetic Proofreading Model used in this manuscript.

D) Comparison of the three models using the rates that best approximate the early spliceosome (K_f = 1, K_s = 0.15). Models 1 and 2 predict a nearly identical relationship between ΔΔG and PSI, where a 10.7-fold difference in dissociation rate between site S2 and S3 is required to produce PSI<0.1. The proofreading model (model 3) is requires only a 4.3-fold difference in dissociation rate.

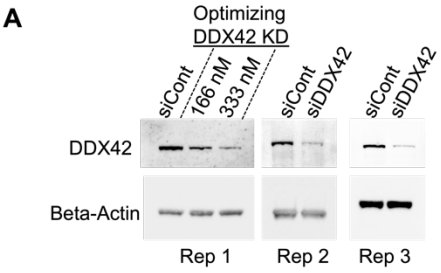
E) Comparing the predicted relationship between ΔΔG and PSI under less stringent conditions (K_f = 1, K_s = 1) reveals some differences between Model 1 and Model 2.

F) Measuring U2AF binding to the AdML 3'SS in the absence and presence of 10% nuclear extract. These are the conditions we used for the RBNS + Nuclear Extract Assay.

G) Scanning U2AF Probound models in a 12 nt window for the 50 nt upstream of the MYH9 intron 1 3'SS. This 3'SS was used in the orbital tracking assay.

H) Comparing the relationship between ΔΔG and PSI for siControl and siSF3B1 at alternative 3' splice sites using RNA-seq data from [28]. The error bars represent the standard deviation from 7 and 2 replicates for siControl and siSF3B1 experiments, respectively.

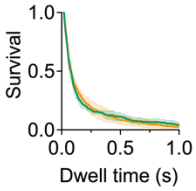
I) Representative SF3B1 knockdowns for imaging U2AF1-HALO and splicing kinetics in Figure 3.



B

Fast-tracking U2AF1-HALO

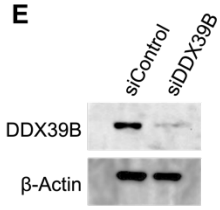
siRNA	siDDX42
# Cells	90
# Tracks	17855
Reps	3



D

Slow-tracking U2AF1-HALO (siDDX42)

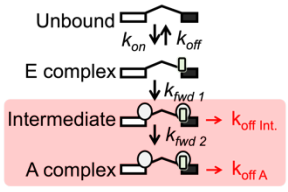
siRNA	siControl	siDDX42
# Cells	22	27
# Tracks	3348	5611
# reps	3	3



Slow-tracking U2AF1-HALO (siDDX39B)

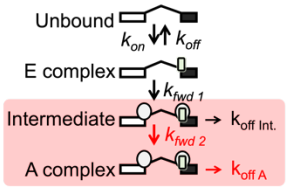
	siControl	siDDX39B
# Cells	5	10
# Tracks	1969	3342
# reps	1	1

F



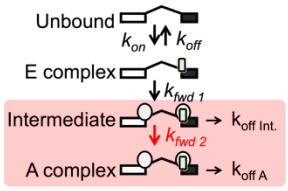
Interpretation: Only $k_{fwd\ 2}$ is affected (same result as PB)

fix $k_{off\ Int}$ and $k_{off\ A}$		
	BIC	
REP	siControl	siDDX42
1	-207	-336
2	-367	-385
3	-229	-361



Interpretation: Only $k_{off\ Int}$ is affected

fix $k_{off\ A}$ and $k_{fwd\ 2}$		
	BIC	
REP	siControl	siDDX42
1	-256	-482
2	-545	-596
3	-314	**-515**

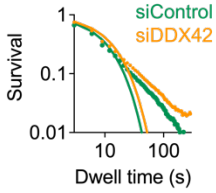


Interpretation: DDX42 accelerates $k_{off\ int.}$ and $k_{off\ A}$

fix $k_{fwd\ 2}$		
	BIC	
REP	siControl	siDDX42
1	**-276**	**-496**
2	**-565**	**-601**
3	**-327**	-503

G

Splice model, $k_{off\ int}$ and $k_{off\ A}$ constrained

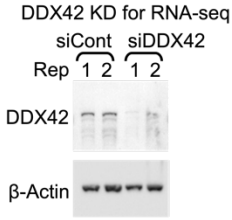


H

Orbital tracking MYH9-MS2 + U2AF1-HALO

	siControl	siDDX42
# Cells	39	11
# Reps	4	2

I

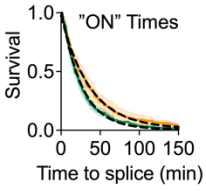


J

Splicing kinetics of Gene Trap Cell Line

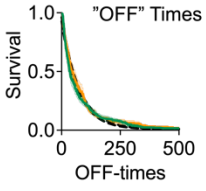
	siControl	siDDX42
# Cells	122	49
# Reps	4	2

K



	$\tau_{1/2\ ON}$
siControl	18 min
siDDX42	26 min

L



	$\tau_{1/2\ OFF}$
siControl	69 min
siDDX42	74 min

1637
1638

Extended Data Fig. 6. Characterizing the role of DDX42 in spliceosome assembly kinetics

A) Western blot of DDX42 knockdown for each replicate of the SMT assay

B) Experimental details for fast tracking of U2AF1-halo after DDX42 knockdown.

C) Dwell times times measured in fast SMT assay for siControl and siDDX42 experiments. DDX42 does not seem to impact the U2AF1-HALO dwell times on these time scales.

D) Experimental details for slow tracking of U2AF1-HALO after DDX42 knockdown.

E) Western blot of DDX39B knockodown and experimental details on slow tracking of U2AF1-HALO after DDX39B knockodown.

F) Model testing for simultaneous fitting of siControl and siDDX42 slow SMT datasets. We first tried the exact same fitting conditions as the PB experiment (left panel). Fitting was improved by constraining $k_{fwd\ 2}$ and $k_{off\ A}$ (middle panel) and further improved by only constraining $k_{fwd\ 2}$ (right panel). Results of model testing are shown in table for each replicate. Best fits (as indicated by Bayesian Information Criterion values) are indicated in bold.

G) Simultaneous fit for siControl and siDDX42 slow SMT datasets to the same model as used for PB experiment. We deemed this fit inadequate. BIC values are shown in the left panel of F.

H) Experimental details for orbital tracking experiment (siControl and siDDX42).

I) Western blot of DDX42 knockdown for sample sent for RNA-sequencing.

J) Experimental details for overnight confocal imaging of gene trap cell line.

K) Comparison of ON-time distributions for siControl and siDDX42 in the polyclonal gene trap cell line. ON-time represents the splicing kinetics.

L) Comparison of OFF-time distributions for siControl and siDDX42 from imaging the polyclonal gene trap cell line. 'OFF-time' distribution represents the time between a splicing event and the start of a transcription event.