

Fractionation iCLIP detects persistent SR protein binding to conserved, retained introns in chromatin, nucleoplasm and cytoplasm

Mattia Brugiolo¹, Valentina Botti¹, Na Liu¹, Michaela Müller-McNicoll² and Karla M. Neugebauer^{1,*}

¹Department of Molecular Biophysics and Biochemistry, Yale University, 333 Cedar St., New Haven, CT 06520, USA and ²RNA Regulation Group, Cluster of Excellence 'Macromolecular Complexes', Goethe-University Frankfurt, Institute of Cell Biology and Neuroscience, Max-von-Laue-Str. 13, 60438 Frankfurt/Main, Germany

Received October 18, 2016; Revised June 26, 2017; Editorial Decision July 20, 2017; Accepted July 20, 2017

ABSTRACT

RNA binding proteins (RBPs) regulate the lives of all RNAs from transcription, processing, and function to decay. How RNA–protein interactions change over time and space to support these roles is poorly understood. Towards this end, we sought to determine how two SR proteins—SRSF3 and SRSF7, regulators of pre-mRNA splicing, nuclear export and translation—interact with RNA in different cellular compartments. To do so, we developed Fractionation iCLIP (Fr-iCLIP), in which chromatin, nucleoplasmic and cytoplasmic fractions are prepared from UV-crosslinked cells and then subjected to iCLIP. As expected, SRSF3 and SRSF7 targets were detected in all fractions, with intron, snoRNA and lncRNA interactions enriched in the nucleus. Cytoplasmically-bound mRNAs reflected distinct functional groupings, suggesting coordinated translation regulation. Surprisingly, hundreds of cytoplasmic intron targets were detected. These cytoplasmic introns were found to be highly conserved and introduced premature termination codons into coding regions. However, many intron-retained mRNAs were not substrates for nonsense-mediated decay (NMD), even though they were detected in polysomes. These findings suggest that intron-retained mRNAs in the cytoplasm have previously uncharacterized functions and/or escape surveillance. Hence, Fr-iCLIP detects the cellular location of RNA–protein interactions and provides insight into co-transcriptional, post-transcriptional and cytoplasmic RBP functions for coding and non-coding RNAs.

INTRODUCTION

RNAs are rarely, if ever, alone in the cell. Most RNA classes are bound by RNA binding proteins (RBPs), thus forming ribonucleoproteins (RNPs). This process begins during transcription and is fundamental for the maturation and stabilization of RNAs (1,2). More than 600 RBPs are annotated in the mammalian genome based on the presence of characterized RNA binding domains, and recent experiments suggest that ~1,000 proteins expressed by cells have RNA binding activity (3,4). RBPs regulate and often catalyze essential steps in the processing and function of coding and non-coding RNA including: 5' end capping, editing, pre-mRNA splicing, 3' end cleavage and polyadenylation, assembly of export-competent RNPs, RNA localization, translation, stability and degradation. Accordingly, RNPs contain different proteins, depending on the RNA class and sequence as well as the stage of maturation. The composition of RNPs thereby determines the fate and function of all RNAs (1).

RNP maturation is likely a dynamic process involving the binding and release of multiple factors that occurs on chromatin, within the nucleoplasm, and in the cytoplasm. Many RBPs bind pre-mRNAs during transcription by RNA Polymerase II (Pol II). This co-transcriptional binding is a fundamental feature in pre-mRNA maturation, which regulates co-transcriptional processing steps like capping and splicing (5,6). Co-transcriptional RNA binding produces nascent RNPs, which lie adjacent to the DNA axis (7). Historically, RNPs containing pre-mRNAs were termed heterogeneous nuclear ribonucleoprotein particles (hnRNPs), which may be expected to include both nascent RNPs and those released from chromatin by polyadenylation cleavage. Splicing continues in the nucleoplasm, where mRNP assembly for export is finalized (8). In the cytoplasm, RBPs regulate mRNA localization, translation, stability, and degradation.

*To whom correspondence should be addressed. Tel: +1 203 785 3322; Email: karla.neugebauer@yale.edu

The serine-arginine rich splicing factors, SR proteins, are a highly conserved family of RBPs that regulate Pol II transcription, pre-mRNA splicing, polyadenylation, nuclear export, translation and stability (9,10). SR proteins bind exonic and intronic splicing enhancers (ESEs and ISEs) to promote the inclusion or exclusion of exons. Recent genome-wide studies have shown that SR proteins preferentially bind exonic sequences—possibly because of the higher abundance of exonic sequences in total cellular RNA—but also have a great number of binding sites in intronic regions (11–15). Consistent with their role in co-transcriptional splicing, SR proteins are present at sites of transcription and can be detected on chromatin by ChIP (13,16–17). Some SR proteins can recruit the nuclear export factor 1 (NXF1) to bind RNAs, leading to the export of mRNA to the cytoplasm (18–20). Consistent with this activity, SR proteins shuttle to the cytoplasm, where they can regulate translation and/or stability (10–11,15–16,21–23). Finally, SR protein interactions with many different ncRNAs, including snoRNAs, 7SK, pri-miRNAs and MALAT1, participate in gene regulatory programs through strictly nuclear activities (12–13,15,24). Thus, SR proteins can perform multiple functions on multiple classes of RNA in both the nucleus and the cytoplasm.

How RBPs, including SR proteins, interact with (pre-)mRNA and/or ncRNA along the pathway of gene expression is poorly understood. Most genome-wide methods are not adapted to the detection of RBP functions in terms of cellular compartments and RNP dynamics. Specifically, ultraviolet (UV) CrossLinking ImmunoPrecipitation (CLIP) combined with deep sequencing is a powerful method for capturing RNA–protein interactions in the whole cells and tissues (25,26). Variations on CLIP, namely HITS-CLIP, PAR-CLIP and iCLIP, allow for specific identification of targets and binding sites of RBPs. Because UV crosslinking induces covalent bonds only at short distances, CLIP has the potential to reveal the dynamics of RNA–protein interaction in different cellular compartments and/or biochemical preparations. For example, two previous studies employed UV-crosslinking to uncover RBP functions in cytoplasm (27,28). Yet, this property has not been fully exploited to comprehensively address RBP function throughout the cell.

Here, we developed a broadly applicable method, Fractionation iCLIP (Fr-iCLIP), to determine RBP targets and binding sites in chromatin, nucleoplasmic and cytoplasmic subcellular fractions. Building on iCLIP, Fr-iCLIP does not require the introduction of modified nucleotides or mutations yet identifies RBP binding sites and their targets with high precision and resolution (29,30). We applied Fr-iCLIP to two SR proteins, SRSF3 and SRSF7, because they are expected to interact with RNA in all three fractions: SRSF3 and SRSF7 are both involved in co-transcriptional splicing and maturation of export-competent mRNPs through recruitment of NXF1 (18,19). Furthermore, both shuttle from the nucleus to the cytoplasm (16,21,23). Indeed, we show that SRSF3 and SRSF7 persist on mRNAs and RNA elements consistent with nuclear and cytoplasmic processing events. We report the unexpected detection of a subset of highly conserved, retained introns in the fraction cytoplasmic and explore their features.

MATERIALS AND METHODS

Cell lines and growth conditions

Recombineering and BAC-transgenesis was used to generate stable P19 cell lines carrying stably integrated alleles encoding SRSF7-GFP and SRSF3-GFP, as described (11). Cells were grown in Dulbecco's Modified Eagle Medium, (Life Technologies). The medium was supplemented with 10% heat-inactivated Fetal Bovine Serum (FBS, Life Technologies) and 100 units/ml (U/ml) Penicillin and 100 µg/ml Streptomycin (Pen-Strep, Life Technologies). Additionally, for BAC-containing cell lines, 500 µg/ml of Geneticin (Life Technologies) was added to the media.

Fractionation iCLIP (Fr-iCLIP)

Cells were grown to confluency ($\sim 20.0 \times 10^6$ cells) and they were then UV crosslinked using a Spectrolinker XL-1500 (Spectronics) with a wavelength of 254 nm and energy of 100 mJ/cm² for 14 s and with the cell plate at 8 cm from the UV source. The cells were then subjected to cell fractionation as follows. The cells were washed with ice cold 1× PBS and detached from the plate by scraping with a cell scraper. The detached cells (in PBS) were transferred to a 15 ml falcon tube and then centrifuged at 180 g for 5 min at 4°C. At this point, the supernatant was removed and the pellet was gently resuspended in 2 ml Hypotonic Buffer (10 mM Tris-HCl pH 7.5, 10 mM KCl, 1.5 mM MgCl₂, 0.5 mM DTT; supplemented with 1× protease inhibitor cocktail (Roche)). The samples were separated into two fresh 1.5 microfuge tubes with 1 ml each that were processed in parallel. The samples were incubated on ice for 15 min and centrifuged at 425 × g for 10 min at 4°C. The supernatant was discarded. Cell pellets were resuspended in 1 ml of Lysis Buffer 0.3 (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM MgCl₂, 0.3% NP-40 (v/v); supplemented with 1× protease inhibitor cocktail (Roche)) and incubated on ice for 10 min before centrifugation at 950 × g for 10 min at 4°C. The supernatant was saved in a clean microfuge tube and was designated the cytoplasmic fraction. The pellet was resuspended with 1 ml Lysis Buffer 0.5 (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM MgCl₂, 0.5% NP-40 (v/v); supplemented with 1× protease inhibitor cocktail (Roche)) and incubated on ice for 10 min before being centrifuged at 950 g for 10 min at 4°C. The supernatant was discarded, and the pellet containing the nuclear sample was fractionated further to obtain nucleoplasm and chromatin (similarly to what was described in (31)). To do so, the nuclear pellet was resuspended in 100 µl of Buffer 1 (50% glycerol (v/v), 20 mM Tris-HCl pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 0.85 mM DTT), followed by 900 µl of Buffer 2A (20 mM HEPES pH 7.6, 300 mM NaCl, 0.2 mM EDTA, 1 mM DTT, 7.5 mM MgCl₂, 1 M urea, 1% NP-40 (v/v), 400 U of RNaseOUT (Invitrogen)). The samples were vortexed for 10 sec and incubated on ice for 10 min. Chromatin was sedimented at 15 000 × g for 5 min at 4°C. The supernatant was transferred to a clean 1.5 ml microfuge tube (nucleoplasmic fraction). Then 100 µl of Buffer 1 was added to the samples with 900 µl of Buffer 2B (20 mM HEPES pH 7.6, 300 mM NaCl, 0.2 mM EDTA, 1 mM DTT, 7.5 mM MgCl₂, 1 M urea, 1.5% NP-40 (v/v), 400 U of RNaseOUT (Invitrogen)). Samples

were vortexed for 10 s and incubated on ice for 10 min. The chromatin was sedimented at $15\,000 \times g$ for 5 min at 4°C . The supernatant was discarded, and the pellets were washed twice by adding $600\ \mu\text{l}$ of Buffer 2A. Finally, the chromatin was sedimented at $15\,000 \times g$ for 5 min at 4°C . This chromatin fraction was resuspended in 1 ml of Buffer 3 (50 mM Tris-HCl pH 7.4, 100 mM NaCl, 0.1% SDS, 0.5% Sodium deoxycholate, 400 U of RNaseOUT (Invitrogen)). To disrupt DNA before immunopurification, the chromatin and nucleoplasmic fractions were sonicated with a Branson digital sonifier (BRANSON) at 30% amplitude, for 30 s total (10 s ON and 20 s OFF). All three fractions were separately centrifuged at $20\,000 \times g$ for 5 min. The supernatants were tested with fraction-specific markers by western blotting using 1/100th of each fraction. Fr-iCLIP samples were then subjected to iCLIP protocol as described in (30). For IP protein G Dynabeads, coupled with goat αEGFP (D. Drechsel, Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG), Dresden). High-throughput sequencing of iCLIP libraries was performed on Illumina HiSeq2000 platform, with single-end 75nt reads.

SDS-PAGE and western blot analysis

For SDS-PAGE, 1–20 μg of total protein samples was denatured with Laemmli loading buffer (Bio-Rad). The samples were run on a pre-casted NuPAGE 4–12% Bis-Tris gel (Invitrogen). After electrophoresis, the proteins were transferred to nitrocellulose (Whatman), which was incubated overnight with the primary antibody at 4°C . The following antibodies were used: mouse αEGFP (Millipore; 1:5000), rabbit αGAPDH (Santa Cruz Biotechnology; 1:1000), rabbit αSRSF7 (Santa Cruz Biotechnology; 1:1000), mouse αSRSF3 (7B4, REF), goat αNXF1 (Santa Cruz Biotechnology; 1:750), rabbit $\alpha\text{Histone H3}$ (Abcam; 1:10 000), rabbit $\alpha\text{RNA Pol II}$ (Santa Cruz Biotechnology; 1:2000), donkey $\alpha\text{-rabbit-HRP}$ (GE-Health care; 1:8000), donkey $\alpha\text{-goat-HRP}$ (Sigma; 1:8000), goat $\alpha\text{-mouse-HRP}$ (Sigma; 1:10 000).

Bioinformatic analysis of Fr-iCLIP-Seq data

Fr-iCLIP-Seq data was uploaded to the bioinformatic tool iCOUNT (<http://icount.fri.uni-lj.si/>) and analyzed using default iCOUNT options and the mm9 reference genome. After analysis of reproducibility, replicates were pooled to allow the definition of the position and score of the significant peaks. Allocation of the Fr-iCLIP-tags to different RNA biotypes and regions within mRNA was performed using ENSEMBLE gene annotations. To plot the SR protein binding distribution (crosslink sites) from the Fr-iCLIP data along exon–intron junctions and surrounding polyA cleavage sites, Fr-iCLIP crosslink sites were mapped within ± 200 nt from the exon–intron junction or $-200/+600$ nt for polyA sites. Each crosslink site was assigned to the closest junction with a score of one, and the resulting signal was normalized to the local maximum within the plot to allow comparison among different fractions and libraries. Junctions for exons shorter than 60 nt and introns shorter than 200 nt were not considered in our analysis.

Intron analysis was performed by intersecting the peak locations obtained from iCOUNT for the cytoplasmic fraction with the genomic coordinates for introns. Reads containing rRNA sequences that mapped to introns were excluded to avoid ambiguity. We tabulated the number of Fr-iCLIP tags for each cytoplasmic intron bound by either SRSF3 or SRSF7. Based on the resulting frequency distribution, 286 introns bound by either SR protein were selected as top hits determined with the criteria of ≥ 19 for SRSF3 and ≥ 24 for SRSF7. The overlap Venn diagram of SRSF3-/SRSF7-binding introns was produced in R. To analyze the conservation of the resulting introns, the PhastCons track from UCSC genome browser (32) was used. The average of conservation scores across the whole intron represent the conservation score of the intron. The same calculation was applied to all genomically encoded introns (mm9 introns) and to previously reported 200nt-long UCEs (33,34) in our list with average conservation scores of 0.65. The 286 cytoplasmic introns were grouped into three categories based on their conservation scores: low (0–0.2), medium (0.2–0.6), and high (>0.6). For size characterization, the coordinates of the resulting introns and their flanking exons (left flanking exon and right flanking exon in the direction of transcript) together with all genomic exons were extracted based on Ensembl database annotation (<http://www.ensembl.org/index.html>). To determine the presence of PTCs in our identified 286 cytoplasmic introns, the protein sequences, exon sequences, intron sequences, mRNA sequences of transcripts were extracted from the UCSC genome browser to generate intron-retained transcript sequences, based on the numbering of the retained intron and estimate the translation start site. If an in-frame stop codon was in the retained intron, this intron-retained transcript was annotated as PTC-containing. To identify potentially new protein products, *in silico* translation of the obtained intron-retained sequences was performed by our in-house translation codes. Then the translation products of the intron-retained transcripts were loaded into the SMART database (<http://smart.embl-heidelberg.de>) for domain annotation analysis and compared to their original protein products for the analysis of domain gain/loss. The cytoplasmic RNA-seq data from ENCODE used in Supplementary Figure S7 is available at GEO under the accession number: GSE30567.

RNA isolation and RT-PCR

RNA was isolated using Trizol (Life Technologies) according to manufactures instructions. RNA was then resuspended in $80\ \mu\text{l}$ of water and treated with $10\ \mu\text{l}$ of $10\times$ TURBO DNase I buffer and $10\ \mu\text{l}$ of TURBO DNase I at 37°C for 30 min. Isolated total RNA was converted to cDNA with Superscript III Reverse Transcriptase (Invitrogen), following manufacturer instructions. Conventional PCR was used for the analysis of cDNA. The reaction was carried out in a total volume of $25\ \mu\text{l}$ which contained $5\ \mu\text{l}$ $5\times$ PhusionTM HF Buffer (Biozyme), $1\ \mu\text{l}$ 10 mM dNTP mix (Invitrogen), $0.5\ \mu\text{l}$ each of 10 μM forward and reverse primer, $1\text{--}2\ \mu\text{l}$ of cDNA, $0.2\ \mu\text{l}$ of Phusion polymerase (Biozyme) and ddH_2O to fill up the reaction. The material

was amplified in an Eppendorf PCR cycler following the manufacturer instructions.

Polysome fractionation

Cells were treated with 100 mg/ml cycloheximide (CHX) for 30 min, trypsinized and pelleted at $1000 \times g$ for 5 min. The cell pellet was washed with PBS, centrifuged at $1000 \times g$ for 5 min and resuspended in lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$, 0.3% NP-40 (v/v), 400 U of RNaseOUT (Invitrogen)) supplemented with 1x protease inhibitor cocktail (Roche) and incubated for 10 min on ice. The cell lysate was centrifuged at $20\,000 \times g$ for 5 min at $4^\circ C$. The resulting supernatant was layered onto a 15–45% linear sucrose gradient, spun down at 40 000 rpm for 2 h at $4^\circ C$ in a Beckmann rotor (SW41Ti) and 44 fractions were collected from the top of the gradient. The absorbance of each fraction was measured at 254 nm. Every 4 fractions were then pooled into 1 for downstream applications (11 pooled fractions in total). From each pooled fraction, the protein content was analyzed by SDS-PAGE and the RNA was extracted using Trizol (according to manufacturer's instructions), followed by ethanol precipitation. In the +EDTA experiment, CHX treatment was omitted and cells were lysed in lysis buffer containing 50 mM EDTA. The samples were then processed as described above.

NMD inhibition

CHX treatment was performed as described (12). For UPF1 knockdown, cells were grown to 25% confluency and were transfected with 70 pmol of siRNA (5'-UCAAGGUUCCUGAUAAUATT-3') using Lipofectamine RNAiMAX (Thermo Fisher). 70 pmol of a scrambled siRNA was used as control. Cells were incubated for 48h and then RNA was extracted using standard Trizol protocol. UPF1 knock down was evaluated by western blot.

RESULTS

Chromatin, nucleoplasmic and cytoplasmic fractions of UV-crosslinked cells

To effectively study SRSF3 and SRSF7 in P19 cells, we used transgenic cell lines in which each protein was tagged at its C-terminus with GFP and expressed from an integrated bacterial artificial chromosome, as previously described (12,16,35). These tagged SR proteins are expressed at physiological levels, complement the effects of knock-down of endogenous SR proteins on gene expression, and undergo nucleocytoplasmic shuttling (12,23,35). To develop Fr-iCLIP, we established a subcellular fractionation protocol for P19 cells after UV-crosslinking. Cytoplasmic, nucleoplasmic and chromatin fractions were subsequently subjected to iCLIP, which allowed the identification of RNA targets and binding sites specific to each fraction (Figure 1A).

Subcellular fractionation requires optimization and modification, depending on the cell lines or tissues used as starting material. Nucleo-cytoplasmic fractionation of P19 cells was previously established (11) and served as a starting point for the fractionation undertaken here after UV-crosslinking. The nuclear fraction was further separated

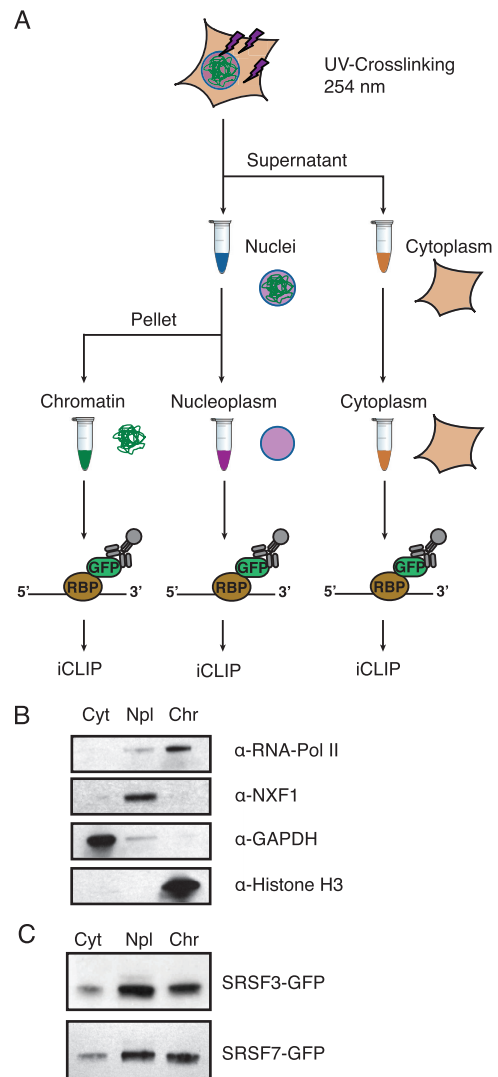


Figure 1. Fr-iCLIP combines RNA–protein crosslinking with subcellular fractionation. (A) Schematic showing workflow of Fr-iCLIP, beginning with UV crosslinking of whole cells and nuclear-cytoplasmic fractionation followed by separation of nuclear fraction (blue) into chromatin (green) and nucleoplasmic (pink) fractions. The cytoplasmic fraction is shown in orange. RNA binding proteins of interest (RBP-GFP) were immunopurified from each of these three fractions independently and subjected to standard iCLIP procedures. (B) Western blot characterization of UV crosslinked subcellular fractions, showing enrichment of Pol II and histone H3 in chromatin (Chr), NXF1 in the nucleoplasm (Npl), and GAPDH in cytoplasm (Cyt). (C) Subcellular distribution of SRSF3-GFP and SRSF7-GFP, using anti-GFP for western blot detection. In B and C, 1% of each fraction was loaded.

into chromatin and nucleoplasm through a series of sedimentation steps and washes (see Materials and Methods). Figure 1B shows the enrichment of specific components in each fraction. Histone H3 and Pol II were highly enriched in the chromatin fraction and GAPDH was cytoplasmic, as expected. Furthermore, we found that nuclear export factor NXF1 was a reliable marker of the nucleoplasmic fraction. Thus, we established markers for each fraction of interest and showed that subcellular fractions can be obtained after UV-crosslinking.

SR proteins are highly enriched in the nucleus (18,19), although the proportions associated with chromatin, nucleoplasm and cytoplasm were previously unknown. To address this, western blotting was performed with α -GFP, reactive with the tag to be used for affinity purification (Figure 1C). Because other antibodies are sensitive to phosphorylation state, which varies among cellular compartments (11,19), the tag provided objective detection of total SRSF3 and SRSF7 proteins in the cellular fractions. SRSF3 and SRSF7 showed strong enrichment in the nuclear fraction, as expected (19). Within the nucleus, SRSF3 and SRSF7 were strongly detected in the chromatin fraction from P19 cells, consistent with high co-transcriptional activity for both SR proteins (13,16–17). Low but significant levels of both SRSF3 and SRSF7 were detected in the cytoplasmic fraction, in accordance with their ability to efficiently shuttle from the nucleus to the cytoplasm (16,21,23).

Fr-iCLIP identifies SRSF3 and SRSF7 targets in three cellular compartments

iCLIP was performed on the three subcellular fractions from SRSF3-GFP and SRSF7-GFP cell lines, obtaining Fr-iCLIP libraries (Supplementary Figure S1) for RNA-Seq on the Illumina platform (75bp, single end reads). The mapped reads from three to four biological replicates were well-correlated (Supplementary Tables S1 and S2), showing reproducibility. The data was then analyzed using iCOUNT (36), yielding datasets denoting significant binding sites (FDR < 0.05) for SRSF3 and SRSF7.

The number and identity of SRSF3 and SRSF7 RNA targets in different subcellular fractions was determined (Figure 2A and B, top panels). Comparison of the set of unique and common mRNA targets between the nucleus (nucleoplasm plus chromatin) and cytoplasm revealed the dynamic behavior of both RBPs. On the one hand, SRSF3 and SRSF7 had 4214 and 2338 mRNA targets uniquely detected in the nucleus and 331 and 1190 targets uniquely in the cytoplasm, respectively, consistent with distinct roles in nuclear and cytoplasm RNA regulation. On the other hand, 1520 and 1395 SRSF3 and SRSF7 mRNA targets were shared between the nucleus and cytoplasm, in line with the function of both SR proteins as major mRNA export adapters that may remain associated with their mRNA cargoes (11). Consistent with this possibility, 5% and 15% of SRSF3 and SRSF7 binding signals, respectively, were present at the same mRNA sites from nucleus to cytoplasm, suggesting a small proportion of persistent interactions. SRSF3, globally the major mRNA export adapter (11), displayed strong overlap of mRNA targets between nucleoplasm and chromatin, with a large number of targets identified only in the chromatin fraction. One possibility is that nucleoplasmic mRNAs are only transiently bound and/or quickly exported to the cytoplasm, resulting in their relatively inefficient crosslinking and detection. Overall, the distinct SRSF3 and SRSF7 binding profiles detected in the nucleus and cytoplasm indicates that many interactions with (pre-)mRNA are compartment-specific.

If Fr-iCLIP data accurately reflect compartmentalized RNA–protein interactions, then the (pre-)mRNA binding regions observed should reflect the expected processing sta-

tus of the RNA detected in that compartment. There are specific expectations for the chromatin fraction, which contains nascent RNA (37,38). First, we expect a bias towards intron binding in the chromatin fraction, because most introns are removed co-transcriptionally (6). Indeed, intron reads were enriched in chromatin and reduced in nucleoplasm (Figure 2A and B, bottom panels), where intronic reads likely reflect delayed splicing and/or RBP interactions with lariat intermediates before degradation (8,39). Second, only the chromatin fraction should contain transcripts that map to gene regions downstream of polyA cleavage sites and before transcription termination. To determine whether SRSF3 and SRSF7 Fr-iCLIP detected these reads in a compartment-specific manner, the density of Fr-iCLIP reads along the 3'UTR-intergenic boundary for all bound 3'UTRs was plotted (Supplementary Figure S2A). Reads downstream of polyA cleavage sites were almost exclusively detected in the chromatin fraction. Overall, these findings confirm that Fr-iCLIP detects compartment-specific (pre-)mRNAs and nascent RNA through the positive selection afforded by RBP immunopurification.

Using standard iCLIP, previous studies have reported SR protein binding to non-coding RNAs, such as snoRNAs (11,12). As expected, high levels of SRSF3-GFP and SRSF7-GFP binding to non-coding RNA (ncRNA) was detected (Figure 2, bottom panels). Analysis of binding sites mapping to different ncRNA classes revealed differences among the three compartments (Figure 3, left panels). Mitochondrial mt-ncRNAs (mt-rRNA and mt-tRNA) represented the ncRNA targets with highest cytoplasmic binding for both SRSF3 and SRSF7 (>55%), whereas it encompassed <5% of the ncRNA reads in the chromatin fraction. Conversely, the most highly represented ncRNA class detected in the nucleus was snoRNAs (>55% of reads), whereas binding in the cytoplasm was almost absent (Figure 3A and B, left panels). This compartmentalized interaction can be appreciated through examination of iCLIP reads mapped to unprocessed protein-coding transcripts that harbor snoRNAs within introns (Figure 3A and B, right panels): both SRSF3 and SRSF7 display binding to exons, snoRNAs, and some introns in the nuclear fractions, whereas predominantly exons are bound in cytoplasm. Binding to introns and intron-encoded snoRNAs likely occurs during splicing and/or downstream processing of snoRNAs from the intron lariat (12,40). Finally, reads mapping to long ncRNAs (lincRNA) displayed a bias towards the nucleus, reflecting the commonly observed nuclear localization of this class (41). Taken together, the interactions of SR proteins with different classes of ncRNA supports unique roles in ncRNA metabolism, particularly in the nucleus, and further validates the compartment specificity of the RNA–protein interactions detected by Fr-iCLIP.

A stringent test of Fr-iCLIP is to determine whether the sum of the reads from all three cellular compartments recapitulates iCLIP from total cell lysates. To test this, the Fr-iCLIP data from the three fractions were pooled for SRSF3 and SRSF7 and compared to our published total-iCLIP data (11). Pooled Fr-iCLIP data overlapped almost completely with total cell iCLIP data for both proteins (Supplementary Figure S2B). Furthermore, the level of bind-

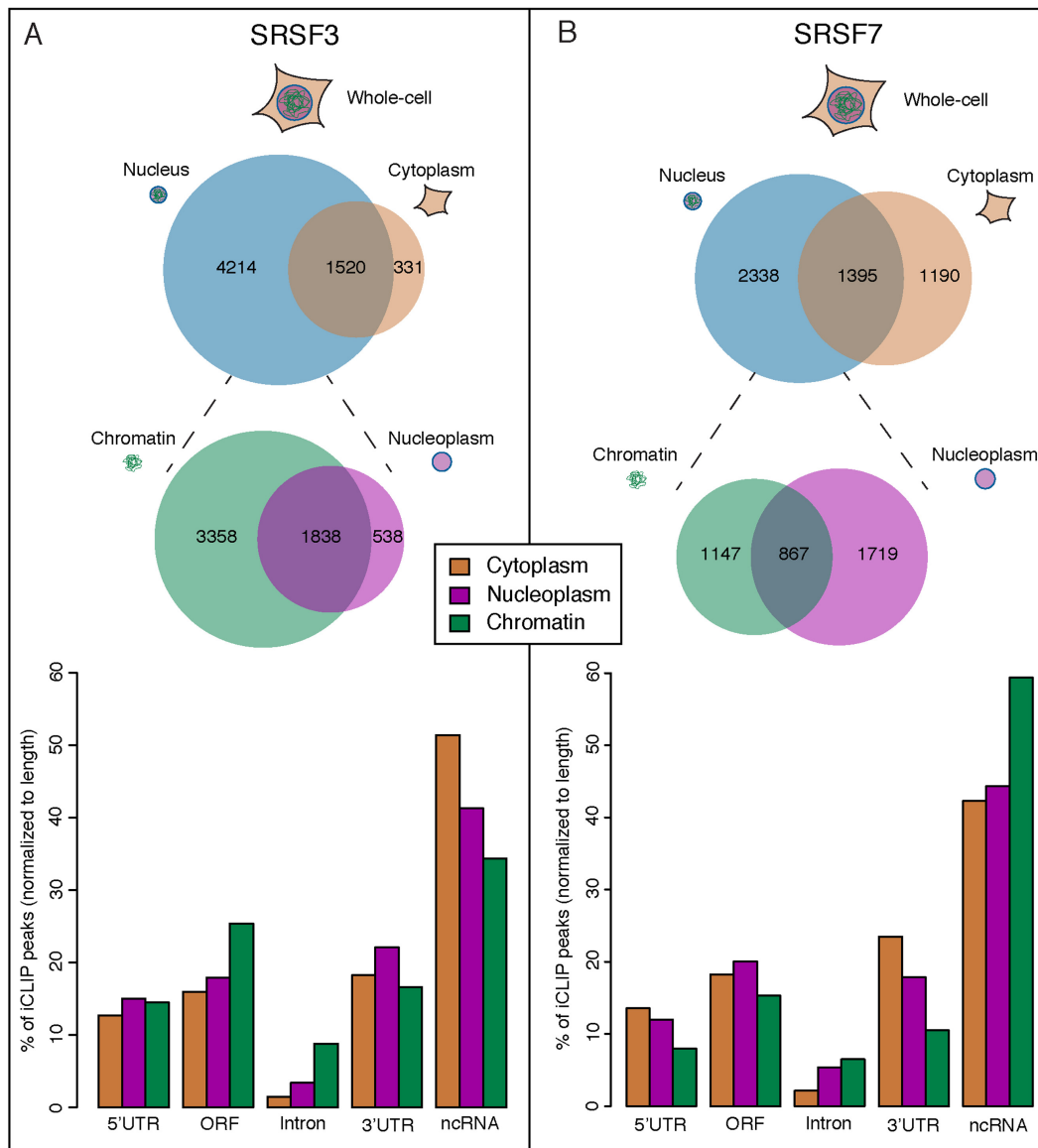


Figure 2. Fr-iCLIP reveals the RNA targets of SRSF3-GFP and SRSF7-GFP in chromatin, nucleoplasm and cytoplasm. (A) Upper panel, Venn diagram representing the number and degree of overlap among Fr-iCLIP mRNA/pre-mRNA targets for SRSF3-GFP in nucleus and cytoplasm (Cyt) and between nucleoplasm (Npl) and chromatin (Chr). Lower panel, distribution of SRSF3-GFP Fr-iCLIP peaks among mRNA regions and ncRNAs. Percent of total identified Fr-iCLIP peaks normalized to feature length is shown for each cellular fraction, other features such as intergenic regions are not shown due to their low level. (B) Fr-iCLIP data for SRSF7-GFP, following the scheme shown in A.

ing to overlapping targets was analyzed and the pooled Fr-iCLIP data was highly correlated with the whole cell iCLIP data (Supplementary Figure S2C). Thus, Fr-iCLIP recapitulates total RBP-RNA interactions obtainable from whole cell iCLIP methods and datasets. Importantly, Fr-iCLIP adds fundamental knowledge regarding the localization of RNA-protein interactions to distinct cellular compartments where different steps in RNA biogenesis and regulation occur.

SRSF3 and SRSF7 bind distinct functional mRNA groups in cytoplasm

One application of compartment-specific analysis of RNA-protein interactions is to address the role of RBPs in nuclear versus cytoplasmic events. To determine whether SRSF3 and SRSF7 regulate nuclear and cytoplasmic mRNAs with different functions, GO-term enrichments for the identified transcripts were determined (Supplementary Tables S3&S4). Transcripts enriched in splicing variants were enriched in all fractions. As previously described, SRSF3 and SRSF7 targets were enriched in RNA-binding or nucleotide binding (11–13). Interestingly, GO-term enrichments for SRSF3 and SRSF7 targets bound uniquely in the cytoplasm include those encoding for proteins containing transmembrane regions. In addition, SRSF7 cytoplasmic targets were enriched in transcripts encoding ER pro-

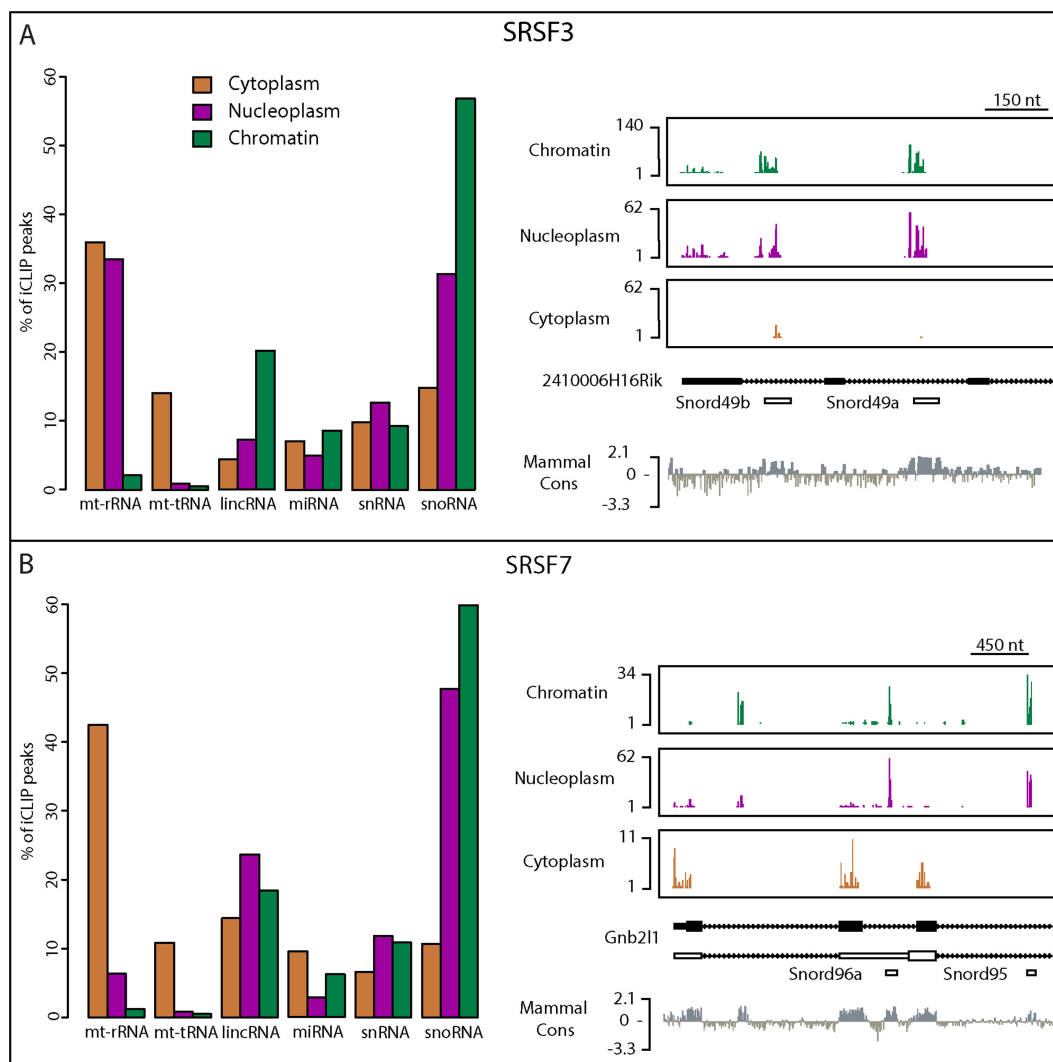


Figure 3. Fr-iCLIP detects SR protein interactions with ncRNAs in specific subcellular compartments. (A) Left panel, distribution of SRSF3-GFP Fr-iCLIP peaks among ncRNA species. Percent of total identified Fr-iCLIP peaks for each cellular fraction are shown. Only ncRNAs with >1% ncRNA binding in at least one fraction were considered in this analysis. Right panel, SRSF3-GFP Fr-iCLIP peaks mapping within 2410006H16Rik, which harbors two snoRNAs in its introns. (B) Left panel, distribution of SRSF7-GFP Fr-iCLIP peaks, following the scheme shown in A. Right panel, SRSF7-GFP peaks mapping within Gnb211, which contains a possible novel snoRNA in intron 1 and two snoRNAs in introns 2 and 3.

teins, whereas SRSF3 cytoplasmic targets were enriched in transcripts encoding intracellular proteins and proteins involved in different metabolic processes. These findings suggest roles for SRSF3 and SRSF7 in the nuclear processing of transcripts encoding RBPs themselves and in the cytoplasmic regulation—possibly translation or stability—of discrete pools of mRNAs encoding proteins with different functions.

Fr-iCLIP detects retained introns in the cytoplasm

Because transcripts are expected to be fully spliced in the nucleus before export to the cytoplasm, intron binding is expected to be nuclear. To address this globally, SRSF3 and SRSF7 signals along exon–intron junctions were analyzed on all bound transcripts (Figure 4). In all fractions, maximum signals peaked in the exon area, whereas intronic sig-

nal varied among fractions. Specifically, SRSF3 and SRSF7 binding to introns was highest in chromatin, lower in nucleoplasm, and lowest in cytoplasm. The decrease in intron binding from chromatin to nucleoplasm to cytoplasm may reflect the range of splicing kinetics for individual introns, because splicing is predominantly co-transcriptional but can continue post-transcriptionally (6,37–38). However, Fr-iCLIP analysis detected low levels of binding to introns in the cytoplasm, raising the possibility that SR proteins may significantly bind some introns in the cytoplasm.

To identify introns that may be significantly bound by SRSF3 and SRSF7 in the cytoplasm and minimize false positives, a signal-based threshold was applied (Supplementary Figure S3A and B) and identified 137 and 243 introns bound by SRSF3 and SRSF7, respectively (Figure 5A). Due to the high degree of overlap, we pooled the 286 cytoplasmic intronic targets of SR proteins (Supple-

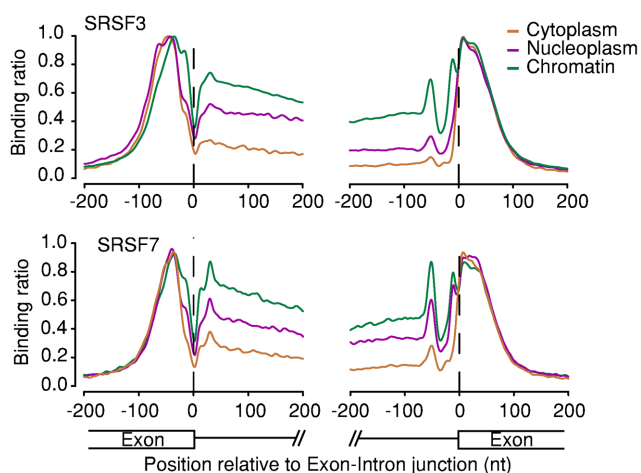


Figure 4. SRSF3 and SRSF7 contact exons in both nucleus and cytoplasm but intron binding is almost exclusively nuclear. Meta-analysis for all SRSF3-GFP and SRSF7 Fr-iCLIP peaks detected along exon-intron junctions (left) and intron-exon junctions (right). The CLIP-tag densities for each protein at exon-intron and intron-exon junctions (± 200 nt) are plotted. Higher intron binding is observed in the nuclear fractions. Y-axes represent the abundance of peaks for the region normalized to local maximum.

mentary Table S5) and queried potentially shared features among them. First, these cytoplasmic introns displayed significantly higher conservation than typical introns in the mouse transcriptome (Figure 5B) (32). Indeed, 17 of 286 introns harbor previously identified ultra-conserved elements (UCEs), which are typically defined as 200 nt sequences with conservation between 80% and 100% between human, rat and mouse (33,34). Furthermore, many of our 286 cytoplasmic introns are highly conserved along their full sequence (Figure 5C). Plotting all 286 introns according to their phastCons conservation score, we divided the cytoplasmic introns into three categories for further analysis: low, medium and high conservation (Figure 5C). Typical mouse introns have a PhastCons score of 0.1 (or 10%), leading us to set the conservation score threshold between low and medium categories 2-fold higher (0.2); the threshold between medium and high conservation (0.6) was chosen, as it is close to the median conservation score observed for UCE-containing introns (Figure 5B and C). Both low and high conservation SRSF3 and SRSF7 binding sites were observed in the three groups (Supplementary Figure S3C). Thus, the cytoplasmic introns detected by Fr-iCLIP are enriched in highly conserved sequences.

Using the cytoplasmic introns grouped into low, medium, and high conservation categories, we asked if particular features of each intron were uniquely correlated. Comparison of median intron size among the groups and to typical murine introns (1,288bp) revealed that cytoplasmic introns in the low group were 6-fold longer (7943 bp), while those in the medium and high groups were not (Figure 5D and Supplementary Figure S4A). In contrast, size differences were not observed for the exons to the right or left of the cytoplasmic introns (Supplementary Figure S4B and C). One explanation for the prevalence of long introns in the low conservation pool is that longer introns, which may be less

efficiently spliced, may have more SR protein binding sites that are each lower in their conservation. Indeed, analysis of binding site conservation revealed that cytoplasmic introns in the low group displayed a prevalence of lowly conserved binding sites, while those in the high group displayed a prevalence of highly conserved binding sites (Supplementary Figure S3C). Thus, intron-retained mRNAs detected by Fr-iCLIP in the cytoplasm are either typical in size with highly conserved binding sites or significantly longer with many lowly conserved binding sites.

The high conservation of cytoplasmic introns suggests that the mRNAs harboring them may have specific biological functions. To address this, GO-term analysis for the three groups was performed (Supplementary Table S6). The GO-term enrichment for the transcripts containing cytoplasmic introns with high and medium conservation shared most biological functions; moreover, most processes enriched in these two classes were gene expression and splicing- and RNA processing-related, in line with the idea that SR proteins can regulate splicing either directly or indirectly by regulating splicing regulators (11,12). In contrast, transcripts containing cytoplasmic introns with low conservation were more enriched in general metabolic and biosynthesis processes; other biological processes including RNA splicing and processing were identified with much lower enrichment and *P*-values.

To further pursue the functional significance of conserved cytoplasmic introns detected in the cytoplasm, we considered the possibility that the corresponding intron-retained mRNAs could be targeted by nonsense mediated decay (NMD), in which transcripts containing premature stop codons (PTCs) are normally degraded in the cytoplasm (42–44). UCE-containing transcripts, such as those encoding the SR proteins themselves, are well known to employ this mechanism for auto-regulation of protein levels (12,33–34,45). To address this, mRNAs containing cytoplasmic introns detected by Fr-iCLIP were analyzed for the frequency of introduction of PTCs into the corresponding host mRNAs. 80% of cytoplasmic introns occurred within annotated coding regions, and all of these introduce at least one PTC (Figure 5E). An alternative hypothesis is that these introns retained within coding regions could, if translated, give rise to new protein domains. Indeed, *in silico* translation into cytoplasmic introns revealed that 18% lead to the addition of potentially new domains, including transmembrane domains and low complexity domains (Supplementary Figure S5 and Supplementary Table S7). These domain types are characterized by highly repetitive amino acid stretches, in line with highly repetitive RNA sequences typical of introns. It is possible that these putative isoforms are produced at low levels or in particular cell types, providing one explanation for why these mRNA isoforms are not currently annotated.

If the intron-retained mRNA isoforms are physiologically relevant, one might expect them to be specific mRNA export targets. To address this, we focused on a distinct subset of the cytoplasmic bound introns were highly conserved along the full intronic sequence (Figure 6 and Supplementary Figure S6). Two of the most highly bound SRSF3 and SRSF7 intron targets in this class were their own transcripts (Figure 6A&B). In the Fr-iCLIP data, we saw that this

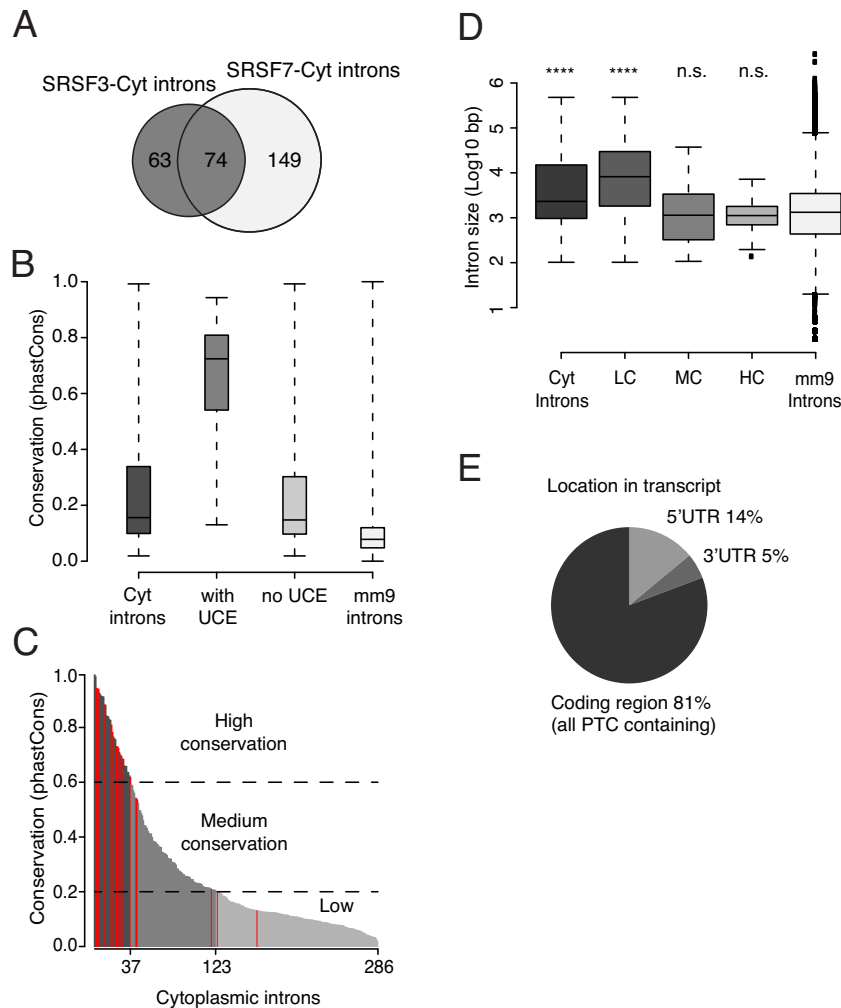


Figure 5. Features of cytoplasmic introns bound by SRSF3 and SRSF7. (A) Venn diagram showing number and overlap of cytoplasmic introns bound by either SRSF3-GFP (SRSF3-Cyt introns) and/or SRSF7-GFP (SRSF7-Cyt introns). (B) Box-plot representation of the PhastCons conservation scores for the introns identified in the cytoplasm by Fr-iCLIP (Cyt introns, $n = 286$), versus all mouse introns (mm9 introns). The subset containing previously characterized UCEs (with UCE, $n = 17$) and those without UCEs (no UCE, $n = 269$) are plotted separately; the UCEs considered are as described (33,34). The median conservation score for each group is significantly higher than mm9 introns (P -value < 0.05). (C) Rank order distribution of each Cyt intron according to conservation score. Introns are grouped as follows for further analysis: High, with conservation scores above 0.6 (dark gray); Medium, with conservation scores 0.2 to 0.6 (gray); low, with conservation scores < 0.2 (light gray). Cyt introns marked in the red contain previously characterized UCEs. (D) Box plot showing the size distribution of all mouse introns (mm9 introns), all cytoplasmic introns detected by Fr-iCLIP (Cyt introns), and cytoplasmic introns with low conservation (LC), medium conservation (MC), and high conservation (HC). Asterisks indicates that these data are significantly different from mm9-introns (P -value $< 2.2e-16$) in a two-tailed t -test. (E) Location of the identified cytoplasmic introns within different transcript regions. All introns detected in coding regions (81%) create at least one PTC.

auto-regulatory binding is maintained during RNA maturation, with the majority of SRSF3 and SRSF7 binding along highly conserved introns (90–97% nucleotide conservation between human, mouse and rat) within their own transcripts. These introns harbor so-called ‘poison cassette’ exons that introduce PTCs and trigger NMD in the cytoplasm (12,33,46). Surprisingly, we could also show that such binding is not restricted to the poison cassette, but extends along the entire intron and is maintained in the cytoplasmic fraction (Figure 6). SR proteins can recruit the nuclear export factor, NXF1, to mRNAs to facilitate their export to the cytoplasm (11). We used our previously published iCLIP data to determine whether NXF1 binds these introns (11). Indeed, NXF1 crosslinks to intronic sequences flank-

ing the poison cassette exons in both *SRSF3* and *SRSF7* (Figure 6A&B, lower panels), while the negative control (NLS-GFP) showed no binding. Furthermore, other highly conserved cytoplasmic introns detected by Fr-iCLIP (Supplementary Figure S6); these include introns in *ARGLU1*, *DDX5* and a highly conserved intron in *HNRNPH1*, which was excluded by our list due to stringent filtering. All cytoplasmic introns analyzed showed NXF1 binding. Taken together, these data suggest that the intron-retained mRNAs detected by Fr-iCLIP could be specifically exported to the cytoplasm by NXF1.

The strong binding to the introns surrounding the poison cassette exons in the cytoplasm suggests that the conserved introns may be included together with the poison cassette

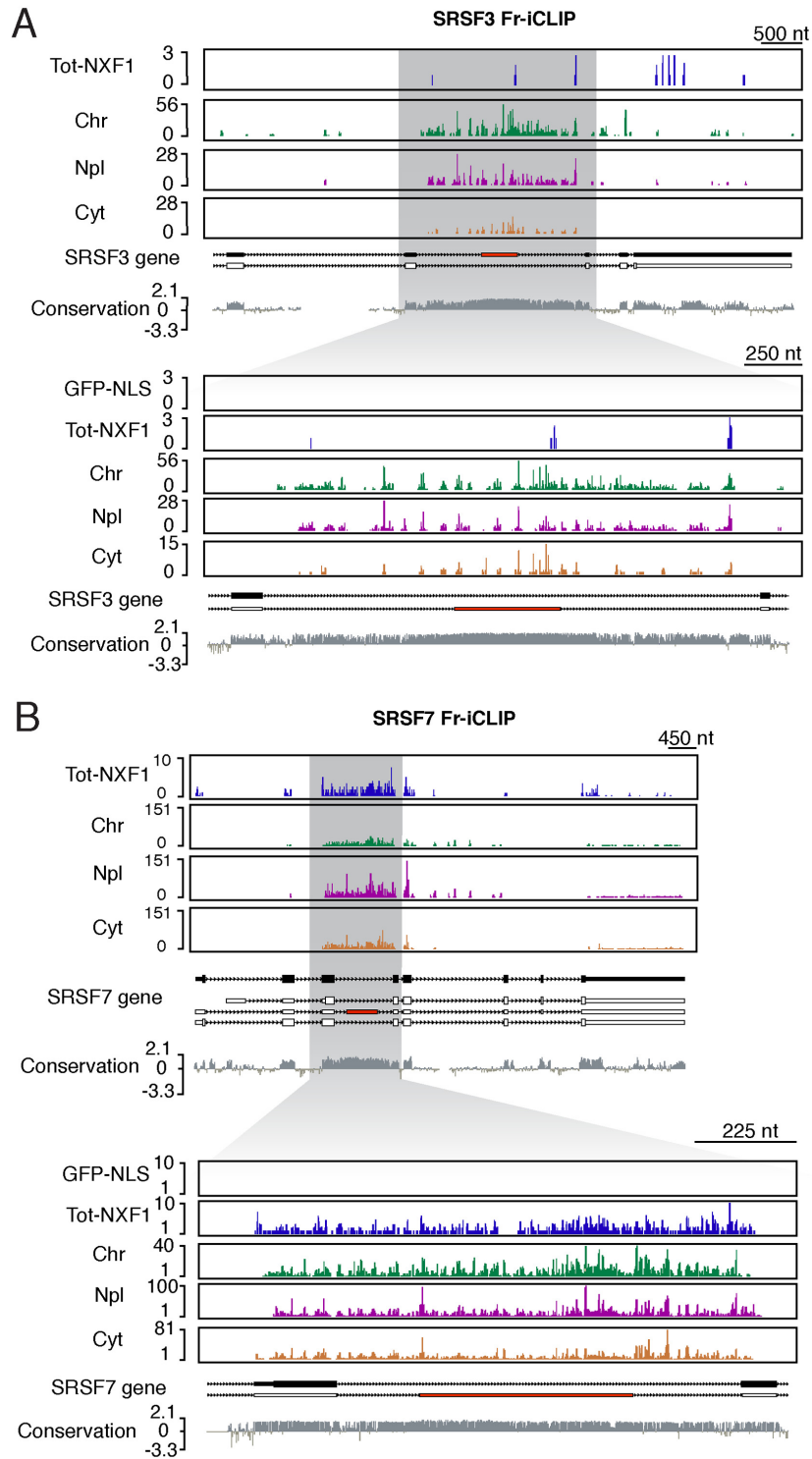


Figure 6. SRSF3 and SRSF7 strongly bind their own transcripts in all fractions, including highly conserved introns. (A) Top panel, distribution of SRSF3-GFP Fr-iCLIP peaks as well as NXF1-GFP total iCLIP peaks along *SRSF3* transcripts for the three fractions. Lower panel, zoom-in on highly conserved third intron of *SRSF3*. (B) Top panel, distribution of SRSF7-GFP Fr-iCLIP peaks as well as NXF1-GFP total iCLIP peaks along *SRSF7* transcripts for the three fractions. Lower panel, zoom-in on highly conserved third intron of *SRSF7*. Total NXF1-GFP iCLIP data is from (11).

exons. To address this, cytoplasmic mRNA was subjected to RT-PCR (Supplementary Figure S6F), validating the inclusion of the highly conserved introns and showing that the poison cassette exons can be included together with the flanking conserved introns. In contrast, intronic signal for *SRRM2* was absent by RT-PCR; the SR protein CLIP tags mapping to the *SRRM2* intron were not detectable in cytoplasm rendering *SRRM2* a negative control (Supplementary Figure S6E). Moreover, publicly available data produced for polyA+ RNA-Seq confirmed elevated levels of these introns excluding *SRRM2* in cytoplasmic mRNAs prepared from numerous cell lines (Supplementary Figure S7) (47). We conclude that intron-retained mRNA isoforms identified by Fr-iCLIP are independently detectable in the cytoplasm of P19 cells and also occur in multiple cell lines.

Intron-retained mRNAs detected in polysomes are not substrates for NMD

Because the intron-retained mRNA isoforms detected by Fr-iCLIP contain PTCs, they may trigger NMD in the cytoplasm. To test whether these mRNAs are translated, we performed polysome profiling and extracted RNA from monosome, early polysome and late polysome fractions (Figure 7A). Note that NMD-sensitive RNAs can be mostly found in the monosomes and early polysome fractions (48). RT-PCR was used to determine whether the intron-retained mRNAs discussed above were present in polysomes (Figure 7B). The intron-retained mRNAs were mostly present in monosome and early polysome fractions; this pattern of migration in the sucrose density gradient was disrupted by EDTA treatment as were polysomes (Supplementary Figure S8A&B), arguing that the presence of intron-retained mRNAs in polysome fractions is not fortuitous. We conclude that intron-retained mRNAs bound by SRSF3 and SRSF7 in the cytoplasm are present on ribosomes and candidates for regulation by NMD.

To test whether the detected intron-retained mRNAs are degraded by NMD, the abundance of intron-retained mRNA was determined under two independent conditions that inhibit NMD: CHX treatment and UPF1 knockdown (42,49). Both conditions increased the levels of the *SRSF3* and *SRSF7* poison cassette isoforms, as expected (Figure 7C and D). However, neither treatment had detectable effects on the levels of any of the other four intron-retained isoforms (*ARGLUI*, *HNRNPH1*, and *SRSF3* and *SRSF7*). Taken together, these data indicate that the intron-retained mRNAs detected by Fr-iCLIP are not substrates for NMD.

DISCUSSION

Here we combined UV-crosslinking, cell fractionation and immunopurification of RBPs (Fr-iCLIP), to obtain sensitive, high resolution RNA-protein interaction data *in vivo*. Fr-iCLIP revealed changes in the RNA binding landscape of SRSF3 and SRSF7 as transcripts proceed from chromatin to nucleoplasm to cytoplasm. Continuous occupancy of some sites, notably in exons, suggests retention of these interactions through cellular compartments and during different regulatory events. Interestingly, persistent binding to conserved introns in all three fractions highlights the un-

appreciated export of intron-retained mRNAs to the cytoplasm. Below, we expand on these points and discuss our evidence that at least some of the detected intron-retained mRNAs may be stable and functional. Given the emerging importance of intron retention in development, cellular proliferation and differentiation pathways (39,50–53), Fr-iCLIP offers a sensitive means of addressing these phenomena and their molecular underpinnings.

The Fr-iCLIP method is general and adapted to current high throughput iCLIP protocols. Two previous studies combined cytoplasmic fractionation with CLIP to detect a limited complexity of transcripts (27,28). Our chief concerns have been leakage among fractions (i.e. nucleoplasm leakage into the cytoplasmic fraction) and potential effects of UV crosslinking on existing fractionation protocols. Our protocol yields well-separated fractions, because SRSF3 and SRSF7 crosslinks to snoRNAs were limited to the nucleus and crosslinks to mitochondrial RNAs to the cytoplasm, as expected for these highly localized RNAs. The biological significance of SR protein binding to these ncRNA classes is currently unknown. SRSF3 and SRSF7 crosslinks to lncRNAs were mostly nuclear, also as expected (24,54). Finally, the sum of the iCLIP reads from all three compartments matched well with whole cell iCLIP, showing that no class or population of RNA-protein interactions was lost during the cellular fractionation steps. We anticipate that Fr-iCLIP will be broadly applicable to other experimental systems, such as cells, tissues, and model organisms.

SRSF3 and SRSF7 Fr-iCLIP revealed class of conserved introns that are retained in cytoplasmic mRNAs. The majority of crosslinking to introns was limited to the nucleus, with the highest signals on chromatin, consistent with the known predominance of co-transcriptional splicing (6). Note that co-transcriptional splicing is lower in mouse than in other species analyzed (55). Given the good separation of the cellular fractions, the detection of a specific population of introns in the cytoplasm should not be interpreted as leakage. Instead, the positive selection of bound transcripts through crosslinking allows identification of low abundance transcripts. Despite their low levels, the presence of selected intron-retained mRNAs was validated by RT-PCR, polysome fractionation, and analysis of cytoplasmic mRNA-Seq datasets. Bioinformatic analysis of 286 high confidence cytoplasmic introns revealed that 50% of these introns are conserved at least two-fold more conserved than typical mouse introns; indeed, a subset contains ultraconserved elements or UCEs (33,34) and 37 are >60% conserved overall. In addition, the cytoplasmic introns tend to be larger, with the low conservation group 6.6kb larger than usual. Highly conserved introns tended to have highly conserved binding sites; lowly conserved, long introns did not. Interestingly, one of the latter group, *CHST11*, is among the transcripts that is predicted to acquire a novel protein domain through intron retention.

Most of the intron-retention events detected by Fr-iCLIP led to the introduction of PTCs into the host mRNAs. Ultra-conserved introns have previously been shown to be auto-regulatory targets of SR proteins; poison cassette exons within these conserved introns contain PTCs and target alternative isoforms for NMD (12,33,45–46). We show here

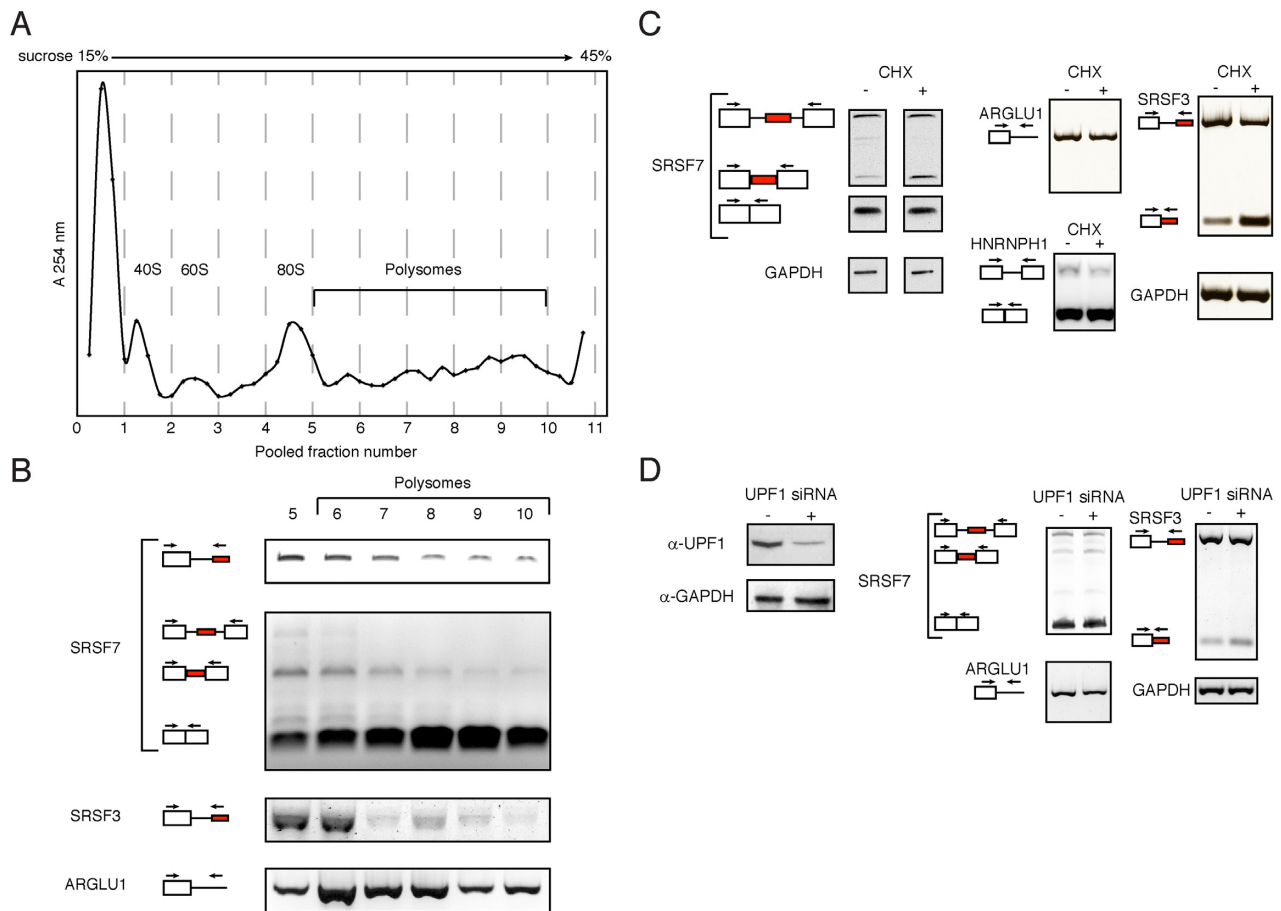


Figure 7. Intron-retained mRNAs detected by Fr-iCLIP are present in early polysomes and monosomes. **(A)** Polysome fractionation of wild-type P19 lysate by sucrose density gradient centrifugation. The cell extract was loaded into a 15–45% sucrose gradient and 44 fractions were collected from the top of the gradient. The absorbance of each fraction was measured at 254 nm and it is represented by a single dot in the profile. Peaks of absorbance of the 40S, 60S and 80S ribosomal subunits and fractions containing polyribosomes are indicated. Subsequent to the absorbance measurement and for downstream applications, every 4 fractions were pooled and numbered from 1 to 11 as indicated in the x-axis. **(B)** Total RNA was extracted from pooled fractions number 5 to 10 and the presence of intron-retained mRNAs was tested by RT-PCR. The positions of the gene-specific PCR primers used are indicated on the left. **(C)** Test of NMD sensitivity for intron-retained and poison cassette isoforms, when present, for *SRSF7*, *ARGLU1*, *SRSF3* and *HNRNPH1* mRNAs; *GAPDH* mRNA served as loading control. NMD was inhibited by treatment with CHX for 3 hours, after which the indicated RT-PCR reactions were performed using total RNA. **(D)** Test of NMD sensitivity for intron-retained and poison cassette isoforms, when present, after knock-down of UPF1. Left panel shows western blot analysis of UPF1 and GAPDH protein levels after control siRNA (–) or UPF1 siRNA (+) treatment for 48 hours. NMD sensitivity was tested by comparing changes in isoform levels by RT-PCR of total RNA as indicated.

that SR proteins and NXF1 crosslink to introns flanking poison cassette exons in *SRSF3* and *SRSF7* transcripts and that these introns are present in cytoplasm. In addition, we detected conserved target introns, such as those in *DDX5* and *HNRNPH1*, which do not contain poison cassette introns and are nevertheless present in cytoplasm. Interestingly, an intron-retained, cytoplasmic isoform of *ARGLU1* mRNA was detected shown to be resistant to NMD. It was recently shown that the conserved intron of *ARGLU1* can also undergo alternative splicing to render the transcript sensitive to NMD, and both intron retention and alternative splicing were linked to the UCE (56). Additionally, we show the presence of NXF1 on conserved intronic regions within these introns. This scenario emphasizes the significance of highly conserved intron sequences, which can have multiple and overlapping regulatory functions in splicing, mRNA export, and mRNA stability.

Taken together, Fr-iCLIP has revealed the nuclear export of a subset of intron-retained targets of *SRSF3* and *SRSF7*. We show that the four intron-retained mRNAs tested are present in monosomes and light polysomes but were not stabilized by UPF1 knockdown or CHX treatment, which would indicate degradation by NMD. These transcripts may be substrates for cytoplasmic degradation by nonsense-mediated translational repression (NMTR), a poorly studied surveillance mechanism that seems to target NMD resistant isoforms and does not use standard NMD factors (43,57). Our observation that these isoforms are not affected by UPF1 depletion and are present in polysomes with similar profiles as NMTR targets (58), would be consistent with this mechanism or the production of truncated protein isoforms. Additionally, these low abundance transcripts could potentially have escaped surveillance for currently unknown reasons. Overall, we can conclude that at

least a fraction of intron-retained mRNAs, previously characterized by others and assumed to be nuclear (39), may indeed be exported to the cytoplasm at low levels. Therefore, Fr-iCLIP has provided insights into rare and specific RNA–protein interactions with different RNAs that occur in a dynamic fashion, from synthesis and processing to translation and decay.

ACCESSION NUMBERS

Data can be accessed at GEO under the accession number: GSE79792.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank members of the Neugebauer Lab for helpful discussions and comments on the manuscript. We are grateful to L. Maquat and M. Popp for advice regarding UPF1 knockdown and S.C. Sridhara for advice on chromatin fractionation.

FUNDING

MPI-CBG; FP7 Marie Curie Initial Training Network project RNPnet [289007]; Deutsche Forschungsgemeinschaft [NE909/3-1 to K.M.N.]; NIGMS [NIH R01GM112766]. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. Funding for open access charge: NIH. *Conflict of interest statement.* None declared.

REFERENCES

- Muller-McNicoll, M. and Neugebauer, K.M. (2013) How cells get the message: dynamic assembly and function of mRNA–protein complexes. *Nat. Rev. Genet.*, **14**, 275–287.
- Singh, G., Pratt, G., Yeo, G.W. and Moore, M.J. (2015) The clothes make the mRNA: past and present trends in mRNP fashion. *Annu. Rev. Biochem.*, **84**, 325–354.
- Baltz, A.G., Munschauer, M., Schwanhauser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M. *et al.* (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell*, **46**, 674–690.
- Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M. *et al.* (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, **149**, 1393–1406.
- Bentley, D.L. (2014) Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.*, **15**, 163–175.
- Brugiolo, M., Herzel, L. and Neugebauer, K.M. (2013) Counting on co-transcriptional splicing. *F1000prime Rep.*, **5**, 9.
- Wetterberg, I., Zhao, J., Masich, S., Wieslander, L. and Skoglund, U. (2001) In situ transcription and splicing in the Balbiani ring 3 gene. *EMBO J.*, **20**, 2564–2574.
- Vargas, D.Y., Shah, K., Batish, M., Levandoski, M., Sinha, S., Marras, S.A., Schedl, P. and Tyagi, S. (2011) Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell*, **147**, 1054–1065.
- Zhong, X.Y., Wang, P., Han, J., Rosenfeld, M.G. and Fu, X.D. (2009) SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Mol. Cell*, **35**, 1–10.
- Anko, M.L. (2014) Regulation of gene expression programmes by serine-arginine rich splicing factors. *Semin. Cell Dev. Biol.*, **32**, 11–21.
- Muller-McNicoll, M., Botti, V., de Jesus Domingues, A.M., Brandl, H., Schwich, O.D., Steiner, M.C., Curk, T., Poser, I., Zarnack, K. and Neugebauer, K.M. (2016) SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes Dev.*, **30**, 553–566.
- Anko, M.L., Muller-McNicoll, M., Brandl, H., Curk, T., Gorup, C., Henry, I., Ule, J. and Neugebauer, K.M. (2012) The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol.*, **13**, R17.
- Ji, X., Zhou, Y., Pandit, S., Huang, J., Li, H., Lin, C.Y., Xiao, R., Burge, C.B. and Fu, X.D. (2013) SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell*, **153**, 855–868.
- Pandit, S., Zhou, Y., Shiue, L., Coutinho-Mansfield, G., Li, H., Qiu, J., Huang, J., Yeo, G.W., Ares, M. Jr and Fu, X.D. (2013) Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Mol. Cell*, **50**, 223–235.
- Sanford, J.R., Wang, X., Mort, M., Vanduyne, N., Cooper, D.N., Mooney, S.D., Edenberg, H.J. and Liu, Y. (2009) Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res.*, **19**, 381–394.
- Sapra, A.K., Anko, M.L., Grishina, I., Lorenz, M., Pabis, M., Poser, I., Rollins, J., Weiland, E.M. and Neugebauer, K.M. (2009) SR protein family members display diverse activities in the formation of nascent and mature mRNPs in vivo. *Mol. Cell*, **34**, 179–190.
- Neugebauer, K.M. and Roth, M.B. (1997) Distribution of pre-mRNA splicing factors at sites of RNA polymerase II transcription. *Genes Dev.*, **11**, 1148–1159.
- Huang, Y., Gattoni, R., Stevenin, J. and Steitz, J.A. (2003) SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Mol. Cell*, **11**, 837–843.
- Huang, Y., Yario, T.A. and Steitz, J.A. (2004) A molecular link between SR protein dephosphorylation and mRNA export. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 9666–9670.
- Lai, M.C. and Tarn, W.Y. (2004) Hypophosphorylated ASF/SF2 binds TAP and is present in messenger ribonucleoproteins. *J. Biol. Chem.*, **279**, 31745–31749.
- Caceres, J.F., Sreaton, G.R. and Krainer, A.R. (1998) A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes Dev.*, **12**, 55–66.
- Sanford, J.R., Gray, N.K., Beckmann, K. and Caceres, J.F. (2004) A novel role for shuttling SR proteins in mRNA translation. *Genes Dev.*, **18**, 755–768.
- Botti, V., McNicoll, F., Steiner, M.C., Richter, F.M., Solovyeva, A., Wegener, M., Schwich, O.D., Poser, I., Zarnack, K., Wittig, I. *et al.* (2017) Cellular differentiation state modulates the mRNA export activity of SR proteins. *J. Cell Biol.*, **216**, 1993–2009.
- Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A. *et al.* (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell*, **39**, 925–938.
- Konig, J., Zarnack, K., Luscombe, N.M. and Ule, J. (2011) Protein-RNA interactions: new genomic technologies and perspectives. *Nat. Rev. Genet.*, **13**, 77–83.
- Anko, M.L. and Neugebauer, K.M. (2012) RNA–protein interactions in vivo: global gets specific. *Trends Biochem. Sci.*, **37**, 255–262.
- Sanford, J.R., Coutinho, P., Hackett, J.A., Wang, X., Ranahan, W. and Caceres, J.F. (2008) Identification of nuclear and cytoplasmic mRNA targets for the shuttling protein SF2/ASF. *PLoS One*, **3**, e3369.
- Kutluay, S.B., Zang, T., Blanco-Melo, D., Powell, C., Jannain, D., Errando, M. and Bieniasz, P.D. (2014) Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. *Cell*, **159**, 1096–1109.
- Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M. and Ule, J. (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.*, **17**, 909–915.
- Huppertz, I., Attig, J., D’Ambrogio, A., Easton, L.E., Sibley, C.R., Sugimoto, Y., Tajnik, M., Konig, J. and Ule, J. (2014) iCLIP: protein-RNA interactions at nucleotide resolution. *Methods*, **65**, 274–287.

31. Wuarin, J. and Schibler, U. (1994) Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. *Mol. Cell Biol.*, **14**, 7219–7225.
32. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
33. Ni, J.Z., Grate, L., Donohue, J.P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., Clark, T.A., Blume, J.E. and Ares, M. Jr (2007) Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.*, **21**, 708–718.
34. Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S. and Haussler, D. (2004) Ultraconserved elements in the human genome. *Science*, **304**, 1321–1325.
35. Anko, M.L., Morales, L., Henry, L., Beyer, A. and Neugebauer, K.M. (2010) Global analysis reveals SRp20- and SRp75-specific mRNPs in cycling and neural cells. *Nat. Struct. Mol. Biol.*, **17**, 962–970.
36. Sugimoto, Y., Konig, J., Hussain, S., Zupan, B., Curk, T., Frye, M. and Ule, J. (2012) Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol.*, **13**, R67.
37. Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakraborty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R. and Guigo, R. (2012) Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.*, **22**, 1616–1625.
38. Carrillo Oesterreich, F., Preibisch, S. and Neugebauer, K.M. Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol. Cell*, **40**, 571–581.
39. Boutz, P.L., Bhutkar, A. and Sharp, P.A. (2015) Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev.*, **29**, 63–80.
40. Kiss, T. (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, **109**, 145–148.
41. Quinn, J.J. and Chang, H.Y. (2016) Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.*, **17**, 47–62.
42. Popp, M.W. and Maquat, L.E. (2013) Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu. Rev. Genet.*, **47**, 139–165.
43. Hwang, J. and Kim, Y.K. (2013) When a ribosome encounters a premature termination codon. *BMB Rep.*, **46**, 9–16.
44. Lejeune, F., Li, X. and Maquat, L.E. (2003) Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylation, and exonucleolytic activities. *Mol. Cell*, **12**, 675–687.
45. Sun, S., Zhang, Z., Sinha, R., Karni, R. and Krainer, A.R. (2010) SF2/ASF autoregulation involves multiple layers of post-transcriptional and translational control. *Nat. Struct. Mol. Biol.*, **17**, 306–312.
46. Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C. and Brenner, S.E. (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*, **446**, 926–929.
47. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
48. Heyer, E.E. and Moore, M.J. (2016) Redefining the translational status of 80S monosomes. *Cell*, **164**, 757–769.
49. Hurt, J.A., Robertson, A.D. and Burge, C.B. (2013) Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay. *Genome Res.*, **23**, 1636–1650.
50. Wong, J.J., Au, A.Y., Ritchie, W. and Rasko, J.E. (2016) Intron retention in mRNA: No longer nonsense: Known and putative roles of intron retention in normal and disease biology. *BioEssays*, **38**, 41–49.
51. Wong, J.J., Ritchie, W., Ebner, O.A., Selbach, M., Wong, J.W., Huang, Y., Gao, D., Pinello, N., Gonzalez, M., Baidya, K. et al. (2013) Orchestrated intron retention regulates normal granulocyte differentiation. *Cell*, **154**, 583–595.
52. Boothby, T.C., Zipper, R.S., van der Weele, C.M. and Wolniak, S.M. (2013) Removal of retained introns regulates translation in the rapidly developing gametophyte of *Marsilea vestita*. *Dev. Cell*, **24**, 517–529.
53. Middleton, R., Gao, D., Thomas, A., Singh, B., Au, A., Wong, J.J., Bomane, A., Cosson, B., Eyraes, E., Rasko, J.E. et al. (2017) IRFinder: assessing the impact of intron retention on mammalian gene expression. *Genome Biol.*, **18**, 51.
54. Tripathi, V., Song, D.Y., Zong, X., Shevtsov, S.P., Hearn, S., Fu, X.D., Dundr, M. and Prasanth, K.V. (2012) SRSF1 regulates the assembly of pre-mRNA processing factors in nuclear speckles. *Mol. Biol. Cell*, **23**, 3694–3706.
55. Zhang, D., Jiang, P., Xu, Q. and Zhang, X. (2011) Arginine and glutamate-rich 1 (ARGLU1) interacts with mediator subunit 1 (MED1) and is required for estrogen receptor-mediated gene transcription and breast cancer cell growth. *J. Biol. Chem.*, **286**, 17746–17754.
56. Pirnie, S.P., Osman, A., Zhu, Y. and Carmichael, G.G. (2017) An Ultraconserved Element (UCE) controls homeostatic splicing of ARGLU1 mRNA. *Nucleic Acids Res.*, **45**, 3473–3486.
57. Lee, H.C., Oh, N., Cho, H., Choe, J. and Kim, Y.K. (2010) Nonsense-mediated translational repression involves exon junction complex downstream of premature translation termination codon. *FEBS Lett.*, **584**, 795–800.
58. You, K.T., Li, L.S., Kim, N.G., Kang, H.J., Koh, K.H., Chwae, Y.J., Kim, K.M., Kim, Y.K., Park, S.M., Jang, S.K. et al. (2007) Selective translational repression of truncated proteins from frameshift mutation-derived mRNAs in tumors. *PLoS Biol.*, **5**, e109.