Same same but different: A comprehensive functional analysis of SRSF3 and SRSF7 in the regulation of alternative polyadenylation within the 3’ untranslated region

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

vorgelegt im Fachbereich 15 Biowissenschaften

der Johann Wolfgang Goethe - Universität

in Frankfurt am Main

von

Oliver Daniel Schwich

aus Bad Arolsen

Frankfurt 2021

(D30)

vom Fachbereich 15 Biowissenschaften der

Johann Wolfgang Goethe - Universität als Dissertation angenommen.

Dekan: Prof. Dr. Sven Kimpel

Gutachter: Prof. Michaela Müller-McNicoll, PhD; Dr. Kathi Zarnack

Datum der Disputation: 29.09.2021

The more you suffer

The more it shows you really care

Brian Keith “Dexter” Holland, PhD, The Offspring

Table of Content

1 Zusammenfassung 18

2 Abstract 25

3 Introduction 28

3.1 Life cycle of messenger RNA 28

3.2 RNA binding proteins 31

3.3 Pre-mRNA splicing and alternative splicing 35

3.4 Cleavage and polyadenylation 39

3.5 SR proteins: A family of essential splicing factors 44

3.6 Alternative polyadenylation 50

3.7 SRSF3 and SRSF7: Two close, but different siblings and their roles in connecting pre-mRNA splicing to APA and mRNA export 64

3.8 Objectives of study 67

4 Material and methods 68

4.1 Equipment 68

4.2 Chemicals and reagents 71

4.3 Biological methods 80

4.4 Databases 103

4.5 Programs 105

4.6 Data availability 110

4.7 Bioinformatics 111

5 Results 117

5.1 Definition of the poly(A)-tome of P19 cells 117

5.2 SRSF3 and SRSF7 have opposite effects on proximal poly(A) site (pPAS) usage 121

5.3 SRSF3 and SRSF7 may compete for binding to proximal PAS 125

5.4 Regulation of the pPAS is splicing-independent and concentration-dependent 126

5.5 SRSF3 and SRSF7 interact differently with cleavage and polyadenylation factors 132

5.6 SRSF3 promotes dPAS usage by preserving high levels of CPSF6 144

5.7 SRSF3-sensitive pPAS usage might be inhibited by CFIm binding 150

5.8 SRSF7 and FIP1 levels decrease during neuronal differentiation resulting in global 3' UTR extension 156
6 Discussion 159
6.1 The first poly(A)-tome of pluripotent P19 mouse cells 159
6.2 SRSF3 and SRSF7 have opposite effects on pPAS usage 162
6.3 SRSF3 and SRSF7 preferentially regulate pPASs and may compete for binding 165
6.4 SRSF3 and SRSF7 regulate 3' UTR-APA independent of splicing, but in a concentration-dependent manner 166
6.5 SRSF7 interacts with CFIm and FIP1 factors independent of RNA 169
6.6 SRSF3 promotes dPAS usage by controlling CFIm expression levels 177
6.7 CFIm binding at SRSF3-sensitive pPASs may inhibit their usage 182
6.8 dPAS usage is promoted in neuronal differentiated P19 cells, while SRSF7 and FIP1 expression decreased 187
7 Conclusion and outlook 190
8 References CXCVI
9 Supplemental CCXXXIX
9.1 List of primers CCXXXIX
9.2 Supplemental figures CCXLIII
9.3 Supplemental tables CCLII
10 Danksagung CCLVI
11 Erklärung CCLIX
12 Versicherung CCLIX

List of figures

Figure 1: The life cycle of eukaryotic messenger RNA (mRNA). *28*

Figure 2: Schematic model of a gene body including two exons and one intron. *35*

Figure 3: Process of pre-mRNA splicing, regulation of spliceosome recruitment and schematic results of different types of alternative splicing. *37*

Figure 4: Schematic overview over the CPA complexes and subunits involved in Pol II-linked CPA. *39*

Figure 5: Crystal and domain structure of CFIm and its subunits. *42*

Figure 6: Domain structure of FIP1. ..... *43*

Figure 7: The domain structures of SR proteins. *45*

Figure 8: Regulation of splice site activation by SR proteins. *46*

Figure 9: The phosphorylation cycle of SR proteins. *47*

Figure 10: Scheme of alternative polyadenylation variants. *51*

Figure 11: Splicing and alternative polyadenylation are closely connected to define the 3' terminal exon by interactions between several subunits. *60*

Figure 12: Scheme of the core factors and complexes involved in m^6A methylation of mRNA. *62*

Figure 13: Phylogenetic tree of all 12 human SR proteins representing evolutionary distance and sub-clusters of distinct proteins. *64*

Figure 14: Scheme of the identification of dynamic alternative polyadenylation usage between control and knockdown samples. *112*

Figure 15: MACE-Seq after KD of Srsf3 or Srsf7. ..... *117*

Figure 16: Separation between samples of different conditions and reproducibility between biological replicates of MACE-Seq samples. *119*

Figure 17: Definition of the P19 poly(A)-tome using MACE-Seq datasets. *119*

Figure 18: Characterization of PASs in APA-positive transcripts. *121*

Figure 19: Western blot to validate the KD efficiency of Srsf3 and Srsf7 after transfection of specific esiRNAs. *121*

Figure 20: Separation between control and KD RNA-Seq samples and reproducibility between biological replicates. *122*

Figure 21: Analysis of changes in APA after KD of Srsf3 and Srsf7 from RNA-Seq data using DaPARS. *123*

Figure 22: Analysis of Tandem 3'UTR changes after KD of Srsf3 and Srsf7 by MISO. *123*

Figure 23: Enriched gene ontology terms of Srsf3 regulated APA targets. *124*

Figure 24: Validation of APA changes of selected targets after KD of Srsf3 and Srsf7 by 3'RACE-PCR. *124*

Figure 25: Binding pattern of SRSF3 and SRSF7 around sPAS, pPAS and dPAS. *125*

Figure 26: Binding of SRSF3 and SRSF7 to SRSF3-responsive PAS. *126*

Figure 27: Browsershots of Ddx21 (A), Anp32e (B) and Rab11a (C). *127*

Figure 28: Validation of the existence of two PASs within the 3' UTR of chosen target genes and the respective reporter genes using 3'RACE-PCR. *127*

Figure 29: KD of Srsf3 and Srsf7 affects pPAS usage of Luc and mCherry reporter constructs. *129*

Figure 30: Expression of increasing amounts of SRSF7 increases pPAS usage. 130
Figure 31: Increasing the binding potential of SRSF7 upstream of the pPAS increases its usage. 131
Figure 32: CPA factors that are present in SRSF3-containing RNPs. 132
Figure 33: Sequence of the RS/RS-like domains in CPSF6 and FIP1, respectively. 132
Figure 34: P19 cell lines expressing GFP-tagged isoforms of CPSF5 and CPSF6 (CFIm) as well as FIP1 from bacterial artificial chromosomes (BACs). 133
Figure 35: SRSF3 and SRSF7 interact differently with cleavage factors CPSF5 and FIP1, depending on RNA and phosphorylation states of the RS domain. 134
Figure 36: Depletion of CFIm subunits does not affect RNA-independent interaction of FIP1 and SRSF7. 136
Figure 37: Generation and validation of tetracycline repressor (TetR) protein-RS domain fusion constructs including phosphomimetic variants. 137
Figure 38: A hypophosphorylated RS domain of SRSF7 is sufficient to mediate interaction with FIP1. 137
Figure 39: Generation and validation of mCherry-tagged phosphomimetic isoforms of SRSF7 and SRSF3. ... 139
Figure 40: CPSF5 and FIP1 prefer different phosphorylation states of SRSF7. 139
Figure 41: SRSF3 and SRSF7 are closely related but show distinct differences on the polypeptide sequence level. 140
Figure 42: Generation of SRSF7 mutant cell lines lacking either the hydrophobic stretch (Δ27aa) or carrying an inactive Zn-knuckle domain (mutZn). 141
Figure 43: Mutation of both SRSF7-specific features decreases RNA-independent interaction with CPSF5, CPSF6 and FIP1. 142
Figure 44: Generation of SRSF3 chimeras containing the Zn-knuckle domain and the hydrophobic stretch derived from SRSF7. 143
Figure 45: Insertion of the hydrophobic stretch increases interaction between SRSF3 and the cleavage factors CPSF5, CPSF6 and FIP1. 144
Figure 46: KD of Srsf3 significantly reduces mRNA levels of Cpsf6. 145
Figure 47: KD of Srsf3 leads to unproductive splicing of Cpsf6. 146
Figure 48: KD of Srsf3 decreases expression of CPSF6 and CPSF5. 147
Figure 49: Validation of Cpsf6 KD by RNAi using specific esiRNAs. 147
Figure 50: Sample distance and reproducibility between biological replicates of RNA-Seq samples upon KD of Cpsf6. 148
Figure 51: Depletion of CPSF6 and SRSF3 *leads to global shortening of 3' UTRs.* 149
Figure 52: CFIm binding motif UGUA is enriched upstream of sPAS and dPAS but has a bimodal distribution around the pPAS. 150
Figure 53: Distribution of UGUA motifs and distances between two adjacent UGUA motifs around sPASs, pPASs and dPASs. 151
Figure 54: Generation of iCLIP libraries from GFP-FIP1 and CPSF5-GFP. 152
Figure 55: CPSF5 and FIP1 binding enriches upstream of PASs. 153
Figure 56: Binding of CPSF5 and FIP1 at pPASs and dPASs is higher in SRSF3-sensitive transcripts. 154
Figure 57: CPSF5 and FIP1 are recruited to pPASs of short and long 3' UTRs. 155

Figure 58: Differentiation of P19 wt cells into neuronal cells using retinoic acid. *156*

Figure 59: 3' UTRs globally elongated after differentiation. *157*

Figure 60: Expression of SRSF7 and FIP1 decreases in neuronal differentiated P19 cells. *158*

Figure 61: Model of the competitive mechanisms how SRSF3 and SRSF7 might regulate 3'UTR-APA at the pPAS. *172*

Figure 62: Model how the SRSF3-dependent expression of circCpsf6 and CFIm could affect 3'UTR-APA. *181*

Figure 63: Model how bipartite and dual UGUA-motifs might inhibit or activate CPA at proximal and distal PAS, respectively. *183*

List of tables

Table 1: Summary of the effects of depletion of core/auxiliary CPA factors on 3' UTR length as comprehensively analyzed by Li et al., 2015. 53
Table 2: List of used equipment stating the intended usage, catalogue number and vendor. 68
Table 3: List of used chemicals, solutions, biologicals, and plastic ware stating name, catalogue number and vendor. 71
Table 4: E. coli strains used in this thesis. 80
Table 5: Murine P19 cell lines used in this thesis. 81
Table 6: Approximate number of P19 cells passaged on fresh cell culture dishes, depending on the size of the cell culture dish. 81
Table 7: List of plasmids used in this thesis. 83
Table 8: Master mix to prepare esiRNA templates by in vitro transcription. 86
Table 9: PCR cycler setting for in vitro transcription of esiRNA templates. 86
Table 10: PCR cycler setting to perform reverse transcription using SuperScript™ III. 90
Table 11: Master mix for reverse transcription using SuperScript™ III. 90
Table 12: PCR cycler settings to perform 3'RACE-PCR using ALLin™ RPH Polymerase. 90
Table 13: Program to run qRT-PCR. 91
Table 14: Master mix to perform SAP treatments. 93
Table 15: Recipe for homemade SDS-Glycine gels. 94
Table 16: Master mix for T4 PNK driven 3' dephosphorylation. 96
Table 17: L3 Adapter ligation master mix. 96
Table 18: PNK master mix for radioactive labelling. 97
Table 19: PCR cycler program for reverse transcription using SuperScript™ IV. 98
Table 20: Master mix, per sample, for reverse transcription using SuperScript™ IV. 98
Table 21: Master mix for circularization of cDNAs using CircLigase II. 99
Table 22: Master mix for CutC4 oligo annealing. 99
Table 23: PCR cycler program to anneal CutC4 oligo and linearize cDNA with BamHI. 100
Table 24: Master mix to amplify iCLIP libraries for Illumina sequencing with AccuPrime™ SuperMix I. 100
Table 25: PCR cycler setting to amplify iCLIP libraries for Illumina sequencing with AccuPrime™ SuperMix I. 100
Table 26: List of used R packages stating the versions used and the source. 108
Table 27: Statistics of MACE-Seq reads after KD of Srsf3 and Srsf7. 118
Table 28: Statistics of RNA-Seq after KD of Srsf3 and Srsf7. 121
Table 29: Statistics of RNA-Seq reads after KD of Cpsf6. 148

List of supplementary figures

SuppFigure 1: Plasmid maps of basic vectors used for various cloning purposes. ..... *CCXLIV*

SuppFigure 2: Plasmid maps of psTet-Resistance-RS-Domain-phosphomimetic constructs. ..... *CCXLIV*

SuppFigure 3: Plasmid maps of pmCherry-SRSF3/SRSF7-phosphomimetic constructs. ..... *CCXLIV*

SuppFigure 4: Plasmid maps of pEGFP-SRSF3, pEGFP-SRSF7 and pEGFP-SRSF3-Chimera constructs. ..... *CCXLV*

SuppFigure 5: Plasmid maps of Luciferase-Reporter-3' UTR constructs. ..... *CCXLVI*

SuppFigure 6: Nucleotide and amino acid sequence alignment of phosphomimetics derived from the RS domain of SRSF3. ..... *CCXLVII*

SuppFigure 7: Nucleotide and amino acid sequence alignment of phosphomimetics derived from the RS domain of SRSF7. ..... *CCXLVIII*

SuppFigure 8: Nucleotide and amino acid sequence alignment of phosphomimetics derived from full length SRSF3. ..... *CCXLIX*

SuppFigure 9: Nucleotide and amino acid sequence alignment of phosphomimetics derived from full length SRSF7. ..... *CCLI*

List of supplementary tables

SuppTable 1: Primers used in this thesis. ..... *CCXXXIX*

SuppTable 2: Genes with significant changes in 3'UTR length upon KD of Srsf3 and Srsf7 quantified with DaPARS. ..... *CCLII*

SuppTable 3: Genes with significant changes in 3'UTR length upon KD of Srsf3 or Srsf7 quantified with MISO. ..... *CCLII*

SuppTable 4: Quantitative interactome of SRSF3-GFP. ..... *CCLII*

SuppTable 5: Nucleotide sequences of RS domains of SRSF3 and SRSF7 and the respective phosphomimetics RA and RD. ..... *CCLII*

SuppTable 6: Nucleotide sequences of SRSF3, SRSF3-RA, SRSF3-RS, SRSF7, SRSF7-RA, SRSF7-RD. ..... *CCLIII*

SuppTable 7: Nucleotide sequences of the SRSF3-Chimera constructs. ..... *CCLV*

SuppTable 8: Genes with significant change in gene expression upon KD of Srsf3 or Srsf7 quantified with DESeq2. ..... *CCLV*

SuppTable 9: Genes with significant changes in 3'UTR length upon KD of Cpsf6 quantified with DaPARS. . *CCLV*

SuppTable 10: Genes with significant changes in 3'UTR length upon neural differentiation quantified with DaPARS. ..... *CCLV*

SuppTable 11: Genes with significant change in gene expression upon neuronal differentiation quantified with DESeq2. ..... *CCLV*

List of abbreviations

(

(v/v) *Volume per volume*

(w/v) *Weight per volume*

°

°C *Degree celsius*

µ

µg M*icrogram*

µL *Microliter*

A

aa *Amino acid(s)*

ALS *Amyotrophic lateral sclerosis*

Ara-C *Cytosine β-D-arabinofuranoside hydrochloride*

ARE *AU-rich element*

ATP *Adenosine triphosphate*

B

BAC *Bacterial artificial chromosomes*

BMLS *Buchmann Institute for Molecular Life Sciences*

bp *Basepairs*

BSA *Bovine serum albumin*

C

CDS *Coding sequence*

CE *Cleavage event*

CFII C*leavage factor II*

CFIm *Mammalian cleavage factor I*

CLL Chronic lymphatic leukemia

ClustalO *Clustal Omega*

cm C*entimeter*

CO_2 *Carbon dioxide*

Co-IPs *Co-Immunoprecipitations*

CPA *Cleavage and polyadenylation*

CPSF *Cleavage and polyadenylation specificity*

Cq *Quantitation cycles*

CS C*leavage site*

CstF *Cleavage stimulation factor*

CT/CGRP Calcitonin/calcitonin gene-related peptide

CTD *C-terminal domain*

CTE C*onstitutive transport element*

CTP *Cytosine triphosphate*

D

DAPT G*amma-Secretase Inhibitor IX*

dATP *Deoxyadenosine triphosphate*

dCTP *Deoxycytoosine triphosphate*

dGTP Deoxyguanosine triphosphate

DI *Detained intron*

DMEM *Dulbecco's modified Eagle's medium*

DMSO *Dimethyl sulfoxide*

DNA *Deoxyribonucleic acid*

dNTP *Desoxyribose nucleoside triphosphate*

dPAS *Distal PAS*

ds *Double-stranded*

dsRBD double-stranded RNA binding domain

DTT *Dithiothreitol*

dTTP *Deoxythymidine triphosphate*

E

E. coli Escherichia coli

E/ISS *Exonic/intronic splicing silencer*

EBI *European Bioinformatics Institute*

EDTA *Ethylenediaminetetraacetic acid*

EGFP *Enhanced green fluorescent protein*

EJC E*xon junction complex*

ESC E*mbryonic stem cell*

ESE *Exonic splicing enhancer*

EST *Expressed sequence tag*

EtOH *Ethanol*

F

FCAM *Frankfurt Center for Advanced Light Microscopy*

FGF *Fibroblast Growth Factor*

FISH *fluorescence in situ hybridization*

FMRP Fragile X mental retardation protein

FXS *Fragile X Syndrome*

G

g *Gram, Gravity*

G Glycine

GEO *Gene Expression Omnibus*

GFP *Green fluorescent protein*

GO *Gene ontology*

GRE *GU-rich element*

GTP *Guanosine triphosphate*

H

h *Hour(s)*

HCl *Hydrochloric acid*

HIV1 *Human immunodeficiency virus 1*

hnRNP *Heterogeneous nuclear ribonucleoprotein*

HRP *Horseradish peroxidase*

I

iCLIP *Individual nucleotide UV-crosslinking and immunoprecipitation*

IGV *Integrative Genomics Viewer*

IRES *Internal ribosome entry side*

ISE *Intronic splicing enhancer*

K

K Lysine

kb *Kilobases*

kDa *Kilo Dalton*

KH *hnRNPK homology*

L

L L*iter*

LC *Low-complexity*

lncRNA *Long non-coding RNA*

M

M *Molar*

m6A *N6-Methyladenosine*

m^7G *7-methylguanylate cap*

MACE *Massive analysis of cDNA ends*

mCi *Milicurie*

mg *Miligram*

min *Minute(s)*

miRNA M*icro RNA*

mJ M*ilijoule*

mL *Mililiter*

mM *Milimolar*

MoRF Molecular recognition feature

MPI-CBG *Max Planck Institute of Molecular Cell Biology and Genetics*

mRNA M*essenger RNA*

mRNP *Messenger ribonucleoprotein*

N

N_2 *Nitrogen*

NaCl *Sodium chloride*

NaOAC *Sodium acetate*

NCBI *National Center for Biotechnology Information*

NEXT *Nuclear Exosome Targeting*

ng N*anogram*

NHGRI *National Human Genome Research Institute*

NLM *National Library of Medicine*

NLS Nuclear localization signal

nm *Nanometer*

NMD N*onsense mediated decay*

NRS *Negative regulator of splicing*

nt *Nucleotide(s)*

NTP N*ucleoside tri-phosphate*

NuRS *nuclear retention signal*

NXF1 Nuclear export factor 1

O

OPMD *Oculopharyngeal muscular dystrophy*

ORF O*pen reading frame*

P

pA *Polyadenosine*

PABPN1 N*uclear polyA binding protein 1*

PAINT *Points Accumulation for Imaging in Nanoscale Topography*

PAP P*olyA polymerase*

PAS *Polyadenylation site*

PAZ *Piwi/Argonaute/Zwille*

PBS *Phosphate Buffered Saline*

PCE *Poison cassette exon*

PCR *Polymerase chain reaction*

PDUI *Percentage of Distal poly(A) site Usage Index*

PEG *Polyethylene glycol*

PET paired-end tag

PIR *Protein Information Resource*

pM *Pikomolar*

PNK *Polynucleotide Kinase*

pPAS *Proximal PAS*

PTC *Premature termination codon*

PTM Post-transcriptional modification

PU *Percent usage*

PUF *Pumilio/FBF*

Q

qPCR *Quantitative PCR*

qRT-PCR *real-time quantitative PCR*

R

R Arginine

RNA R*ibonucleic acid*

rpm *Revolution per minute*

RRM RNA recognition motif

RRMH RNA recognition motif homologue

rRNA R*ibosomal RNA*

RS Arginine-serine, *Arginine-serine*

RSV *Rous Sarcoma Virus*

RT *Room temperature*

S

S Serine

SAP *Shrimp-alkaline phosphatase*

SDS *Sodium dodecylsulfate*

sec *Second(s)*

seq *Sequencing*

SIB *Swiss Institute of Bioinformatics*

siRNA *Small interfering RNA*

SLiM Short linear motif

snRNA *Small nuclear RNA*

SRPK SRSF protein kinase

ss *Single-stranded*

SUMO *Small ubiquitine-like modifications*

T

TBE *Tris-Borate-EDTA*

TEMED *Tetramethylethylene¬diamine*

TPM *transcripts-per-million*

TPP *Thiamine pyrophosphate*

tRNA T*ransfer RNA*

TSS T*ranscription start site*

U

U *Units*

UCSC *University of California, Santa Cruz*

UMI *Unique molecular identifier*

USER *Uracil-specific excision reagent*

UTP *Uracil triphosphate*

UTR *Untranslated region*

W

WTTS *Whole transcriptome termini site*

Z

ZnMut *Zn-knuckle mutant*

Contribution to figures

The work and figures presented in this thesis were generated by me under the supervision of Prof. Michaela Müller-McNicoll, PhD and support of Dr. Kathi Zarnack (Buchmann Institute for Molecular Life Sciences, Frankfurt, Germany) unless not stated otherwise. All external contributions are referenced accordingly.

The following data and figures were contributed by colleagues as stated below.

Affiliations: [1]Goethe Universität Frankfurt, Germany; [2]Buchmann Institute for Molecular Life Sciences, Frankfurt, Germany

Figure 17: Definition of the P19 poly(A)-tome using MACE-Seq datasets. A) Samarth Thonta Setty[2] prepared the scheme of the MACE-Seq analysis pipeline to analyze the MACE-Seq datasets which were prepared from my samples at GenXpro. B-E) Mario Keller[2] ran the analysis of the MACE-Seq data and prepared the figures. My contribution was the planning of the experiments, preparation of sequencing samples and initial mapping of sequenced reads to the genome providing the datasets for the downstream analysis.

Figure 18: Characterization of PASs in APA-positive transcripts. A, C, D) Mario Keller[2] prepared the figures. I contributed the mapped datasets derived from MACE-Seq.

Figure 21: Analysis of changes in APA after KD of Srsf3 and Srsf7 from RNA-Seq data using DaPARS. B&D) Mario Keller[2] prepared the figures based on mapped RNA-Seq datasets and DaPARS results files I contributed.

Figure 25: Binding pattern of SRSF3 and SRSF7 around sPAS, pPAS and dPAS. Mario Keller[2] prepared these figures based on iCLIP crosslink datasets provided by Michaela Müller-McNicoll[1] and PAS positions identified from MACE-Seq datasets I provided.

Figure 26: Binding of SRSF3 and SRSF7 to SRSF3-responsive PAS. A) Mario Keller[2] prepared the scheme. B) Mario Keller[2] prepared these figures based on iCLIP crosslink datasets provided by Michaela Müller-McNicoll[1] and PAS positions identified from MACE-Seq datasets and DaPARS results I provided.

Figure 27: Browsershots of Ddx21 (A), Anp32e (B) and Rab11a (C). The browsershots were prepared by Mario Keller[2] using iCLIP datasets provided by Michaela Müller-McNicoll[1] and MACE-Seq and RNA-Seq datasets I contributed.

Figure 32: CPA factors that are present in SRSF3-containing RNPs. The lab of Christian Münch[1] performed the quantitative MS-experiment and ran the analysis. I prepared the figure based on the results we received from the Münch lab.

Figure 34: P19 cell lines expressing GFP-tagged isoforms of CPSF5 and CPSF6 (CFIm) as well as FIP1 from bacterial artificial chromosomes (BACs). E) Nicole Blümel[1] prepared and arranged the confocal microscopy images. I generated and provided the GFP-tagged CPSF5, CPSF6 and FIP1 cell lines used in this experiment.

Figure 36: Depletion of CFIm subunits does not affect RNA-independent interaction of FIP1 and SRSF7. B) The Co-IPs were prepared and arranged by Nicole Blümel[1]. I provided the knockdown samples of CPSF6 and CPSF5 in the GFP-tagged FIP1 cell lines and supervised the Co-IP experiments.

Figure 37: Generation and validation of tetracycline repressor (TetR) protein-RS domain fusion constructs including phosphomimetic variants. B&C) Nicole Blümel[1] ran the SAP-treatment experiment and prepared the western blot images. I generated and provided the TetR-constructs and supervised the experiments.

Figure 38: A hypophosphorylated RS domain of SRSF7 is sufficient to mediate interaction with FIP1. The Co-IP experiments were prepared and run, and the results were arranged by Nicole Blümel[1]. I provided the GFP-tagged FIP1 and CPSF5 cell lines and the transfected TetR-constructs and supervised the experiments.

Figure 39: Generation and validation of mCherry-tagged phosphomimetic isoforms of SRSF7 and SRSF3. B&C) The Western blot and confocal microscopy images were prepared and arranged by Nicole Blümel[1]. I generated and provided the used GFP-FIP1 cell line and the transfected mCherry-phosphomimetic-plasmids and supervised the experiment.

Figure 40: CPSF5 and FIP1 prefer different phosphorylation states of SRSF7. Nicole Blümel[1] ran the Co-IP experiments and arranged the figures. I supervised the experiment and provided the cell lines and transfected mCherry-phosphomimetic-plasmids.

Figure 42: Generation of SRSF7 mutant cell lines lacking either the hydrophobic stretch (Δ27aa) or carrying an inactive Zn-knuckle domain (mutZn). Nicole Blümel[1] generated the SRSF7-mutant-BAC-constructs, ran the SAP-treatment experiment and prepared and arranged the Western Blot images as well as the confocal microscopy images. Michaela Müller-McNicoll[1] provided the SRSF7-GFP BAC and the SRSF7-GFP cell line. I supervised the planning and cloning and supervised the downstream experiments.

Figure 43: Mutation of both SRSF7-specific features decreases RNA-independent interaction with CPSF5, CPSF6 and FIP1. Nicole Blümel[1] ran the Co-IPs and arranged the figure. I supervised the experiments.

Figure 44: Generation of SRSF3 chimeras containing the Zn-knuckle domain and the hydrophobic stretch derived from SRSF7. Nicole Blümel[1] generated the SRSF3-plasmid-chimaera-constructs, ran the SAP-treatment experiment and prepared and arranged the Western Blot images as well as the confocal microscopy images. I provided the SRSF3-GFP plasmid and Michaela Müller-McNicoll[1] provided the SRSF3-GFP cell line. I supervised the planning and cloning and supervised the downstream experiments.

Figure 45: Insertion of the hydrophobic stretch increases interaction between SRSF3 and the cleavage factors CPSF5, CPSF6 and FIP1. Nicole Blümel[1] ran the Co-IPs and arranged the figure. I supervised the experiments.

Figure 51: Depletion of CPSF6 and SRSF3 leads to global shortening of 3' UTRs. D&E) Mario Keller[2] prepared the analysis and figures based on DaPARS results I provided.

Figure 52: CFIm binding motif UGUA is enriched upstream of sPAS and dPAS but has a bimodal distribution around the pPAS. Mario Keller[2] ran the motif-enrichment analysis and prepared the figures based on MACE-Seq datasets and DaPARS results I provided.

Figure 53: Distribution of UGUA motifs and distances between two adjacent UGUA motifs around sPASs, pPASs and dPASs. Mario Keller[2] ran the motif-enrichment analysis and prepared the figures based on MACE-Seq datasets and DaPARS results I provided.

Figure 54: Generation of iCLIP libraries from GFP-FIP1 and CPSF5-GFP. E-G) Mario Keller[2] ran the motif-enrichment analysis and prepared the figures based on iCLIP datasets of CPSF5 and FIP1 I provided.

Figure 55: CPSF5 and FIP1 binding enriches upstream of PASs. Mario Keller[2] ran the analysis and prepared the figures based on CPSF5 and FIP1 iCLIP datasets and MACE-Seq datasets I provided.

Figure 56: Binding of CPSF5 and FIP1 at pPASs and dPASs is higher in SRSF3-sensitive transcripts. Mario Keller[2] ran the analysis and prepared the figures based on CPSF5 and FIP1 iCLIP datasets, MACE-Seq datasets and DaPARS results I provided.

Figure 57: CPSF5 and FIP1 are recruited to pPASs of short and long 3' UTRs. Mario Keller[2] ran the analysis and prepared the figures based on CPSF5 and FIP1 iCLIP datasets, MACE-Seq datasets and DaPARS results I provided.

Figure 60: Expression of SRSF7 and FIP1 decreases in neuronal differentiated P19 cells. D&E) Mario Keller[2] ran the analysis and prepared the figures based on SRSF3 and SRSF7 iCLIP datasets provided by Michaela Müller-McNicoll[1] and MACE-Seq datasets and DaPARS results I provided.

1 Zusammenfassung

Das klassische zentrale Dogma der Biologie beschreibt die Synthese funktionaler Proteine basierend auf den Informationen, die in der DNA kodiert sind. In einem notwendigen Zwischenschritt wird zunächst die entsprechende DNA-Sequenz in ein messenger-RNA (mRNA) Molekül abgeschrieben (transkribiert), bevor diese RNA-Sequenz durch Ribosomen in das finale Protein übersetzt (translatiert) werden kann. In Eukaryoten sind die Transkription und Translation durch eine Kompartimentierung der Zelle in den Zellkern und das Zytosol örtlich und zeitlich voneinander getrennt. Diese Trennung ermöglicht eine eingehende Qualitätskontrolle der gereiften mRNA im Zellkern, bevor diese durch einen aktiven Prozess in das Zytoplasma exportiert wird. In Eukaryoten liegen die Informationen für die Proteine fragmentiert vor. Kodierende Sequenzen (Exons) werden unterbrochen von nicht-kodierenden Abschnitten (Introns), welche zunächst beide abgeschrieben werden und die prä-mRNA bilden. Diese initiale RNA-Sequenz muss im Anschluss prozessiert werden, um die Introns zu entfernen und die Exons miteinander zu legieren (Spleißen). Die entstehende neue prä-mRNA wird sofort an ihrem 5‘-Ende methyliert, um sie vor dem Verdau durch 5’-Exonukleasen zu schützen (5‘Capping). Abschließend wird die Transkription terminiert, und um das 3‘-Ende ebenfalls vor einem möglichen Abbau zu schützen, erhalten die Transkripte einen so genannten poly(A)-Schwanz, eine Sequenz aus Adenosinen, die nicht in der DNA-Matrize vorgegeben sind (Polyadenylierung).

Diese Prozesse werden durch verschiedene Multi-Protein-RNA-Komplexe im Zusammenspiel mit spezifischen RNA-bindenden Proteinen (RBPs) katalysiert. Das Spleißen wird vom Spliceosom durchgeführt, welches mittels zweier aufeinanderfolgender Umesterungen das Intron zwischen zwei Exons entfernt und die Exons miteinander ligiert. Hierbei können auch ein oder mehrere Exons übersprungen werden. Dieses alternative Spleißen (AS) ermöglicht die Expression alternativer Protein-Isoformen aus demselben Gen. Zusätzlich können durch AS aber auch alternative, toxische Exons in die reife mRNA integriert werden, welche die Stabilität des Transkripts negativ beeinflussen und somit eine Möglichkeit zur Regulation der Proteinexpression bieten. Die Assemblierung des Spliceosoms an der prä-mRNA wird durch die Präsenz von RNA-bindenden Spleiß-Aktivatoren oder -Inhibitoren beeinflusst. Eine bekannte Familie von Spleiß-Aktivatoren ist die der Serin/Arginin-reichen Proteine (SR-Proteine). Diese binden spezifische Sequenzen in Exons und fördern die Assemblierung des Spliceosoms an den jeweiligen Spleißstellen und somit die Inklusion der gebundenen Exons. Dem entgegen wirken Inhibitoren, wie die Proteine aus der hnRNP-Familie, die vorzugsweise in Introns binden.

Die Transkription einer neuen prä-mRNA wird durch eine hydrolytische Spaltung in der 3‘-untranslatierten Region (UTR) beendet und das neu entstandene 3‘-Ende dieser prä-mRNA wird durch die neue Synthese eines poly(A)-Schwanzes vor der vorzeitigen Degradation geschützt. Diese

zusammenhängenden Prozesse werden von vier Multi-Protein-Komplexen (CFIm, CFIIm, CPSF und CsTF) und der Poly(A)-Polymerase (PAP) katalysiert. Die Adenosin-reiche Sequenz wird durch die Bindung des Poly(A)-bindenden Proteins (PABPN1) stabilisiert wodurch die Aktivität von PAP weiter stimuliert wird. Wie Spleißen ist auch die endonukleolytische Spaltung und Polyadenylierung sequenzspezifisch und abhängig von RBPs, die diese Sequenzen erkennen. Das zentrale Erkennungsmotiv ist das Hexamer ‚AAUAAA' sowie bestimmte Varianten dieses Motivs. Dieses so genannte Poly(A)-Signal wird durch die spezifischen Untereinheiten WDR33 und CPSF30 des CPSF-Komplexes erkannt. Die Assemblierung der gesamten Polyadenylierungsmaschinerie wird unterstützt durch den CFIm-Komplex, der UGUA-Motive oberhalb des Poly(A)-Signals bindet sowie durch den CsTF-Komplex, der U/GU-reiche Sequenzen unterhalb des Poly(A)-Signals bindet. Analog zum Spleißen ist auch die Polyadenylierung in den meisten eukaryotischen Genen (bei humanen/murinen Zellen in bis zu 70% der Gene) an mehreren Positionen möglich (alternative Polyadenylierung, APA). Abhängig von der Position der alternativen Polyadenylierungsstellen entstehen dadurch entweder Transkripte mit alternativen terminalen Exons, falls diese Stelle in einem Intron liegt (CDS-APA), oder Transkripte mit unterschiedlich langen 3'UTRs aber identischer kodierender Sequenz, wenn die alternativen Poly(A)-Signale in der 3'UTR liegen (3'UTR-APA). In Abhängigkeit von der Entfernung zum vorherigen STOP-Codon wird die erste Polyadenylierungsstelle (PAS) als ‚proximal' (pPAS) und die am weitesten entfernte als ‚distal' (dPAS) betitelt. Die Länge der 3'UTR hat Auswirkungen auf die Stabilität, Exporteffizienz, subzelluläre Lokalisation, Translationsrate und lokale Translation der entsprechenden mRNA-Isoform. Einzelne Polyadenylierungsfaktoren wurden mit der Expression bestimmter APA-Isoformen in Verbindung gebracht. Die Reduktion von CFIm führte zur vermehrten Expression von Transkripten mit verkürzten 3'UTRs, wohingegen verringerte Expressionen von CsTF-Komponenten und von FIP1 (Untereinheit des CPSF-Komplexes) die Expression von Transkripten mit langen 3'UTRs förderte. Bisher sind die Komponenten und Funktionen der einzelnen Polyadenylierungsfaktoren umfassend erforscht, dennoch ist die Regulation der alternativen Polyadenylierung – die Entscheidung, ob die proximale oder distale PAS benutzt wird – weniger entschlüsselt und benötigt zusätzliche Studien.

Vereinzelt wurden bereits individuelle SR-Proteine mit der Regulation von APA in Verbindung gebracht. Besonders hervorzuheben ist hier ein durch SRSF3 reguliertes CDS-APA-Event im *calcitonin-related polypeptide-α gene* (*Ctrg*), welches in direktem Zusammenhang mit AS steht: Normale Expression von SRSF3 stabilisiert die Assemblierung des Spliceosoms durch die Interaktion zwischen SRSF3 und U1 snRNP. Dieses führt zunächst dazu, dass das betroffene Exon4 in die prä-mRNA inkludiert und anschließend die darin vorhandene PAS aktiviert wird. Reduzierte Expression von SRSF3 hingegen führt dazu, dass dieses Exon übersprungen wird. Dadurch wird ein alternatives terminales Exon in *Ctrg* definiert, was schlussendlich zur Expression einer alternativen Protein-Isoform führt. Darüber hinaus

wurde mittels GST-Pull-Down Experimenten gezeigt, dass SRSF3, SRSF7 und SRSF10 mit CPSF6 (Untereinheit vom CFIm-Komplex) interagieren können, was eine Verbindung zwischen dieser Familie der Spleiß-Faktoren und Polyadenylierung andeutet.

Unsere Arbeitsgruppe konnte in murinen P19-Zellen zum ersten Mal eine direkte Verknüpfung zwischen alternativem 3'End Prozessieren durch SRSF3 und SRSF7 und dem Export gereifter mRNA aus dem Nukleus nachweisen und den Mechanismus weiter entschlüsseln. Im Kern steht hierbei die direkte Rekrutierung des Exportfaktors NXF1 durch SRSF3 und SRSF7 an Transkripte mit unterschiedlich langen 3'UTRs. Obwohl die RNA-Seq Datensätze dieser Studie Hinweise auf eine APA-Regulation durch alle Mitglieder der SR-Protein-Familie gaben, stachen besonders SRSF3 und SRF7 hervor, da sie eine entgegengesetzte Auswirkung auf die Länge der untersuchten 3'UTRs hatten: SRSF3 förderte die Expression langer 3'UTRs und SRSF7 die Expression kurzer 3'UTRs.

Das vorrangige Ziel der hier präsentierten Thesis war, es die Funktion von SRSF3 und SRSF7 in der Regulation von 3'UTR-APA zu entschlüsseln und den grundlegenden Mechanismus zu bestimmen. Dazu wurden verschiedene genom-weite Methoden, wie beispielsweise RNA-Seq, MACE-Seq und iCLIP-Seq integriert und die Erkenntnisse durch Reportergen und Mutationsstudien untermauert.

Da es bisher keinen Datensatz mit allen aktiven Polyadenylierungsstellen in pluripotenten murinen P19-Zellen gab, und sich diese stark unterscheiden in Abhängigkeit von der Zellart und dem Differenzierungsstatus, haben wir zuerst das poly(A)-tome dieser Zelllinie mittels MACE-Seq und einer maßgeschneiderten Bioinformatik-Pipeline definiert. Der überwiegende Teil der rund 16.000 Polyadenylierungsstellen lag in den 3'UTRs protein-kodierender Transkripte, welche zumeist das kanonische Poly(A)-Motiv ‚AAUAAA' oder die Variante ‚AUUAAA' enthielten. Etwas mehr als 50% aller untersuchten Gene nutzten nur eine PAS und demnach gab es hier nur eine 3'UTR-Isoform. Alle anderen Gene nutzten zwei oder mehr PAS und bildeten verschiedene 3'UTR Transkript-Isoformen aus. Bei den dPASs wurde verstärkt das kanonische Poly(A)-Motiv gefunden, wohingegen an den pPAS häufiger alternative Sequenzmotive identifiziert wurden. Dies deutete darauf hin, dass APA vor allem über die pPASs-Nutzung reguliert wird und die dPAS eher als Standard-Schnitt- und Polyadenylierungsstelle fungieren, um eine korrekte Reifung der mRNA zu gewährleisten.

Mithilfe eines spezifischen Algorithmus (DaPARS) welcher APA-Events basierend auf RNA-Seq Datensätzen analysiert bestätigten wir, dass beim SRSF3-Knockdown deutlich mehr Transkripte beeinflusst als SRSF7 wobei SRSF3 die Entstehung von Transkripten mit langen 3'UTRs fördert und SRSF7 die mit kurzen 3'UTRs. Ein Vergleich der identifizierten Transkripte ergab, dass beide SR-Proteine nicht die gleichen Transkripte regulieren, es aber eine kleine Gruppe an antagonistisch regulierten Transkripten gibt (17). Die Ergebnisse der DaPARS-Analyse wurden mittels semi-quantitativer 3'RACE-

PCRs für fünf der antagonistisch regulierten APA-Targets validiert. Die Integration der MACE-seq-Daten mit publizierten iCLIP-Datensätzen von SRSF3 und SRSF7 in P19-Zellen zeigte eine verstärkte Bindung von SRSF3 und SRSF7 oberhalb der pPASs, was auf eine potenzielle Konkurrenz zwischen den beiden SR-Proteinen um die Bindung an der regulierbaren proximalen PAS hindeutet.

Da die Familie der SR-Proteine essenzielle Spleiß-Faktoren darstellt, erschien es notwendig zu überprüfen, ob und inwieweit die Regulierung von APA durch SRSF3 und SRSF7 an die physische Spleiß-Reaktion gekoppelt ist. Zu diesem Zweck klonierten wir die vollständigen 3'UTRs der zuvor verifizierten fünf Transkripte hinter zwei verschiedene, intron-freie Reportergene (*mCherry*, *luciferase*). Transfektionen der Reporter-Konstrukte und Knockdown von *Srsf3* bzw. *Srsf7* führte erneut zur Expression von verkürzten bzw. verlängerten Transkripten, wie bereits zuvor, unabhängig vom fehlenden Intron in den Reportern, gezeigt. Dieses bedeutet, dass die SRSF3- und SRSF7-bedingte Regulation von 3'UTR-APA mechanistisch unabhängig vom Spleiß-Prozess ist. Weitergehende Experimente mit den *Ddx21*-basierten Konstrukten deuteten zudem an, dass die Regulation von APA durch SRSF7 konzentrationsabhängig ist. Je stärker SRSF7 von einem co-transfizierten Plasmid überexprimiert wurde, desto mehr wurde die proximale Polyadenylierungsstelle aktiviert. Die Mutation SRSF3-spezifischer Bindestellen zu SRSF7-spezifischen Motiven und *vice versa* im Bereich bis zu 150 nt oberhalb der proximalen Polyadenylierungsstelle hatte einen limitierten Effekt auf die alternative Polyadenylierung der *Ddx21*-Konstrukte, was die vorherige Vermutung einer möglichen Konkurrenz zwischen den beiden SR-Proteinen unterstützt. Analog zu den aktuellen strukturellen Kenntnissen zur spezifischen Erkennung des Poly(A)-Motivs durch WDR33 and CPSF30 führten stärkende/schwächende Punktmutationen des endogenen proximalen Poly(A)-Motivs von ‚AGTAAA' zu ‚A**A**UAAA' bzw. ‚AGTAA**G**' zu drastischen Änderungen der Polyadenylierung. Dieses bestätigt die Relevanz dieser Sequenz als Kern-Element der Transkriptionstermination und nachfolgender Reifung und Prüfung der entstandenen mRNA.

Da die 3'UTR-APA-Regulation durch SRSF3 und SRSF7 unabhängig von einer vorhergehenden Spleiß-Reaktion ist, aber beide SR-Proteine bevorzugt oberhalb der pPAS binden, wollten wir als nächstes mögliche Interaktionen mit Polyadenylierungsfaktoren identifizieren. Ein quantitativer Massenspektrometrie-Datensatz von aufgereinigtem SRSF3-GFP identifizierte Interaktionen mit CPSF33, WDR33, CPSF2, FIP1, CPSF5 und PABPN1. Da die SRSF3-GFP-Aufreinigung ohne RNA-Verdau durchgeführt wurde, konnten diese Interaktionen sowohl direkt als auch indirekt (beide Proteine binden dasselbe Transkript) sein. Um dies zu überprüfen, erstellten wir P19-Zelllinien, die stabil GFP-markiertes CPSF5, CPSF6 (beide Untereinheiten des CFIm-Komplexes) und FIP1 exprimierten. Diese Polyadenylierungsfaktoren wurden ausgewählt, weil sie genau wie SR-Proteine, eine Serin/Arginin-reiche Domäne (RS-Domäne) enthalten, mit der sie direkte Protein-Protein-Wechselwirkungen

ausbilden können. Co-Immunopräzipitationen, basierend auf GFP-Pull-Downs zeigten überraschend, dass lediglich SRSF7 unabhängig von RNA mit dem CFIm und FIP1 interagieren kann. Darüber hinaus konnten wir feststellen, dass die Interaktion mit CFIm im hyperphosphorylierten Zustand von SRSF7 passiert, wohingegen die Interaktion mit FIP1 den hypophosphorylierten Zustand der RS-Domäne bevorzugt.

Ein Vergleich von SRSF3 und SRSF7 auf Ebene der Proteinsequenz ergab, dass die beiden Proteine in ihrer N-terminalen RNA-bindenden-Domäne nahezu identisch sind (~80%), allerdings enthält SRSF7 zwei individuelle Regionen, die nicht in SRSF3 zu finden sind. Zum einen eine CCHC-Zink-Fingerdomäne vor der C-terminalen RS-Domäne und zum anderen eine hydrophobe, 27 Aminosäuren lange Region inmitten der RS-Domäne. Zusätzliche Co-IPs mit Mutationen/Deletionen dieser Regionen in SRSF7 und das Einbringen dieser Regionen in SRSF3 deuten darauf hin, dass beide Regionen gemeinsam für die Ausbildung der SRSF7-spezifischen Interaktion und APA-Regulation notwendig sind. Weitergehende Experimente sind hier jedoch noch notwendig, um die individuellen Funktionen dieser Regionen genauer zu bestimmen.

Die Reduktion von SRSF3 führt zur vermehrten Expression von Transkripten mit kürzeren 3'UTRs. Allerdings konnte eine direkte Interaktion von SRSF3 mit Polyadenylierungsfaktoren ausgeschlossen werden. Interessanterweise fanden wir aber eine signifikante Reduktion der Expression von *Cpsf6* nach *Srsf3*-Knockdown auf RNA- und Proteinebene. Die verringerte Expression von *Cpsf6* konnte auf das Überspringen von Exon 6 im *Cpsf6*-Transkript zurückgeführt werden. Das Fehlen dieses Exons führt zu einer mRNA-Isoform mit einem vorzeitigen Stop-Codon, welches als Signal für den Abbau durch den Nonsense-mediated mRNA Decay (NMD) Abbauweg fungieren kann. RNA-Seq gefolgt von einer DaPARS-Analyse bestätigte, dass ein Knockdown von *Cpsf6* zur vermehrten Expression von Transkripten mit kurzer 3'UTR führt. Ein Vergleich der identifizierten Zielgene nach den Knockdowns von *Srsf3* und *Cpsf6* ergab eine deutliche Überlappung, was darauf hindeutet, dass SRSF3 3'UTR-APA auch indirekt über die Mengen von CPSF6 reguliert. Zusätzlich konnten wir zeigen, dass die Reduktion von CPSF6 auch die Expression von CPSF5 reduziert, und demnach den CFIm-Komplex destabilisiert.

Um den Mechanismus der CFIm-abhängigen 3'UTR-APA-Regulation SRSF3-sensitiver Transkripte an der pPAS genauer zu untersuchen analysierten wir zunächst die Anreicherung des CFIm-spezifischen Binde-Motivs ‚UGUA'. Wie zu erwarten, fanden wir die meisten ‚UGUA'-Motive oberhalb der in P19-Zellen identifizierten PASs. Interessanterweise fanden wir aber insbesondere im Bereich der pPASs bimodale ‚UGUA'-Motive ober- sowie unterhalb der PAS. Vor dem Hintergrund, dass jedes CFIm-Halo-Molekül aus sowohl zwei identischen CPSF5- und CPSF6/7-Untereinheiten besteht und demnach zwei ‚UGUA'-Motive binden kann, bietet dieses die Möglichkeit, dass die Bindung von CFIm diese PASs maskieren, und somit die Polyadenylierung inhibieren könnte, indem das Poly(A)-Signal und/oder

spezifische Downstream Sequence Elements (DSEs) um das CFIm-Molekül gebogen werden. Um zu überprüfen, ob diese Motive tatsächlich vom CFIm-Komplex erkannt werden und dieses eventuell zu einer nicht-funktionalen Rekrutierung von FIP1 führen kann, wurden die genutzten Bindungsstellen von CPSF5 und FIP1 mittels iCLIP-Seq identifiziert. Neben den starken Anreicherungen oberhalb der PASs fanden wir auch eine verstärkte Anreicherung von CPSF5 unterhalb der pPAS, welche in SRSF3-sensitiven Transkripten noch stärker ausgeprägt war als in der Gesamtheit der analysierten Transkripte. Dieses unterstreicht die Möglichkeit, dass die pPAS SRSF3-sensitiver Transkripte zusätzlich durch inaktive Bindung vom CFIm geblockt und nach dem Knockdown von *Srsf3* durch die SRSF3-bedingte Reduktion von CPSF5 und CPSF6 aktiviert werden kann.

Abschließend interessierte uns die mögliche Verbindung der SR-Protein-bedingten Regulation von APA während der neuronalen Differenzierung, da insbesondere in Neuronen die Expression sehr langer 3'UTRs beobachtet wird. Diese ist mit einem APA-bedingten Wechsel von Transkripten mit kurzen zu langen 3'UTRs verknüpft. Interessanterweise können die pluripotenten P19-Zellen in einem einfachen Verfahren mittels der Zugabe von Retinolsäure ins Medium in neuronale Zellen differenziert werden. Eine DaPARS-Analyse von RNA-Seq-Daten nach der Differenzierung von P19-Zellen bestätigte die genomweite Verlängerung der 3'UTRs nach erfolgreicher Differenzierung. Eine vergleichende Analyse der Proteinkonzentrationen ergab reduzierte SRSF7- und FIP1-Proteinlevel in den differenzierten Zellen, wobei die mRNA-Expression beider zugrundeliegender Transkripte nicht signifikant verändert war. Dieses deutet darauf hin, dass SRSF7 und FIP1 konzentrationsabhängig die Nutzung der pPASs in proliferierenden, pluripotenten Zellen fördern. Diese initialen Ergebnisse sind vielversprechend für nachfolgende Experimente, da die Expression APA-abhängiger Transkriptisoformen in Neuronen verstärkt mit lokaler Translation in Verbindung gebracht wird. Hier könnten explizit SRSF3 und SRSF7 durch die direkte Verbindung zur Regulation des nukleären Exports spezifischer Transkripte mittels NXF1 eine bestimmende Rolle einnehmen.

Basierend auf den durchgeführten Experimenten und Analysen der verschiedenen Datensätze präsentieren wir ein Modell, wie die Nutzung der proximalen Polyadenylierungsstelle (in)direkt von SRSF3 und direkt von SRSF7 sowie vom CFIm-Komplex in Abhängigkeit ihrer Expressionslevel reguliert werden kann. SRSF3 kann direkt mit SRSF7 um die Bindung im Bereich der pPASs konkurrieren und die Rekrutierung der Polyadenylierungsmaschinerie blockieren, sowie indirekt die Expression des dPAS-Aktivators CFIm regulieren. In Gegenwart von SRSF3 wird die Expression von Transkripten mit einer langen 3'UTR begünstigt. SRSF7 hingegen kann durch die direkte Interaktion mit FIP1 den Polyadenylierungsmakrokomplex rekrutieren und stabilisieren und somit die Nutzung des pPAS aktivieren, welches die Expression von Transkripten mit kurzen 3'UTRs zur Folge hat. Vor dem Hintergrund, dass SRSF3 und SRSF7 durch die direkte Rekrutierung des nukleären Exportfaktors NXF1

den Export von gebundenen mRNAs in das Zytosol steuern können, eröffnet die Bindung und APA-Regulation an der pPAS somit die Möglichkeit, Transkripte spezifisch nach der Länge der jeweiligen 3’UTR-Isoform für nachfolgende zytosolische Prozesse (bspw. zur lokalen Translation) zu markieren und zu sortieren.

2 Abstract

The central dogma of biology is based on the concatenated transfer of information from DNA, via transcribed mRNA, to the translated protein. In eukaryotes, transcription and translation are separated locally as well as temporally by cellular compartmentalization. Prior to active export factor-dependent transport from the nucleus to the cytosol, the newly formed pre-mRNA must mature. This involves 5'capping, splicing, and endonucleolytic cleavage and polyadenylation (CPA).

Transcription of a new pre-mRNA is terminated by hydrolytic cleavage in the 3'-UTR, and the newly formed 3'-end is protected from premature degradation by synthesis of a poly(A) tail. These processes are catalyzed by four multi-protein complexes (CFIm, CFIIm, CPSF, and CsTF) and poly(A) polymerase (PAP). CPA is sequence-specific and dependent on RNA-binding proteins (RBPs). APA-specific sequences include the poly(A) motif ('AAUAAA' and certain motif variants), the UGUA motif, and U/GU-rich sequences upstream and downstream of the poly(A) signal, respectively. About 70% of mammalian genes have more than one polyadenylation site (PAS) and express transcripts of different lengths by a mechanism called alternative polyadenylation (APA). This can affect the length of the 3'UTR (3'UTR-APA) or the coding sequence of the transcript (CDS-APA) if the alternative PAS is upstream of the STOP codon. The length of the 3'UTR affects the stability, export efficiency, subcellular localization, translation rate, and local translation of the nascent transcript. 3'UTR-APA is regulated in the interplay of the *cis*-elements (poly(A) motif, UGUA and U/GU) and *trans*-elements (expression of CPA factors). In this context, the functions of the individual cis and trans elements have been extensively studied, yet the regulation of alternative polyadenylation-the decision whether to use the proximal or distal PAS-is less deciphered and requires additional study.

In murine P19 cells, we were able to demonstrate for the first time a direct link between 3'UTR-APA and nuclear export of mature mRNA by the splicing factors SRSF3 and SRSF7 and decipher the mechanism. At the core here is the direct recruitment of the export factor NXF1 by SRSF3 and SRSF7 to transcripts with 3'UTRs of different lengths.

The primary goal of the thesis presented here was to decipher the function of SRSF3 and SRSF7 in the regulation of 3'UTR-APA and to determine the basic mechanism. For this purpose, various genome-wide methods, such as RNA-Seq, MACE-Seq, and iCLIP-Seq, were integrated and the findings were supported by reporter gene and mutation studies.

Initial determination of the poly(A)-tome in P19 cells by MACE-Seq yielded approximately 16,000 PAS and showed that slightly less than 50% of all genes used two or more PAS and expressed alternative 3'UTR isoforms. Further DaPARS analyses after knockdown of *Srsf3* or *Srsf7* confirmed that SRSF3 affected more transcripts than SRSF7 and led primarily to the expression of long 3'UTRs, whereas

SRSF7 promoted the expression of short 3'UTRs. Integration of SRSF3- and SRSF7-specific iCLIP data suggested a possible competition between SRSF3 and SRSF7 at the proximal PAS (pPAS), which could thus act as a hotspot of 3'UTR regulation.

Experiments with intron-free reporter genes revealed that SRSF3- and SRSF7-dependent regulation of 3'UTR-APA is independent of splicing. With respect to SRSF7, a concentration dependence was demonstrated. Mutation experiments involving the SRSF3- and SRSF7-specific binding motifs in the 3'UTR also confirmed the hypothesis of competition between the two SR proteins.

Extensive Co-IP experiments clearly demonstrated that only SRSF7, but not SRSF3, can interact with CFIm and FIP1 (a subunit from the CPSF complex) in an RNA-independent manner. In addition, we showed that these interactions exhibited some phosphorylation dependence, such that the interaction to FIP1 arose primarily in the semi- to hypophosphorylated state of SRSF7. Whereas the interaction to CFIm was mainly detected in the hyperphosphorylated state. The differential affinity between SRSF3 and SRSF7 for polyadenylation factors could be attributed to two SRSF7-specific domains in subsequent mutation experiments: A CCHC-type Zn finger between the RRM and the RS domain, and a hydrophobic 27 amino acid long region in the middle of the RS domain. Together, this suggested that SRSF3 could block the utilization of pPAS, whereas SRSF7 could activate it by directly recruiting polyadenylation factors.

Interestingly, we showed that knockdown of *Srsf3* also negatively regulates the expression of *Cpsf6* (a subunit of CFIm) through alternative splicing, which subsequently leads to decreased expression of CPSF6 and of CFIm. Reduction of CFIm led to increased expression of transcripts with short 3'UTR, analogous to knockdown of *Srsf3*. This mirrors the results of previous studies. A direct comparison between SRSF3- and CPSF6-specific transcripts revealed that not all targets were congruent. In addition, we found preliminary evidence for CFIm-related masking of essential *cis*-pPAS elements by bimodal UGUA motifs at the pPAS. In summary, we present a novel mechanism of indirect 3'UTR-APA regulation through SRSF3-conditional expression of the CFIm subunit CPSF6.

In conclusion, we began to investigate the regulation of the physiologically observed elongation of 3'UTRs during neuronal differentiation of murine P19 cells and found that, in particular, the expression of the pPAS-activating factors SRSF7 and FIP1 was reduced in the differentiated cells. However, initial bioinformatic analyses showed that mRNA expression of both proteins was unchanged after differentiation. The exact mechanism and effect on, for example, local translation needs further investigation.

In summary, we present a model of how the utilization of pPAS can be regulated (in)directly by SRSF3 and directly by SRSF7 as well as by the CFIm complex depending on their expression levels. SRSF3

can directly compete with SRSF7 for binding to pPASs and block recruitment of the polyadenylation machinery, and indirectly regulate expression of the dPAS activator CFIm, favoring expression of transcripts with a long 3'UTR. SRSF7, on the other hand, can activate the utilization of pPAS through direct interaction with FIP1, which leads to the expression of transcripts with short 3'UTRs. Given that SRSF3 and SRSF7 can control the export of bound mRNAs to the cytosol through direct recruitment of the nuclear export factor NXF1, their binding together with APA regulation at the pPAS opens the possibility to specifically label and sort transcripts according to the length of the 3'UTR for subsequent cytosolic processes (e.g., for local translation).

3 Introduction

3.1 Life cycle of messenger RNA

The life cycle of eukaryotic messenger RNA (mRNA, **Figure 1**) begins with its transcription from genomic deoxyribonucleic acid (DNA) in the nucleus. The primary transcript (pre-mRNA) needs extensive processing to mature, until the mRNA can be exported into the cytoplasm for subsequent translation into protein according to the Central Dogma of Biology - DNA ↔ RNA → protein, as proclaimed by Beadle and Tatum in 1941 (Beadle & Tatum, 1941). Finally, the mRNA molecule is degraded.

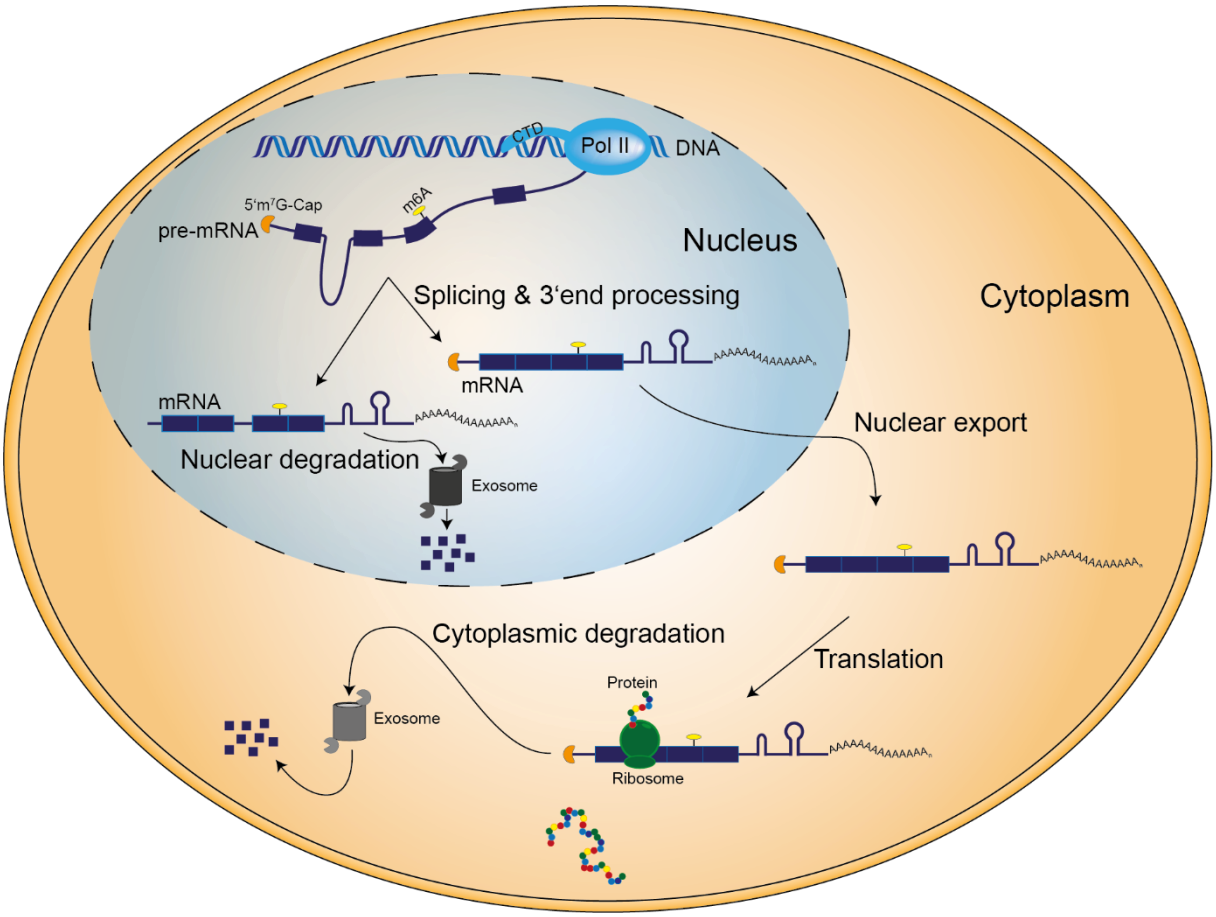


Figure 1: The life cycle of eukaryotic messenger RNA (mRNA). *Pre-mRNAs are transcribed inside the nucleus from DNA by the RNA Polymerase II (Pol II, light blue) and co-transcriptionally processed. The processing and maturation involve protection of the 5' end by m^7G capping (orange), modification of individual nucleotides (yellow), e.g., adenosine-methylation, removal of non-coding (thin line), intronic, sequences and ligation of coding (thick boxes), exonic, sequences by splicing as well as cleavage and polyadenylation. Incorrect, e.g., uncapped, or mis-spliced mRNAs are degraded by the nuclear exosome (dark grey), while mature mRNA is actively exported into the cytoplasm. In the cytoplasm, mRNA is recognized and bound by ribosomes (green) translating the encoded information into proteins. Once the lifetime of an mRNA is over, the mRNA is degraded by the cytoplasmic exosome (light grey).*

In more detail, the genetic information stored within the DNA needs to be copied into portable mRNA units to leave the nucleus, separating the locus of transcription and translation, which contrasts with prokaryotic organisms. For this, the DNA-directed RNA polymerase II (Pol II) transcribes genes from DNA into pre-mRNA. In addition, Pol II transcribes small nuclear RNAs (snRNAs) (Egloff et al., 2008), small interfering RNAs (siRNAs) and micro RNAs (miRNAs) (Lee et al., 2004). Other kinds of transcripts like larger ribosomal RNAs (rRNAs), transfer RNAs (tRNAs) or small rRNAs are transcribed by RNA polymerases I and III (Pol I and Pol III), respectively (WILLIS, 1993; Grummt, 1998). Transcription is divided into three steps: initiation, elongation, and termination. During initiation, Pol II and auxiliary factors bind to promoter regions in the DNA and unwind the double-stranded (ds) DNA, followed by the recognition of a transcription start site (TSS) and binding of the first nucleoside tri-phosphate (NTP). Subsequently, Pol II travels along the single-stranded (ss) DNA template in 3' → 5' direction base pairing additional NTPs complementary to the respective template nucleotide to elongate the nascent RNA transcript. Finally, transcription is terminated by cleavage and polyadenylation (CPA) of the nascent transcript, releasing Pol II from the DNA template. The C-terminal domain (CTD, **Figure 1**, light blue) of Pol II is unique among the three eukaryotic RNA polymerases and is well conserved (Corden, 1990). It is a highly unstructured region representing an interaction hub for many proteins that are involved in RNA maturation events, beginning with transcription initiation until cleavage of the newly transcribed mRNAs. The activity of Pol II and its auxiliary factors is regulated by phosphorylation of specific amino acids within the CTD. It consists of 26 (yeast) to 52 (mammals) repeats of a heptameric residue sequence (YSPTSPS) of which the serine, threonine and tyrosine residues have the potential to be phosphorylated (Corden, 1990; Millhouse & Manley, 2005). Especially two serine residues (Ser2 and Ser5) were identified as key phospho-group acceptors being involved in elongation or initiation, 5' capping and 3' end processing, respectively (Zhang & Corden, 1991). To protect the nascent mRNA transcript from degradation by exonucleases, the 5' end is modified by the addition of a guanine nucleotide and subsequent methylation at position 7 of the respective guanine (m^7G, **Figure 1**, orange), thereby capping the mRNA (Rottman et al., 1974). The elongating pre-mRNA is further processed, while transcription continues. For example, single nucleotides can be chemically modified by addition of small side groups (e.g., methylation of adenosines, **Figure 1**, yellow) (Roundtree et al., 2017a; Zhao et al., 2017) or deamination of adenosines creating inosines (A-to-I-editing) (Nishikura, 2010), expanding the epitranscriptome of each cell. The following major step during RNA maturation is the removal of non-coding introns and the ligation of coding exons by pre-mRNA splicing, which is described in more detail in chapter 3.3. In the last step of mRNA processing, the nascent pre-mRNA is cleaved and the 3' end is completed with a 150-200 nt long non-templated polyadenosine (pA) tail (Chang et al., 2014). Subsequently, correct mature mRNAs are compacted and actively exported into the cytoplasm by the main export factor NXF1 (Strässer et al., 2002). All these maturation processes

are constantly controlled as part of the nuclear surveillance to ensure that only correctly matured mRNA transcripts are exported. Defective mRNAs, e.g., uncapped or insufficiently spliced transcripts, are sensed and labeled for degradation by the nuclear exosome. In the cytoplasm, first the small and subsequently the big subunits of the ribosomes (**Figure 1**, green) assemble on the mRNA molecule to synthesize proteins translating the encoded information from the mRNA. The ribosome consists of ribosomal proteins and rRNAs, which are transcribed by Pol I and Pol III. Like transcription, translation is divided into the three steps: initiation, elongation, and termination. During initiation, the ribosome assembles on the target mRNA and attaches the first methionine-loaded tRNA at the start codon. Subsequently, in the elongation phase, the next residue-loaded tRNA matching the downstream codon joins the ribosome complex. A peptide bond between both residues is formed breaking the bond between the first residue and its tRNA. This process continues as the ribosome translocases along the mRNA until it reaches a termination codon, which does not encode for any amino acid. This leads to the termination of translation, release of the polypeptide, as well as release and recycling of the ribosome subunits. During repeated rounds of translation, the polyA tail is steadily shortened until it is too short to protect the transcript from exonuclease degradation (Meyer et al., 2004; Parker & Song, 2004). The mRNA is then degraded by the cytoplasmic exosome in a 3' → 5' direction (Figure 1, light grey).

During the synthesis of nascent pre-mRNA, the RNA molecules are immediately bound by many different TNA binding proteins (RBPs) forming large messenger ribonucleoproteins (mRNPs). RBPs protect mRNAs from degradation and mediate all maturation steps within the life cycle of each mRNA molecule. The composition of mRNPs is constantly remodeled to facilitate and connect the different processes. Despite decades of research the details of these networks, the composition of mRNPs and the different functions of RBPs are not well understood, leaving much space to investigate and identify new functions and interactions of RBPs, as well as new regulatory pathways that influence the fate of the mRNA.

3.2 RNA binding proteins

Directly from its biogenesis, any kind of RNA is accompanied and clothed by specific proteins capable to bind RNAs directly forming RNPs (Singh et al., 2015). During evolution and the concomitant increase in the number of introns, the number of RNA-binding proteins increased. Recent analyses have revealed the existence of nearly 2,000 RBPs in human cells, representing a large fraction of all translated proteins (Beckmann et al., 2016). RBPs have functions in all aspects during the life cycle of RNAs ranging from processing, modification, localization, nuclear export, translation, RNA turnover and stability. Interestingly, the composition of RNPs are constantly remodeled to facilitate the various steps during RNA maturation changing the fate of the bound RNA (Müller-McNicoll & Neugebauer, 2013).

Until recently, RBPs were mainly defined by the presence of globular RNA-binding domains (RBDs), which are associated with either binding to specific motifs or structural elements within the targeted transcripts. Most RBDs bind single-stranded RNA like RNA recognition motifs (RRMs), hnRNPK homology (KH) domains, DEAD-box helicase domains, Zn-finger domains, cold-shock domains, Pumilio/FBF (PUF) domains or Piwi/Argonaute/Zwille (PAZ) domains (Wang et al., 2002; Brown, 2005; Lunde et al., 2007; Cléry et al., 2008; Valverde et al., 2008; Linder & Jankowsky, 2011). Double-stranded RNAs are bound by the double-stranded RNA binding domain (dsRBD) (Banerjee & Barraud, 2014). These RBPs were termed conventional RBPs and can contain multiple copies of certain RBDs or varying combinations increasing the specificity and affinity of each RBP individually. Interestingly, the diversity in functions and targets is represented by the fact that the three most abundant RBDs are found in only 20% of all RBPs compared to the less diverse family of DNA binding proteins (DBPs), where the three most used DNA binding domains (DBDs) are present in 80% of all DBPs (Vaquerizas et al., 2009; Gerstberger et al., 2014). Moreover, not only the RBDs were shown to be important for mediating the interaction between protein and RNA, but the region in between two adjacent RBDs also plays a crucial role in defining the affinity between both molecules (Finger et al., 2004; Lunde et al., 2007).

Interestingly, the globular RBDs were found in only half of the described RBPs. Recent findings attribute intrinsic disordered regions (IDRs) with RNA binding potential. These domains were found in 55% of all RBPs, termed unconventional RBPs (Beckmann et al., 2016). IDRs are characterized by a high local enrichment of polar and/or charged residues, which often appear in repeats (Romero et al., 2001; Williams et al., 2001; Jorda et al., 2010). By nature, IDRs show extended conformations, lacking structural components. Therefore, it is assumed that they bind target sequences with reduced affinity, which might be compensated by a higher specificity, due to the increased binding surface (Liu & Huang, 2014). Yet, being unconventional, unstructured domains IDRs binding RNA can be recognized by a couple disordered RNA-binding motifs of residue-patterns. These motifs contain RG[G] repeats, RS/RG-

rich domains, K/R patches, short linear motifs (SLiMs), molecular recognition features (MoRFs) and low-complexity (LC) sequences. RG[G] repeats are enriched in arginine (R) and glycine (G) residues (Kiledjian & Dreyfuss, 1992). To resemble the motif, at least three of these repeats need to be present, segregated by ten residues in between. RS/RG-rich domains are enriched in arginine and serine (S) or glycine. Especially RS repeats are present within a special protein family termed SR-proteins which will be introduced in chapter 3.5. Similar to the previous motifs, K/R patches are sequences of four to eight lysine (K) or arginine residues often found in a context of conventional RBDs (Castello et al., 2012). While the basic motifs range from two to eight residues SLiMs, MoRFs and LC sequences represent bigger motifs with SLiMs being composed of up to 10 residues, MoRFs ranging from 25-50 amino acids and LC sequences comprising up to 100 residues (Balcerak et al., 2019). As conventional RBDs, IDRs are usually organized modularly containing multiple disordered motifs and combinations of those or they appear in combination with conventional RBDs.

RBPs can influence the fate of their targeted RNAs either directly by containing additional catalytic active domains themselves, or indirectly by bridging additional proteins via protein-protein-interactions. An example for the first kind is the dsRNA binding ADAR2, which contains a dsRBD in combination with a deaminase domain that is responsible for A-to-I editing of RNA molecules (Valente & Nishikura, 2005). ssRNA-binding RPBs with catalytic active domains are often RNA endo- or exonucleases degrading the targeted transcripts. Yet, most events during processing, nuclear export and translation of transcripts are mediated via interactions between various RBPs with platform proteins that specifically bind targets and recruit effector proteins mediating distinct action. Examples for this are splicing factors, such as the family of SR-proteins, which recruit and stabilize subunits of the spliceosome, facilitating intron removal and exon ligation. Even if the family of RBPs comprises nearly 2,000 proteins, only a small subset of around 70 RBPs are ubiquitously active functioning as ‘housekeeping’ RBPs. The time and spatial-dependent activity of RBPs is regulated by post-translational modification, distinct co-factors, protein-protein-interactions, and the availability of target RNAs. RBPs were shown to be modified by phosphorylation (Tacke et al., 1997; Arif et al., 2011; Zhou & Fu, 2013), methylation of arginine residues (Shen et al., 1998; Yu et al., 2004; Bressan et al., 2009; Sinha et al., 2010), acetylation (Edmond et al., 2011; Yin et al., 2018) and small ubiquitin-like modifications (SUMO) (Vassileva & Matunis, 2004; Kota et al., 2018).

RBPs are present within the whole cell, enriching within the nucleus. While many RBPs are limited to one of the two major cellular compartments, some RBPs, such as members of the SR-protein family, have the ability to shuttle between the nucleus and the cytoplasm to facilitate nuclear export downstream processing of the bound transcripts (Müller-McNicoll et al., 2016; Botti et al., 2017). Moreover, some RBPs were shown to be organized or stored in membrane-less granules or cell bodies,

such as nuclear Cajal-bodies, the nucleolus and paraspeckles or cytoplasmic processing (P)-bodies and stress-granules (Anderson & Kedersha, 2009; Buchan, 2014). These organelles are formed by liquid-liquid-phase separation due to condensation of cellular components. These granules usually consist of distinct sets of RNAs and proteins, locally organizing proteins and RNAs in a dynamic way, enabling rearrangements and diffusion of components (Boeynaems et al., 2018). Moreover, these organelles can increase cellular processes by locally enriching substrates and effectors, as for example in the cytoplasmic P-bodies that are enriched in enzymes connected to mRNA degradation (Cougot et al., 2004).

As RBPs are intertwined with every step during RNA metabolism (**Figure 1**), alterations in their expression or mutation-driven loss of RNA affinity, aggregation or sequestration from the native target have critical impact on the cell's fate and have related to diseases. RBP-mediated defects have been shown to enrich especially in neurological disorders, atrophies and cancers (Lukong et al., 2008; Nussbacher et al., 2019). A well-studied example of a neurological disorder is the Fragile X syndrome (FXS), which is caused by CGG-triplet expansions within the 5' UTR of the *fragile X mental retardation protein* (*FMRP*) leading to a hypermethylation-induced gene silencing (Verkerk et al., 1991; Yu et al., 1991). In addition, a mutation of Isoleucine304 to asparagine within the second KH-domain of FMRP was shown to cause severe cases of FXS (Boulle et al., 1993). FMRP functions in the transport and repression of synaptic translation of bound mRNAs and therefore is important for neuronal integrity and synaptic function (Li et al., 2001b; Mazroui et al., 2002; Davidovic et al., 2007; Dictenberg et al., 2008). A slightly different example is the gain-of-function of the RRM containing Poly(A)-binding Protein (PABPN1) by an extension of GCG repeats within the first exon. The native gene contains six of these repeats coding for alanine. However, the extension to up to 13 repeats enhances the homopolymeric stretch leading to an aggregation of the protein within skeletal muscle fibers causing Oculopharyngeal muscular dystrophy (OPMD) (Brais et al., 1998; Abu-Baker & Rouleau, 2007). Interestingly, with the expansion of RBPs by including proteins containing RNA-binding IDRs, recent analysis of monogenetic diseases revealed that most mutations causing these diseases are located within RNA-binding regions, especially within IDRs rather than the globular RBDs (Castello et al., 2013; Castello et al., 2016). In difference to these monogenetic diseases, cancer is connected to multiple genetic changes switching oncogenes on and tumor suppressors off. Especially RBPs involved in RNA processing, such as splicing and cleavage and 3' end formation, exert high potential to induce cancer when their expression is altered as those are closely connected to transcript stability (Desterro et al., 2019). A comprehensive analysis including various cancer types identified that alternative splicing events increased by up to 30% in cancer cells compared to normal tissue (Kahles et al., 2018). In addition, that study showed that cancer cells contain splicing isoforms that are absent in healthy tissue. Another recent comprehensive analysis showed that most RBPs were downregulated in human

cancers cells, when compared to normal cells underlining the close connection of RBPs and tumorigenesis (Wang et al., 2018). Accordingly, aberrant cleavage and polyadenylation has been shown to be relevant in around 2,000 events within 17 cancer types (Xiang et al., 2018). Usually, the cancer-related events expressed shorter 3' UTRs compared to healthy cells, which is connected to an increased stability of the respective transcript (Sandberg et al., 2008; Mayr & Bartel, 2009; Park et al., 2018). Interestingly, aberrant expression of cleavage and polyadenylation factors leads to the development of neuronal cancers, such as glio- or neuroblastomas, connecting the two major RBP-connected disease patterns (Masamha et al., 2014; Ogorodnikov et al., 2018; Chu et al., 2019).

This chapter briefly introduced the recently expanded family of RBPs, including information about the defining differences between conventional and unconventional RBPs. The broad range of RBPs is essential for cells as they are in involved in every metabolism. Hence, divergent expression of RBPs can cause severe diseases, often connected with neurological phenotypes. In the following two chapters (3.3 and 3.4), two important RNA processing mechanisms, splicing and cleavage and polyadenylation, involving distinct subsets of RBPs are introduced in detail. Moreover, one family of essential splicing factors, the SR proteins, are introduced in chapter 3.5.

3.3 Pre-mRNA splicing and alternative splicing

Eukaryotic precursor mRNAs are composed of coding and non-coding sequences termed exons and introns, where the introns are much longer than the exons, exceeding the average exon length (320 nt) by 23 times (Lee & Rio, 2015). To mature, the non-coding introns need to be excised from the pre-mRNA and the exons need to be ligated precisely to maintain the correct open reading frame (ORF). For this, splicing-regulating factors are deployed from the CTD of Pol II during transcription, which recognize specific *cis*-sequence elements, such as 5' and 3' splice sites (SS), a polypyrimidine (pY) tract and the branch point (BP) that distinguish introns from exons (**Figure 2**) (Yuryev et al., 1996; Kim et al., 1997). Both splice sites define the borders between exons and introns and are recognized by specific *trans*-factors of the spliceosome. The nucleotide composition of the SSs is variable, but the GU and AG dinucleotide motifs are enriched, respectively. The BP and pY tract are located inside the introns upstream the 3' SS. The BP consists of a single adenosine, which drives the first nucleophilic attack on the 5' SS to form the lariat structure. The pY tract is enriched in pyrimidine-residues, especially uracil. These two *cis*-elements are recognized by different RBPs, such as U2AF, SF1 and the polypyrimidine tract-binding protein (PTB) and promote the assembly of the spliceosome (Krämer, 1996; Berglund et al., 1998). In contrast to the other RBPs, PTB is a negative splicing regulator repressing the inclusion of exons with weak splice sites (Wagner & Garcia-Blanco, 2001).

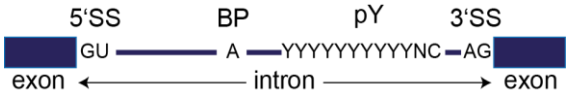


Figure 2: Schematic model of a gene body including two exons and one intron. *Exons and introns are distinguished by specific cis-sequence elements, such as the 5' and 3' splice sites (SS) at the exon-intron borders, the branch point (BP) and the polypyrimidine (pY) tract.*

The splicing reaction at each intron is catalyzed by the spliceosome, a mega-Dalton RNA-protein complex that consists of five snRNPs (U1, U2, U4, U5 and U6) and approximately 200 auxiliary proteins, in a multi-step process (**Figure 3**A) (Scheres & Nagai, 2017). First, as soon as the intron emerges from Pol II, the 5' SS is recognized and bound by the U1 (green) by base pairing. This is supported by auxiliary factors like SF1 and U2AF that bind to the BP and pY tract, respectively. Subsequently, the U2 snRNP, facilitated by the DEAD-box proteins Prp5 and Sub2/UAP56, associates with the BP forming the pre-spliceosome. Thereby, a short U2 snRNP-BP duplex is made presenting the BP's adenosine for the first catalytic step of splicing. In a second step during spliceosome assembly, the pre-catalytic spliceosome is formed by the recruitment of the preassembled tri-snRNP, consisting of U4/U6 and U5 (yellow). Due to these interactions, the intron is looped out and the adjacent exons are moved into proximity. Moreover, rearrangements within the RNA and between the RNA and the bound RNPs, driven by UAP56 (Shen et al., 2008), lead to the destabilization of the U1 and U4 snRNPs. The U1 snRNP is released from the 5' SS and leaves the spliceosome together with the U4 snRNP. During splicing catalysis U1 is replaced by U6 at the 5' SS, facilitated by the DEAD-box protein Prp28 and the

DExD/H-box protein Brr2 (Staley & Guthrie, 1999; Maeder et al., 2008). U6, supported by Prp2 (Warkocki et al., 2009), then interacts tightly with the BP-bound U2 snRNP to activate the oxygen atom of the BP adenosine, which now performs a nucleophilic attack of the phosphate group at the 5' SS resulting in the first transesterification reaction. The 5' SS is cleaved and the 3' end of the intron is ligated to the branch point, resulting in a structure called the lariat. Next, after some additional rearrangements within the spliceosome and driven by additional auxiliary factors, such as Prp16 and Prp22, the free hydroxyl-group at the 3' end of the upstream exon attacks the 3' end of the intron, resulting in the second transesterification reaction, cleavage of the intron and ligation of both exons. Subsequently, the remaining snRNPs U2, U5 and U6 are released from the lariat intron and are recycled, while the naked intron is rapidly degraded (Moore & Sharp, 1993; Will & Lührmann, 2011).

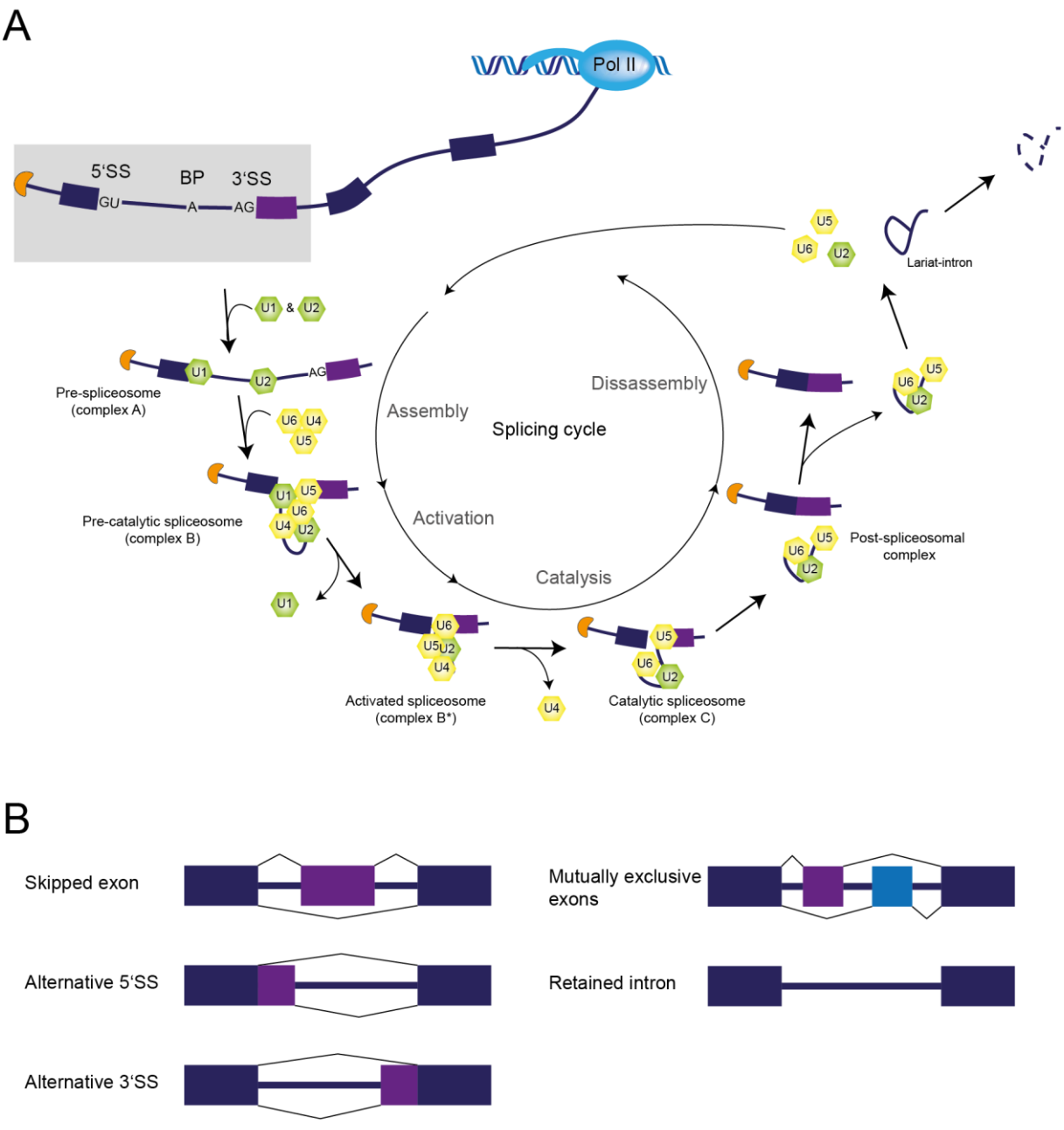


Figure 3: Process of pre-mRNA splicing, regulation of spliceosome recruitment and schematic results of different types of alternative splicing. A) *Nascent pre-mRNA, transcribed by the RNA Pol II, needs to be spliced to mature. This process is catalyzed by a multimeric RNP, the spliceosome, whose composition is remodeled constantly. Introns harbor specific cis-elements being recognized by the spliceosome: the 5' splice site (5' SS), branch point (BP) and the 3' splice site (3' SS). First, the snRNPs U1 and U2 (green) bind to the 5' SS and 3' SS forming the pre-spliceosome. In the following step, the assembly of the spliceosome is completed by the entry of the U4/U5/U6 tri-snRNP (yellow). Subsequently, the spliceosome is activated, while U1 is released from the transcript and the excision of the intron and the ligation of the two adjacent exons is catalyzed, while the U4 snRNP is released as well. Finally, the lariat-intron, bound by U2, U5 and U6 is released. The snRNPs dissociate and are recycled, while the intron is degraded. Modified from (Will & Lührmann, 2011).* ***B)*** *The tight competition between splicing enhancers and splicing inhibitors, as well as the presence of alternative splice sites can result in a variety of splicing products increasing the coding potential of any given pre-mRNA. Constitutive exons in dark purple, alternatively spliced exons in light purple or blue. Modified from (Park et al., 2013).*

The average human cell can express up to approximately 75,000 different proteins. These proteins are encoded by only approximately 21,000 genes (taken from UniProt, human proteome, latest update 11th of June 2019). This discrepancy between the size of the genome and the proteome can be ascribed to an increase of the transcriptome through alternative splicing (AS) (Yeo et al., 2005; Ben-Dov et al., 2008). The various possibilities of AS are summarized schematically in Figure 3B.

Splice sites can vary in their strength, dependent on the surrounding nucleotide compositions that influence their recognition and the usage of alternative 5' or 3' SS. In addition, the inclusion or skipping of individual exons is regulated through the differential binding and the crosstalk between splicing enhancers, e.g., SR proteins, and splicing inhibitors, e.g., heterogeneous nuclear ribonucleoproteins (hnRNPs). Furthermore, the decision which splice sites are activated is linked to the transcription rate, RNA secondary structures and post-transcriptional nucleotide modifications (Shenasa & Hertel, 2019). Slower transcription rates enable spliceosome assembly on weaker splice sites, compared to fast elongation rates, when strong splice sites are favored (La Mata et al., 2003; Ip et al., 2011). Alternative splicing can be modulated by secondary structures of the transcript by forming double stranded RNA sections masking potential SSs, negative splicing regulators, trans-acting binding sites of snRNAs or RBPs or decreasing the distance between splice sites (Singh et al., 2007; Jin et al., 2011; Yadegari et al., 2016). A special kind of secondary structures are so called riboswitches, which are RNA sequences that can bind metabolites leading to changes in the secondary structure (Sashital & Butcher, 2006). One example of an AS-connected riboswitch is the thiamine pyrophosphate (TPP) sensitive riboswitch (Cheah et al., 2007). In conditions of low TPP levels, the riboswitch adapts a conformation masking the close by 5' SS inhibiting splicing at this SS. Increasing amount of TPP result in the binding of the TPP within the aptamer and a conformational change rendering the 5' SS accessible increasing splicing at this site. Moreover, secondary structures can modulate AS in combination with RBPs. HnRNPA1 and PTB were shown to bind distinct RNA targets form dimers looping out the RNA sequences in between resulting in the activation of upstream 5' SS (Nasim et al., 2002; Lamichhane et al., 2010). In addition, various post-transcriptional nucleotide modifications were shown to influence splicing via distinct mechanisms. For example, m6A modifications were shown to

increase exon inclusion via recruitment of the splicing enhancer SRSF2 (Zhao et al., 2014). Moreover, these methylations were shown to disrupt secondary structures making the sequence accessible for ssRNA-binding proteins like hnRNPC and hnRNPG (Liu et al., 2015; Liu et al., 2017). Besides methylation, adenosines were shown to be modified into inosines by deamination mediated by the adenosine deaminase acting on RNAs (ADAR) protein, what has been termed RNA editing (Rueter et al., 1999). As Inosine is structurally related to guanine the deamination of the second adenosine in an AA-sequence leads to the reassembly of an AG-like 3' SS (Solomon et al., 2013).

AS alters the coding potential of the various transcripts resulting in different protein isoforms being translated. For example, AS enables the expression of different protein isoforms in different circumstances, such as different tissues, cell types, cell cycle or stress conditions. Besides increasing the transcriptome, AS can also result in intron retention or inclusion of exons containing premature termination codons (PTCs) that would result in truncated proteins (Hwang & Maquat, 2011). Intron retention is an important regulatory mechanism of fine-tuned gene expression, which can lead to the expression of truncated proteins or proteins containing additional amino acids. Transcripts containing retained introns are either exported to the cytoplasm where they are translated or are more likely degraded by the nonsense mediated decay (NMD) pathway. In addition, recently transcripts with retained introns were shown to be enriched in the nucleus to prevent pre-mature translation escaping NMD-mediated degradation and were termed detained introns (DIs) (Boutz et al., 2015). DIs are either stored until final processing or degraded by the nuclear exosome (Fan et al., 2017). Exons containing PTCs are termed poison cassette exons (PCEs) and are important to regulate protein homeostasis. Translation of transcripts containing PTCs can result in the expression of toxic proteins or the decrease of native, correct protein leading to a negative environment for the cell. To protect the cell from these products, transcripts containing PTCs are usually identified during translation and degraded by NMD (Kurosaki et al., 2019).

3.4 Cleavage and polyadenylation

Transcription needs to be terminated properly to release Pol II and to recycle it for a new round of transcription and to prevent read-through into the next gene. Termination of Pol II-transcribing mRNAs and long non-coding RNAs (lncRNAs) is closely linked to cleavage and polyadenylation (CPA) of the nascent pre-mRNA at polyadenylation sites (PAS). CPA is a two-step process and is catalyzed by four multimeric complexes consisting of more than 20 core proteins (Proudfoot, 1989; Birse et al., 1998; Yonaha & Proudfoot, 2000; Shi et al., 2009). Transcription of Pol I is terminated by a strong terminator sequence downstream of the rRNA precursors (Kuhn & Grummt, 1989; Lang & Reeder, 1995), while the transcription of Pol III is terminated by a T-rich stretch, located in close proximity of the transcript's 3' end (Cozzarelli et al., 1983). In comparison to the Pol I and III derived transcripts, CPA of Pol II is more complicated, and is orchestrated by a conserved, tight network of well-defined *cis*- and *trans*-acting factors, described below (**Figure 4**).

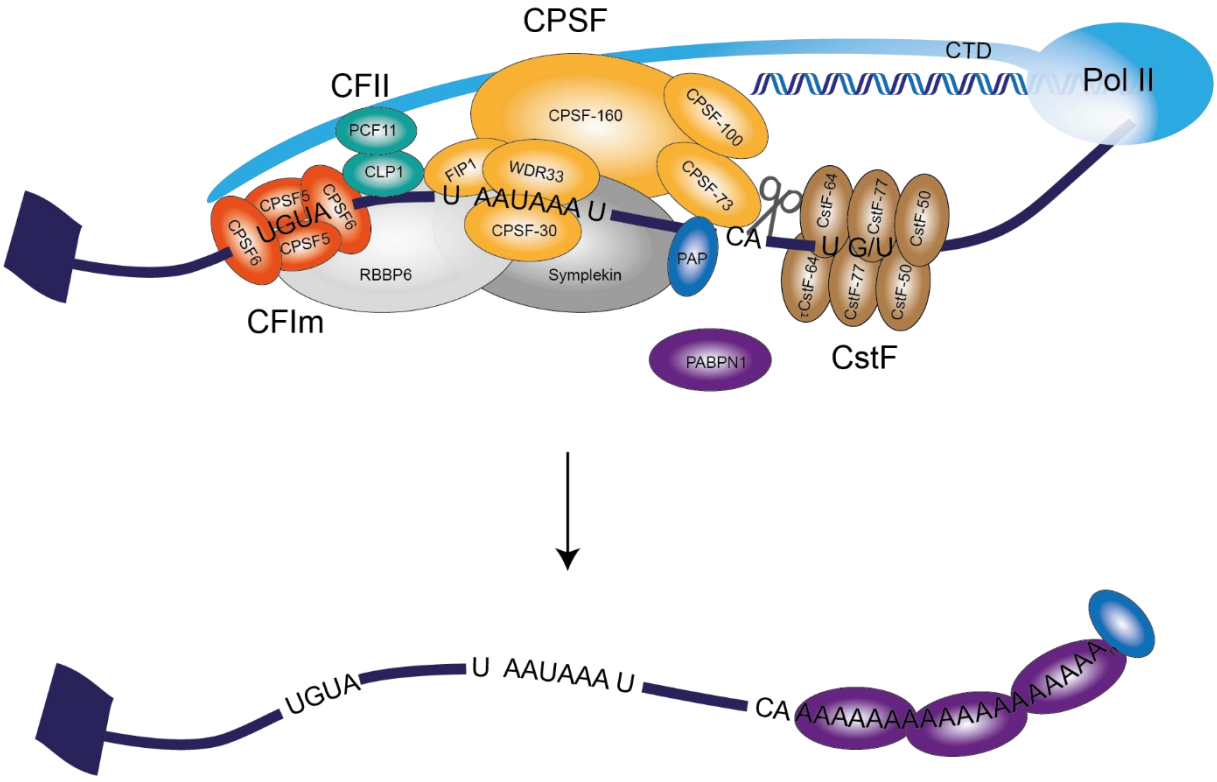


Figure 4: Schematic overview over the CPA complexes and subunits involved in Pol II-linked CPA. *The nascent transcript is transcribed by Pol II (light blue). More than 20 highly conserved proteins form complexes orchestrating the CPA process in connection with conserved cis-elements. The CPSF5 subunit of the CFIm (orange) recognizes the UGUA motif, WDR33 and CPSF-30 of the CPSF complex (yellow) bind to the poly(A) signal motif AAUAAA, while the DSE U/GU-rich sequence is bound by the CstF complex (brown). The nascent transcript is cleaved by CPSF-73 at the cleavage site (CA, grey scissor). PAP (blue) synthesises a non-templated polyA-tail at the 3' end of the cleaved transcript, which is bound by multiple PABPN1 proteins (purple). Symplekin and RBBP6 (grey) are scaffolding proteins and the CFIIm (turquoise) bridges the CFIm and the CPSF complex.*

CPA of mRNAs occurs co-transcriptional and is closely linked to other processing steps via the CTD of Pol II (Hirose & Manley, 1998; Fong & Bentley, 2001). The process is regulated by three *cis*-acting sequence elements located up- and downstream of the cleavage site (CS): the UGUA motif and the polyadenylation sequence represent the upstream sequence elements (USEs), and a U/GU-rich sequence downstream of the CA di-nucleotide of the cleavage site, represent the downstream sequence element (DSE, (Hu et al., 2005; Takagaki & Manley, 1997). The center of the *cis*-elements is the canonical poly(A) signal motif AAUAAA located 10-30 nt upstream of the CS (grey scissor). It is recognized by the WDR33 and CPSF-30 subunits of the cleavage and polyadenylation specificity factor (CPSF, yellow) complex (Schönemann et al., 2014; Clerici et al., 2017; Clerici et al., 2018). The CPSF also contains FIP1 and CPSF-73, which is the endonuclease responsible for the cleavage of the mature transcript (Ryan et al., 2004; Mandel et al., 2006). Nearly 20 sequence variants of the poly(A) signal motif were found across the genome, but they are much less frequent and are being processed less efficiently (Gruber et al., 2016; Beaudoing et al., 2000). Recently, the group of Martin Jinek was able to explain the importance of the invariant residues U3 and A6 in the poly(A) signal, which needs to form a Hogsteen base pair to be efficiently recognized by the CPSF complex (Clerici et al., 2018). The UGUA motif, which is upstream of the poly(A) signal is recognized by the heterotetrameric mammalian cleavage factor I (CFIm, orange, described more detailed in chapter 3.4.1) (Venkataraman et al., 2005), which was implicated in poly(A) signal selection and recruitment of CPA complexes (Rüegsegger et al., 1998; Brown & Gilmartin, 2003). The DSE, a less conserved sequence enriched in G/GU nucleotides located approximately 30 nt downstream of the CS (Hu et al., 2005) is recognized by the heterotrimeric cleavage stimulation factor (CstF, brown) (Takagaki & Manley, 1997). The cleavage factor II (CFII, turquoise) consists of two subunits, CLPI and PCF11. CLPI was shown to interact with CFIm and CPSF (Vries et al., 2000), while PCF11 has been recently connected to early transcription termination (Kamieniarz-Gdula et al., 2019; Wang et al., 2019). In addition, the single proteins Symplekin (grey), PolyA Polymerase (PAP, blue) and Nuclear PolyA Binding Protein 1 (PABPN1, purple) are associated with the CPA complexes. Symplekin functions as a scaffold protein (Takagaki & Manley, 2000), while PAP synthesizes the non-templated poly(A) tail at the 3’ end of the transcript after it was cleaved by CPSF-73, to stabilize the mature mRNA molecule (Dreyfus & Régnier, 2002). PAP was shown to be recruited and stimulated via interaction with the CPSF subunit FIP1 (Kaufmann et al., 2004). The emerging poly(A) tail is bound and protected from exonucleolytic degradation by PABPN1. Furthermore, PABPN1 stimulates PAP processivity until the poly(A) tail reaches a length of approximately 81-114 nt (Minvielle-Sebastia et al., 1997; Kühn et al., 2009; Legnini et al., 2019).

The compositions of the mammalian core complexes, necessary for CPA and transcription termination are well known by now (Shi et al., 2009). Since then various groups try to gain insights into the structure of the CPA complex, which is nearly as big as the spliceosome A complex, and its subunits.

So far individual structures of the CFIm (Yang et al., 2011a), CFIIm (Schäfer et al., 2018), CstF (Yang et al., 2018) were published. Yet the structure of the CPSF complex, the largest one, was not solved. So far, the lab of Martin Jinek was able to reconstitute a core complex consisting of WDR33 and CPSF-30 and CPSF-160 discriminating the molecular basis of the AAUAAA polyadenylation signal recognition (Clerici et al., 2018). Just recently, the lab of Lori Passmore succeeded to define a minimal core complex competent to perform CPA and resolve the structure in yeast, suggesting the *de novo* assembly of the CPA machinery by the four sub-complex at each CPA event undergoing dynamic rearrangements during, similar to splicing (Hill et al., 2019). Many groups are constantly working to decipher the functions und further analyze structures of individual subunits to understand the mechanism in more and more detail. Still, the number, nature and impact of potential auxiliary or regulatory factors are less well understood.

3.4.1 The mammalian cleavage factor I (CFIm)

The highly conserved mammalian cleavage factor I is one of three major multimeric protein complexes regulating CPA (**Figure 4**) by binding to USEs. The complex recognizes a specific 4-mer motif UGUA (Brown & Gilmartin, 2003; Yang et al., 2010). It was recently shown that CFIm functions as an enhancer of 3' end processing and that the optimal distance between the UGUA motif and the CS is between 40-60 nt (Zhu et al., 2018). The heterotetramer (**Figure 5**A) is composed of two small subunits of 25 kDa and two larger subunits of either 59 kDa, 68 kDa or 72 kDa (Rüegsegger et al., 1996; Venkataraman et al., 2005; Kim et al., 2010; Yang et al., 2011a). The smaller subunit, CPSF5, is encoded by the *Nudt21* gene, while the 59 kDa and the 68 kDa subunits are encoded by two paralogue genes *CPSF7* and *CPSF6*, respectively. The 72 kDa subunit is an alternative splice variant of *CPSF6* including a small alternative exon (Neve et al., 2017). CPSF5 contains a Nudix hydrolase domain, while CPSF6 and CPSF7 both contain an N-terminal RRM, followed by a proline-glycine rich center and an RS-like domain composed of RE/RD/RS di-nucleotides at the C-terminus (**Figure 5**B) (Martin et al., 2010). Therefore, they are highly similar to SR proteins and are considered SR-like proteins (Rüegsegger et al., 1998). The RS-like domain was shown to be involved in protein-protein interactions with other proteins containing similar domains like FIP1 or U2AF65 and to target the protein into nuclear speckles (Cardinale et al., 2007; Zhu et al., 2018; Zhou et al., 2002; Rappsilber et al., 2002).

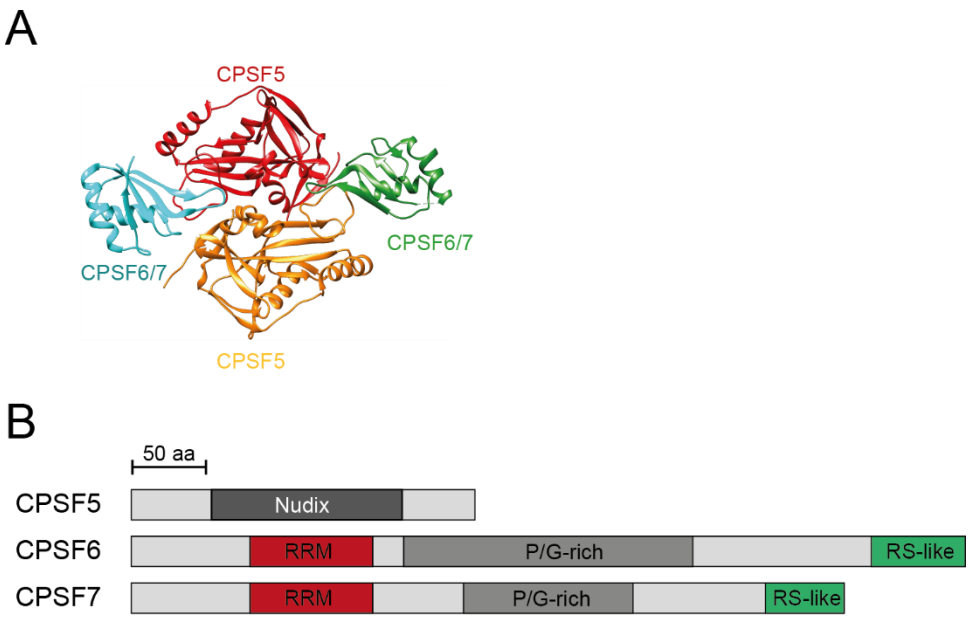


Figure 5: Crystal and domain structure of CFIm and its subunits. A) *Crystal structure of CFIm (PBD ID: 3P5T) with the two core subunits of CPSF5 (red and orange) and the two additional subunits CPSF6 or CPSF7 (cyan and green).* ***B)*** *Domain structure of CFIm subunits CPSF5 and the paralogues CPSF6 and CPSF7. CPSF5 contains an inactive Nudix hydrolase domain, which binds to the UGUA motif. CPSF6 and CPSF7 contain an N-terminal RRM, a central prolin/glycine-rich region and a C-terminal RS-like domain composed of RE/RD/RS di-nucleotide, like SR proteins.*

Surprisingly, the UGUA motif is recognized by the CPSF5 subunits, despite the fact that CPSF6/7 contain RRMs (Brown & Gilmartin, 2003; Yang et al., 2010). The binding of CPSF5 to RNA increases, when the larger subunits are present, but alone they bind RNA only with very low affinities (Dettwiler et al., 2004). It was shown that CPSF5 contains a unique α-helix loop blocking the Nudix-hydrolase activity and simultaneously forming a platform for RNA interaction (Yang et al., 2010). In contrast, a C-terminal α-helix following the RRM of CPSF6 was found to decrease its affinity towards RNA. In addition, one of three aromatic residues, essential for RNA binding, is mutated to leucine at position 128 (Cléry et al., 2008). Instead, the RRM is necessary for the interaction between the large and small subunits of the CFIm (Yang et al., 2011b). Interestingly, each of the larger subunits contact both small 25 kDa subunits, via hydrophobic interactions mediated by lysine residues.

It was shown that depletion of CPSF5 and CPSF6, but not CPSF7, led to a shortening of 3’ UTR length (Li et al., 2015; Gruber et al., 2012; Martin et al., 2012). This fits well with a general enrichment of UGUA motifs at the distal end of 3’ UTRs and the current model suggesting the CFIm as an UGUA motif dependent activator of the furthest downstream located distal PAS (dPAS). Accordingly, the CFIm containing CPSF6 binds to UGUA motifs upstream of the dPAS and recruits the CPSF complex directly by an interaction between the RS and RS-like domains of CPSF6 and FIP1. In addition, the CstF complex is recruited indirectly. The different activities of CPSF6 and CPSF7 were derived from slightly different RS domains with CPSF6 containing higher amounts of RS/E/D dipeptides. Shortening of target 3’ UTRs

upon depletion of the CFIm was explained by decreased activation of the dPAS, which did not affect the UGUA-depleted proximal PAS (pPAS, closest to the upstream stop codon), leading to a shift towards pPAS usage (Zhu et al., 2018). Yet, having defined CFIm as a *cis*-sequence dependent dPAS activator it remains an open question if similar regulators targeting the pPAS exists. Furthermore, this indicates that CPSF6 and CPSF7 are not redundant despite being highly similar in their structure. Otherwise, the effect of depletion of CPSF6 would not be as drastic and the same result would be expected upon depletion of CPSF7. This suggests that CPSF7 is required for different functions, which are not described yet.

3.4.2 FIP1

FIP1 is a conserved subunit of the CPSF complex and encoded by the *Fip1l1* gene. It was described to bind to USEs recognizing U-rich sequences via an arginine-rich RNA binding motif (Martin et al., 2012; Kaufmann et al., 2004). The N-terminus constitutes the PAP interaction domain (PID), followed by regions linked to interactions towards CPSF and CstF subunits, concluding with the C-terminus, which is enriched in RE/RD/RS di-nucleotides, resembling an RS-like domain (**Figure 6**). FIP1 was shown to interact with various CPA *trans*-factors like PAP, CFIm and CPSF-160 (Kaufmann et al., 2004; Zhu et al., 2018).

Figure 6: Domain structure of FIP1.

The interaction with PAP is essential for the polyadenylation process after the nascent transcript was cleaved. PAP alone was shown to bind to RNA weakly and nonspecifically and attaches adenosines inefficiently, because it dissociates after each round of catalysis (distributive mode, (Wahle, 1991). The recruitment by FIP1 stabilizes PAP and increases its specificity and efficiency, switching PAP to the processive mode in which it synthesizes the full-length poly(A) tail (Wahle, 1991; Keller et al., 1991; Kaufmann et al., 2004; Preker et al., 1997). By depletion studies, FIP1 was shown to be essential for self-renewal in murine embryonic stem cells (ESCs) and somatic cell programming (Ding et al., 2009; Hu et al., 2009; Lackford et al., 2014).

3.5 SR proteins: A family of essential splicing factors

The SR proteins are a conserved, multifunctional family of essential constitutive and alternative splicing regulators (Zahler et al., 1992; Zahler et al., 1993). Beside their function during splicing, individual SR proteins also perform distinct functions during 3' end processing, mRNA export, translation and NMD (Müller-McNicoll et al., 2016; Howard & Sanford, 2015). Members of the SR protein family share a highly conserved, characteristic domain structure, which was defined by Manley and Krainer in 2010 (Manley & Krainer, 2010). At the N-terminus they contain at least one RNA recognition motif (RRM), which is followed by a glycine/arginine-rich linker and at the C-terminus they contain a region enriched in arginine (R) and serine (S) di-peptides, called RS domain. To be classified as an RS domain this region must be at least 50 residues long and contain at least 40% RS di-peptides. In addition, SR proteins can have supplementary domains between the RRM and the RS domain, e.g., a second, noncanonical RRM, called RNA recognition motif homologue (RRMH) or a Zinc-knuckle (Zn-knuckle) domain. Twelve canonical SR proteins are members of the SR protein family, named SRSF1-12 (**Figure 7**) that contain RS domains of different lengths. Several other RBPs contain RS domains, but either their RRMs and RS domains are in the wrong orientation, or the proteins contain additional domains, or they contain RE/RD (arginine-aspartate/arginine-glutamic-acid) domains instead. Those RBPs are termed RS-like proteins, but they likely interact with SR proteins or have similar functions.

The RRM enables the interaction of SR protein with single stranded RNA, the linker region was shown to interact with the nuclear export factor 1 (NXF1), while the RS domain facilitates protein-protein interactions and represents a nuclear localization signal (NLS) (Neugebauer et al., 1995). The additional RRMH or Zn-knuckle domain, as well as the compositions of the linker region and RS domain contribute to the binding specificity and determine the distinct sequence each individual SR protein recognizes as a target (Wegener & Müller-McNicoll, 2019).

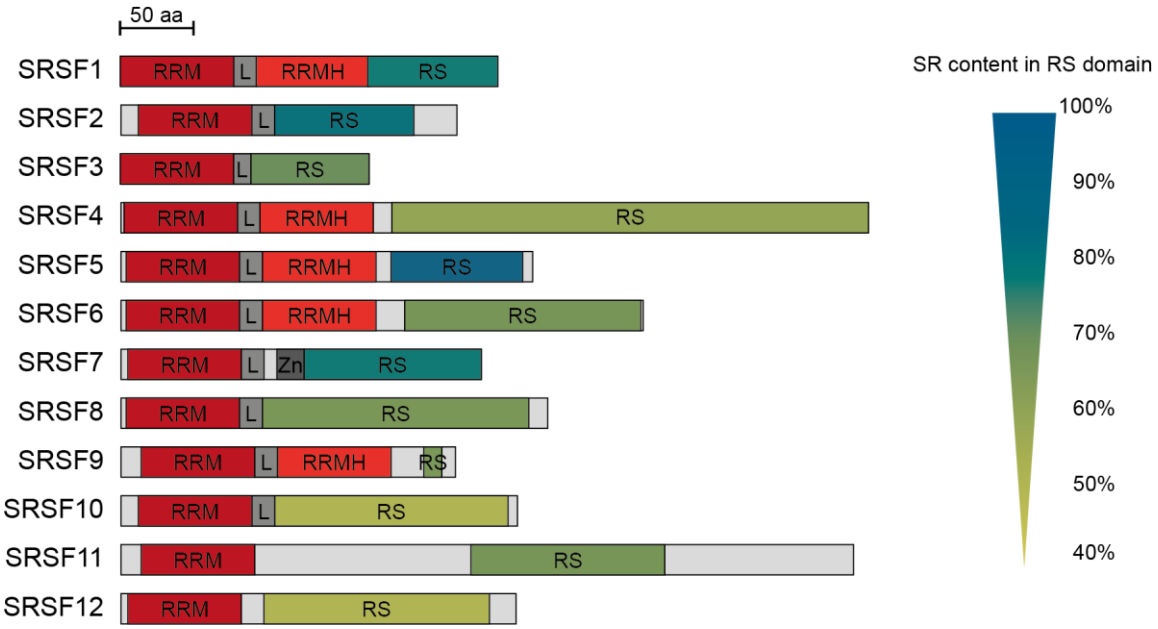


Figure 7: The domain structures of SR proteins. *The SR protein family consists of 12 proteins named SRSF1-12. These proteins share a highly conserved domain structure, starting with an N-terminal RRM (red), followed by a linker sequence (grey). Some SR proteins have one additional RNA recognition motifs, called RRMH (light red) or a Zn-knuckle (dark grey) in front of their RS domains (green to blue). The RS domain is enriched for arginine-serine di-nucleotides. The content (in%) of RS per RS domain is indicated by a color code ranging from 40% (light green) to 100% (dark blue). Modified from (Wegener & Müller-McNicoll, 2019)*

As the spliceosome assembles *de novo* at each intron to be spliced, auxiliary factors contribute to the correct recognition of the splice sites. One family of splicing factors, the SR proteins, contribute to the definition of exon-intron borders by binding to short (4-8 nt) intronic splicing enhancer (ISE) sequences within the intron or exonic splicing enhancers (ESE) within neighboring exons (**Figure 8**) (Zhu et al., 2001; Conti et al., 2013). The SR proteins stabilize the binding of the U1 snRNP component U170K at the 5' SS, as well as the binding of U2AF65 at the 3' SS. Furthermore, the binding of U2 and U6 is enhanced, promoting the assembly of a catalytic active spliceosome (Shen & Green, 2006). These interactions are mediated via arginine-serine (RS) rich region of the participating proteins. Hence, binding of SR proteins facilitates the recognition of bound exons. This can be antagonized by the binding of proteins from the hnRNP family, which bind to short exonic and intronic splicing silencer (E/ISS) sequences, inhibiting the assembly of the spliceosome subunits e.g., by extending U1-exon interactions or blocking U2 association (Tange et al., 2001; Cáceres & Kornblihtt, 2002; Chiou et al., 2013).

SR proteins are redundant to each other in constitutive splicing, but they display non-overlapping functionalities in alternative splicing where they bind to ESEs within alternative exons (Änkö et al., 2012). In general, binding of SR proteins facilitates the inclusion of bound exons. Two SR proteins (SRSF10 and SRSF12) act in the opposite way, as they were shown to be potent inducible splicing inhibitors of bound target exons (Shin & Manley, 2002; Simard & Chabot, 2002).

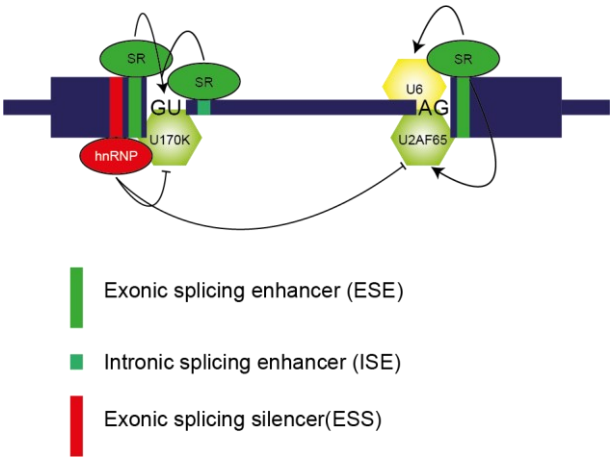


Figure 8: Regulation of splice site activation by SR proteins. *SR proteins (green) bind to exonic or intronic splicing enhancer (E/ISE) sequences near the 5' SS and 3' SS recruiting and stabilizing the U1, U2 and U6 snRNPs via specific protein-protein interactions. Thereby, SR proteins facilitate splicing of the surrounded introns. Proteins of the hnRNP family (red) bind to exonic splicing silencer (ESS) sequences antagonizing the effect of SR proteins by blocking recruitment of the spliceosome subunits.*

The activity of SR proteins is regulated via post-translational modifications (PTMs), especially by phosphorylation of serine residues within the RS domain (Zhou & Fu, 2013). The extent of RS-phosphorylation is tightly controlled by the interplay of SR protein specific kinases and phosphatases (**Figure 9**). SR proteins are stored in nuclear speckles in an intermediate phosphorylated state. Nuclear speckles are interchromatin granule nuclear bodies found as storage space of gene expression regulators of all processing steps from transcription initiation until nuclear export. Moreover, nuclear speckles consist of the lncRNA MALAT1 and snRNAs, suggesting a partial preassembly of the spliceosome within nuclear speckles (Cáceres et al., 1997; Galganski et al., 2017). Upon transcription activation, SR proteins are hyper-phosphorylated by the nuclear CDC2-like kinase CLK1/4 leading to their release from nuclear speckles and their recruitment to nascent pre-mRNA (Misteli et al., 1998). Subsequently, the spliceosome is recruited to the pre-mRNA and splicing is initiated. During splicing, SR proteins are de-phosphorylated by the nuclear phosphatases PP1 (Mermoud et al., 1994; Ma et al., 2010) and PP2A (Shi et al., 2006) to release the splicing machinery and to recruit NXF1, when splicing is completed and the RS proteins are hypophosphorylated.

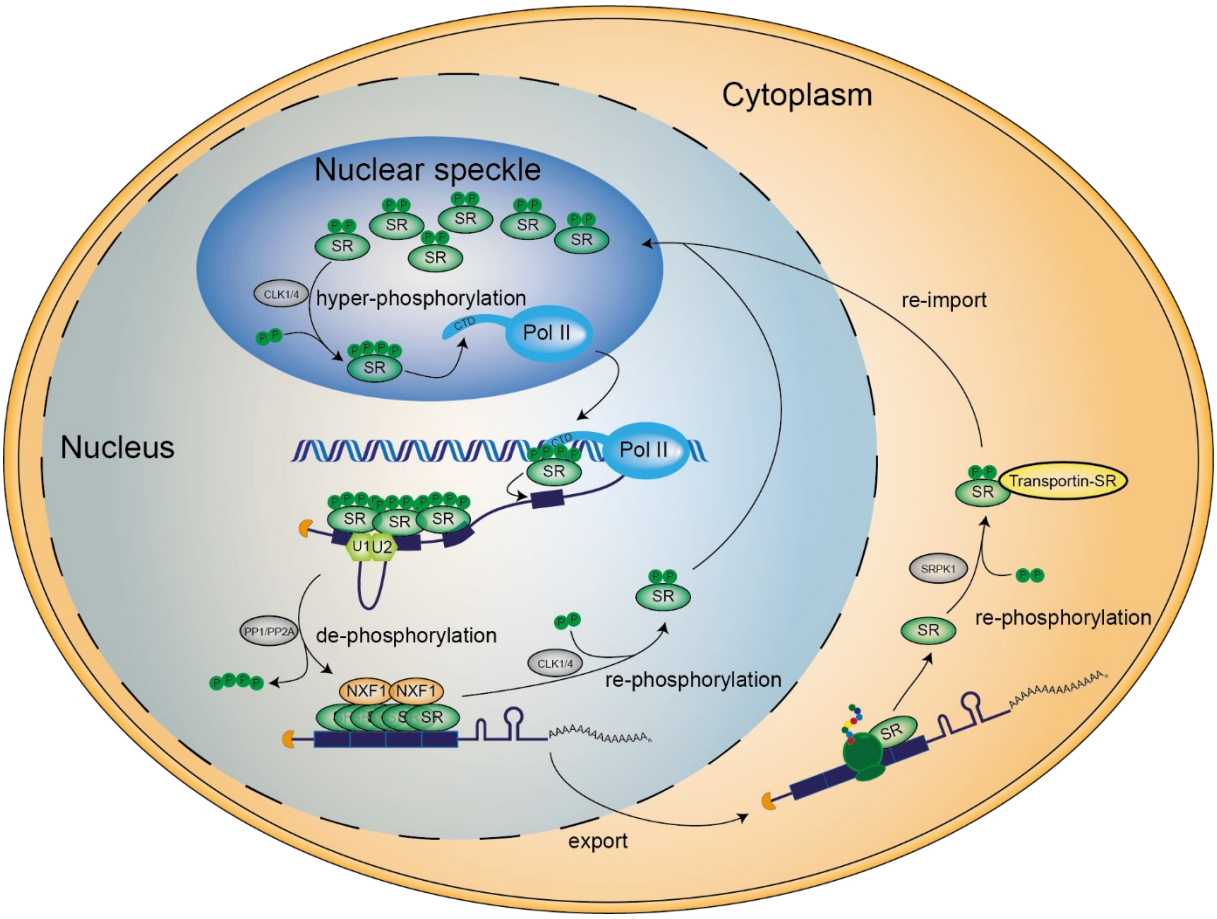


Figure 9: The phosphorylation cycle of SR proteins. *The activity and localization of SR proteins is regulated by PTMs, mainly by phosphorylation of serine residues. The phosphorylation cycle begins in nuclear speckles (dark blue), where intermediate phosphorylated SR proteins (green) and other RNA processing factors are stored. Upon transcription initiation, SR proteins are completely phosphorylated by CLK1/4 (grey), released from nuclear speckles, and loaded onto the CTD of Pol II. Once splice sites emerge in the nascent pre-mRNA, they are deployed on the exons binding ESEs and recruit the spliceosome (light green). During splicing, SR proteins are de-phosphorylated by PP1 and PP2A (grey), to complete splicing catalysis and in their hypophosphorylated state, they recruit NXF1 (orange) for the export of the mature mRNP to the cytoplasm. In the cytoplasm, SR proteins bound to the mRNA are stripped off during the initial round of translation. SR proteins are partially re-phosphorylated by SRPK1 (grey) and re-imported into the nucleus by the interaction with TNPO1 (yellow). Alternatively, SR proteins can be partially re-phosphorylated by CLK1/4 in the nucleus to be recycled immediately. Modified from (Wegener & Müller-McNicoll, 2019).*

Due to the interesting fact that SR proteins bind mostly to exons in comparison to other splicing factors, they stay bound to the mature transcript and engage in mRNP compaction and nuclear export into the cytoplasm (Müller-McNicoll & Neugebauer, 2013). Once in the cytosol, SR proteins are stripped off from the mRNA during the initial round of translation by moving ribosomes and are subsequently partially re-phosphorylated by the SR protein kinase 1/2 (SRPK1/2). This promotes re-import of SR proteins into the nucleus by Transportin-SR and re-localization to nuclear speckles for another round of splicing (Kataoka et al., 1999; Lai et al., 2000; Yun & Fu, 2000; Lai et al., 2001; Long et al., 2018). This way, most SR proteins shuttle from the nucleus to the cytoplasm and back again (Botti et al., 2017). Alternatively, non-shuttling SR proteins, such as SRSF2, are partially re-phosphorylated by CLK1/4 in the nucleus, released from the mature mRNA prior to export and recycled for subsequent splicing reactions (Lin et al., 2005). In summary, RS domain phosphorylation

influences the protein interactions, sub-nuclear localization, RNA binding affinity, nuclear export, and re-import of SR proteins.

In addition to their essential function in fine-tuning gene expression by regulating alternative splicing, individual SR proteins are involved in additional steps of RNA biogenesis and processing turning them into key players of gene regulation. First, SR proteins were reported to be recruited from the paraspeckles to sites of active transcription, mediated by the CTD of PolII in an RNA-dependent manner regulating the speed of transcription and being loaded onto nascent transcripts to mediate slicing (Yuryev et al., 1996; Misteli et al., 1997; Sapra et al., 2009). SRSF1 and SRSF2 have active functions in the regulation of transcription elongation rates (Lin et al., 2008; Ji et al., 2013). Both SR proteins switch PolII from its pause into an active state. SRSF2 and P-TEFb, which is important to release PolII, are stored as parts of the 7 SK RNA complex, which is associated with genomic DNA. Initial assembly results in an early stretch of nascent RNA containing an SRSF2-specific ESE. Subsequently, SRSF2 is recruited from the 7 SK complex to the emerging ESE within the 5' UTR, delivering P-TEFb. Depletion of SRSF2 resulted in the accumulation of PolII at transcription start sites decreasing the speed of transcription. Moreover, truncation of the CTD lead to subsequent inhibition of pre-mRNA splicing due to loss of recruitment of splicing factors to the active sites of transcription (Du & Warren, 1997). Following transcription and splicing, SR proteins were also shown to influence 3' end processing for few targets (Lou et al., 1998; Bradley et al., 2015). In these cases, depletion of SR proteins yielded transcripts using alternative, upstream 3' end processing sites expressing shorter transcript isoforms. Yet, the full picture and underlying mechanisms are unknown. Downstream, shuttling SR proteins also facilitate nuclear export of spliced and viral un-spliced transcripts by recruiting the nuclear RNA export factor 1 (NXF1) (Huang et al., 2003; Lai & Tarn, 2004; Escudero-Paunetto et al., 2010; Müller-McNicoll et al., 2016; Botti et al., 2017). With bound target transcripts translocated to the cytoplasm SRSF1, was found in polysomes being involved in translation (Sanford et al., 2004). SRSF1 does that by stimulating the phosphorylation of 4E-BP via association with the mTOR kinase (Michlewski et al., 2008; Maslon et al., 2014). Phosphorylation of 4E-BP leads to the dissociation of this factor from eIF4E increasing the cap-dependent translation initiation. SRSF3 and SRSF7, on the other hand, were connected to cap-independent translation of viral transcripts mediated via IRES or constitutive transport elements (CTEs) (Bedard et al., 2007; Swartz et al., 2007). In addition, SRSF5 and SRSF6 enhance translation of unspliced HIV-1 RNA (Swanson et al., 2010). Interestingly, similar to its opposite function in alternative splicing SRSF10 was shown to inhibit splicing (Liu & Harland, 2005).

However, the functional spectrum of SR proteins is not limited to affect the expression levels of mRNAs on many levels. SRSF1 is involved in the splicing-independent biogenesis of miRNAs from the primary-miR-7 transcript by promoting cleavage of mature miRNA by DROSHA (Wu et al., 2010). In

addition, SRSF1 seems to bind to additional primary-miRNAs as do other shuttling SR proteins, which also associate with distinct subsets of miRNA precursor transcripts (Sanford et al., 2009; Änkö et al., 2012).

Aberrant expression of SR proteins has major effects on the cell viability and is connected to several severe diseases such as cancer and neurological diseases, like mis-expression of other RBPs. As an example, SRSF1 is the key oncogene for small cell lung cancer, a highly aggressive type of lung cancer (Karni et al., 2007; Jiang et al., 2016). In this case, increase of SRSF1 genomic copy numbers and yielding over-expression were connected to poor survival rates due to a downstream activation of PI3K/AKT and MEK/ERK pathways. Depletion of SRSF1 reduced the amount of 3D cell spheroids compared to non-target siRNAs indicating a potential therapeutic approach. In general, aberrant alternative splicing is associated with progression of cancer (Venables et al., 2009). As an example SRSF1, SRSF3 and SRSF7 facilitate alternative inclusion of exon v9 of adhesion molecule CD44 (Galiana-Arnoux et al., 2003). This exon is spliced in together with exons v8 and v10 and exon v8 was only found in RNA samples derived from tumor cells. Another example is the expression of the PK-M2 isoform that is generated by SRSF3 mediated alternative splicing of pyruvate kinase M, which also was connected to tumor growth (Wang et al., 2012). Besides cancer, SR protein-derived alternative splicing have implications on neuronal diseases. For example, SRSF3-mediated AS of *Tropomyosin receptor kinase B* (*TRKB*) yields the expression of TRKB-Shc that is involved in Alzheimer's disease (Wong et al., 2012). Moreover, SRSF3 mediates the early-to-late stage switch during human papillomavirus infections by regulating the gene expression of viral RNAs containing A/C-rich sequence elements (Jia et al., 2009).

Hence, SR proteins are not limited to being essential for splicing only. Moreover, they are involved in all steps of an mRNA's life cycle and crucial for fine-tuning gene expression. Aberrant expression causes several severe diseases of different kinds, representing the importance of this protein family. Yet, the full functional spectrum and respective mechanisms of SR proteins are not entangled and need to be investigated in detail.

3.6 Alternative polyadenylation

The complexity of the eukaryotic transcriptome is not only increased through the variable connection of alternative exons by AS, as described before in chapter 3.3, it is also increased by a process termed alternative polyadenylation (APA). Since the mid-1980s it is known that genes can possess more than one PAS and can give rise to different transcripts dependent on the PAS that is used by the CPA machinery. By now it is known that approximately 70% of all mammalian mRNAs contain multiple PAS and can undergo APA (Derti et al., 2012; Hoque et al., 2013) to regulate or fine-tune gene expression during development or in different tissues (Ji et al., 2009; Derti et al., 2012). Most of the PAS are found within the 3' untranslated regions (3' UTRs) of a transcript (Tian & Manley, 2017), but PAS can also be located in exons and introns (**Figure 10**A) (Tian et al., 2007; Hoque et al., 2013). PAS usage of PAS located in 5' end introns is inhibited by U1 snRNP and telescripting, while PAS located in the last intron are inhibited by CPSF5 and CPSF6 (Kaida et al., 2010; Li et al., 2015). The PAS furthest downstream within the terminal exon of each gene is referred to as the distal PAS (dPAS), while any PAS upstream of the dPAS is called proximal PAS (pPAS). 3' UTRs play important roles in the regulation of biological complexity and consistently their size increased during evolution (Chen et al., 2012; Mayr, 2016, 2018). APA events that occur within the 3' UTR are termed 3' UTR-APA, which gives rise to transcripts with shorter or longer 3' UTRs but with identical coding potential. APA upstream of the terminal exon is termed coding sequence APA (CDS-APA) and here the coding potential is altered.

Genome-wide analysis of mouse transcripts found that the length between short and long 3' UTR-APA isoforms differs on average by 7-fold (Hoque et al., 2013). As 3' UTRs contain various *cis*-regulatory elements, the length and composition of 3' UTRs affects mRNA stability, translation, nuclear export, cellular localization and even protein localization (Tian & Manley, 2017). Most of these effects are terminated by *trans*-acting factors, such as RBPs, miRNAs and lncRNAs (**Figure 10**B). miRNAs are small RNAs of 22 nt length, which bind to target mRNAs and either inhibit their translation or induce their degradation thereby decreasing their expression and accessibility for translation. miRNA binding sites are enriched in 3' UTRs (Sandberg et al., 2008; Ji et al., 2009) and lengthening of the 3' UTRs includes more potential miRNA target sequences. Extending the 3' UTRs can also include other destabilizing *cis*-acting elements such as AU-rich elements (AREs) and GU-rich elements (GREs, (Garneau et al., 2007). Furthermore, *Alu* elements can be included in 3' UTRs which often form double stranded RNA regions with lncRNAs that contain complementary *Alu* sequences inducing STAU1-mediated decay of these transcripts (Gong & Maquat, 2011). Finally, it was shown that very long 3' UTRs can trigger degradation via the NMD pathway (Hogg & Goff, 2010). Studies on the sub-cellular localization of transcripts revealed a differential distribution of transcripts with short or long 3' UTRs in the cell, whereby short 3' UTRs were enriched in the cytoplasm and long 3' UTRs in the nucleus,

indicating a connection between APA and nuclear export/nuclear retention (Djebali et al., 2012; Neve et al., 2016).

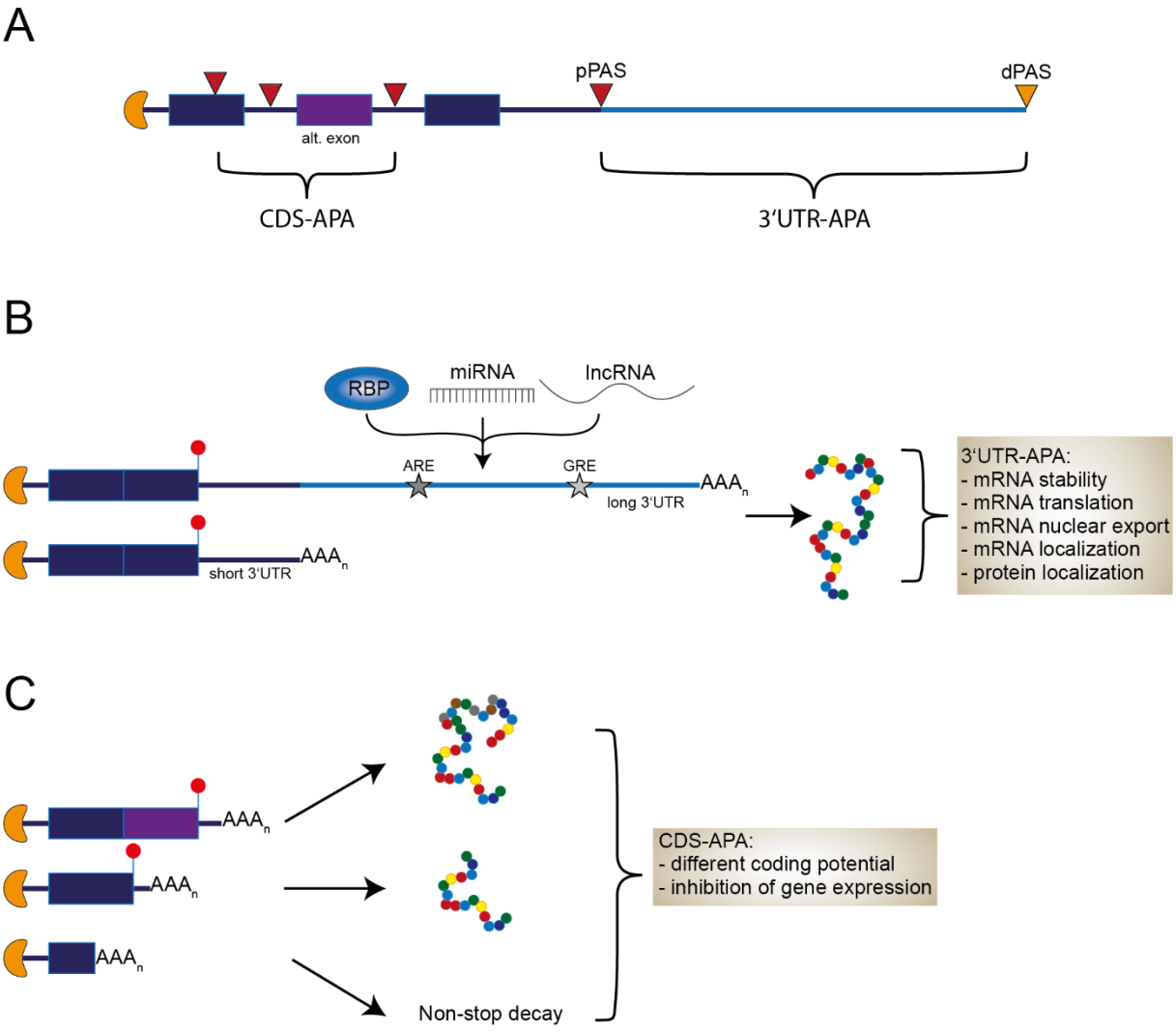


Figure 10: Scheme of alternative polyadenylation variants. A) *Scheme of a capped pre-mRNA. Exons (thick boxes) are separated by introns (thin lines). An alternative terminal exon is highlighted in purple. PAS are marked by triangles. The 3' most PAS is termed distal PAS (dPAS, orange), while all upstream PAS are termed proximal PAS (pPAS, red). Depending on the PAS used, two kinds of APA can be discriminated: coding-sequence APA (CDS-APA) or 3' UTR-APA. Both kinds are shown in detail in B) and C).* ***B)*** *Scheme of 3' UTR-APA. 3' UTR-APA gives rise to transcript isoforms with identical coding sequences but vary in the length of their 3' UTR. The longer part of the 3' UTR (light blue) is more likely to contain destabilizing sequence elements, e.g., AU- and GU-rich elements (ARE/GRE), RBP, miRNA or lncRNA binding sites leading to mRNA decay. Aspects influenced by 3' UTR-APA are summarized in the box. STOP-codons are marked by red lollipops.* ***C)*** *CDS-APA changes the information of the coding sequence and the attached 3' UTR giving rise to distinct protein isoforms with individual functions. This can be achieved either by the inclusion of an alternative last exon (top), inhibition of a weak 5' SS and following read-through into the adjacent intron containing an intronic PAS (composite exon, middle) or the activation of an exonic PAS leading to a transcript without in-frame stop codon (bottom). The last isoform will be degraded by the non-stop decay pathway, while the others will express distinct protein isoforms.*

The length of 3' UTRs was also shown to affect the sub-cellular localization of distinct transcript isoforms resulting in the localized translation of the respective protein isoforms (Martin & Ephrussi, 2009). This has been well studied in neuronal cells, where it was shown for several genes, that the short 3' UTR isoforms localized to the cell body, while the long 3' UTR isoform was translated in the

dendrites and axons (An et al., 2008; Yudin et al., 2008). Furthermore, the lab of Christine Mayr has proven that distinct 3' UTRs can also impact protein localization independent of the CDS by recruiting distinct adapter proteins (Berkovits & Mayr, 2015; Mayr, 2018). The long 3' UTR of *CD47* was found to contain HuR-binding motifs, recruiting the respective protein to the newly translated CD47 protein initiating its sorting to the cell membrane. These motifs are missing when the shorter alternative transcript is expressed resulting in endoplasmic reticulum-bound translation.

In addition to multiple PAS within the 3' UTR of genes, PAS can also be located upstream of the terminal exon, in introns or downstream of alternative last exons. Such PAS were found in 40% of all mouse genes (Hoque et al., 2013) and they have a major impact on the encoded proteins as the CDS and 3' UTR are changed. Depending on the site of the alternative upstream PAS different kinds of CDS-APA can be discriminated. The first one is the inclusion of an alternative terminal exon in the mature transcript by AS (**Figure 10**C top). An example for this kind is the tissue-specific expression of *calcitonin-related polypeptide-α gene* derived transcripts (Lou et al., 1998). The native isoform including the 3' terminal exon is translated into the calcitonin gene-related peptide 1, which is enriched in the hypothalamus. Insertion of an upstream, alternative last exon and the activation of the adjacent PAS gives rise to the calcitonin peptide, which is enriched in thyroid cells. Also connected to splicing is the composite exon, where a weak 5' SS is not recognized leading to retention of the adjacent intron, which contains an intronic PAS, whose usage leads to transcription termination and extension of the exon (**Figure 10**C middle). The first physiological example for the usage of a composite exon was already described in 1980. During the activation of B cells, the expression of the *immunoglobulin M heavy gene* switches from the dPAS to the usage of an upstream PAS that causes a composite exon and a shortened transcript missing the C-terminal exons encoding trans-membrane anchor sequences. This leads to the switch from membrane bound to secreted IgM antibodies (Alt et al., 1980). Until now, an activation of proximal CDS-APA PAS and composite exons were found in more than 300 mouse gene products (Davis et al., 2006). More recently, the lab of Christine Mayr identified the activation of intronic PAS and the resulting expression of truncated protein isoforms of more than 300 genes, with predominately tumor-suppressing functions, as a key driver of chronic lymphatic leukemia (CLL) in malignant B cells (Lee et al., 2018). The intronic-PAS derived truncated protein isoforms lacked their tumor suppressing functions turning several genes into oncogenes, driving CLL, rather than accumulated genetic mutations. These three CDS-APA types result in the expression of shorter protein isoforms with distinct functions than the full-length proteins.

Besides the production of alternative functional proteins, CDS-APA can give rise to truncated protein isoform without functions. An example for this is APA at an intronic PAS within the *CstF77* transcript (**Figure 4**). At high cellular levels of the CstF-77 protein the upstream PAS is activated to

reduce functional CstF-77 levels and maintain CstF-77 homeostasis by a negative, auto-regulatory feedback-loop (Pan et al., 2006). PAS are located sometimes within exons and their activation gives rise to truncated transcripts likely lacking an in-frame stop codon (**Figure 10**C bottom). Those transcripts are subsequently subjected to degradation via the non-stop decay pathway (Vasudevan et al., 2002). In addition, similar to 3' UTR-APA, cellular proliferation is connected to the expression of shortened transcripts by activation of proximal PAS increased expression of the respective genes (Yao et al., 2012; Elkon et al., 2012; Hoque et al., 2013). These examples show the relevance of different APA types on expanding the transcriptome complexity and possibilities to fine-tune the expression of distinct transcripts and protein isoforms.

The CPA core components and their functions in (alternative) cleavage and polyadenylation are well described. However, the regulation of APA, e.g., how certain PAS are activated or blocked is not well understood. It was shown that the PAS choice is influenced directly by the strength of the poly(A) signal and all auxiliary *cis*-acting elements in a manner similar to the strength of splice sites (Cheng et al., 2006b). Comparing multiple PAS sites, it was found that poly(A)signals upstream of pPASs are usually weaker than their corresponding dPAS (Tian et al., 2005). This suggests that the pPASs are subject to regulation whereas the dPAS is part of a fail-safe mechanism to guarantee transcription termination at the 3' most end of the transcribed gene. Moreover, it was shown that the expression level of core CPA factors influences PAS choice as summarized in **Table 1**.

Table 1: Summary of the effects of depletion of core/auxiliary CPA factors on 3' UTR length as comprehensively analyzed by Li et al., 2015.

Protein	Complex	Motif	Effect of depletion on 3' UTR	Additional citations
CPSF5	CFIm	UGUA	shortening	(Gruber et al., 2012; Martin et al., 2012)
CPSF6				
CLP1	CFII		shortening	
PCF11	CFII	G-rich	lengthening	(Wang et al., 2019)
CPSF-160	CPSF		lengthening	
CPSF-100	CPSF		shortening	
CPSF-73	CPSF		lengthening	
CPSF-30	CPSF	AAUAAA	shortening	
WDR33	CPSF	AAUAAA	shortening	
FIP1	CPSF	U-rich	lengthening	
CstF-50	CstF	G/U-rich	lengthening	
(τ)CstF-64	CstF	G/U-rich	lengthening	(Yao et al., 2013)
CstF-77	CstF	G/U-rich	lengthening	

Symplekin			lengthening	
RBBP6			lengthening	(Di Giammartino et al., 2014)
PABPN1			shortening	(Jenal et al., 2012; Klerk et al., 2012)

Overexpression of CSTF64, FIP1, RBBP6 and PCF11 were shown to activate upstream pPAS (Takagaki et al., 1996; Lackford et al., 2014; Li et al., 2015; Wang et al., 2019), while CPSF5, CPSF6 and PABPN1 increased the usage of dPAS (Gruber et al., 2012; Martin et al., 2012; Jenal et al., 2012; Klerk et al., 2012; Li et al., 2015; Zhu et al., 2018). Furthermore, additional *trans*-acting RBPs were shown to regulate PAS choice in distinct transcripts, many of them being splicing factors this way connecting both co-transcriptional processes. The first described example was NOVA2, which was found to regulate pPAS in a distance-dependent manner (Licatalosi et al., 2008). The targeted pPAS was inhibited when NOVA2 bound near it but activated when NOVA2 binding sites were more distant. ELAV was shown to inhibit dPAS usage in *Drosophila melanogaster*, resulting in 3' UTR extension beyond the annotated 3' end in a neuron-specific manner (Hilgers et al., 2012). FUS, a splicing factor linked to amyotrophic lateral sclerosis (ALS), was also found to regulate APA of specific target transcripts in a position-dependent manner (Masuda et al., 2015). The protein was shown to interact with the CPSF and CstF complex, recruiting them to upstream pPAS to activate those sites, leading to the expression of shortened transcripts. Another RBP found to activate or repress PASs in a distance-dependent manner is TDP-43, which is also linked to ALS (Ratti & Buratti, 2016; Rot et al., 2017). When TDP-43 bound to GU-rich sequences in close proximity to a PAS it inhibited its usage, while it increased the PAS usage, when the TDP-43 binding site was further downstream, similar to its regulatory mechanism during AS. A highly versatile function in regulating CPA was assigned to the 5' SS binding factor U1, which was shown to interact with the CFIm complex and CPSF-160 protein inhibiting premature CPA at cryptic PASs within 5' introns (Awasthi & Alwine, 2003; Kaida et al., 2010; Devany et al., 2016). Besides selecting alternative PASs to be used, RBPs and known alternative splicing factors such as RBM10 have splicing-independent functions in generally activation of CPA (Mohan et al., 2018). RBM10, a protein mainly described as AS factor related to apoptosis and inflammation, was shown to recruit and stabilize Star-PAP to a certain subset of mRNAs related to cardiac functions, to express anti-hypotrophic factors during heart failure. Finally, PAS usage is also regulated by the transcription rate of PolII. PolII was shown to slow down and pause in G-rich regions and thereby facilitate CPA at nearby PAS (Yonaha & Proudfoot, 1999). In addition, transcription factors were found to recruit CPA factors to increase pPAS usage (Rosonina et al., 2003; Glover-Cutter et al., 2008; Ji et al., 2011).

APA gives rise to distinct transcript isoforms with specific functions and is tightly regulated to ensure the correct spatial and temporal expression of encoded proteins. Mis-regulation of APA has been linked to various diseases and types of cancer, especially when 3' UTRs are shortened in differentiated cells (Mayr & Bartel, 2009; Xia et al., 2014). The overexpression of CPSF5 was linked to the suppression of glioblastoma tumours, while changing the copy numbers by depletion or duplication of *Nudt21* was linked to neuropsychiatric disorder (Masamha et al., 2014; Gennarino et al., 2015). Another example of cancer-connected APA is a CDS-APA derived truncated isoform of RBBP6 that was found to be decreased in several tumour tissues (Mbita et al., 2012; Di Giammartino et al., 2014). Interestingly, this alternative, truncated protein isoform competes with full-length RBBP6 for interactions with CPA factors inhibiting 3' end processing.

3.6.1 The evolution of APA analysis

In the beginning, identification and analysis of alternative polyadenylation events was tedious as it was based on single gene approaches using RT-PCR (Alt et al., 1980; Lou et al., 1998). The generation of large numbers of expressed sequence tags (ESTs) enabled first genome-wide analysis and characterizations of APA events (Gautheret et al., 1998; Tian et al., 2005). Being limited to 3' end derived datasets these approaches helped defining the poly(A) signal and its distribution upstream of the cleavage site, as well as first auxiliary *cis*-sequence elements e.g., the UGUA motif. Moreover, this approach pointed the differential strengths of pPAS vs. dPAS out. Yet, EST derived datasets have strong limitations as they represent only small regions, which are sequences individually and hence are error prone.

Following the ESTs, microarrays emerged as a widely used tool to analyse alternative polyadenylation, despite the fact of being biased due to the design of each chip (Sandberg et al., 2008; Flavell et al., 2008; Ji et al., 2009; Ji & Tian, 2009). Microarrays enabled a handy analysis of differential usage of tandem PAS within 3' UTRs comparing the pPAS/dPAS usage ratios. Thereby, Rickard Sandberg was able to describe the global shortening of 3' UTRs in proliferating cells compared to differentiated cells. Being more agile than the tedious analysis of EST datasets, the analysis of alternative polyadenylation events by microarrays has some drawbacks as it is limited by the number of probes, as well as the design of each chip. In addition, analysis of genes with more than two alternative APA-derived isoforms were difficult to separate. Moreover, the design of the chip was still based on EST-derived findings. Hence, no new potential targets could be identified *de novo*.

Subsequently, paired-end tags (PET) sequencing was established for APA analysis in the transition towards multiplexed, high-throughput sequencing techniques (Ng et al., 2005; Ng et al.,

2006). This method is based on sequencing short, unique DNA sequences at the 5’ and 3’ end of gene fragments to identify the sequence composition in-between relying on full-genome sequencing data to map back the sequenced tags. The advantage of this technique was that multiple PETs could be concatenated within a final library to be sequenced at once utilizing the long reads of Sanger sequencing. Later, with the development of the first generation of next-generation sequencing techniques like 454 sequencing multiplexing became available, enabling more efficient analysis of many samples.

With further advances within the field of next-generation sequencing and the broad development of RNA-based sequencing techniques the increase in the amount and quality of information gave new power to the analysis of the expanding transcriptome and RNA processing related questions, such as alternative splicing and alternative polyadenylation. With the new power, new challenges in the experimental and library design arose to increase the efficiency of RNA-Seq based APA analyses. On the one hand the reads are limited in length up to a maximum of around 300 nt per read with shorter reads being superior regarding the single read quality. Moreover, APA-derived transcript isoforms needed to be identified by present untemplated stretches of adenosines. This was improved by selecting for poly(A)$^+$ prior to the library preparation. Still, only a small fraction of the RNA-Seq derived reads mapped to the most 3’ end of the transcript, which is of particular interest in APA research. Hence, a high variety of protocols enriching for the 3’ end of polyadenylated transcripts were developed in the recent years to deal with the short read lengths, internal priming issues, the difficulties of sequencing homonucleotide stretches, such as the poly(A) tail and 3’ -5’ vs 5’ -3’ sequencing direction. Those will be explained more detailed in the following. In general, three kinds of approaches can be differentiated either focusing on an oligo(dT) priming strategy, or RNA manipulation or direct RNA sequencing.

The oligo(dT) based approaches utilize the nature of cleaved and polyadenylated transcripts by targeting the poly(A) tail as a platform for reverse transcription using oligo(dT) primers. This is a simple, straightforward approach enabling the introduction of additional sequence elements necessary to generate libraries for sequencing. Moreover, different samples can be prepared in parallel for multiplexed analysis by introducing individual experimental identifiers, such as molecular, nucleotide-based barcodes (Fox-Walsh et al., 2011). The various oligo(dT)-based sequencing techniques focusing on the 3’ end of transcripts vary in their respective strategies of introducing the necessary 5’ and 3’ adapters and fragmentation of mRNA to enrich for 3’ terminal RNA fragments (Chen et al., 2017). In general, random fragmentation by heat and/or chemical shearing is preferred over enzymatic cleavage to reduce any bias. A current issue with these 3’ end sequencing strategies is the potential of internal priming towards encoded A-rich stretches and the resulting generation of false-

positive PAS annotations in the downstream analysis or the task to remove those computationally. One method termed whole transcriptome termini site (WTTS) sequencing tried to reduce internal priming events by using specific PCR primers during library amplification containing additional T-residues at the 3' end (Zhou et al., 2016). Yet, this method was set up to utilize the Ion Torrent sequencing platform, which is less commonly used as the Illumina platform. The biggest limitation of oligo(dT) based enrichment of 3' end RNA fragments and subsequent sequencing from the 3' end is the presence of the oligo(dT) stretch, which retained during the library preparation. The base-calling quality of the most widely used Illumina sequencing platform has been shown drop after reading through stretches of homonucleotides (Wilkening et al., 2013). The obvious solution of shortening the oligo(dT) stretch would increase the chances of internal priming, hence different strategies were implemented by distinct 3' end sequencing protocols. In PA-Seq the RT primer was modified by introducing a single dUTP nucleotide in the beginning of the oligo(dT) primer which allowed to shorten the T-rich stretch after poly(A)$^+$ enrichment using an uracil-specific excision reagent (USER) cleaving after uracil nucleotides (Ni et al., 2013). Another protocol, termed A-Seq, introduced a very long RT primer containing the 3' adapter sequence and its complementary sequence to form a hairpin structure, splitting the oligo(dT) stretch into two parts, reducing the number of Ts, which were subsequently read during sequencing (Martin et al., 2012). But not only can the RT primer be modified to reduce the sequencing of the T-rich stretch. Stefan Wilkening suggested to fill the corresponding poly(A) stretch with unlabelled dTTPs before the actual sequencing run (Wilkening et al., 2013).

Besides the oligo(dT)-based approaches to specifically sequence 3' ends of RNAs, additional methods were developed to circumvent internal priming. Usually those methods enrich for poly(A)$^+$ RNAs, followed by a ligation of the 3' adapter, which is used as the platform for subsequent reverse transcription instead of an oligo(dT) primer. Subsequently, the 5' and 3' adapters are ligated, and samples amplified by PCR to sequence the libraries. A widely used protocol of this kind is the 3' READS+ approach, which is a further development of the original 3' READS protocol (Hoque et al., 2013; Zheng et al., 2016). The protocol begins with the enrichment of poly(A)$^+$ RNAs using an oligo(dT)$_{25}$ primer and subsequent enzyme-based fragmentation. Subsequently, the 5' adapter is already ligated, followed by a second poly(A)$^+$ enrichment using a specific oligo(dT) primer containing five locked thymidine nucleotides which form stable interactions with RNA nucleotides, but are insensitive to the subsequent RNase H digestion, which degrades DNA:RNA hybrid sequences. Thereby, the poly(A) tail gets shortened before the 3' adapter is ligated. This adapter is then used as the platform for reverse transcription and subsequent library amplification enabling sequencing from the 3' end without reading through too many homonucleotides.

Last, direct RNA sequencing (DRD) was used for APA studies (Ozsolak et al., 2010; Wu et al., 2011; Mata, 2013). This method enabled the sequencing without any manipulation by hybridizing the poly(A)$^+$ RNA to a flowcell covered with oligo(dT) anchors. Subsequently, the free A-stretches were filled with dTTP prior to the actual sequencing. Unfortunately, this method is not available anymore as the required Helicos sequencing platforms went bankrupt.

Besides approaches to generate 3' end-derived datasets to increase the number of usable reads to improve APA analysis compared to RNA-Seq, specific algorithms and tools are generated to identify and quantify PAS usage from RNA-Seq datasets. Usually, those tools rely on already annotated PAS to quantify reads in those regions and compare the usage between different conditions (e.g., Roar and QAPA) (Grassi et al., 2016; Ha et al., 2018). Other algorithms like DaPARS and TAPAS can annotate PAS *de novo* by using a change-point model of the read coverage (Xia et al., 2014; Wang et al., 2014a; Arefeen et al., 2018).

In the recent years, comprehensive library preparation and analysis based on the 3' end of transcripts became of a higher interest so that now also 3' end sequencing kits are commercially available. On the one hand massive analysis of cDNA ends (MACE) Seq, developed by GenXPro, and on the other hand the QuantSeq 3' mRNA-Seq Library Prep Kit by Lexogen are available on the market and promise a straightforward library preparation. Both protocols are derived from the oligo(dT)-based approaches and aim to sequence the libraries in the 5' -3' direction to avoid sequencing through a T-rich stretch in the beginning. Yet, this might result in too short reads, which lack the information of the untemplated poly(A) tail to identify the read as a correct transcript end. During the MACE-Seq protocol each transcript is linked with a unique molecular identifier (UMI) that allows the removal of PCR-derived duplicates to increase the power of transcript quantification.

Yet, all known methods have specific drawbacks, such as long and tedious procedures, high requirements of starting material, internal priming, issues with sequencing through an oligo(dT) stretch in the 3' -5' direction or application of less widely used sequencing platforms. Therefore, existing methods are improved, and new protocols are established constantly, so far lacking the optimal solution. Moreover, each protocol needs individual computational solutions for the downstream analysis to identify peaks and map PAS comprehensively.

3.6.2 Splicing and CPA are interconnected

Splicing and CPA are both co-transcriptional processing steps to mature pre-mRNA molecules by mega-Dalton, multimeric protein complexes. As the sizes of introns increased during evolution, splicing regulation switched from an intronic-based to an exon-based mode to define exons (Ram & Ast, 2007; McGuire et al., 2008). In this mode, exons are defined by a network of splicing factors containing U2, U1 and SR proteins that recognize 3' SS, downstream 5' SS and ESEs, spanning the complete exon. This raised the question how the 3' terminal exon is defined as it is not bound by 5' SS splicing factors, but rather by CPA factors in the 3' UTR. A potential interconnection between splicing and CPA was first described using an *in vitro* system showing that an upstream intron is necessary to stimulate CPA of labeled transcripts and that a crosstalk between splicing and CPA-factors might help to define the 3' terminal exon (Niwa et al., 1990; Berget, 1995). CPA factors, including CFIm, were found to be present in spliceosomal purifications further supporting the crosstalk theory (Rappsilber et al., 2002; Zhou et al., 2002). The importance of these interactions is supported by the fact that CFIm is absent in lower eukaryotes that use an intron-based splicing mechanism (Rüegsegger et al., 1998; Yang et al., 2010). In addition, U2AF65 interacts with another CPA factor, PAP, connecting splicing to CPA and activating splicing of the terminal intron (Vagner et al., 2000). The CFIm is not the only CPA factor interacting with splicing factors. CPSF, which binds to upstream sequence elements, was reported to interact with the U2 snRNP and depletion of CPSF resulted in decreased splicing of the upstream intron (Li et al., 2001a; Kyburz et al., 2006). Furthermore, CPSF160 and CPSF73 interact with the splicing co-activator SRm160 (McCracken et al., 2002). Usually SRm160 is involved in intron definition and its RNA binding is supported by its interaction with SR proteins, due to its own low RNA affinity (Eldridge et al., 1999; Longman et al., 2001). Interestingly, its non-specific PWI motif was shown to be necessary for the enhancement of CPA independent of its interaction with SR proteins (Szymczyna et al., 2003; McCracken et al., 2003). These summarized interactions of different splicing factors with two major CPA complexes indicate the presence of two pre-cleavage complexes interconnecting these processing steps (**Figure 11**).

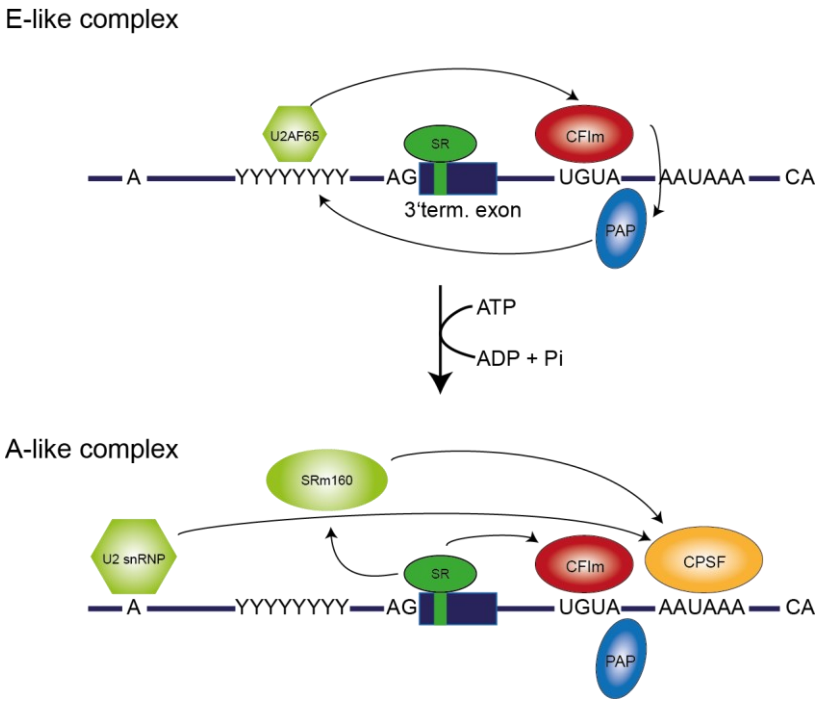


Figure 11: Splicing and alternative polyadenylation are closely connected to define the 3' terminal exon by interactions between several subunits. *First, an E-like complex consisting of U2AF65, bound to the poly-pyrimidine tract, and SR proteins, bound to ESEs, recruit the CFIm complex. In addition, PAP is recruited by CFIm and interacts with U2AF65. As a result, the complex actively transitions into an A-like complex. The U2 snRNP binds to the branch point and interacts with the CPSF complex. The splicing-coactivator SRm160 is stabilized by SR proteins and interacts with the CPSF complex as well. Modified from (Martinson, 2011).*

First, an E-like complex consisting of U2AF65 and distinct SR proteins interacts with CFIm and PAP to define the 3' terminal exon and to initiate CPA. Subsequently, the complex is activated and switches to an A-like complex replacing U2AF65 by the U2 snRNP and introducing SRm160 to interact with the CPSF complex to enhance CPA activity. Besides activating CPA, the crosstalk also enhances splicing of the upstream terminal intron. All the previously discussed connections between CPA and splicing were mediated via factors binding the 3' SS and defining the 3' -terminal exons. However, as described in previous chapters PAS can be located within introns (**Figure 10**A). Activation of those intronic PAS, leading to the expression of composite exons, is mediated via competition between splicing and CPA and more specifically a competition between factors binding the 5' SS and CPA factors binding the PAS for the upstream 3' SS (Tian et al., 2007). Weak 5' SS or high levels of CPA factors result in the activation of intronic PAS (Peterson & Perry, 1989; Takagaki et al., 1996). In contrast, on the one hand, strong binding of U1 snRNP to the 5' SS suppresses CPA by inhibiting PAP (Gunderson et al., 1998) and suppresses the usage of promotor-proximal PAS (Kaida et al., 2010). On the other hand, U1 activates proximal PAS, when its complementary to the 5' SS sequence is low (Berg et al., 2012). These examples illustrate that splicing and CPA are intimately linked to each other through the direct interactions between distinct *trans*-acting factors of both processes defining the 3' -terminal exon and

regulating the competition between splicing and intronic CPA. However, it remains unknown whether the regulation of APA by splicing factors is dependent on their splicing function.

3.6.3 Alternative polyadenylation is connected to nuclear mRNA export

In eukaryotic cells, the sites of transcription and translation are separated by the nuclear envelope. To be translated into functional proteins, mature mRNA needs to be exported from the nucleus into the cytoplasm. Macromolecules larger than 40 kDa, like mRNPs, need to pass through the nuclear pores in an active, energy-consuming manner. For this, mRNPs need to be compacted and bound by specific export adapters implying a complex signaling pathway to initiate nuclear export. Several proteins act as export adapters to recruit the NXF1:NXT1 complex to mature mRNAs, such as Exon Junction Complex (EJC) that is loaded on each splice junction upon successful ligation, or the TREX complex, ALYREF or SR proteins (Grüter et al., 1998; Zhou et al., 2000; Strässer et al., 2002; Cheng et al., 2006a; Müller-McNicoll et al., 2016). Besides its link to splicing, nuclear export is also connected to CPA. Correct polyadenylation is an important signal that a transcript is mature. The key regulator is PABPN1, which binds to the emerging poly(A) tail, stabilizing PAP and controlling the correct length of the poly(A) tail (Kühn & Wahle, 2004; Kühn et al., 2009). Furthermore, PABPN1 was shown to be necessary for the export of mRNPs and affects APA (Lemay et al., 2010; Jenal et al., 2012; Klerk et al., 2012). In addition, some export adapters interact directly with CPA factors. One of them is THOC5, a subunit of the TREX complex, which recruits CPSF-100 to certain target genes and activates the PAS nearby Tran et al., 2014. Furthermore, THOC5 was shown to interact with the CFIm (Katahira et al., 2013).This interaction is mainly mediated by the CFIm subunit CPSF6 and depletion of THOC5 resulted in shortening of targeted transcripts. Moreover, additional export factors have been associated with APA recently. One of them is CHTOP, a co-adaptor of the TREX mediated RNA-export complex that is enriched within the 3’ UTR promoting usage of the pPAS and decreasing usage of the pPAS (Viphakone et al., 2019). As THOC5 and CHTOP, two adaptors recruiting the export receptor NXF1, were shown to modulate APA, NXF1 itself was identified as an APA regulating factor (Chen et al., 2019). Decreased levels of NXF1 resulted in reduced transcriptional speed of PolII and increased usage of the pPAS. These shortened transcripts were shown to be exported independent of a CFIm-NXFI interaction. Vice versa, the cleavage factor CPSF6 was found to be an adaptor for NXF1 (Ruepp et al., 2009; Katahira et al., 2013). The CFIm-mediated export of transcripts via CPSF6 was shown to facilitate via NXF1 recruitment using a GAR motif in the linker domain between the N-terminal RRM and the C-terminal RS-like domain (**Figure 5**B) (Ruepp et al., 2009; Ruepp et al., 2011). These examples illustrate the connections of the different mRNA-processing steps within the life cycle of mRNAs. More and more of these tightly

controlled networks are discovered raising new questions about the underlying mechanisms that fine-tune gene expression and quality control of maturing transcripts.

3.6.4 RNA modifications and APA

Beside the major processing steps, mRNAs are also extensively modified at the single nucleotide level by various chemical modifications. The most prevalent modification in mammalian mRNAs is the methylation of adenosines at the 6' position of the amino group (m^6A). Methylations are controlled by three types of enzymes: writers, readers and erasers (**Figure 12**) (Meyer & Jaffrey, 2017).

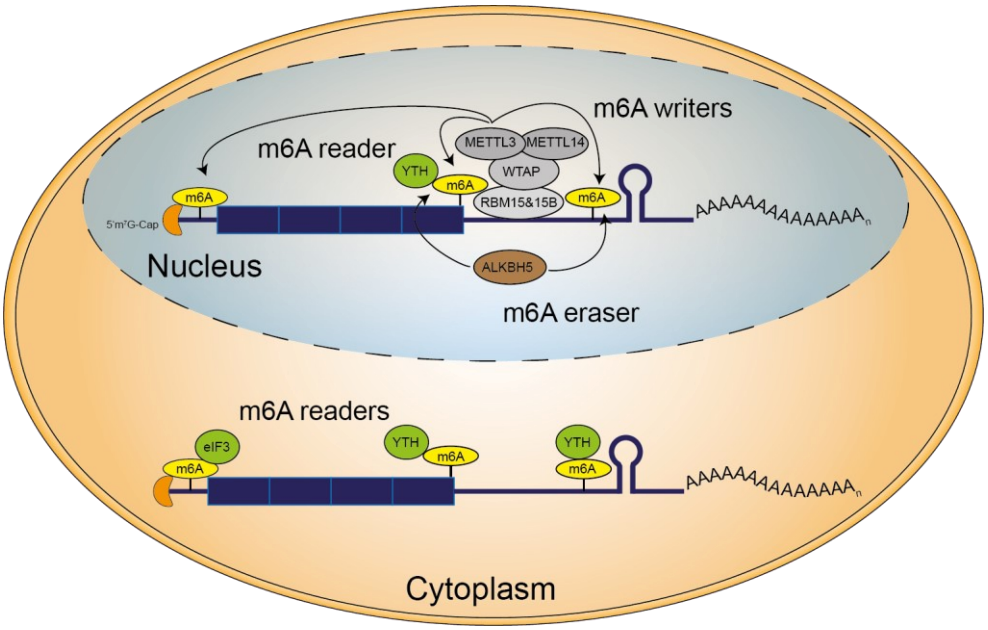


Figure 12: Scheme of the core factors and complexes involved in m⁶A methylation of mRNA. *Adenosine residues are methylated (yellow) in the nucleus by a complex of m⁶A writers (grey) consisting of specificity factors RBM15/15B, adapter protein WTAP, methyltransferase METTL3 and activity factorMETTL14. These modifications can be demethylated by the nuclear m⁶A eraser AKBH5 (brown). In the nucleus and the cytoplasm, the m⁶A-modification can be read by specific m⁶A reader proteins from the YTH family or eIF3 (green) influencing the fate of the modified transcript.*

m^6A modifications are added by a nuclear, multimeric complex – the m^6A writers complex – consisting of four major proteins. RBM15 and its homologue RBM15B mediate specificity to the complex by binding to U-rich regions in close proximity of the modified nucleotide and with the important hub protein WTAP, which interconnects all m^6A writer subunits (Agarwala et al., 2012; Horiuchi et al., 2013; Patil et al., 2016). The modification is introduced by a dimer consisting of METTL3 and METTL14, where METTL3 is the active SAM-dependent methyltransferase and METTL14 is a necessary adapter (Bokar et al., 1997; Zhong et al., 2008; Ping et al., 2014; Liu et al., 2014; Śledź & Jinek, 2016; Wang et al., 2016). Interestingly, METTL3 modifies nearly all RNAs, but not the U6 snRNA and ribosomal RNAs (Shimba et al., 1995). This is achieved by METTL16, another member of the methyltransferase family (Pendleton et al., 2017). The methyl-modification can be removed by a m^6A

eraser, ALKBH5, which is highly specific to m^{6}A (Zheng et al., 2013; Mauer et al., 2017). Interestingly, ALKBH5 localizes to the nucleus meaning that the adenosine modifications are set and erased only in the nucleus during the biogenesis of mRNA. The modified adenosine residues can be recognized by several proteins, the m^{6}A readers, which then influence the subsequent fate of the respective mRNA. The main proteins recognizing m^{6}A-modification belong to the YTH family (Dominissini et al., 2012). This family consists of five proteins sharing the defining YTH-domain and a highly disordered remaining domain. These proteins can be further separated into the cytosolic DF proteins termed YTHDF1-3, the nucleoplasmic YTHDC2 and the nuclear YTHDC1. YTHDF1/2 and YTHDC1 have been figured out to be the only physiological active readers with known functions so far (Patil et al., 2016). YTHDF1 is able to recruit the translation initiation factor eIF3 (Wang et al., 2015), while YTHDF2 was connected to decreased mRNA stability by mediating interactions with P-bodies that are important locations for mRNA degradation (Wang et al., 2014c). A bit later all three YTHDF proteins have been implicated to function more redundantly, reducing the half-life time of mRNAs (Du et al., 2016). The nuclear YTHDC1 was first known to be an alternative splicing factor (Zhang et al., 2010). The association between the YTH domain and the m^{6}A-modification led to the finding that YTHDC1 is a m^{6}A-dependent factor regulating alternative splicing (Xiao et al., 2016). Interestingly, eIF3 was shown to interact with m^{6}A located in the 5’ UTR to promote cap-independent translation (Meyer et al., 2015).

Extensive research and high-throughput technologies have found m^{6}A modifications to be enriched in 3’ terminal exons and 3’ UTRs indicating potential functions in pre-mRNA processing or gene expression regulation (Batista et al., 2014; Ke et al., 2015; Molinie et al., 2016; Bartosovic et al., 2017). For example the artificial reduction of m^{6}A resulted in a slower nuclear export rate of affected transcripts (Camper et al., 1984; Fustin et al., 2013; Zheng et al., 2013). Accordingly, an interaction between the m^{6}A writer complex and several subunits of the TREX complex (ALYREF, UAP65, THOC5 and CHTOP) was published (Lesbirel et al., 2018). Interactions with other mRNA-processing factors were also described. Especially the m^{6}A reader YTHDC1 was shown to have various functions in connecting methylation to other RNA metabolic processes functioning as an adapter protein. For example, YTHDC1 was found to recruit SRSF3 to methylated exons and thereby facilitated their inclusion. Simultaneously it was blocking SRSF10, which would result in the skipping of these exons (Xiao et al., 2016). The interaction between YTHDC1 and SRSF3 was also shown to increase export of target transcripts by the subsequent recruitment of NXF1 via SRSF3 (Roundtree et al., 2017b). Furthermore, enrichment of m^{6}A in long last exons was shown to influence APA. Curiously, two independent studies provided opposing evidence that m^{6}A enrichments increased or decreased the usage of pPAS (Ke et al., 2015; Molinie et al., 2016). This suggests that m^{6}A-based regulation might be different for specific targets, tissues, cell-types, or developmental stages and likely depend on different

trans-acting factors that interact with these reader proteins. Further research will help to disentangle the crosstalk between the methylation machinery, modifications, and mRNA-processing steps.

3.7 SRSF3 and SRSF7: Two close, but different siblings and their roles in connecting pre-mRNA splicing to APA and mRNA export

So far, it became clear that SR proteins have multiple additional functions beyond their essential functions in the regulation of alternative splicing. The core SR protein family consists of 12 members, structurally related proteins that share redundant functions in constitutive splicing. Early research showed that SR proteins can replace each other for constitutive splicing. However, in recent years it became clear that individual SR proteins also perform unique functions from alternative splicing to other steps in the mRNA life cycle. Two very interesting SR proteins are SRSF3 and SRSF7. SRSF3 is the smallest of all SR proteins and its RS domain makes up nearly 50% of the total protein. SRSF7 is the only SR protein that contains an additional CCHC-type Zinc-knuckle in front of its RS domain (**Figure 7**). Interestingly, both proteins cluster together and form their own sub-family when compared to all other SR proteins, based on the human protein sequences (**Figure 13**) (Busch & Hertel, 2012). This is mostly due to their very similar RRM and indicates that they have a common origin. Both proteins are oncogenic proteins (Fu & Wang, 2018; DeLigio et al., 2019; Jia et al., 2019) and share a few similarities, but they also display major differences with SRSF3 being much more studied than SRSF7. Apart from being important regulators of various RNA-processing events in the nucleus, both proteins have in common that they both shuttle robustly between nucleus and cytoplasm, and efficiently recruit NXF1 to their target transcripts for nuclear export (Müller-McNicoll et al., 2016; Botti et al., 2017). Their function in export was first described for intron-less histones, where it was shown that the NXF1 recruitment is mediated by the linker domain of de-phosphorylated SR proteins (Huang & Steitz, 2001; Huang et al., 2003; Huang et al., 2004; Hargous et al., 2006). In addition, both SR proteins were linked with the export of herpes simplex virus 1 transcripts (Escudero-Paunetto et al., 2010). SRSF3 and SRSF7 were also associated with translation regulation. SRSF3 was shown to initiate internal ribosome entry side (IRES) mediated translation of viral RNAs (Bedard et al., 2007), but was also shown to repress translation of the

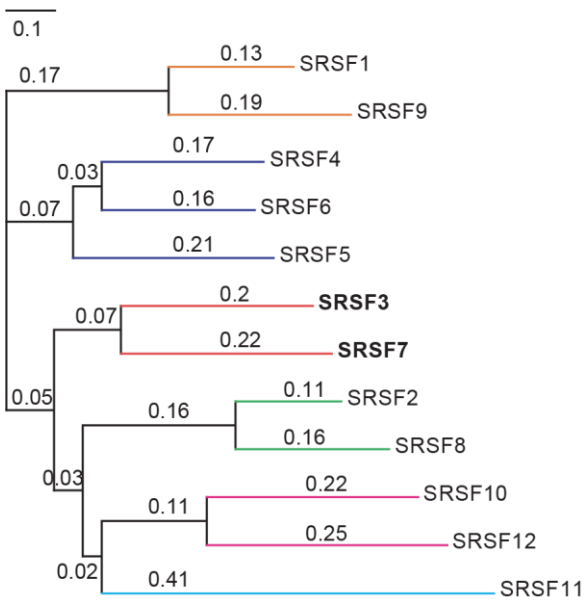


Figure 13: Phylogenetic tree of all 12 human SR proteins representing evolutionary distance and sub-clusters of distinct proteins. *The numbers on the bars indicate the similarity score. The SRSF3/7 sub-cluster is indicated in bold and a red line. Modified from (Busch & Hertel, 2012).*

programmed cell death 4 gene (Kim et al., 2014; Park & Jeong, 2016). SRSF7 was shown to initiate translation of un-spliced viral RNAs containing constitutive transport elements (CTEs) and reporter genes (Swartz et al., 2007; Mo et al., 2013).

Despite their common functions, the RNA binding motifs, binding patterns, and RNA targets of SRSF3 and SRSF7 are different. Using individual nucleotide UV-crosslinking and immuno-precipitation (iCLIP) it was found that SRSF3 prefers binding to a pyrimidine-rich 4-mer (CNYC), while SRSF7 prefers binding to a purine-rich sequence (GAYGAY) (Müller-McNicoll et al., 2016). RNA binding specificity of SRSF3 is exclusively mediated by its RRM, while for SRSF7 the RRM and the Zn-knuckle contribute to its distinct specificity, whereby the Zn-knuckle mediates binding to purines (Cavaloc et al., 1999; Königs et al., 2020). SRSF3 was shown to depend on the CTD of PolII to be active in alternative splicing (La Mata & Kornblihtt, 2006). Furthermore, SRSF3 and SRSF7 are both able to regulate splicing of their own mRNA, and auto-regulate their protein levels. They promote the inclusion of a highly conserved PCE, which leads to the rapid degradation of the transcript by NMD and reduced protein levels (Lareau et al., 2007; Ni et al., 2007; Änkö et al., 2012; Königs et al., 2020).

SRSF3 is also known for its large interaction network with auxiliary RBPs and processing factors, such as *trans*-factors chemically modifying nucleotides. The most enriched modification found in RNAs is the methylation of adenosines at the nitrogen-6 position, termed m^6A, which is directed by specific methyltransferases and can be recognized by special m^6A-reading proteins. Recently, it was shown that SRSF3 is recruited by the m^6A-reader YTHDC1 to exons that contain methylated adenosines (m^6A) either to include them into the mature transcript (Xiao et al., 2016) or to promote the recruitment of NXF1 for a rapid export of m^6A-containing transcripts (Roundtree et al., 2017a). Furthermore, SRSF3 was connected to the nuclear exosome via the Nuclear Exosome Targeting (NEXT) complex to initiate degradation of intron-less Epstein Barr Virus mRNAs (Mure et al., 2018).

Most interestingly, SRSF3 and SRSF7 have been suggested to be involved in alternative 3' end processing. It was first described in 1998 that SRSF3 regulates the tissue-specific alternative polyadenylation (APA) of the human *calcitonin/calcitonin gene-related peptide* (*CT/CGRP*) gene (Lou et al., 1998). SRSF7 was shown to be involved in the 3' end processing of human immunodeficiency virus 1 (HIV1) mRNA (Valente et al., 2009). Extensive research by the McNally laboratory identified the splicing-independent involvement of all core SR proteins, except SRSF4, in the activation of a weak PAS of the mostly unspliced Rous Sarcoma Virus transcript(RSV)(McNally & McNally, 1996). The PAS is weakened as the DSEs are missing and, in addition, the USEs are weak (Maciolek & McNally, 2008). Interestingly, despite the weak PAS *in vivo* analysis still revealed a high level of CPA (~85%), while the usage *in vitro* was decreased until stimulation by the NRS or artificial SR protein binding sites (Maciolek & McNally, 2007, 2008). Efficient CPA is mediated by an activation of the PAS via a network of SR

proteins juxtapositioning two certain regions, the negative regulator of splicing (NRS) and the 3' SS complex at the *env* enhancer, and subsequent promotion of CFIm binding (Hudson et al., 2016). The NRS is enriched in strong SR protein-binding sites tethering these proteins there, circumventing efficient splicing, yet favoring cleavage and polyadenylation (McNally & McNally, 1996, 1998). Experiments showed that the NSR can be replaced by SRSF1 and SRSF3 binding sites, however the CPA activation was weaker compared to the native NRS, which contains binding sites for more than just these two SR proteins, indicating the importance of other SR protein family members activating the RSV's PAS (McNally & McNally, 1996). Interestingly, recent research strengthened the potential activation of the weak PAS via a SRSF1 or SRSF7-dependent recruitment of CFIm, when the RSV substrate was enriched in artificial SRSF1 and SRSF7 binding sites (Hudson et al., 2016). Yet, most of this research focused on SRSF1, indicating a rather indirect role of this particular SR protein in the activation of the RSV's PAS, including other SR proteins like SRSF7 as a side-notion (Hudson et al., 2016). Therefore, the potential direct function of SRSF7 in recruiting CFIm (Dettwiler et al., 2004) to the weak PAS needs further research to be understood in detail.

Furthermore, recently a physiological connection between SRSF3-knockdown and senescence phenotype was described by Shen et al., 2019. They show that depletion of SRSF3 increases the usage of pPASs leading to the expression of shorter transcript isoforms. The study found that affected transcripts enrich for senescence-associated pathways. Consequently, depletion of SRSF3 related to senescence-related phenotypes, which was traced back to increased protein levels and activities of those shortened transcripts e.g., PTEN. The underlying molecular mechanism remains unclear, but Shen and colleagues speculate that the preferred binding of SRSF3 at pPASs might inhibit their activation, while depletion of SRSF3 renders their accessibility to facilitate CPA at pPASs.

In recent years, more and more articles were published connecting individual SR proteins with functions in alternative 3' end processing and involved proteins (Dettwiler et al., 2004; Kasowitz et al., 2018; Shen et al., 2019) indicating specific protein-protein interactions via RS domains. A comprehensive study from our group investigated alternative 3' end processing after depletion of individual SR proteins. Individual SR proteins were depleted by RNAi and total RNA subjected for RNA-Sequencing. Subsequently, APA events were analyzed using the MISO pipeline, which contains a defined set of known APA targets. These analyses revealed that all SR proteins showed stronger or weaker effects in APA on distinct subsets of targets, increasing or decreasing pPAS usage. However, interestingly, SRSF3 and SRSF7 emerged as the most potent regulators of alternative 3' end processing in opposite directions (Müller-McNicoll et al., 2016). The underlying mechanism and interactors are so far unknown and need further investigation.

3.8 Objectives of study

SRSF3 and SRSF7 have been found to regulate alternative 3' end processing in opposite directions and connect this to mRNA export. This study was conducted in pluripotent P19 cells. However, it is not clear what the extent of this regulation in P19 cells is, how these two splicing factors manage to affect alternative polyadenylation, whether this function is dependent on splicing, how the opposing outcome of APA is achieved, which role m^6A modifications might play in this interplay and whether SRSF3 and SRSF7 also affect APA in differentiated cells. For cleavage and polyadenylation, four major complexes consisting of more than 20 proteins are well described in their function and their targeted *cis*-elements are highly conserved. Additional *trans*-acting factors were found to direct the CPA machinery and to trigger activation of alternative PAS. Yet, the regulation of the usage of alternative, upstream proximal PAS is less well understood. This study aims to identify a distinct mechanism of regulation of the 3' UTR length by SR proteins. The objectives were as following:

1. Characterize the global PA-tome in P19 cells for the first time.
2. Identify the global extent of depletion of SRSF3 and SRSF7 on PAS usage.
3. Identify binding patterns of SRSF3 and SRSF7 around PAS in affected and unaffected transcripts.
4. Clarify the splicing-dependency of SRSF3 and SRSF7 on the regulation of CPA.
5. Identify and characterize potential interactions between SRSF3, SRSF7 and core CPA factors.
6. Identify the individual mechanisms by which SRSF3 and SRSF7 actively influence CPA of targeted PAS.
7. Analyze the extent of expression of SRSF3 and SRSF7 affecting APA during neuronal differentiation of P19 cells.
8. Formulate a mechanistic model on how SRSF3 and SRSF7 regulate 3' UTR-APA.

4 Material and methods

4.1 Equipment

All equipment used to perform this study is summarized in Table 2 including the name, vendor and catalogue number.

Table 2: List of used equipment stating the intended usage, catalogue number and vendor.

Name	Usage	Catalogue number	Vendor
102-C Converter	Converter for sonification		Branson
12-Tube Magnetic Separation Rack	Magnetic rack	S1509S	New England Biolabs Inc.
2100 Bioanalyzer	Automated electrophoresis	G2939BA	Agilent
24x1.5/2.0mL Rotor with ClickSeal	Rotor	10544733	Fisher Scientific
450 Digital Sonifier	Sonifier	B450	Branson
Accu-jet pro	Pipette controller	26300	Brand
AE31	Inverted microscope		Motic
BlueLight Table	LED Transilluminator	LED001	Serva
Bomann MWG 2211 U CB	Microwave	622111	
CanoScan 9000F Mark II	Scanner		Canon
ChemiDoc XRS+	Gel imaging system	1708265	Bio-Rad
CL-1000 Ultraviolet Crosslinker	UV crosslinker	732-4302	VWR
Cover glass, Ø13 mm, No. 1	Cover slips	631-0149	VWR
CP 3813	Fridge and freezer		Liebherr
Digital Heatblock	Heatblock		VWR
EMB 6000-1	Precision balance		Kern & Sohn GmbH
Eppendorf 5430 R	Centrifuge	521-2646	VWR
Eppendorf F-35-6-30	Rotor	05-401-512	Fisher Scientific
Eppendorf Mastercycler nexus Gradient	PCR cycler	71003-562	VWR
Eppendorf Mastercycler nexus X2e	PCR cycler	10119-850	VWR
Eppendorf ThermoMixer C	Thermomixer	5382000252	Eppendorf
Eppendorf Thermomixer comfort	Thermomixer		VWR
Evolution 60	UV-Visible Spectrophotometer		Thermo Scientific
FE20 Benchtop pH Meter	pH meter	10526655	Fisher Scientific
G 7883 CD	Dishwasher		Miele Professional
GCUC100	Fridge		Gastro-Cool

GENi	Gel documentation system		Syngene
GP 1376	-20°C freezer		Liebherr
GP 4013	-20°C freezer		Liebherr
Heracell 150i	CO_2 incubator	50116047	ThermoFischer Scientific
HERAfreeze HFU T Series	-80°C freezer	11650823	Fisher Scientific
Herasafe KSP	Safety cabinet	10123603	Fisher Scientific
Ice Maker	Ice machine	FM-80KE	Hoshizaki
IKA KS 4000 i control	Incubating shaker	444-0298	VWR
INCU-Line IL23	Digital incubator	390-0482	VWR
Laboport® N86 KN.18	Mini vacuum pump	AP86	A.Hartenstein GmbH
LSM780	Confocal laser scanning microscope		Zeiss
Mercury lamp power supply	Power supply	MHG-100B	Motic
Micro Star 17	Microcentrifuge	521-1646	VWR
Micro Star 17R	Microcentrifuge	521-1647	VWR
Mini Blot Mixer	Orbital shaker	95057-436	VWR
Mini-PROTREAN Short Plates	Gel casting system	1653308	Bio-Rad
Mini-PROTREAN Spacer Plates with 1.0 mm Integrated Spacers	Gel casting system	1653311	Bio-Rad
Mini-PROTREAN Tetra Cell Casting Stand Clamps	Gel casting system	1658050	Bio-Rad
Mini-PROTREAN Tetra Vertical Electrophoresis Cell	Electrophoresis system	1658004	Bio-Rad
Mitsubishi K65HM-CE	High density paper	KP65HM-CE	Griebel Medizin- und Modelltechnik
Mitsubishi P95	Digital printer	P95DE	Griebel Medizin- und Modelltechnik
Moticam 3	Camera		Motic
MS 3 basic vortex mixer with bioanalyzer chip	Automated electrophoresis	IKAA3617036	VWR
Nanodrop 2000	Spectrophotometer		Eppendorf
neoLab 3-Speed Mini Centrifuge	Microcentrifuge	D-6015	neolab
neoLab Vortex Mixer	Vortex mixer	7-2020	neolab
neoMag Digital Magnetic Stirrer with heater	Magnetic stirrer	D-6010	neolab
neoMag Magnetic stirrer without heater	Magnetic stirrer	D-6011	neolab
PCB 250-3	Precision balance		Kern & Sohn GmbH
PikoReal 96	Real-time PCR system	12675885	Fisher Scientific
PlateFuge	Microcentifuge	C2000*	Benchmark Scientific

PowerPac Basic	Power supply	1645050	Bio-Rad
Precision cover slips, Ø12 mm	Cover slips	YX02.1	Carl Roth
Rocking Platform	Rocking platform		VWR
RS-DS 5	Orbital shaker		Phoenix Instrument
Stuart SB2	Rotator	445-2101	VWR
Sunlab Mini Centrifuge	Microcentrifuge	D-8550	neolab
Sunlab mini Vortex mixer	Vortex mixer	D-8900	neolab
Systec VX-150	Autoclave		Systec
Thermomixer MT-100	Thermomixer		Universal Labortechnik
Typhoon Phospho Imager	Phospho Imager		GE Healtchare
Vortex mixer	Vortex mixer	444-1372	VWR
VWB2 12	Water bath	462-0557	VWR
VWR® Microscopy Slides, Cut Colour Frosted Yellow	Microscopy slides	631-1557	VWR
VWR® Microscopy Slides, Ground Edges Frosted	Microscopy slides	631-1553	VWR
XCell SureLock and XCell II Blot module	Electrophoresis system	EI0002	ThermoFisher Scientific

4.2 Chemicals and reagents

All chemicals, reagents and biologicals used to perform this study are summarized in **Table 3** including the name, vendor, and catalogue number.

Table 3: List of used chemicals, solutions, biologicals, and plastic ware stating name, catalogue number and vendor. *Sorted by functional categories.*

Name	Catalogue number	Vendor	Comment
Chemicals/Solutions			
0.5 M EDTA pH 8.0	AM9260G	ThermoFischer Scientific	
1 M $MgCl_2$	AM9530G	ThermoFischer Scientific	
10 mM dNTP Mix	R0192	ThermoFischer Scientific	
2-Mercaptoethanol	M6250	Sigma-Aldrich	
2-Propanol	I9516	Sigma-Aldrich	
2X RNA loading dye	R0641	ThermoFischer Scientific	
6X Orange DNA loading dye	R0631	ThermoFischer Scientific	
Acetone	179973	Sigma-Aldrich	
Agarose	A6539	Sigma-Aldrich	
Ambion® Phenol:Chloroform:Isoamyl alcohol (25:24:1), pH 6.7	AM9722	ThermoFischer Scientific	
Ambion® RNase I	AM2295	ThermoFischer Scientific	
Ambion® TRIzol® reagent	15596018	ThermoFischer Scientific	
AmershamECL™ Primer Western blotting detection reagents	RPN2232	Sigma-Aldrich	
Amersham™ Protran™ 0.1 µm nitrocellulose	10600000	Sigma-Aldrich	
Amersham™ Protran™ 0.45 µm nitrocellulose	10600002	Sigma-Aldrich	
Ampicillin sodium salt	A9518	Sigma-Aldrich	
Amresco® Agar, bacteriological	J637	VWR	
Amresco® HEPES free acid	0511	VWR	
Amresco® Ponceau S stain	K793	VWR	
Amresco® TBE (Tris-Borate-EDTA) buffer powder	0478	VWR	
Amresco® TWEEN® 20	M147	VWR	
Amresco® X-Gal	0428	VWR	
Bad Stabil®	1-6095	neoLab	
Bicine	B3876	Sigma-Aldrich	
Bis-Tris	B9754	Sigma-Aldrich	
Boric acid	B6768	Sigma-Aldrich	
Bovine serum albumin	A3059	Sigma-Aldrich	
Bromphenol blue	15375	Serva	
Calcium chloride	CN93.2	Carl Roth	

Chloramphenicol	C0378	Sigma-Aldrich	
cOmplete™ EDTA-free protease inhibitor cocktail	5056489001	Sigma-Aldrich	
Corning® Costar® Spin-X® centrifuge tube filters	CLS8161	Sigma-Aldrich	
Cycloheximide from microbial source	C7698	Sigma-Aldrich	
dATP solution (100 mM)	R0141	ThermoFischer Scientific	
dCTP solution (100 mM)	R0151	ThermoFischer Scientific	
dGTP solution (100 mM)	R0161	ThermoFischer Scientific	
DL-Dithiothreitol (DTT)	43819	Sigma-Aldrich	
Dodecylsulfate (SDS) Na salt	20765	Serva	
DTT	R0861	ThermoFischer Scientific	
dTTP solution (100 mM)	R0171	ThermoFischer Scientific	
Dulbecco's Phosphate Buffered Saline (PBS)	D8537	Sigma-Aldrich	
Effectene® transfection reagent	301425	QIAGEN GmbH	
Epicentre® CircLigase™ II	131406	Biozym Scientific GmbH	
Ethanol (EtOH) 70% (v/v) Euro denatured	85825.440	VWR	
Ethanol absolute	20821.321	VWR	
Ethylenediaminetetetraacetic acid (EDTA)	E9884	Sigma-Aldrich	
Fisherbrand™ Badstabilisator	10707061	ThermoFischer Scientific	discontinued
GammaBind™ G Sepharose™	17-0885-01	Sigma-Aldrich	
Gibco™ HEPES (1 M)	15630080	ThermoFischer Scientific	
Glycerol	G2025	Sigma-Aldrich	
Glycerol	6967.1	Carl Roth	
Glycine	10119CU	VWR	
Hydrochloric acid	H1758	Sigma-Aldrich	
IGEPAL® CA-630	I8896	Sigma-Aldrich	
Invitrogen™ DNAzol™ Reagent	10503027	ThermoFischer Scientific	
Invitrogen™ 0.1 M DTT	Y00147	ThermoFischer Scientific	
Invitrogen™ 15% TBE-Urea gel	EC68852BOX	ThermoFischer Scientific	
Invitrogen™ 6% TBE-Urea gel	EC68652BOX	ThermoFischer Scientific	
Invitrogen™ Dynabeads™ Protein G	10004D	ThermoFischer Scientific	
Invitrogen™ 1 M Tris pH 7.0	AM9850G	ThermoFischer Scientific	
Invitrogen™ 1 M Tris pH 8.0	AM9855G	ThermoFischer Scientific	
Invitrogen™ 10X TBE buffer	AM9863	ThermoFischer Scientific	
Invitrogen™ 3 M Sodium acetate pH 5.5	AM9740	ThermoFischer Scientific	
Invitrogen™ 5 M NaCl	AM9760G	ThermoFischer Scientific	

Invitrogen™ GlycoBlue™ ™ Coprecipitant (15 mg/mL)	AM9515	ThermoFischer Scientific	
Invitrogen™ ProLong™ Diamond Antifade Mountant	P36965	ThermoFischer Scientific	
Invitrogen™ SYBR™ Gold Nucleic Acid Gel Stain	S11494	ThermoFischer Scientific	
Invitrogen™ UltraPure™ 1 M Tris-HCl Buffer, pH 7.5	15567027	ThermoFischer Scientific	
IPTG, dioxane-free	R0392	ThermoFischer Scientific	
Isopropanol, technical	444250050	ThermoFischer Scientific	
jetOPTIMUS® DNA transfection reagent	117-15	VWR	
jetPRIME® transfection reagent	114-15	VWR	
Kanamycin sulfate from Streptomyces kanamyceticus	K40000	Sigma-Aldrich	
LB Broth	L3022	Sigma-Aldrich	
Lipofectamine™ 2000	11668019	ThermoFischer Scientific	
Lithium chloride	L121	ThermoFischer Scientific	
Lithium dodecyl sulfate	L9781	Sigma-Aldrich	
Magnesium chloride	M33	ThermoFischer Scientific	
Magnesium chloride solution (1 M)	A3888	AppliChem	discontinued
Methanol (MeOH)	20847.320	VWR	
Micro Bio-Spin™ columns	7326008	Bio-Rad	
Milk powder	A0830	AppliChem	
N,N,N',N'-Tetramethylethylene-diamine (TEMED)	T9281	Sigma-Aldrich	
NAD	A1124	AppliChem	
Nonidet™ P 40 Substitute	74385	Sigma-Aldrich	
Novex® TBE-Urea sample buffer (2X)	LC6876	ThermoFischer Scientific	
NuPAGE® LDS sample buffer (4X)	NP0007	ThermoFischer Scientific	
NuPAGE® MOPS SDS running buffer (20X)	NP0001	ThermoFischer Scientific	
NuPAGE® sample reducing agent (10X)	NP0009	ThermoFischer Scientific	
NuPAGE® transfer buffer (20X)	NP0006-1	ThermoFischer Scientific	
NuPAGE™ 4-12% Bis-Tris gels, 10 well	NP0321BOX	ThermoFischer Scientific	
NuPAGE™ 4-12% Bis-Tris gels, 12 well	NP0322BOX	ThermoFischer Scientific	
Invitrogen™ Novex™ TBE-Urea Gels, 6%, 10 well	EC6865BOX	ThermoFischer Scientific	
Invitrogen™ Novex™ TBE Gels, 6%, 10 well	EC6265BOX	ThermoFischer Scientific	

NuPAGE™® antioxidant	NP0005	ThermoFischer Scientific	
Paraformaldehyde	158127	Sigma-Aldrich	
Paraformaldehyde	P6148	Sigma-Aldrich	
Paraformaldehyde	A3813	AppliChem	
Phase Lock Gel™ Heavy	733-2478	VWR	
Phosphate buffered saline (10X) (PBS)	P5493	Sigma-Aldrich	
PIPES	P6757	Sigma-Aldrich	
Polyethylene glycol 8000	P2139	Sigma-Aldrich	
Polyethylene glycol 1500	11775800	Sigma-Aldrich	
Polyethylene glycol 400	81172	Fluka® Analytical	
Potassium chloride	P017.1	Carl Roth	
Potassium chloride	P217-3	ThermoFischer Scientific	
Potassium hydroxide	P747.1	Carl Roth	
Pufferan® HEPES	9105.3	Carl Roth	
Pufferan® Tris	4855.2	Carl Roth	
Q Sepharose™ FastFlow	17-0510-10	Sigma-Aldrich	
Quick Start™ Bradford 1x dye reagent	500-0205	Bio-Rad	
RedSafe™ Nucleic Acid Staining Solution	21141	HiSS Diagnostics GmbH	
RNaseZAP™	R2020	Sigma-Aldrich	
Rotiphorese® Gel 30 (37.5:1)	3029.5	Carl Roth	
Rotisolv® Trichloromethane/Chloroform	T901.1	Carl Roth	
SDS solution 10% (w/v)	161-0416	Bio-Rad	
Sodium acetate	S2889	Sigma-Aldrich	
Sodium chloride	S7653	Sigma-Aldrich	
Sodium deoxycholate	D6750	Sigma-Aldrich	
Sodium dodecyl sulfate (SDS)	L6026	Sigma-Aldrich	
Sodium hydroxide	6771.1	Carl Roth	
T4 Polynucleotide Kinase	M0201	New England Biolabs Inc.	
T4 RNA Ligase	M0204	New England Biolabs Inc.	
T5 Exonuclease	M0363	New England Biolabs Inc.	
Taq DNA Ligase	M0208	New England Biolabs Inc.	
Thermo Scientific™ Proteinase K, recombinant, PCR grade	EO0491	ThermoFischer Scientific	
Tris-EDTA buffer solution	93302	Sigma-Aldrich	
Triton X-100	3051.4	Carl Roth	
Triton® X-100	H5142	Promega	

Trizma® base	T1503	Sigma-Aldrich	
Trizma® base	T6066	Sigma-Aldrich	
Urea	108487	Merck Millipore	
Water	W4502	Sigma-Aldrich	
Whatman® glass microfiber filters, Grade GF/D	WHA1823010	Sigma-Aldrich	
X-Gal	XG1134C433	VWR	discontinued
β-Glycerophosphate disodium salt hydrate	G5422	Sigma-Aldrich	
γ-P^{32}ATP, 800 Ci/mM, 10 mCi/mL	SRP-801	Hartmann Analytic	
Cell culture media/Solutions			
Cycloheximide solution, 100 mg/mL in DMSO	C4859	Sigma-Aldrich	
Cytosine β-D-arabinofuranoside hydrochloride (Ara-C)	C6645	Sigma-Aldrich	
Dimethyl sulfoxide (DMSO)	D2650	Sigma-Aldrich	
Dulbecco's Phosphate Buffered Saline	D8537	Sigma-Aldrich	
Fibroblast Growth Factor (FGF-8), human	SRP4053	Sigma-Aldrich	
Gelatin solution, Type B, 2% in H_2O	G1393	Sigma-Aldrich	
Gibco™ B-27™ Supplement (50X), serum free	17504044	ThermoFischer Scientific	
Gibco™ CTS™ Neurobasal® Medium	A1371201	ThermoFischer Scientific	
Gibco™ DMEM, high glucose, GluataMAX™ Supplement, pyruvate	31966047	ThermoFischer Scientific	
Gibco™ DMEM/F-12, GlutaMAX™ Supplement	31331028	ThermoFischer Scientific	
Gibco™ Fetal Bovine Serum, heat inactivated	10500064	ThermoFischer Scientific	
Gibco™ N-2 Supplement (100X)	17502048	ThermoFischer Scientific	
Gibco™ Opti-MEM, Serum Reduced Medium	31985070	ThermoFischer Scientific	
Gibco™ Penicillin-Streptomycin (10,000 U/mL	15140122	ThermoFischer Scientific	
Gibco™ Trypsin-EDTA (0.25%), phenol-red	25200054	ThermoFischer Scientific	
Laminin from Engelbreth-Holm-Swarm murine sarcoma basement membrane	L2020	Sigma-Aldrich	
Retinoic acid	R2625	Sigma-Aldrich	
γ-Secretase Inhibitor IX (DAPT)	D5942	Sigma-Aldrich	

Antibodies			
Mouse anti-7B4 (SRSF3)	-/-	Neugebauer, Dresden (Germany)	1:2 in 3% BSA
Rabbit anti-Beta catenin (CTNNB)	ab2365	Abcam	1:1000 in 3% BSA
Mouse anti-Nudt21 (CPSF5)	sc-81109	Santa Cruz Biotechnology	1:500 in 3% BSA
Rabbit anti-CPSF6	ab99347	Abcam	1:1000 in 3% BSA
Mouse anti-FIP1	sc-398392	Santa Cruz Biotechnology	1:1000 in 3% BSA
Rabbit anti-SFRS7	C18943	Assay biotech	1:1000 in 3% BSA
Goat anti-GFP		D. Drechsel, MPI Dresden	1:3000 in 3% BSA
Rabbit anti-PABPN1	ab75855	Abcam	1:1000 in 3% BSA
Rabbit anti-mCherry	PA5-34974	ThermoFischer Scientific	1:1000 in 3% BSA
Rabbit anti-Tet-Repressor	TET01	MoBiTec	1:1000 in 5% milk
Mouse anti-mAb104	-/-	Neugebauer, Dresden (Germany)	1:3 in 3% BSA; IgM
Mouse anti-GAPDH	sc-32233	Santa Cruz Biotechnology	1:400 in 3% BSA
Rabbit anti-alpha Tubulin	ab176560	Abcam	1:1000 in 3% BSA
Goat anti-mouse-IgG	A9917	Sigma-Aldrich	
Donkey anti-mouse-IgG-HRP	AP192P	Sigma-Aldrich	1:10.000 in 3% BSA, secondary antibody
Donkey anti-rabbit-HRP	AP182P	Merck Millipore	1:10.000 in 3% BSA, secondary antibody

Donkey anti-goat-HRP	AB324P	Sigma-Aldrich	1:10.000 in 3% BSA, secondary antibody
Goat anti-mouse-IgM-HRP	A8786	Sigma- Aldrich	1:10.000 in 3% BSA, secondary antibody
Restriction enzymes			
*Bam*HI-HF	R3136L	New England Biolabs Inc.	
*Kpn*I-HF	R3142L	New England Biolabs Inc.	
*Nhe*I-HF	R3131L	New England Biolabs Inc.	
*Xma*I-HF	R0180L	New England Biolabs Inc.	
Polymerases			
ALLin™ RPH Polymerase, 5 u/µL	HLE0101	highQu GmbH	
HiScribe™ T7 High Yield RNA Synthesis Kit	E2040S	New England Biolabs Inc.	
Invitrogen™ AccuPrime™ SuperMix I	12342010	ThermoFischer Scientific	
Invitrogen™ SuperScript™ III Reverse Transcriptase	18080044	ThermoFischer Scientific	
Invitrogen™ SuperScript™ IV Reverse Transcriptase	18090010	ThermoFischer Scientific	
ORA™ qPCR Green ROX H Mix, 2X	QPD0205	highQu GmbH	
Phusion High Fidelity DNA Polymerase	M0530L	New England Biolabs Inc.	
Q5 High Fidelity DNA Polymerase	M0491L	New England Biolabs Inc.	
Taq DNA Polymerase	M0267L	New England Biolabs Inc.	
DNases/RNases			
Invitrogen™ Ambion™ RNase I, cloned, 100 U/µL	AM2295	ThermoFischer Scientific	
Invitrogen™ TURBO™ DNase (2 U/µL)	AM2238	ThermoFischer Scientific	
Lucigen RNase A, 5 µg/µL	160630	Biozym Scientific GmbH	
RNase III		MPI-CBG, Dresden	
DNA and protein ladders			
Thermo Scientific™ GeneRuler 1 kb DNA Ladder, ready-to-use	SM0314	ThermoFischer Scientific	
Thermo Scientific™ GeneRuler 1 kb Plus DNA Ladder, ready-to-use	SM1331	ThermoFischer Scientific	

Thermo Scientific™ GeneRuler 50 bp DNA Ladder, ready-to-use	SM0373	ThermoFischer Scientific	
Thermo Scientific™ GeneRuler Ultra Low Range DNA Ladder, ready-to-use	SM1213	ThermoFischer Scientific	
Thermo Scientific™ PageRuler™ Prestained Protein Ladder, 10 to 180 kDa	26616	ThermoFischer Scientific	
DNA isolation kits			
DNA Clean & Concentrator™-25	D4033	Zymo Research	
GenElute™ Gel Extraction Kit	NA1111-1KT	Sigma-Aldrich	
GenElute™ HP Plasmid Miniprep Kit	NA0160-1KT	Sigma-Aldrich	
NucleoBond® PC 20	740571.100	Macherey-Nagel	
NucleoBond® Xtra Midi EF	740420.50	Macherey-Nagel	
Wizard® SV Gel and PCR Clean-Up System	A9282	Promega	
Zymoclean™ Gel DNA Recovery Kit	D4007	Zymo Research	
Plastic ware			
Biosphere® Filter Tips 0.1-20 µL	70.1114.210	Sarstedt	
Biosphere® Filter Tips 100-1000 µL	70.762.211	Sarstedt	
Biosphere® Filter Tips 20-200 µL	70.765.210	Sarstedt	
Biosphere® Filter Tips 2-20 µL	70.760.213	Sarstedt	
Corning® 15 mL centrifuge tubes	CLS430791	Sigma-Aldrich	
Corning® 50 mL centrifuge tubes	CLS430829	Sigma-Aldrich	
Cryogenic vials	479-1238	VWR	
Micro tube 0.5 mL DNA LowBind	72.704.700	Sarstedt	
Micro tube 1.5 mL	72.690.001	Sarstedt	
Micro tube 1.5 mL DNA LowBind	72.706.700	Sarstedt	
Micro tube 1.5 mL Protein LowBind	72.706.600	Sarstedt	
Multiply®-µStrip Pro 8-strip	72.991.002	Sarstedt	
Nunc™ Cell Culture/Petri Dishes, 6/10/15 cm	10111351 10508921 168381	ThermoFischer Scientific	
Pipette tips, Bevel Point 1-200 µL	613-0732P	VWR	
Pipette tips, Extended length, with Tubegard™ ring, 0.1-10 µL	613-0735P	VWR	
Pipette tips, UltraFine™, FlexTop™, extended, 100-1250 µL	613-0739	VWR	
Scraper	83.1830	Sarstedt	
Stripette® Serological pipettes, 10mL	734-1693	VWR	
Stripette® Serological pipettes, 25 mL	734-1695	VWR	
Stripette® Serological pipettes, 5 mL	734-1691	VWR	

Stripette® Serological pipettes, 50 mL	734-1743	VWR	

4.3 Biological methods

4.3.1 Bacteria and bacteria culture

For this thesis, TOP10 and DH10B *Escherichia coli* (*E. coli*) strains were used for cloning and for BAC recombineering, respectively. Detailed information about the two strains is summarized in **Table 4**.

Table 4: E. coli strains used in this thesis. *Summarized are the strain of E. coli, the corresponding genotype, and the source.*

E. coli strain	Genotype	Source
Invitrogen™ One Shot™ TOP10	*F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(StrR) endA1 λ-*	ThermoFischer Scientific
DH10B	*F– Δ(ara-leu)7697[Δ(rapA'-cra')] Δ(lac)X74[Δ('yahH-mhpE)] duplication(514341-627601)[nmpC-gltI] galK16 galE15 e14– (icdWT mcrA) φ80dlacZΔM15 recA1 relA1 endA1 Tn10.10 nupG rpsL150(StrR) rph+ spoT1 Δ(mrr-hsdRMS-mcrBC) λ– Missense(dnaA glmS glyQ lpxK mreC murA) Nonsense(chiA gatZ fhuA? yigA ygcG) Frameshift(flhC mglA fruB)*	ThermoFischer Scientific

Bacteria were grown in 1x Lennox LB Broth liquid cultures at 37°C. For cultivation on solid LB Broth the medium was supplemented with 15g/L bacteriological Agar. Ampicillin, Chloramphenicol and Kanamycin were used for selection at final concentrations of 100 ng/mL, 50 ng/mL, and 50 ng/mL, respectively.

Chemical competent TOP10 cells were transformed via heat shock. For this, the cells were mixed with plasmid DNA (10:1 (v/v)) and incubated on ice for 30 min. Next, the cells were placed into a thermomixer set at 42°C for 30-45 sec before placing them on ice for 2 min. Subsequently, the cells were mixed with 1 mL pre-warmed LB medium without antibiotics and incubated for 1 h in an incubator at 37°C before plating them on antibiotic-containing, selective LB-Agar plates. To select colonies by blue-white screening, 100 µL of 100 mM IPTG and 20 µL of 50mg/mL X-Gal were spread onto the surface of LB-Agar plates and pre-incubated for 30 min at 37°C.

For short-term storage, LB-Agar plates were sealed with Parafilm (Sigma-Aldrich) and kept at 4°C. For long-term storage 600 µL from a 5 mL overnight culture were mixed with sterile 80% glycerol and stored at -80°C.

4.3.2 Cell lines and cell culture

This thesis used murine, pluripotent P19 cells in all cell culture related experiments. Stable cell lines are summarized in **Table 5**.

Table 5: Murine P19 cell lines used in this thesis.

Cell line	Vendor/Generated by	Antibiotic resistance	Citation
P19 wild type (CRL-1825™)	ATCC®	-/-	(McBurney & Rogers, 1982)
P19 CPSF5-GFP	Oliver Schwich	Neomycin	
P19 CPSF6-GFP	Oliver Schwich	Neomycin	
P19 GFP-FIP1	Oliver Schwich	Neomycin	
P19 SRSF3-GFP	Michaela Müller-McNicoll	Neomycin	(Müller-McNicoll et al., 2016)
P19 SRSF7-GFP	Michaela Müller-McNicoll	Neomycin	(Müller-McNicoll et al., 2016)
P19 SRSF7-ZnMut-GFP	Anfisa Soloyeva/Benjamin Arnold	Neomycin	(Königs et al., 2019)
P19 SRSF7-Δ27aa-GFP	Benjamin Arnold	Neomycin	(Königs et al., 2019)

P19 cells were grown in cell culture dishes coated with 0.1% gelatin (v/v) supplemented phosphate buffered saline (PBS) solution using DMEM, high glucose, GlutaMAX™ medium, supplemented with heat-inactivated 10% (v/v) fetal bovine serum and 100 U/mL penicillin-streptomycin under humidified condition at 5% CO_2 at 37°C. P19 cell lines stably expressing GFP-tagged proteins from BACs integrated into the genome were cultured in complete medium supplemented with 450 µg/mL Geneticin® (G418).

Cells were passaged when 10-cm cell culture dishes were grown to 90% confluent. For this, the medium was removed, and cells were washed once with PBS. Cells were detached using trypsin-EDTA and the reaction was stopped by adding 2X(v/v) completed medium before passaging a certain number of cells onto the respective fresh cell culture dishes as summarized in **Table 6**.

Table 6: Approximate number of P19 cells passaged on fresh cell culture dishes, depending on the size of the cell culture dish.

Size cell culture dish	Dilution	Approximate number of cells
6-cm	1:50	$3x10^5$
10-cm	1:15	$7x10^5$
14-cm	1:4.5	$2.2x10^6$

Aliquots of approximately $2.5x10^6$ P19 cell lines were frozen in completed medium supplemented with 10% (v/v) DMSO using Mr. Frosty™ freezing container and placed at -80°C short term or in a liquid N_2 atmosphere for long term storage. To thaw cell lines, cells were pelleted by centrifugation at 1,000x g for 5 min before the supernatant was removed and cells were resuspended and cultured in DMSO-free complete medium without Geneticin.

4.3.3 P19 neuronal differentiation

P19 wild type cells were differentiated into neuronal cells using retinoic acid as described by Nakayama et al., 2014. Briefly, 10 cm culture dishes were coated with 10 µg laminin diluted in 4 mL PBS overnight in the incubator. The next day, the laminin solution was removed, and these dishes were washed three times with 1X PBS. 300 µL P19 wild type cells were passaged from a confluent 10 cm culture dish to the plates, which were prepared with 10 mL Gibco™ DMEM/F-12, GlutaMAX™ Supplement, supplemented with Gibco™ N-2 Supplement (100X) and 100 U/mL penicillin-streptomycin. To start the neuronal induction, the medium was supplemented with 10 ng/mL FGF8, 10 µM DAPT and 500 mM retinoic acid. Cells were grown for two days under humidified condition at 5% CO_2 at 37°C before refreshing the medium. After additional cultivation for two days, synaptogenesis was induced by replacing the medium with 10 mL Gibco™ CTS™ Neurobasal® Medium supplemented with 1X Gibco™ B-27™ Supplement and 100 U/mL penicillin-streptomycin. To remove all dividing cells, 8 µM Cytosine β-D-arabinofuranoside hydrochloride (Ara-C) was added to the cultures. As before, cells were grown for four days with a medium exchange after two days, until the differentiated cells were harvested at day 8. The differentiation progress was monitored by bright field microscopy every second day.

4.3.4 Microscopy

To control growth and integrity of cell cultures, they were constantly checked using an AE31 inverted bright field microscope. To control expression of fluorescently labeled proteins the AE31 microscope was used in combination with a mercury lamp powered by an MHG-100B mercury lamp power supply. The fluorescence attachment was supplied with three filter cubes to detect HOECHST, GFP and mCherry fluorescence, respectively. The Moticam 3 was used to image cells in combination with the supplied software Motic Image Plus 3.0.

For advanced microscopy tasks and subcellular localization studies a confocal laser-scanning microscope LSM780 (Zeiss; Frankfurt Center for Advanced Light Microscopy (FCAM) facility, Buchmann Institute for Molecular Life Sciences (BMLS), Frankfurt, Germany) was used.

To prepare coverslips, cells were grown in coated cell culture dishes containing coverslips as described before until approximately 90% confluence. Cells were washed once with PBS and fixed for 30 min at room temperature (RT) using 4% Paraformaldehyde in 1X PBS. Next, the cells were washed once more with PBS before adding 200 pM HOECHST dissolved in TBST-buffer pH 7.6 (50 mM Tris-HCl, 150 mM NaCl, 0.05% (v/v) Tween® 20) and incubation for 30 min at RT. Subsequently, the HOCHEST solution was removed and the coverslips were washed once with PBS before mounting them upside down on a microscope slide supplemented with 5 µL ProLong® Diamond Antifade Mountant. The slides were set to dry before imaging and stored at 4°C.

4.3.5 Plasmids

For this thesis, a variety of plasmids were generated and used. They are summarized in **Table 7** and described in detail below. Plasmid maps are provided in **SuppFigure 1**-**SuppFigure 5**.

Table 7: List of plasmids used in this thesis.

Plasmid	Cat. number	Vendor/Generated by	Resistance	Usage
pEGFP-N1	6085-1	Clontech	Kanamycin	Reporter gene
pmCherry-N1	632523	Clontech	Kanamycin	Reporter gene
pGEM®-T Easy	A1360	Promega	Ampicillin	Sub cloning
pGL4.51[*luc2*/CMV/Neo]	E132A	Promega	Ampicillin	Reporter gene
psTet-Repressor	-/-	AK Suess, TU Darmstadt, Darmstadt, Germany	Ampicillin	Reporter gene
psTet- Repressor -RS3	-/-	Nicole Blümel	Ampicillin	Reporter gene
psTet- Repressor -RA3	-/-	Nicole Blümel	Ampicillin	Reporter gene
psTet- Repressor -RD3	-/-	Nicole Blümel	Ampicillin	Reporter gene
psTet- Repressor -RS7	-/-	Nicole Blümel	Ampicillin	Reporter gene
psTet- Repressor -RA7	-/-	Nicole Blümel	Ampicillin	Reporter gene
psTet- Repressor -RD7	-/-	Nicole Blümel	Ampicillin	Reporter gene
pmCherry-SRSF3	-/-	Oliver Schwich	Kanamycin	Reporter gene
pmCherry-SRSF3[RA]	-/-	Oliver Schwich	Kanamycin	Reporter gene
pmCherry-SRSF3[RD]	-/-	Oliver Schwich	Kanamycin	Reporter gene
pmCherry-SRSF7	-/-	Oliver Schwich	Kanamycin	Reporter gene
pmCherry-SRSF7[RA]	-/-	Oliver Schwich	Kanamycin	Reporter gene
pmCherry-SRSF7[RD]	-/-	Oliver Schwich	Kanamycin	Reporter gene
pEGFP-SRSF3	-/-	Francois McNicoll	Kanamycin	Reporter gene
pEGFP-SRSF7	-/-	Francois McNicoll	Kanamycin	Reporter gene
pEGFP-SRSF3-Zn	-/-	Nicole Blümel, Benjamin Arnold, Oliver Schwich	Kanamycin	Reporter gene
pEGFP-SRSF3-27aa	-/-	Nicole Blümel, Benjamin Arnold, Oliver Schwich	Kanamycin	Reporter gene

pEGFP-SRSF3-Zn-27aa	-/-	Nicole Blümel, Benjamin Arnold, Oliver Schwich	Kanamycin	Reporter gene
pGL4.51-Ddx21	-/-	Oliver Schwich	Ampicillin	Reporter gene
pGL4.51-Ddx21-strongpPAS	-/-	Oliver Schwich	Ampicillin	Reporter gene
pGL4.51-Ddx21-inactivepPAS	-/-	Oliver Schwich	Ampicillin	Reporter gene
pGL4.51-Ddx21-allSRSF3	-/-	Oliver Schwich	Ampicillin	Reporter gene
pGL4.51-Ddx21-allSRSF7	-/-	Oliver Schwich	Ampicillin	Reporter gene
pGL4.51-Anp32e	-/-	Oliver Schwich	Ampicillin	Reporter gene
pGL4.51-Rab11a	-/-	Oliver Schwich	Ampicillin	Reporter gene
pmCherry-Ddx21	-/-	Oliver Schwich	Kanamycin	Reporter gene
pmCherry-Anp32e	-/-	Oliver Schwich	Kanamycin	Reporter gene
pmCherry-Rab11a	-/-	Oliver Schwich	Kanamycin	Reporter gene

To isolate small amounts of plasmid DNA from bacterial cultures, the GenElute™ Plasmid Miniprep Kit was used as instructed by the manufacturer. To extract up to 400 µg of plasmid and BAC DNA the NucleoBond® Xtra Midi prep kit was used as instructed by the manufacturer. To isolate DNA from BAC-plasmids the NucleoBond® PC 20 kit was used as instructed by the manufacturer. The concentration of isolated plasmids was quantified on a NanoDrop 2000 and plasmids were stored at -20°C.

4.3.6 Primers

Lyophilized, desalted short custom DNA oligo nucleotides for PCRs and qRT-PCRs were ordered from Sigma-Aldrich in single tubes on a 0.025 µmole scale. Primers were reconstituted in nuclease-free water to a concentration of 100 µM and stored at -20°C. For general PCR purposes, primers were diluted to 10 µM using nuclease-free water. All primers used in this thesis are summarized in SuppTable 1 stating the individual ID, name, 5’ -3’ sequence and purpose of each primer, as well as the concentration used in qRT-PCR experiments.

4.3.7 BAC and plasmid transfection

4.3.7.1 BAC transfection and selection of stable transgenic cell lines

Desired BACs were transfected using Effectene® Transfection Reagent as described by Poser et al., 2008. Briefly, approximately $3.5x10^5$ P19 cells per well were passaged in fresh 6-well cell culture dishes and cultured over-night as described before. The next day, the medium was changed and one 1.5 mL reaction tube per transfection was prepared. 500 ng BAC-DNA were transferred to the reaction

tube and mixed with buffer EC to a final volume of 150 µL. Then, 4 µL Enhancer reagent were added and mixed briefly before supplying 25 µL Effectene reagent. The mixture was incubated for 10 min at RT and then added in drops onto the cells. The cells were incubated over-night before 375 µg/mL Geneticin were added. Subsequently, the cells were grown, and the medium was exchanged every 2 days constantly increasing the Geneticin concentration to a maximum of 450 µg/µL. Finally, once confluent, the stable transgenic cell lines were passaged onto fresh 10 cm cell culture dishes supplemented with 450 µg/mL Geneticin to increase cell numbers for long-term storage.

4.3.7.2 Plasmid transfection

Approximately $7x10^5$ P19 cells were passaged on fresh 10 cm culture dishes and cultured overnight as described before. The next day 5 µg of the respective plasmid DNA were transfected using jetOPTIMUS® DNA Transfection Reagent in a 1:1 ratio as described in the manufacturer's instructions. The cells were grown for 24-36 h before harvesting.

Alternatively, plasmids were transfected using Lipofectamine™ 2000 Transfection Reagent. For this, cells were passaged as described before. Next, 5 µg plasmid DNA were diluted in 250 µL Gibco™ Opti-MEM™ Reduced Serum medium and 10 µL Lipofectamine™ 2000 was diluted in 240 µL Opti-MEM™. Both were mixed by inverting and incubated for 5 min at RT. Subsequently, the diluted plasmids and Lipofectamine™ 2000 mixtures were combined, mixed carefully, and incubated for an additional 20 min at RT. Meanwhile, the medium of the cell culture was replaced with fresh completed DMEM before the transfection mix was added in drops. The cells were grown for 24-36 h before harvesting.

4.3.8 esiRNA design, preparation, and transfection

4.3.8.1 esiRNA design and preparation

Murine annotations of genes of interest were used to pick regions of approximately 400 nt with minimal off-target or cross-silencing using the DEQOR3 software (Surendranath et al., 2013) for subsequent esiRNA preparation. esiRNAs were produced as described in Kittler et al., 2005. Briefly, primers flanking the picked window were designed including 5' overhangs containing the T7-Promotor sequence (5' -TAATACGACTCACTATAGGGAGA-3') and templates were prepared by PCR using suitable cDNA and Phusion High-Fidelity Polymerase according to the manufacturer's instructions. Subsequently, the PCR product was used for *in vitro* transcription using HiScribe™ T7 High Yield RNA Synthesis Kit. For this, reactions were prepared as summarized in **Table 8** and transcribed on a PCR cycler using the program summarized in **Table 9**. Double stranded esiRNA templates were stored at RT until digestion.

Table 8: Master mix to prepare esiRNA templates by in vitro transcription.

Component	Volume [µL]
10x T7 Reaction Buffer	1
100 mM ATP	1
100 mM UTP	1
100 mM CTP	1
100 mM GTP	1
100 mM DTT	1
T7 RNA Polymerase Mix	1
DNA template	4
Total	11

Table 9: PCR cycler setting for in vitro transcription of esiRNA templates.

Step	Temperature [°C]	Time [min]
Transcription	37°C	720 min
Denaturation	90°C	3 min
Annealing	Ramp to 70°C with 0.1°C/sec	
	70°C	3 min
	Ramp to 50°C with 0.1°C/sec	
	50°C	3 min
	Ramp to 20°C with 0.1°C/sec	
	20°C	Hold

Next, the *in vitro* transcribed products were mixed with 90 µL Digestion buffer (20 mM Tris-HCl pH 8.1, 0.5 mM EDTA, 5 mM $MgCl_2$, 1 mM DTT, 140 mM NaCl, 2.7 mM KCl, 5% (v/v) glycerol; set to pH 7.9) and digested using 6 µg RNase III (MPI-CBG, Dresden, Germany) at 37°C at 1100 rpm for 2 h. Afterwards 2 µL of digested dsRNAs were mixed with 2 µL nuclease-free water and 4 µL 2x RNA loading dye and run on 3% agarose gels to validate correct digestion. Digested esiRNAs are expected to be between 21-30 nt long. For this, they were run beside Thermo Scientific™ GeneRuler Ultra Low Range DNA Ladder and Thermo Scientific™ GeneRuler 50 bp DNA Ladder. Next, samples were stored on ice or at -80°C for long-term storage until purification. Therefore, Mini Bio-Spin® Chromatography Columns were supplied with 200 µL GE Healthcare Q-Sepharose™ Fast Flow Anion Exchange Media. The purification is based on anion exchange chromatography using the advantage that negatively charged RNA sticks stronger to the surface. Increasing salt concentration will elute proteins first,

followed by small RNAs until longer RNA transcripts are eluted. The selective elution of small RNAs, e.g., the digested esiRNAS, is fine-tuned by the sodium chloride concentration in the elution buffer. For this, sepharose was washed twice with 500 µL Equilibration buffer (20 mM Tris-HCl pH 7.4, 1 mM EDTA, 300 mM NaCl, pH 7.5) by centrifugation at 1,000x g for 1 min before one or two identical esiRNA samples were loaded onto the columns and incubated for 5 min at RT. The samples were centrifuged once at 1,000x g for 1 min and then washed once with 500 µL Washing buffer (20 mM Tris-HCl pH 7.4, 1 mM EDTA, 400 mM NaCl, pH 7.5), again by centrifugation as before. Next, the spin columns were transferred to fresh 1.5 mL reaction tubes and purified esiRNAs were eluted in two rounds using 300 µL Elution buffer each (20 mM Tris-HCl pH 7.4, 1 mM EDTA, 520 mM NaCl, pH 7.5). Subsequently, the purified esiRNAs were precipitated using 500 µL isopropanol and kept on ice for 30 – 60 min before pelleting at 16,000x g at 4°C for 15 min. The supernatants were removed, and the pellets washed twice with 500 µL ice-cold 80% ethanol (EtOH) at 16.000 x g and 4°C for 4 min. Finally, esiRNA pellets were air-dried at the bench for up to 15 min and resuspended in 50 µL TE buffer (10 mM Tris-HCl pH 8.1, 1 mM EDTA, pH 7.9), aliquoted and stored at -80°C. esiRNA concentrations were measured using a NanoDrop 2000.

4.3.8.2 esiRNA and esiRNA/plasmid (co)-transfection

Approximately 7x10^5 P19 cells were passaged on fresh 10 cm culture dishes and cultured overnight as described before. The next morning 5 µg of the respective endoribonuclease-prepared siRNA (esiRNA) or 5µg esiRNA and 5 µg plasmid DNA were transfected using jetPRIME® DNA/siRNA transfection reagent in a 1:2 ratio as described in the manufacturer's instructions. The cells were grown for 36-48 h before harvesting.

4.3.9 Cloning

4.3.9.1 Restriction based

To clone phosphomimetic RS domains, RA and RD, of SRSF3 and SRSF7 downstream of the single-chain Tet-Repressor protein (psTet- Repressor, AK Suess Lab, University Darmstadt, Germany) the respective gene fragments (sequences summarized in **SuppTable 5**) were designed using SnapGene® and ordered as gBlocks® Gene Fragments (Integrated DNA Technologies). Gene fragments were reconstituted in nuclease-free water as instructed by the manufacturer and stored at -20°C. To amplify the inserts and add XmaI restriction sites suitable primers (see **SuppTable 1**) were designed and fragments amplified using ALLin™ RPH Polymerase as instructed by the manufacturer. For long-term storage, the amplicons were purified using the Wizard® SV Gel and PCR Clean-Up System. Subsequently, they were ligated

into the pGEM®-T Easy vector system overnight at 4°C as instructed by the manufacturer. E. coli TOP10 cells were transformed with the ligation mixes as described before (in chapter 4.3.1) and stored as glycerol stocks at -80°C. For cloning, 1 µg PCR-amplified inserts were purified and digested using XmaI-HF before a second round of purification. Meanwhile, recipient plasmids were subjected to XmaI-HF restriction using 5 µg plasmid DNA following the manufacturer's instructions. Digested backbones were separated on 0.8% agarose gels, cut and purified using a Zymoclean Gel DNA Recovery Kit as instructed by the manufacturer. Subsequently, backbone and inserts were ligated in a 1:2 ratio using T4 DNA Ligase and T4 DNA Ligase Buffer (10X) in a total volume of 20 µL at 4°C overnight before transformation into TOP10 cells as described before (see chapter 4.3.1). Positive clones were identified by Colony-PCR using Taq-Polymerase with ThermoPol® Buffer, saved as glycerol stocks and subjected to Sanger sequencing to verify the inserts (Eurofins Genomics).

The same cloning strategy was used to fuse complete phosphomimetic versions of SRSF3[RA/RD] and SRSF7[RA/RD] (**SuppFigure 3**, **SuppTable 6**) as well as SRSF3-Chimera constructs (Zn, 27aa, Zn+27aa, **SuppFigure 4**, **SuppTable 7**) downstream of the mCherry or EGFP reporter genes utilizing pmCherry-N1 and pEGFP-N1 plasmids (Clontech) and *Nhe*I-HF and *Kpn*I-HF restriction enzymes.

4.3.9.2 Gibson Assembly® cloning

To replace the *SV40* 3' UTR of the reporter genes *Luciferase* and *mCherry* with endogenous 3' UTRs of interest and their respective mutants, Gibson Assembly® cloning was used. The insertion of one fragment was planned and primer were designed using SnapGene® with the respective settings: Target T_m for PCR primers 63°C and 15-25 overlapping bases with a target T_m for overlap of 53°C. Backbones and inserts were linearized and amplified from plasmid and genomic DNA respectively, using Q5® High-Fidelity DNA Polymerase as instructed by the manufacturer. 15 µL aliquots of Assembly master mix (100 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 200 pM dNTP mix, 10 mM DTT, 5% (w/v) PEG-8000, 1 mM NAD, 0.1U T5 exonuclease, 0.5U Phusion DNA Polymerase, 0.08U *Taq* DNA Ligase) were thawed on ice. Meanwhile, 50 ng of each vector and insert were mixed in a total volume of 5 µL and a thermoshaker was pre-heated to 50°C. Next, the Assembly master mix was placed into the thermoshaker and the backbone/insert-mix was added and incubated for 60 min. Subsequently, the ligation products were transformed into Top10 cells and positive clones were identified by Colony-PCR using *Taq*-Polymerase with ThermoPol® Buffer, saved as glycerol stocks and subjected to Sanger sequencing to verify the inserts (Eurofins Genomics).

4.3.10 RNA extraction

To isolate total RNA from P19 cells, cells grown on 10 cm cell culture dishes were washed once with 5 mL ice-cold PBS and then scraped into 1.5 mL ice-cold PBS. Cells were pelleted by centrifugation at 1,000x g for 5 min at 4°C, the supernatant was removed, and the cells were resuspended in 1 mL TRIzol™ and stored overnight at -80°C. Next, the samples were thawed on ice, supplemented with 200 µL chloroform and shaken vigorously before centrifugation at 17,000x g for 15 min at 4°C. The aqueous phase was transferred to a fresh reaction tube and mixed with one volume isopropanol, mixed well, and incubated for 60 min on ice to precipitate the RNA. The RNA was pelleted by centrifugation at 17,000x g for 25 min at 4°C and washed once with nuclease-free 70% EtOH. The supernatant was removed, and the pellet was air dried before resuspension in 88 µL nuclease-free water. To remove DNA contaminations, the samples were supplemented with 10 µL 10x TURBO™ DNase buffer and 2 µL TURBO™ DNase and incubated at 37°C for 30 min. The RNA was precipitated overnight at -80°C after adding 100 µL nuclease-free water, 20 µL 3 M sodium acetate (NaOAC) pH 5.5 and 550 µL 100% EtOH. Subsequently, the RNA was centrifuged for 25 min at 17,000x g and 4°C and washed once with nuclease-free 70% EtOH. The supernatant was removed, and the pellet air-dried at RT before dissolving in nuclease-free water. The RNA concentration was measured using a NanoDrop 2000 and the RNA was stored at -80°C.

4.3.11 Reverse transcription

mRNA was reverse transcribed into cDNA using SuperScript™ III Reverse Transcriptase according to the manufacturer's instructions. Briefly, 1 µg total RNA was diluted in a total volume of 10 µL with nuclease-free water and supplemented with 1 µL 10 mM dNTP Mix (10 mM each dATP, dGTP, dCTP, dTTP) and 1 µL 100 mM oligodT(T18) primer. One additional sample was included as –RT control. Next, the mixtures were denatured as shown in Table 10 and subsequently set on ice immediately. Meanwhile, per sample the transcription master mix was prepared as summarized in Table 11. Next, 7 µL of the master mix were mixed with the –RT sample before the SuperScript™ enzyme was added to the remaining master mix. Finally, 8 µL of the final mix were added to the remaining samples and the program was continued with the transcription step.

Table 10: PCR cycler setting to perform reverse transcription using SuperScript™ III.

Step	Temperature	Time
Denaturation	65°C	5 min
Annealing	4°C	hold
Transcription	50°C	60 min
Inactivation	75°C	15 min
	4°C	hold

Table 11: Master mix for reverse transcription using SuperScript™ III.

Component	Volume
5x First-Strand buffer	4 µL
0.1 M DTT	2 µL
RiboLock RNase Inhibitor	1 µL
After supplying –RT sample	
SuperScript™ III	1 µL

Subsequently, the success of the reverse transcription was validated by *Taq*-PCR and electrophoresis using primers targeted against the murine *calreticulin* (*CalR*) gene. The cDNA was stored at -20°C and was used to perform PCRs and qRT-PCRs.

4.3.12 Rapid amplification of 3’ cDNA ends (3’RACE)

To amplify the 3’ cDNA ends of transcripts; total RNA was reverse transcribed using SuperScript™ III as described before in combination with an anchored oligodT primer containing a 3’ overhang complementary to a reverse primer in subsequent PCRs. Following the reverse transcription and positive test PCR, individual transcript isoforms of a gene of interest were amplified by PCR using ALLin™ RPH Polymerase as instructed by the manufacturer with the setting summarized in **Table 12**, using a gene-specific forward primer and an overhang-specific 3’RACE reverse primer.

Table 12: PCR cycler settings to perform 3’RACE-PCR using ALLin™ RPH Polymerase.

Step	Temperature	Time	Cycles
Initial denaturation	95°C	5 min	
Denaturation	95°C	15 sec	28x
Annealing	68°C	30 sec	
Elongation	72°C	2 min	
Final elongation	72°C	10 min	
	10°C	hold	

The PCR products were separated on 1.5% agarose-TBE gels supplemented with 1X RedSafe™ Nucleic Acid Staining Solution for 40 min at 130V. The gels were imaged and the ratios of individual transcripts were quantified using Fiji/ImageJ (Schindelin et al., 2012; Rueden et al., 2017).

4.3.13 Real-time quantitative PCR (qRT-PCR)

Before real-time quantitative PCRs (qRT-PCRs) were performed, all primers were tested for specificity by *Taq*-Polymerase PCR. In addition, primer efficiency tests were run as technical replicates using variable primer concentrations (250 nm – 1000 nM) and serial cDNA dilutions ranging from 1:10-1:80. Per reaction 1.5 µL diluted forward and reverse primer were mixed with 2.5 µL 2X ORA™ qPCR Green ROX H Mix and 4 µL of the mix were pipetted per well in a 96-well Piko PCR plate. 1 µL of the diluted cDNA was added and the qPCR was run on a PikoReal™ Real-Time PCR System. In general, the program summarized in **Table 13** was used. The resulting data were evaluated and exported using the PikoReal Software (Version 2.2). Quantitation cycles (Cq) values were plotted against the dilution and the respective slope was used to assess primer efficiencies. The data from the melting curve was used so confirm the specificity of the tested primer pairs.

Table 13: Program to run qRT-PCR.

Step	Temperature	Time	Cycles
Initial denaturation	95°C	2min	
Denaturation	95°C	20 sec	35x
Annealing	60°C	20 sec	
Extension (incl. data acquisition)	72°C	30 sec	
Final elongation	72°C	5 min	
Melt curve (incl. data acquisition)	60-95°C		
	25°C	hold	

To perform qRT-PCRs, per reaction 1.5 µL diluted primer pairs (**SuppTable 1**) were mixed with 2.5 µL 2X ORA™ qPCR Green ROX H Mix. 4 µL of the mix were pipetted into each well of a 96-well Piko PCR plate and supplemented with cDNA diluted 1:10 in nuclease-free water. The qRT-PCR was performed as described before using technical replicates and biological triplicates.

4.3.14 RNA-Seq

To generate RNA-Seq datasets, total RNA was isolated as described before. 1 µg of total RNA was diluted with nuclease-free water in a total volume of 10 µL and submitted to Novogene (Hong Kong). At the Novogene facility the samples were controlled for their quality before libraries were

prepared and sequenced on an Illumina HiSeq 4000 platform yielding 75 bp single end or 150 bp paired end reads.

4.3.15 MACE-Seq

MACE libraries were prepared and sequenced at GenXPro GmbH (Frankfurt am Main, Germany) as described by Müller et al., 2014; Zawada et al., 2014. Total RNA was isolated as described before and submitted to GenXPro for downstream procedures. Briefly, first polyA+ RNA was isolated from total RNA using the Invitrogen™ Dynabeads™ mRNA Purification Kit (ThermoFischer Scientific) followed by reverse transcription into cDNA using the Invitrogen™ SuperScript™ Double-Stranded cDNA Synthesis Kit (ThermoFischer Scientific) using an anchored biotinylated poly(dT) primer. Next, the cDNA was randomly fragmented to an average size of 250bp by sonication using a Bioruptor (Diagenode). The biotinylated cDNA ends were captured by Invitrogen™ Dynabeads™ M-270 Streptavidin Beads (ThermoFischer Scientific) and ligated with T4 DNA Ligase 1 (New England Biolabs Inc.) to modified TrueQuant adapters (GenXPro). Libraries were amplified using KAPA HiFi Hot-Start Polymerase (KAPA Biosystems), followed by purification using Agencourt AMPure XP beads (Beckman Coulter). The libraries were sequenced on a HiSeq2000 (Illumina) platform yielding 75 bp single end reads.

Raw sequencing files were downloaded from GenXPro. Downstream preparation and analysis were performed by Samarth Thonta Setty and Mario Keller (Zarnack Lab, Buchman Institute for Molecular Life Sciences, Frankfurt).

4.3.16 Protein extraction

To extract total proteins from P19 cells, cells were harvested as described for RNA extraction. The cell pellets were snap-frozen in liquid nitrogen and stored at -80°C until protein isolation. For this, pellets were thawed on ice and resuspended in up to 400 µL ice-cold NET-2 buffer (150 mM NaCl, 0.05% (v/v) NP-40, 50 mM Tris-HCl pH 7.5) supplemented with 1X cOmplete™, EDTA-free Protease Inhibitor Cocktail and 10 mM β-Phosphoglycerate. The cells were lysed for 10 min on ice before additional sonication for 30 sec at 20% amplitude in intervals of 10 sec “On”/20 sec “Off”. Subsequent, the lysates were cleared by centrifugation at 17,000x g for 10 min at 4°C. The supernatant was transferred to fresh 1.5 mL reaction tubes. To measure protein concentrations, the Quick Start™ Bradford protein assay (Bio-Rad) was used in combination with a NanoDrop 2000 as instructed by the manufacturer. Briefly, 5 µL of protein lysate were mixed with 250 µL 1X Quick Start™ Bradford Dye Reagent and incubated for 5 min in the dark at RT. If the protein concentration exceeded the limits of

the standards, 2.5 µL protein lysate was mixed with 2.5 µL sterile water and mixed with dye reagent. The dilution was considered during the analysis. The protein lysates were supplemented with 5X Laemmli sample buffer (312 mM Tris-HCl pH 6.8, 10% (w/v) SDS, 25% (v/v) β-mercaptoethanol, 0.05% (w/v) bromophenol blue, 50% (v/v) glycerol) in a 4:1 ratio and denatured at 95°C for 5 min. Samples were stored at -20°C until usage.

4.3.17 Shrimp-Alkaline phosphatase treatment

Shrimp-Alkaline phosphatase (SAP) treatments were performed by Nicole Blümel.

To perform shrimp-alkaline phosphatase (SAP) treatments, P19 cells were harvested as described for total protein extraction. Frozen cell pellets were thawed on ice for 10 min before resuspension in 400 µL NET-2 buffer supplemented with 1X cOmplete™, EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich) and 10 mM $MgCl_2$. Cells were lysed, cleared, and quantified as described before. Next, 10 µg of total protein were transferred into two 1.5 mL reaction tubes and supplemented with additional reagents as summarized in **Table 14**.

Table 14: Master mix to perform SAP treatments.

Component	–SAP	+SAP
10 µg total Protein	X µL	X µL
10X SAP buffer (New England Biolabs Inc.)	2 µL	2 µL
100 mM β-phosphoglycerate	2 µL	-/-
rSAP (1U/µL, New England Biolabs Inc.)	-/-	5 µL
NET-2 buffer +10 mM $MgCl_2$	Up to 20 µL	Up to 20 µL

All samples were incubated at 37°C for 30 min shaking at 300 rpm before adding 5 µL 5X Laemmli buffer to the –SAP sample and 5 µL 5X Laemmli buffer (without glycerol) to the +SAP sample. Proteins were denatured at 90°C for 5 min before Western blotting.

4.3.18 Polyacrylamide-Gel-Electrophoresis and Western blotting

10-20 µg total proteins were separated on purchased NuPAGE™ 4-12% Bis-Tris gels or homemade 10% SDS-Glycine gels (Recipe summarized in **Table 15**).

Table 15: Recipe for homemade SDS-Glycine gels.

Component	10% Resolving gel	4% Stacking gel
Water (ddH_2O)	1.65 mL	1.22 mL
30% Bis-acrylamide (37.5:1)	2 mL	260 µL
1 M Tris-HCl pH 8.8	2.25 mL	-/-
1 M Tris-HCl pH 6.8	-/-	500 µL
10% SDS (w/v)	60 µL	20 µL
10% Ammonium persulphate	30 µL	25 µL
TEMED	10 µL	2.5 µL

NuPAGE™ gels were run using 1X NuPAGE™ MOPS SDS Running buffer in XCell SureLock™ Mini-Cell electrophoresis chambers. Proteins were blotted onto 0.1 µm Amersham™ Protran® nitrocellulose membranes using 1X transfer buffer (25 mM Bicine, 25 mM Bis-Tris (free base), 1 mM EDTA).

Homemade gels were run using 1X SDS-Glycine buffer (192 mM Glycine, 25 mM Tris-Base (free base), 0.1% (w/v) SDS) in Mini-PROTEAN Tetra Cells. Proteins were blotted onto 0.1 µm nitrocellulose membranes using 1X Towbin buffer (192 mM Glycine, 25 mM Tris-Base (free base), 20% (v/v) Methanol).

Successful transfer was validated by staining the membranes with Ponceau S for 5 min with agitation at RT. The membranes were washed with ddH_2O until residual Ponceau S solution was removed and subsequent imaged on a CanoScan 9000F Mark II. Next, the membranes were destained with 1X TBST for 10 min at RT. Subsequently, the blots were blocked using 10 mL 1X TBST supplemented with 3% (w/v) bovine serum albumin for 90 min at RT or overnight at 4°C. Subsequently, the blots were washed two times using 1X TBST for 15 min each, before incubation with primary antibodies (**Table 3**). The membranes were washed again twice with 1X TBST before incubation with HRP-conjugated secondary antibodies (**Table 3**). After another round of washing, the blots were developed using ECL™ Prime Western Blotting reagents on a ChemiDoc™ XRS+ system in combination with the Image Lab™ software.

4.3.19 Co-Immunoprecipitation

Co-Immunoprecipitations (Co-IPs) were performed by Nicole Blümel.

To prepare co-immunoprecipitation (co-IP) experiments, approximately $4x10^7$ P19 cells were treated, washed them twice with 1X ice-cold PBS, harvested and stored at -80°C until usage. First, 500 µL GammaBind™ G Sepharose™ beads were blocked at 4°C for 90 min using 500 µL 0.2 mg/mL bovine serum albumin in 1X PBS solution. The beads were pelleted by centrifugation at 1,000x g for 5

min at 4°C, resuspended in 500 µL 1X PBS and stored at 4°C. Subsequently, per co-IP four 1.5 mL reaction tubes were supplemented with 50 µL blocked beads, which were washed once with NET-2 buffer (150 mM NaCl, 0.05% (v/v) NP-40, 50 mM Tris-HCl pH 7.5) supplemented with 1X cOmplete™, EDTA-free Protease Inhibitor Cocktail and 10 mM β-Phosphoglycerate. Respectively, two tubes each were mixed with 10 µg goat anti-GFP (provided by D. Drechsel, MPI-CBG, Dresden, Germany) or goat anti-mouse-IgG in a total volume of 500 µL NET-2 buffer and coupled for 2h at 4°C on a rotating wheel. Subsequently, the beads were pelleted by centrifugation for 2 min at 1,000x g and 4°C. The beads were washed three times using 900 µL NET-2 buffer and stored at 4°C until the protein lysates were ready. P19 cells were thawed on ice and lysed using 1 mL NET-2 buffer and sonication as described before. After clearing lysates by centrifugation at 17,000x g for 10 min at 4°C, the supernatant was split into two 1.5 mL reaction tubes, 500 µL each. To remove bound RNA, one of the tubes was supplemented with 0.2 µg RNaseA and both tubes were incubated at 21°C for 20 min shaking at 800 rpm. Both lysates were brought to a final volume of 1 mL using NET-2 buffer. Input samples were prepared from both tubes by mixing 22.5 µL 1X NuPAGE® LDS sample buffer, supplemented with 1:10 (v/v) NuPAGE® sample reducing agent with 2.5 µL of each sample, followed by boiling the input samples for 5 min at 90°C. These samples were kept on ice until later use. Next, 490 µL of each protein lysate were mixed with the blocked and coupled beads and incubated for 90 min at 4°C on a rotating wheel. Subsequently, the beads were washed four times using 900 µL NET-2 buffer as before. After the last washing the buffer was removed and the beads were resuspended in 25 µL 1.32X NuPAGE® LDS sample buffer, supplemented with 1:10 (v/v) NuPAGE® sample reducing agent and denatured as described before. Co-IP and input samples were separated on NuPAGE™ 4-12% Bis-Tris gels and subjected to Western blotting using specific antibodies as described before.

4.3.20 Individual nucleotide UV-crosslinking and immunoprecipitation (iCLIP)

Approximately $4x10^7$ P19 cells were washed twice with ice-cold 1X PBS and irradiated once with 300 mJ/cm² UV light at 254 nm (CL-1000, UVP) and harvested on ice. As a control, a non-cross-linked (-UV) sample was included for each experiment. The cell pellets were stored at -80°C until iCLIP was performed as described by Huppertz et al., 2014 with minor modifications. Briefly, GFP-tagged proteins of interest were immunoprecipitated. For this, per sample, 100 µL Dynabeads Protein G were washed twice with 800 µL Lysis buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1% (v/v) IGEPAL®, 0.1% (w/v) SDS, 0.5% (w/v) Sodium deoxycholate). Next, the beads were resuspended in 200 µL Lysis buffer, coupled with 12 µg goat anti-GFP antibody (provided by D. Drechsel, MPI-CBG, Dresden, Germany) for 90 min at 4°C on a rotating wheel. Meanwhile, the lysates were prepared by thawing the pellets and resuspension in 1 mL Lysis buffer, supplemented with 1X cOmplete™, EDTA-free Protease Inhibitor

Cocktail, and 800 U RNAseOUT™. The cells were lysed for 10 min on ice before sonication and clearing as described before. To digest unprotected RNA, RNase I was diluted 1:200 (v/v) in Lysis buffer, supplemented with protease inhibitors and kept at RT. The lysates were then supplemented with 4 U Turbo™ DNase and 10 µL of the RNase I dilution before incubation at 37°C and 1100 rpm for exactly 3 min. Subsequently, the samples were placed on ice and kept until the beads were ready. To continue, the beads were washed once with 800 µL High Salt Washing buffer (50 mM Tris-HCl pH 7.5, 1 M NaCl, 1 mM EDTA, 1% (v/v) IGEPAL®, 0.1% (w/v) SDS, 0.5% (w/v) sodium deoxycholate) and twice with 800 µL Lysis buffer without protease inhibitors. The buffer was removed from the beads and they were resuspended using the RNase I digested lysates. The beads/lysate mixture was incubated for 90 min at 4°C on a rotating wheel before washing twice with 800 µL High Salt Wash buffer, followed by washing twice and resuspension with 800 µL PNK Wash buffer (20 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 0.2% (v/v) Tween-20). Subsequently, the 3' end of the RNA needs to be dephosphorylated to ligate the 3' adapter. For this, a T4 Polynucleotide Kinase (PNK) master mix was prepared as summarized in **Table 16**. The supernatant was removed from the beads and each sample was resuspended in 20 µL T4 PNK mix followed by an incubation at 37°C for 30 min at 1250 rpm.

Table 16: Master mix for T4 PNK driven 3' dephosphorylation.

Component	Volume
5x PNK buffer pH 6.5*	4 µL
T4 PNK	0.5 µL
RNaseOUT™	0.5 µL
Nuclease-free water	15 µL

**350 mM Tris-HCl pH 6.5, 50 mM $MgCl_2$, 5 mM DTT*

Subsequently, the beads were washed once with 800 µL PNK Wash buffer, twice with 800 µL High Salt Wash buffer and again twice with 800 µL PNK Wash buffer. Next, the pre-adenylated L3 Adapter (**SuppTable 1**) was ligated to the bead-bound RNA fragments in 20 µL master mix containing T4 RNA ligase and in the presence of polyethylene glycol (PEG) 8000 as summarized in **Table 17** overnight at 16°C.

Table 17: L3 Adapter ligation master mix.

Component	Volume
L3 Adapter (20 µM)	1.5 µL
10X RNA ligase buffer	2 µL
PEG 8000	4 µL
T4 RNA ligase	1 µL
RNaseOUT™	0.5 µL
Nuclease-free water	11 µL

The next day, 5' ends of the RNA fragments were labeled radioactively with P^{32}-ATP. For this, the beads were washed once with 800 µL PNK Wash buffer, twice with 800 µL High Salt Wash buffer and resuspended in 1 mL PNK Wash buffer. Next, each sample was split in a 200 µL and an 800 µL aliquot and kept on ice. The supernatant of the 200 µL aliquots were removed and each sample mixed with 7.2 µL PNK master mix as summarized in **Table 18**, before 0.8 µL P^{32}-ATP were supplemented per sample.

Table 18: PNK master mix for radioactive labelling.

Component	Volume
10X PNK buffer pH 7.6	0.8 µL
T4 PNK	0.4 µL
Nuclease-free water	6 µL

The beads were incubated at 37°C for 5 min at 1100 rpm and then placed on ice. To select Protein-RNA complexes of the desired size the samples were separated by PAGE and transferred onto nitrocellulose membranes. For this, the supernatant of both aliquots of the same sample was removed and first the beads containing the radioactively labeled RNA, then the remaining beads were resuspended in 25 µL 1.5X NuPAGE® LDS sample buffer supplemented with 10% (v/v) NuPAGE® Sample Reducing Agent. All samples were denatured at 71°C for 10 min before separating them on NuPAGE™ 4-12% Bis-Tris gels as described before beside 5 µL PageRuler™ Prestained Protein Ladder, supplementing the Running buffer with 0.1% (v/v) NuPAGE™ Antioxidant. Subsequently, the labeled Protein-RNA complexes were blotted onto nitrocellulose membranes as described previously and the radiation level was checked using a Geiger counter. To detect and visualize radioactive signals, the membrane was exposed to a phosphoimager screen (GE Healthcare) overnight.

The next day, the signals were detected using a Typhoon PhosphoImager (Entian Lab, University Frankfurt). To isolate Protein-bound RNA fragments from the membrane, small pieces of the desired molecular weight were cut using a scalpel and treated with Proteinase K solution. For this, the membrane pieces were chopped to smaller pieces and transferred into 1.5 mL reaction tubes, supplemented with 200 µL Proteinase K buffer (100 mM Tris-HCl pH 7.5, 50 mM NaCl, 10 mM EDTA). Next, 0.2 U Proteinase K were added and incubated at 37°C for 20 min at 1100 rpm. To increase RNA recovery, 200 µL Proteinase K/Urea buffer (7M Urea, 100 mM Tris-HCl pH 7.5, 50 mM NaCl, 10 mM EDTA) were added and incubated for another 20 min. Subsequently, the RNA was precipitated using PhaseLock Heavy gel tubes and 400 µL neutral Phenol:Chloroform:Isoamyl alcohol (25:24:1), pH 6.7 as instructed by the manufacturer. The aqueous phase was transferred into fresh 1.5 mL reaction tubes and mixed with 400 µL chloroform to further extract RNA fragments. The samples were centrifuged

for 5 min at RT, before the aqueous phase was collected, mixed with 40 µL 3M NaOAC and 1 µL GlycoBlue™, followed by precipitation with 1 mL ice-cold 100% EtOH overnight at -80C.

Next, the purified fragments were reverse transcribed with SuperScript™ IV and barcoded RT primers, including all necessary sequences for later library amplification by PCR, and size selected to prepare the library amplification. For this, the samples were thawed on ice and pelleted by centrifugation at 17,000x g for 20 min at 4°C and washed once with 800 µL 80% nuclease-free EtOH. The pellets were resuspended in 5 µL nuclease-free water and placed on ice. PCR tubes, one per sample, were prepared by adding 1 µL of the respective, freshly diluted 0.5 µM RT-primer (**SuppTable 1**) and 1 µL 10 mM dNTP mix before mixing in the RNA samples, followed by denaturation as summarized in **Table 19**.

Table 19: PCR cycler program for reverse transcription using SuperScript™ IV.

Step	Temperature	Time
Denaturation	70°C	5 min
Annealing	25°C	hold
Transcription	50°C	40 min
Inactivation	80°C	5 min
	4°C	hold

In the meantime, a reverse transcription master mix was prepared as summarized in **Table 20**. After denaturation, each sample was supplemented with 13 µL of the master mix before the PCR program was continued.

Table 20: Master mix, per sample, for reverse transcription using SuperScript™ IV.

Component	Volume
5X First-strand buffer	4 µL
0.1 M DTT	1 µL
RNaseOUT™	0.5 µL
SuperScript™ IV	0.5 µL
Nuclease-free water	7 µL

Subsequently, the template RNA was destroyed by supplementing 1.65 µL 1M sodium hydroxide (NaOH) and incubation at 98°C for 20 min. Then, the samples were neutralized with 20 µL 1M HEPES-NaOH pH 7.3 before precipitation with 350 µL TE buffer, 40 µL 3M NaOAC pH 5.5 and 1 µL GlycoBlue™ in 1 mL ice-cold 100% EtOH overnight at -20°C.

The following day, the cDNAs were pelleted by centrifugation at 17,000x g for 20 min at 4°C and washed once with 400 µL nuclease-free 80% EtOH, before resuspension in 8 µL nuclease-free

water. To separate the cDNAs and unused RT-primers, the samples were run on a purchased 6% Novex™ TBE-Urea Gels. For this, each sample was mixed with 8 µL 2X Novex™ Urea Loading buffer and denatured at 80°C for 5 min, before separation. When the gel was run, pieces corresponding the size between 80-150 nt were excised using scalpels. Gel pieces were crushed to improve cDNA isolation with 400 µL TE buffer for 2 h at 37°C and 1100 rpm. Gel residue and isolated cDNA were separated using Corning® Costar® Spin-X® centrifuge columns supplemented with Whatman® glass microfiber filters before Phenol:Chloroform:Isoamyl alcohol, chloroform extraction and precipitation as described before. Subsequently, to prepare the samples for library amplification, the cDNA was circularized and re-linearized to reposition the PCR primer binding sites to the 5' - and 3' -end of the cDNAs. For this, the precipitated cDNAs were purified as before and resuspended in 8 µL Ligation mix, as summarized in **Table 21**, before incubation for 60 min at 60°C.

Table 21: Master mix for circularization of cDNAs using CircLigase II.

Component	Volume
10X Circligase buffer II	0.8 µL
50 mM $MnCl_2$	0.4 µL
CircLigase II	0.3 µL
Nuclease-free water	6.5 µL

Subsequently, the samples were supplemented with 1 µL 0.45 mM ATP and incubated for another 60 min at 60°C, before adding 30 µL CutC4-oligo Annealing mix (**Table 22**).

Table 22: Master mix for CutC4 oligo annealing.

Component	Volume
10x CutSmart buffer	3 µL
10 µM CutC4 oligo (SuppTable 1)	1 µL
Nuclease-free water	26 µL

The oligo was annealed using the cycle program summarized in **Table 23**. After annealing, 40 U *Bam*HI-HF were added to each sample and the cycler program was continued to linearize the circular cDNAs. Next, the cDNAs were precipitated as described before.

Table 23: PCR cycler program to anneal CutC4 oligo and linearize cDNA with BamHI.

Step	Temperature	Time	Cycles
Denaturation	95°C	2 min	
Annealing	Ramp to 25°C, reducing by 1°C per cycle	30 sec	70x
Hold	25°C	hold	
Add *Bam*HI-HF			
Linearization	37°C	30 min	
Inactivation	80°C	5 min	
	4°C	hold	

Finally, the cDNAs were purified as described before and resuspended in 21 µL nuclease-free water. To amplify the final libraries adding the Illumina-specific sequences PCRs with the Solexa primers (**SuppTable 1**), P3 and P5, and AccuPrime™ SuperMix I were performed by mixing 5 µL cDNAs with 35 µL master mix and PCR cycler settings described in **Table 24** and **Table 25**.

Table 24: Master mix to amplify iCLIP libraries for Illumina sequencing with AccuPrime™ SuperMix I.

Component	Volume
5 µM P3 primer	0.5 µL
5 µM P5 primer	0.5 µL
AccuPrime™ SuperMix I	20 µL
Nuclease-free water	14 µL

Table 25: PCR cycler setting to amplify iCLIP libraries for Illumina sequencing with AccuPrime™ SuperMix I.

Step	Temperature	Time	Cycles
Initial denaturation	94°C	2 min	
Denaturation	94°C	15 sec	25x
Annealing	65°C	30 sec	
Elongation	68°C	30 sec	
Final elongation	68°C	3 min	
	4°C	hold	

To validate the iCLIP libraries after PCR amplification, 10 µL of each library were mixed with 6X DNA loading dye and separated on a 6% Novex™ TBE Gel beside 2.5 µL of Thermo Scientific™ O'GeneRuler Ultra Low Range DNA Ladder and Thermo Scientific™ O'GeneRuler 50 bp DNA Ladder. The final libraries were subjected to Illumina sequencing on a HighSeq2500 machine (single-end, 75 nt reads) aiming for 20 million reads per sample at the Deep Sequencing Facility (A. Dahl, Center for Molecular and Cellular Bioengineering, TU Dresden, Germany).

iCLIP sequencing data was analyzed using the iCOUNT package (chapter 4.5.8) with default options as defined by Müller-McNicoll and Ruiz de Los Mozos. Briefly, PCR duplicates were removed

utilizing the individual barcodes before raw reads were trimmed off barcodes and adapters before mapping against the murine genome (mm9, version Ensembl59) using Bowtie version 0.12.7. Replicates were merged using all uniquely mapping reads, if they clustered well together, to increase the number of statistically significant crosslinks (x-links). X-link sites, representing the first nucleotide of each read, were extracted and significant events (false discovery rate (FDR) < 0.5) were calculated as described by König et al., 2010; Wang et al., 2010; Yeo et al., 2009, based on normalized amounts of x-link events. To quantify significant x-link events within genes, those events were counted into transcript regions using the iCOUNT annotate and segments utilities based on murine transcript coordinates (mm9, Ensembl59).

RNA metaprofiles of obtained crosslink sites around identified PASs were generated by extracting the crosslink positions within a window of -400 to 100 nt around the respective PAS, followed by a two-step normalization for each PAS-type and each protein investigated (SRSF3, SRSF7, CPSF5 and FIP1). First, the summarized crosslink sites were divided by the total number of PASs (pPASs, dPASs or sPASs, respectively) to enable comparison between these three categories. In addition, the datasets were normalized to the total number of crosslinks derived from each iCLIP library to enable comparison between different iCLIP libraries. Finally, the binding signals were smoothened by the `LOESS` function.

To compare Metaprofiles around PASs containing alterable usage in KD or neuronal differentiation samples with Metaprofiles of unchanged PASs matching background sets were generated and analyzed. Briefly, for each category's background set the same quantity of unchanged PASs was sampled randomly 100 times extracting the iCLIP signal of the protein of interest. Following, the mean and standard deviation of each position was calculated for any of the 100 background profiles, which were further used to determine the z-score for each position in the proteins binding profile within the set of changed PASs usage. Subsequently, the z-scores were calculated into *P* values and corrected by the Benjamini-Hochberg procedure. Finally, z-scores with a False-detection-rate ≤ 5% were identified as iCLIP signals at the respective subset of PASs with significant difference between changed and unchanged PASs usage. Those events were plotted.

To identify motifs around PASs, enriched k-mers were analyzed based on the z-score (Wang et al., 2010). Briefly, sequences around significant iCLIP x-links were extended by 30 nt applying two windows ranging from -30 nt to -5 nt upstream and 5 nt to 30 nt downstream of each crosslink site. Within each interval any k-mers were counted and weighted. As a control, the significant x-link events within the same genes were randomly shuffled 100 times. Z-scores were calculated using both datasets (https://github.com/tomazc/iCount) and the top 25 enriched pentamers were aligned to obtain the *in*

vivo consensus binding motif. The on-line tool WebLogo was used to generate sequence logos (https://weblogo.berkeley.edu/logo.cgi).

4.4 Databases

4.4.1 Ensembl

The Ensembl database (https://www.ensembl.org/index.html, (Zerbino et al., 2018) is a genome browser collecting genome data of vertebrates to support genomic research run by the European Bioinformatics Institute (EBI) and the Wellcome Trust Sanger Institute. The murine assembly GRCm38.p6 (database version 96.38) was used as reference to align high-throughput sequencing data and to match gene/transcript IDs with gene names.

4.4.2 GENCODE

The GENCODE database (https://www.gencodegenes.org/, (Frankish et al., 2019) is a database mainly run by the Wellcome Trust Sanger Institute identifying and mapping all protein encoding genes. It was launched by The National Human Genome Research Institute (NHGRI) containing datasets for human and mouse. The chromosome comprehensive gene annotation file of the mouse releases M18 and M21 were used to align high-throughput sequencing data.

4.4.3 PubMed

The PubMed database of the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/pubmed/, CITATION) comprises abstracts, full-texts, and links of more than 29 million biomedical articles including related scientific fields. In addition, it connects to relevant websites, molecular biology resources and tools of the NCBI. It was used to find relevant literature and DNA/protein sequences.

4.4.4 RNA modification database RMBase v2.0

The RMBase database (Version 2.0, http://rna.sysu.edu.cn/rmbase/index.php) (Xuan et al., 2018) connects data derived from epitranscriptome sequencing to comprehensively explore post-transcriptionally modified RNAs. The dataset expands over 13 species from 566 sample datasets summarizing around 1.5 million modification sites including more than 100 RNA modifications. The RMBase was used to perform an initial screen for N6-Methyladenosine (m6A) modification sites in mouse transcripts' 3' UTRs based on the mm10 assembly.

4.4.5 UniProt

The UniProt database (https://www.uniprot.org/) (The UniProt Consortium, 2019) collects information about proteins of all life forms including viruses. The database is run by the European Bioinformatics Institute (EBI), Swiss Institute of Bioinformatics (SIB) and Protein Information Resource (PIR). It contains comprehensive information about peptide sequences, structures, and functions. UniProt was used to retrieve protein sequences and to screen CPA factors for potential RS and RS-like domains.

4.5 Programs

4.5.1 Adobe Photoshop CC

Photoshop CC by Adobe (Version: 2015, www.adobe.com) is a raster graphic editor program, which can process many types of input data and offers many tools for image processing. Photoshop CC was used to crop and invert TIF and JPEG files of agarose gels.

4.5.2 Adobe Illustrator CC

Illustrator CC by Adobe (Version: 2019, www.adobe.com) is a vector based graphic program, which enables the generation of high-quality figures with adjustable size without quality loss. The program was used to generate all figures in this thesis.

4.5.3 Citavi6

Citavi6 (Version 6.3.0.0, https://www.citavi.com/de) is a program to manage and organize literature and references in individual projects. The program also enables to share personal libraries with colleagues and collaborators. It was used to generate and manage the bibliography of this thesis.

4.5.4 Clustal Omega

Clustal Omega (ClustalO, Version 1.2.4, https://www.ebi.ac.uk/Tools/msa/clustalo/) (Madeira et al., 2019) is a multiple sequence alignment program run by the EBI to align three to 4,000 protein, DNA or RNA sequences at the same time. Therefore, it uses seeded guide trees and a Hidden Markov model engine. ClustalO was used to align protein sequences of SR proteins.

4.5.5 Database for Annotation, Visualization, and Integrated Discovery

The Database for Annotation, Visualization and Integrated Discovery (DAVID, Version 6.7, https://david-d.ncifcrf.gov/) (Huang et al., 2009b, 2009a) is an online collection of tools to functionally annotate and investigate large gene lists from various outputs. It enables the identification of enriched gene ontology (GO) terms as well as the discovery of enriched functional-related groups. DAVID was used for functional annotation and GO term analysis of differential expressed transcripts identified by DaPARS.

4.5.6 DEQOR

DEQOR (Version 3, http://144.76.155.9/deqor_new/input.html) (Surendranath et al., 2013) is an on-line based tool to select target regions with minimum cross-regulation and off-target effects for esiRNA preparation. The algorithm was used to choose target regions to perform efficient knockdowns of *Srsf3*, *Srsf7*, *Nudt21*, *CPSF6* and *FIP1l1* based on the respective DNA sequences derived from mus musculus Refseq annotations.

4.5.7 Fiji/ImageJ

Fiji (Fiji is just ImageJ, Version 1.52n, https://fiji.sc/) (Schindelin et al., 2012; Rueden et al., 2017) is an image processing tool, which can be used on all operating systems. Fiji is an image-processing package with many plugins added on top of ImageJ. The program was used to quantify expression of transcript 3' UTR isoforms after 3'RACE-PCR from agarose gels pictures.

4.5.8 iCount

iCount analysis was performed by Romas Curk, Igor Ruiz de Los Mozos, Michaela Müller-McNicoll and Oliver Schwich.

iCount (https://github.com/tomazc/iCount) (Curk et al., 2019) is a python-based tool enabling to process iCLIP data and analyze transcriptome-wide protein-RNA interactions. Furthermore, it identifies and quantifies the sites of protein-RNA interaction on bound RNA. This script was used to analyze the iCLIP datasets of FIP1 and CPSF5, generated in this thesis as well as SRSF7D27aa, SRSF7ZnMut, SRSF3 and SRSF7, that have been generated previously (Müller-McNicoll et al., 2016; Königs et al., 2020).

4.5.9 Integrative Genomics Viewer

The Integrative Genomics Viewer (IGV, Version 2.3, https://software.broadinstitute.org/software/igv/) (Robinson et al., 2011) is a high-performance tool by the Broad Institute to visualize genomic datasets. It enables the exploration of large datasets from different data types. The IGV was used to visualize BAM/BED files and produce browser shots and sashimi plots (Katz et al., 2014) from RNA-Seq, iCLIP and MACE-Seq datasets.

4.5.10 Image Lab

Image Lab (Version 5.2) is a software package developed by Bio-Rad to acquire, analyze and export gel and blot images. The program was used in combination with a Gel Doc XR+ to image and export Western blots.

4.5.11 Motic Image Plus 3.0

The Motic Image Plus 3.0 software enables imaging, measurement, and quantification of cells in combination with a Motic Cam. The software was used to image P19 cell lines.

4.5.12 PikoReal™ Software 2.2

The PikoReal™ program (Version 2.2) by Thermo Scientific enables PC-driven experiments in combination with the PikoReal™ Real-Time PCR System. It is used to configure the instrument, generate, and manage programs and to store and finally analyze the data. The program was used to define and run protocols. Results were exported and processed with Microsoft Office Excel.

4.5.13 Primer-BLAST

The Primer-BLAST tool by the National Library of Medicine (NLM) is hosted at the NCBI homepage (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). It uses Primer3, followed by an automated BLAST to design specific primers according to sequence templates and individual parameters. The tool was used to design qRT-PCR and 3'RACE forward primers BLASTing against the murine Refseq mRNA database.

4.5.14 R and RStudio

R (Version 3.5.1, https://cran.r-project.org/) is a program language used in statistical calculation and graphics, which can be run in a terminal. To improve usability, the graphical user interface environment RStudio (Version 1.1.463, https://www.rstudio.com/) was used. In addition, R can be supplemented with many open-source packages to suit individual needs. For this thesis, R and RStudio were used to analyze and to generate plots based on RNA-Seq datasets using packages summarized in **Table 26**. Custom R scripts were written to run DESeq2 analysis and plot figures.

Table 26: List of used R packages stating the versions used and the source.

Package	Version	Source	Author/Citation
biomaRt	2.38.0	http://bioconductor.org/packages/release/bioc/html/biomaRt.html	(Durinck et al., 2005; Durinck et al., 2009)
DESeq2	1.22.2	http://bioconductor.org/packages/release/bioc/html/DESeq2.html	(Love et al., 2014)
genefilter	1.64.0	http://bioconductor.org/packages/release/bioc/html/genefilter.html	(Gentleman et al., 2019)
GenomicRanges	1.34.0	http://bioconductor.org/packages/release/bioc/html/GenomicRanges.html	(Lawrence et al., 2013)
RColorBrewer	1.1-2	https://cran.r-project.org/web/packages/RColorBrewer/index.html	Erich Neuwirth
rtracklayer	1.42.1	http://bioconductor.org/packages/release/bioc/html/rtracklayer.html	(Lawrence et al., 2009)
pheatmap	1.0.12	https://cran.r-project.org/web/packages/pheatmap/index.html	Raivo Kolde

4.5.15 SnapGene®

SnapGene® (Version 4.1.9, https://www.snapgene.com/) is a genetic engineering software to visualize genetic procedures and Sanger sequencing results. In addition, the program enables *in situ* planning and documentation of cloning experiments by supplying useful resources including common cloning methods and plasmid information. The tool was used to visualize plasmid data, validate Sanger sequencing results, and design Gibson Assembly cloning experiments.

4.5.16 UCSC Genome Browser

The UCSC Genome Browser (http://genome.ucsc.edu/index.html) (Kent et al., 2002) is run by the University of California, Santa Cruz (UCSC) enabling online access to diverse genome sequence data from many vertebrate and invertebrate species. Furthermore, it allows uploading custom tracks and enables the export of high-quality browser shots. The UCSC Genome Browser was used to analyze transcript isoforms of genes of interest and to scan iCLIP datasets including the download of browser shots.

4.5.17 USCF Chimera

UCSF Chimera (Version 1.1.12, https://www.cgl.ucsf.edu/chimera/) (Pettersen et al., 2004) is a potent tool for the interactive visualization and analysis of molecular structures. It is developed by

the Resource for Biocomputing, Visualization and Informatics and supported by the NIH. It was used to visualize the structure of CFIm.

4.5.18 ZEISS Efficient Navigation

The ZEISS Efficient Navigation (ZEN, Version 2.3, https://www.zeiss.de/mikroskopie/produkte/mikroskopsoftware/zen.html) software enables operating, imaging and image processing using ZEISS microscopes. The software was used in combination with a ZEISS LSM780 to image P19 cell lines.

4.6 Data availability

The complete set of RNA-Seq, MACE-Seq and iCLIP datasets which were generated during this thesis were submitted to Gene Expression Omnibus (GEO) under the reference SuperSeries accession GSE151724.

The mass spectrometry data has been deposited under the reference PXD018090 at the ProteomeXChange Consortium using the PRIDE partner archive.

4.7 Bioinformatics

4.7.1 BedTools

BedTools (Version 2.27.1, https://github.com/arq5x/bedtools2) (Quinlan & Hall, 2010) is a collection of a wide range of genomic analysis tools. For example, it allows simple tasks like marking, counting, or intersecting datasets present in various used file formats. Furthermore, BedTools enables transformation of BAM files into other file formats as for example bedgraph files. BedTools was used to generate bedgraph files from BAM files of mapped RNA-Seq datasets to proceed with DaPARS analysis.

4.7.2 Cutadapt

Cutadapt (Version 1.16, https://cutadapt.readthedocs.io/en/stable/index.html) (Martin, 2011) is a tool to find and remove unwanted sequences from high-throughput sequencing reads, for example adapter sequences or poly(A) tails. Furthermore, the tool can be used to trim the reads to identical lengths and remove nucleotide-based reads with bad quality. Cutadapt was used to trim raw reads derived from MACE-Seq using the following options:

```
cutadapt -q 25 -O 8 -a "A{100}" --minimum-length 25
```

4.7.3 Dynamic analysis of Alternative PolyAdenylation from RNA-Seq (DaPARS)

The dynamic analysis of alternative polyadenylation from RNA-Seq (DaPARS, Version 0.9.1, https://github.com/ZhengXia/dapars) (Masamha et al., 2014; Xia et al., 2014) algorithm is a python2.7 based tool to directly extract and quantify *de novo* dynamic alternative polyadenylation usage from standard RNA-Seq datasets in a two-step approach. DaPARS relies on a given annotated gene model which was generated during the first step on the GENCODE vM18 annotation using BASH on Ubuntu on Windows shell and the following command line:

```
python DaPARS_Extract_Anno.py -b vM18_gencode_whole_gene.bed -s genecode-genesymbol-mm10.txt -o mm10_gencode_extracted_3UTR.bed
```

For each transcript, DaPARS identifies *de novo* the distal PAS based on a continuous signal independent approach in the first step (**Figure 14**). Next, the algorithm assumes an alternative upstream proximal PAS. To *de novo* define this site, DaPARS uses a linear regression model to identify an optimal fitting point (red box) representing the local read-density change between control and knockdown samples. The alternative polyadenylation usage is summarized in the Percentage of Distal poly(A) site Usage Index (ΔPDUI), identifying 3' UTR lengthening as positive values and 3' UTR shortening as negative values.

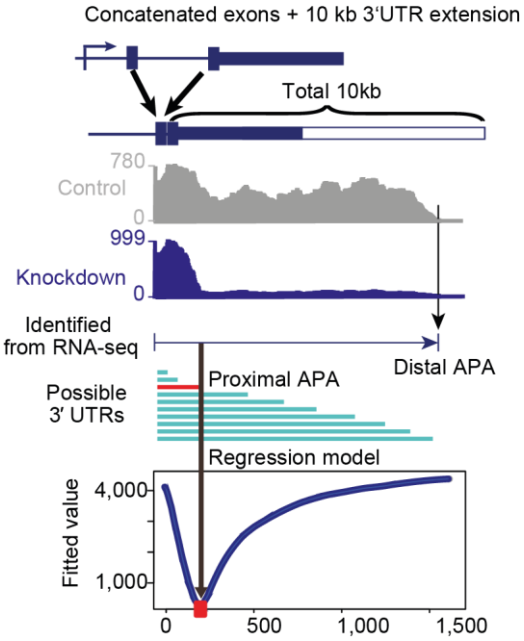


Figure 14: Scheme of the identification of dynamic alternative polyadenylation usage between control and knockdown samples. *The dPAS was identified from combined RNA-Seq data, while the pPAS was extracted via the minimum fitted value of the fitting value using a regression model. Modified from (Xia et al., 2014).*

DaPARS requires bedgraph files as input. Therefore, mapped RNA-Seq bam-files where generated using BedTools using the following command line:

```
genomeCoverageBed -bg -ibam sample_sorted.bam -g mm10chromsize.txt -split > sample.bedgraph
```

Next, the result tables were generated using a configure file setting all relevant parameters and calling the required input files, comparing RNA-Seq datasets after depletion of SRSF3 or SRSF7 against control samples. Finally, the result tables were explored, and figures were prepared using R and RStudio. DaPARS was used to analyze dynamic changes in alternative polyadenylation from RNA-Seq datasets after depletion of SRSF3, SRSF7 compared to control samples and between neuronal differentiated P19 cells compared to P19 wild type cells.

4.7.4 DESeq2

DESeq2 (**Table 26**) is an R package to analyze and explore differential gene expression between RNA-Seq experiments using count-based datasets. The tool normalizes the read counts and tests for differential expression using negative binomial linear models. Furthermore, DESeq2 includes possibilities to assess data and replicate quality by sample clustering and visualization. DESeq2 was used to analyze differential expressed genes after depletion of SRSF3 or SRSF7 compared to control samples and to access the quality between biological replicates.

4.7.5 FASTQC

FastQC (Version 0.11.5, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) is a Java-based tool for quality control of raw high-throughput sequencing data derived from a variety of sequencing platforms. It comprises a set of different analysis modules to gain brief insight into various aspects of RNA-Seq data quality, such as per base sequencing quality, sequence length distribution and overrepresented sequences e.g., not trimmed adapters. FastQC was used to assess the quality of raw data derived from RNA- and MACE-Seq experiments.

4.7.6 MACE-Seq data analysis, identification of PASs from MACE-Seq datasets and motif analysis around PASs

4.7.6.1 MACE-Seq data analysis

MACE-Seq libraries and raw-data were produced by GenX-Pro, Frankfurt, Germany and MACE-Seq data analysis was performed by Mario Keller, Buchman Institute for Molecular Life Sciences, Germany as described in Müller et al., 2014. Briefly, low-quality regions with a Phred-score <16 were trimmed from both ends. Following, Poly(A)-tail (reads of at least 10 adenosines) were detected by using a 5-nt sliding window detecting trailing A's while allowing one nucleotide other than A per window to accept minor sequencing errors. Finally, Poly(A)-tails were removed as well, and the remaining reads were mapped against the mouse reference genome (mm10) using Novacraft's commercial aligner Novoalign (http://www.novocraft.com/products/novoalign/). To only receive uniquely mapped reads the `-r none` option was used and soft clipping was disengaged by `-o FULLNW`. To exclude reads derived from internal priming at intergenic poly(A)-stretches the Novoalign derived reads were mapped again against the mouse reference genome using Bowtie (Langmead, 2010) while permitting two errors (`-v 2, -r all`). Subsequent, the files were transformed to BED-files merging protruding intervals using BedTools (chapter 4.7.1). All reads beside homopolymeric poly(A)-stretched were removed.

4.7.6.2 Identification of PASs from MACE-Seq datasets

PASs from MACE-Seq datasets were identified by the work of Mario Keller and Samarth Thonta Setty, Buchman Institute for Molecular Life Sciences, Frankfurt, Germany.

First, to identify clusters of reads the 3'end coordinates derived from the cleaned datasets from all three sets of samples (WT, KD *Srsf3* and KD *Srsf7*, duplicates each) were combined. Clusters were defined by repeatedly extending the cluster to the successive 3'end coordinate when this was ≤25 nt apart from the median 3'end coordinate of the cluster. To precisely map PASs all cleavage events defined by the position +1 nt downstream of the 3'end coordinate derived from the cleaned

reads were summed up and the resulting cluster were resized to a window of 15 nt centered on mode by cleavage events per nucleotide inside the respective cluster. Following, cleavage events (CE) of each sample were recounted into the 15-nt windows of these defined clusters (CE_{window}). Respective replicates were averaged ($CE_{average}$). Subsequent, the percent usage (PU) for each 15-nt window was calculated by:

$$PU = \frac{CE_{window}}{\sum CE_{Gene}} \times 100$$

As before, PU of the respective replicates were averaged ($PU_{average}$). Last the transcripts-per-million (TPM)-metric for each 15-nt window was calculated by:

$$TPM = \frac{CE_{window}}{\frac{Length_{Window}[kb]}{1x10^7\ reads}}$$

$TPM_{average}$ was calculated to the respective replicates, which was used to determine the maximum across all windows present in the respective gene ($TPM_{maximum}$). In the following the CE_{window}, $PU_{average}$ and $TPM_{maximum}$ values were used to apply the following filters to all three conditions connected by the logical operator `AND` as summarized below:

$$CE_{window} > 4\ CEs\ AND\ PU_{average} > 5\%\ AND\ TPM_{maximum} > 0.25$$

Concluding, the positions of the detected PASs were annotated based on GENCODE vM18. Here, 15-nt windows overlapping more than one gene or which were located more than 50 nt downstream of a gene were excluded. The center of the respective 15-nt windows were defined as cleavage site further on.

Subsequent these defined PASs were further categorized by assigning them to transcript regions. Therefore, each PAS was analyzed for location within 5'UTR, intronic, CDS or 3' UTR regions as included to the GENNCODE annotation. All PASs found in more than one region were categorized as "Other". This was not the case if the PAS was found in the 3' UTR or introns. Furthermore, PASs located up to 50 nt downstream of any gene were associated to the respective 3' UTR.

pPAS, dPAS or intermediate/other PAS (oPAS) were further categorized by designating the most upstream PAS as pPAS and the furthest downstream PAS as dPAS. Any remaining PAS was defined as oPAS. As a special case any sole PAS within a given gene was designated as single PAS (sPAS).

4.7.6.3 Motif analysis around PASs

Motif analysis around PAS was done by Mario Keller, Buchman Institute for Molecular Life Sciences, Frankfurt, Germany.

All CSE variants as reported by Gruber et al., 2016 were looked up within a window of 50 nt upstream of each PAS. Subsequent, a single CSE was chosen by applying the following pecking order, based on strength associated with the respective motif:

AAUAAA >> AUUAAA > remaining hexamers

PAS coordinates derived from MACE-Seq, as described above, were integrated with APA changes, determined by DaPARS, by matching MACE-Seq coordinates with PAS reported from DaPARS. The match was rules valid when the DaPARS PAS located within a window of -250 nt upstream or +50 nt downstream of the respective MACE-Seq PAS coordinate. Furthermore, the type of PAS (either pPAS or dPAS) had to be the same between both datasets for the respective PAS. If DaPARS presented multiple PASs matching the same MACE-Seq-PAS the one being closest to the MACE-Seq coordinate was chosen for the downstream analysis.

In addition, UGUA motifs were analyzed. Therefore, UGUA motifs were identified for any PAS within a window of +/- 300 nt up- and downstream of the PAS coordinate. Following, fractions of PASs containing UGUA motifs were calculated based on their precise distance to the respective PAS followed by loess smoothing to generate Metaprofiles of UGUA motif distributions. To further determine the fraction of PASs either containing two UGUA motifs upstream of the PAS (UGUA-UGUA-PAS), being enclosed by tandem UGUA motifs (UGUA-PAS-UGUA) or containing two UGUA motifs downstream of the PAS (PAS-UGUA-UGUA) windows of incrementing size (Win_{size}, increasing by 1-nt steps) were applied. For PAS containing two UGUA motifs upstream or downstream, the Win_{size} was between 1-150 nt. The UGUA motifs upstream had to begin in a range of -Win_{size} nt up to – 4 nt upstream of the PAs, while the downstream UGUA motifs must be between 1 nt and Win_{size}-3 nt downstream of the PAS. PAS enclosed by an upstream and a downstream UGUA motif had to be within the following ranges: upstream [-Win_{size} nt; -4nt] and downstream [-3 nt; Win_{size}-3 nt]. The minimal distance between two UGUA motifs has been reported for each PAS.

4.7.7 Mixture-of-Isoforms

Mixture-of-Isoforms (MISO, Version 0.5.4, https://miso.readthedocs.io/en/fastmiso/) (Katz et al., 2010), is a python/C based probabilistic tool quantifying expression levels of various alternative transcript isoforms, generated b alternative splicing, alternative transcription start sites or alternative polyadenylation, from RNA-Seq data. MISO uses Bayesian interference to predict probabilities and is based on a Markov Chain Monte Carlo simulation. In difference to DaPARS, MISO cannot handle biological replicate datasets. Therefore, replicate datasets were merged before running MISO. The tool was used to analyze tandem 3’ UTR alternative polyadenylation in combination with supplied

annotations of mouse genome (mm10) alternative events version 1.0 (https://miso.readthedocs.io/en/fastmiso/annotation.html). The 'TandemUTR' events originated from the PolyA database (PolyA_DB, http://exon.umdnj.edu/polya_db/) (Zhang et al., 2005a; Lee et al., 2007).

4.7.8 Samtools

Samtools (Version 1.6, https://github.com/samtools/samtools) (Li et al., 2009) is a collection of useful functions to manipulate sequence alignment files by e.g., indexing, sorting and merging. It was used to index BAM files derived from mapping RNA- and MACE-Seq data files using the STAR aligner and for generating BAI files. In addition, the merge option was used to merge RNA-Seq annotation files from replicates to be used in downstream MISO analyses.

4.7.9 Spliced Transcripts Alignment to a Reference

The Spliced Transcripts Alignment to a Reference alignment tool (STAR, Version 2.6.1d, https://github.com/alexdobin/STAR) (Dobin et al., 2013) is a GCC C++ run ultrafast universal RNA-Seq aligner with many additional options. For example, STAR can generate gene count tables to bypass downstream usage of other tools like HTSeq. STAR was run on a computational cluster, operated by Ingo Ebersberger (Uni Frankfurt, Germany). Before mapping raw RNA-Seq datasets, genomic index files must be generated. Here, reference index files based on the GENCODE vM18 were generated using the `–runMode genomeGenerate` option with additional settings `—sjdbOverhang 100 --limitGenomeGenerateRAM 94498394837`. RNA- and MACE-Seq datasets were mapped against the vM18 reference using the following options `--outSAMattributes All --outSAMtype BAM SortedByCoordinate --runThreadN 2 --outFilterMismatchNmax 2 --readFilesCommand zcat --quantMode GeneCounts`

5 Results

Recently, a study from our group discovered that knockdown (KD) of SRSF3 and SRSF7 affected the length of 3' UTRs in hundreds of transcripts in pluripotent P19 cells (Müller-McNicoll et al., 2016). Those findings were based on RNA-Seq data analyzed with two tools that quantify changes in alternative splicing - DEXSeq and MISO (Anders et al., 2012; Katz et al., 2010). Both tools are less powerful for the quantification of changes in alternative polyadenylation (APA), therefore the experiments were repeated and RNA-Seq was analyzed with DaPARS (Masamha et al., 2014; Xia et al., 2014). To further characterize the poly(A)-tome, the set of all transcribed poly(A) sites, of P19 cells a 3'end-Seq analysis (MACE) was included. Moreover, the underlying mechanism(s) of 3' UTR-APA regulation by SRSF3 and SRSF7 were investigated. The results are presented in the following sections.

5.1 Definition of the poly(A)-tome of P19 cells

In the first part of this thesis, we aimed to map all used PASs in the P19 transcriptome, termed the poly(A)-tome, to investigate how SRSF3 and SRSF7 affect PAS usage. To do so, total RNA samples were submitted to 3'-end sequencing by MACE-Seq (**Figure 15**A). To include all SRSF3 and SRSF7-dependent changes in PAS usage, total RNA samples after KD of *Srsf3* and *Srsf7* using RNAi were used (**Figure 15**B).

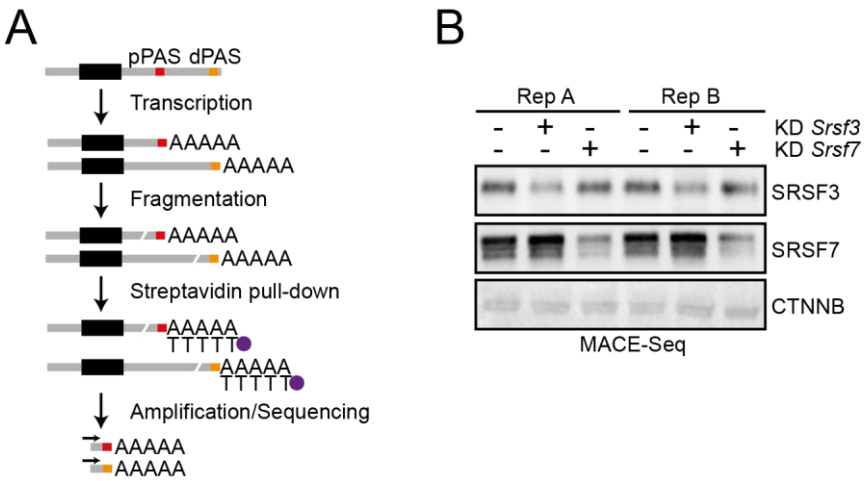


Figure 15: MACE-Seq after KD of Srsf3 or Srsf7. A) *Scheme of the MACE-Seq procedure. Transcripts are randomly fragmented and the most 3'ends purified using streptavidin-tagged oligo-dT oligonucleotides. Subsequently, high-throughput sequencing libraries are generated using these fragments and submitted for sequencing.* ***B)*** *Western blot validation of the depletion of SRSF3 and SRSF7 using specific esiRNAs. CTNNB was used as loading control.*

Five of the six MACE-Seq libraries delivered relatively equal numbers of reads, between 3.4 to 3.9 million, after UMI duplicates were removed and additional quality filtering, and only the second

replicate of *Srsf3* KD delivered more than twice as many reads (**Table 27**). To compare KD and control samples and the replicates with each other, the reads were mapped to the mouse genome (mm10) and tables with read counts per gene were generated simultaneously using the STAR aligner (Dobin et al., 2013). Using this STAR output a data quality assessment was performed with DESeq2 (Love et al., 2014) to test the data for suitability for further analysis. Briefly, all genes without raw counts were removed and remaining raw counts of the replicates were transformed using the regularized logarithm and plotted against each other. Furthermore, the correlation between replicates was calculated using the Spearman's rank correlation method followed by calculations and hierarchically clustering of the distance between the replicates of each sample. This was summarized by plotting heatmaps and revealed that all biological replicates behave similarly and the KD and Control samples were well separated (**Figure 16**A&B).

Table 27: Statistics of MACE-Seq reads after KD of Srsf3 and Srsf7.

Sample	Replicate	Number of reads	Number of uniquely mapped reads	% uniquely mapped reads
KD *Srsf3*	A	3876796	2103224	54.25
	B	8756862	4902181	55.98
KD *Srsf7*	A	3501881	1912095	54.60
	B	3473287	1896403	54.60
Ctrl	A	3376907	1816124	53.78
	B	3384179	1858756	54.92

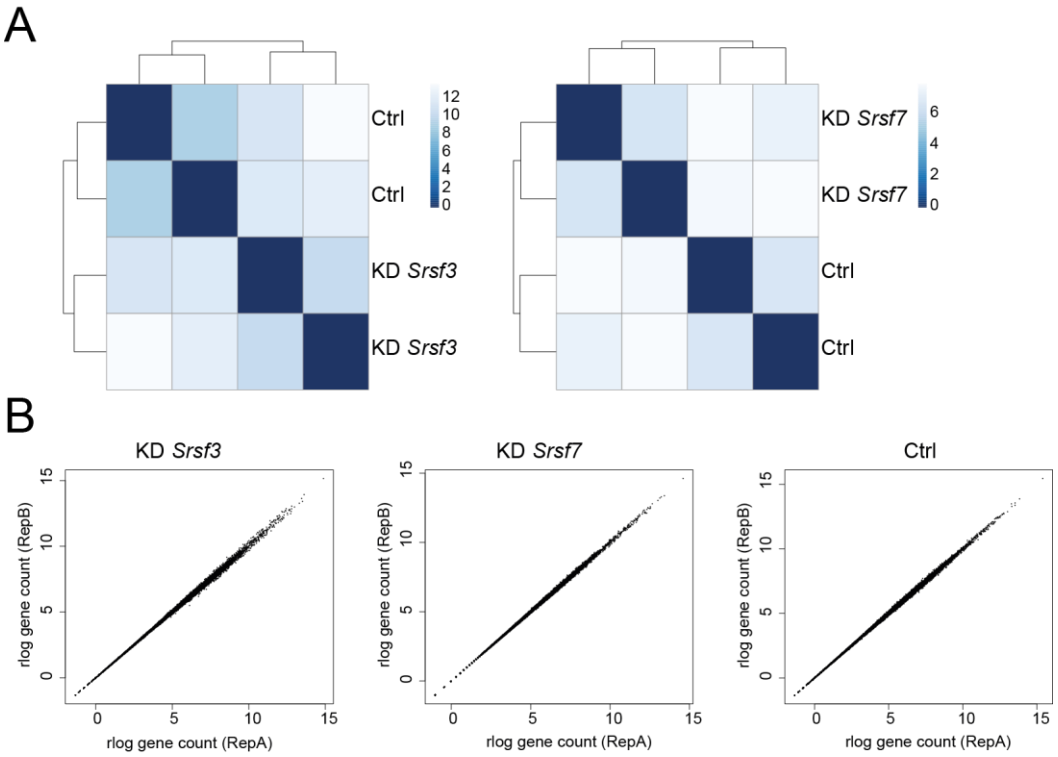


Figure 16: Separation between samples of different conditions and reproducibility between biological replicates of MACE-Seq samples. A) *Heat maps of the Euclidean samples distance after rlog transformation for the different conditions including both corresponding replicates.* ***B)*** *Scatter plots of gene counts (rlog transformed) for both corresponding replicates of the respective conditions.*

To identify and map all used PASs within the P19 transcriptome, the MACE reads of control and KD samples were merged and filtered, removing too low abundant PASs, using a custom pipeline (**Figure 17**A). In total, 15,886 PASs mapping to 9,148 genes were identified. 88% of those PASs mapped to protein coding genes and were enriched in the 3' UTRs (**Figure 17**B&C). The determined PAS positions agreed well with existing GENCODE annotations, with the majority of PAS (61%, 9,659/15,866) mapping within 25 nt of annotated transcripts ends (**Figure 17**D). An analysis of enriched sequence motifs in a window from 5 to 50 nt upstream of the identified PAS revealed a strong enrichment of the hexameric sequence 'AAUAAA' 41% (9,445/23,016), representing the canonical poly(A) signal (**Figure 17**E). This is followed by the major variant poly(A) signal motif 'AUUAAA', which is three times less abundant (14%; 3,163/23,016) than the canonical motif. Other variants of hexameric motifs in this region were much less frequent. Interestingly, only 4.5% (1,035/23,016) of PASs had no poly(A) signal motif.

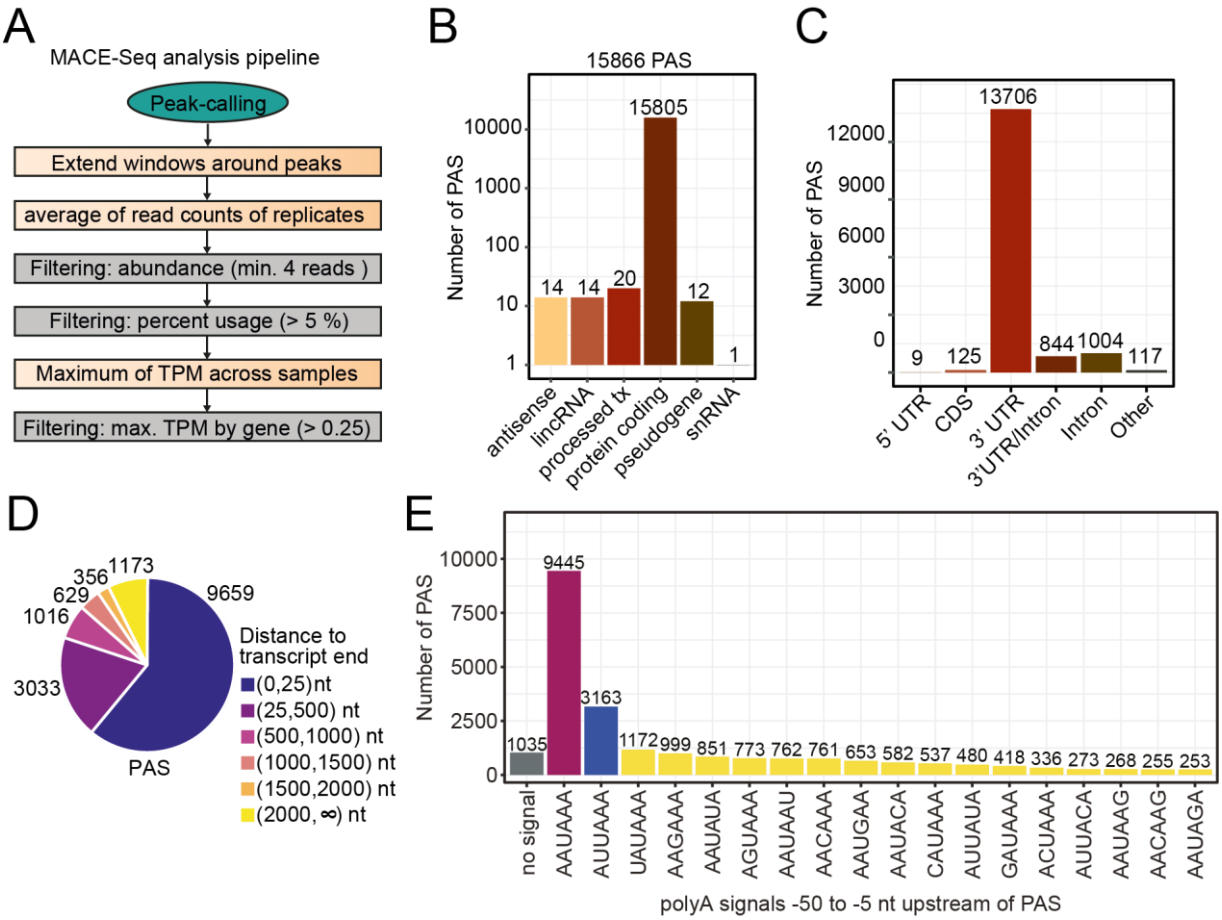


Figure 17: Definition of the P19 poly(A)-tome using MACE-Seq datasets. A) *Scheme of the custom pipeline to map used PAS sites from MACE-Seq datasets using multiple filtering steps.* ***B)*** *Number of PAS mapping to different RNA-categories.* ***C)*** *Number of PAS mapping different regions within a transcript.* ***D)*** *Distribution of PAS sorted by the distance to GENCODE*

annotated end of the transcript. ***E)*** *Number of PAS containing specific PAS signals within a window from 5 to 50 nt upstream of mapped PAS.*

Further inspection of PASs revealed that surprisingly more than half of all protein coding genes in murine P19 cells employ only one PAS (55.4%, 5,069/9,148 genes) (**Figure 18**A). In the following the PASs derived from these genes were termed single PASs (sPASs). The remaining 2,430; 1,026 and 623 protein coding genes showed clear evidence for APA containing two, three or more than three PASs, respectively. In the following PASs were defined as proximal PASs (pPASs), distal PASs (dPASs) or other PASs (oPASs) based on their relative location within the 3' UTR from the end of the coding sequence (**Figure 18**B). Comparison of locations from pPASs, dPASs and sPASs showed that the majority of pPASs are 500 to 2,000 nt upstream of dPASs, with an enrichment around 1,300 nt upstream of the respective dPAS (**Figure 18**C). Interestingly, analysis of hexameric PAS signal motifs in genes that undergo APA showed that dPASs are enriched in the AAUAAA consensus motif (67%), while pPASs and oPASs show higher frequencies of alternative motifs (~ 40%) (**Figure 18**D). The canonical polyadenylation signal is associated with the strongest cleavage and polyadenylation efficiency as it is perfectly recognized by the CPSF subunits WDR-33 and CPSF-30 (Schönemann et al., 2014; Clerici et al., 2017; Clerici et al., 2018). Alternative polyadenylation signals, which are more frequent around pPASs, are thought to be weaker, enabling the regulated expression of distinct transcript isoforms. In comparison, both major PAS signals (AAUAAA and AUUAAA) were found in more than 90% of sPAS genes, suggesting that sPAS are generally stronger.

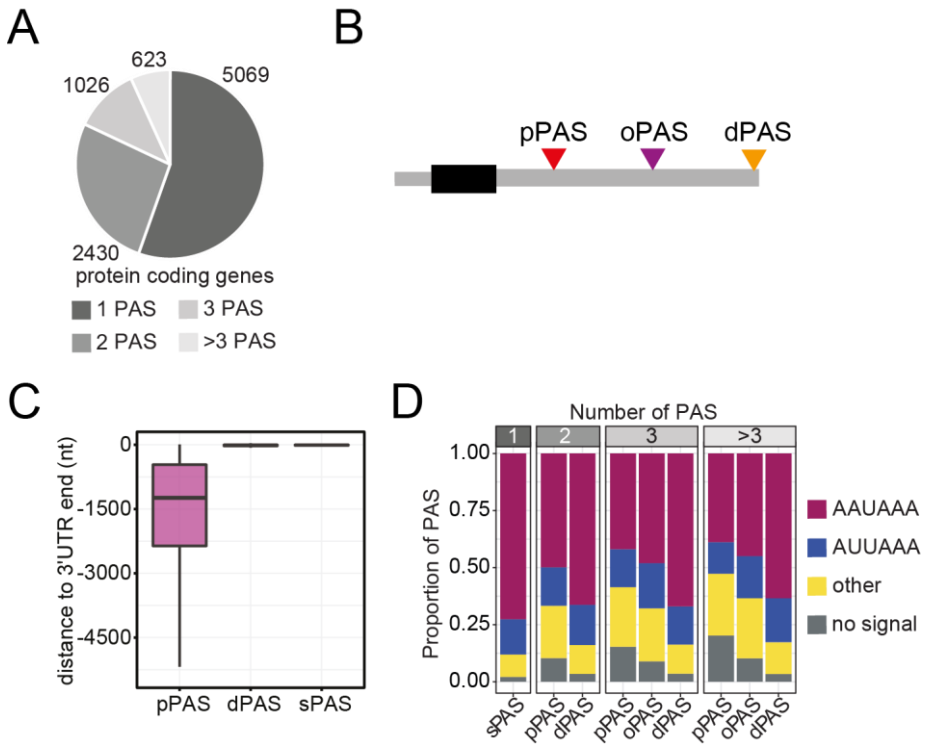


Figure 18: Characterization of PASs in APA-positive transcripts. A) *Proportion of protein coding genes containing one, two, three or more PASs.* ***B)*** *Scheme of the classification of multiple PASs within the same 3' UTR depending on their position relative to the end of the coding sequence (black box).* ***C)*** *Boxplot of the distances of pPASs, dPASs or sPASs to the annotated 3' UTR ends.* ***D)*** *Proportions of PAS motifs in transcripts separated by the number and type of PAS.*

The poly(A)-tome of P19 cells, determined by MACE-Seq, highlighted that nearly half of the transcripts contain more than one PAS giving rise to distinct transcript isoforms with different 3' UTRs. Moreover, pPASs and oPASs are hotspots for APA regulation as their accompanied polyadenylation signal motifs often vary from the consensus sequence compared to strong motifs at dPASs and sPASs. In competition with the stronger default dPASs, weaker pPASs need to be supported by *trans*-factors to facilitate cleavage at these sites.

5.2 SRSF3 and SRSF7 have opposite effects on proximal poly(A) site (pPAS) usage

To study the influence of SRSF3 and SRSF7 on PAS usage, *Srsf3* and *Srsf7* were depleted from P19 wt cells by RNAi using specific esiRNAs. Total RNA from two biological replicates was send for RNA sequencing (**Figure 19**). Sequencing reads were mapped to the mouse genome (mm10, **Table 28**) and quality assessment verified the separation of control and KD conditions (**Figure 20**A) and the high reproducibility between replicates (**Figure 20**B). Interestingly, when both KD samples were compared, they separated even more strongly from each other, suggesting that distinct sets of genes are affected by the depletion of SRSF3 and SRSF7.

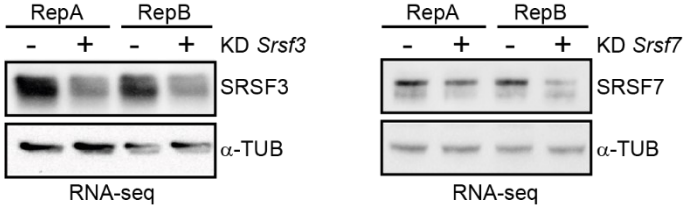


Figure 19: Western blot to validate the KD efficiency of Srsf3 and Srsf7 after transfection of specific esiRNAs. *esiRNA targeting Luciferase was used as control. SRSF3 and SRSF7 were detected using specific antibodies and expression of α-Tubulin was used as a loading control.*

Table 28: Statistics of RNA-Seq after KD of Srsf3 and Srsf7.

Sample	Replicate	Number of reads	Number of uniquely mapped reads	% uniquely mapped reads
Ctrl	A	52078735	41419011	79.53
	B	53905108	42995141	79.76
KD *Srsf3*	A	54860236	43663411	79.59
	B	64128358	50926197	79.41
KD *Srsf7*	A	51491577	41066456	79.75
	B	61083637	48758490	79.82

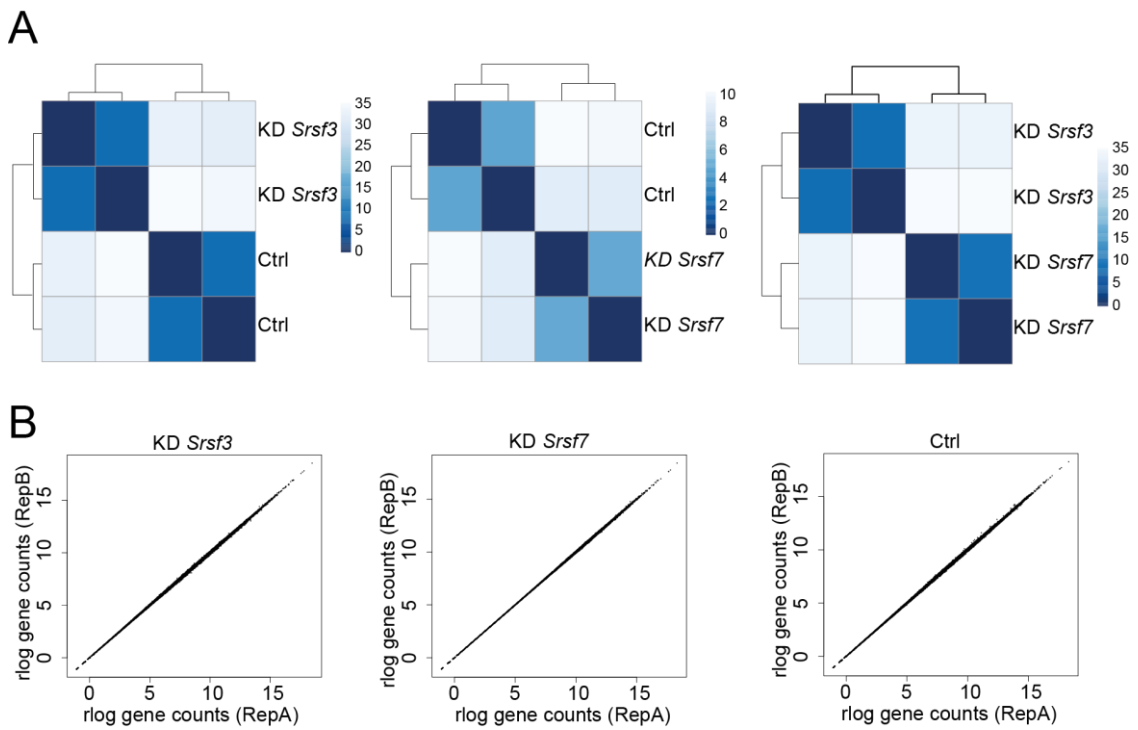


Figure 20: Separation between control and KD RNA-Seq samples and reproducibility between biological replicates. ***A)*** *Heatmaps of the Euclidean samples distance after rlog transformation for the different conditions and replicates.* ***B)*** *Scatter plots of gene counts (rlog transformed) for two replicates of the respective conditions.*

To quantify changes in APA upon SRSF3 and SRSF7 depletion from the RNA-Seq data we used a more suitable analysis tool called DaPARS (Masamha et al., 2014; Xia et al., 2014). DaPARS predicts pPAS *de novo* based on changes in read coverage using a linear regression model after identifying the dPAS and quantifies dPAS usage with the output PDUI (percentage of the distal poly(A) site usage index). For this, the mapped files were first transformed into bedgraph files and then subjected to the DaPARS algorithm. In line with results from our previous study (Müller-McNicoll et al., 2016), depletion of SRSF3 resulted in more significant APA changes (p.adjusted < 0.1), compared to depletion of SRSF7 (686 vs. 138 events), highlighted in red and blue, respectively (**Figure 21**A, **SuppTable 2**). To identify a global trend in 3' UTR length changes, the differential usage of dPASs between control and KD (ΔPDUI) was plotted (**Figure 21**B). This confirmed that KD of *Srsf3* led to a shortening of 3' UTRs (ΔPDUI < 0; 579 hits), while KD of *Srsf7* had the opposite effect (ΔPDUI > 0, 90 hits), albeit to a lesser extent. Although the PCA suggested that genes affected by both KD experiments were distinct, surprisingly, SRSF7 and SRSF3 APA targets showed a fair overlap (55/134) (**Figure 21**C). However, a more detailed analysis, including shortening or lengthening of 3' UTRs, showed that most APA events were unique, and that 17 targets showed antagonistic effects, whereby the 3' UTR was shortened after KD of *Srsf3* and extended after KD of *Srsf7* (**Figure 21**D).

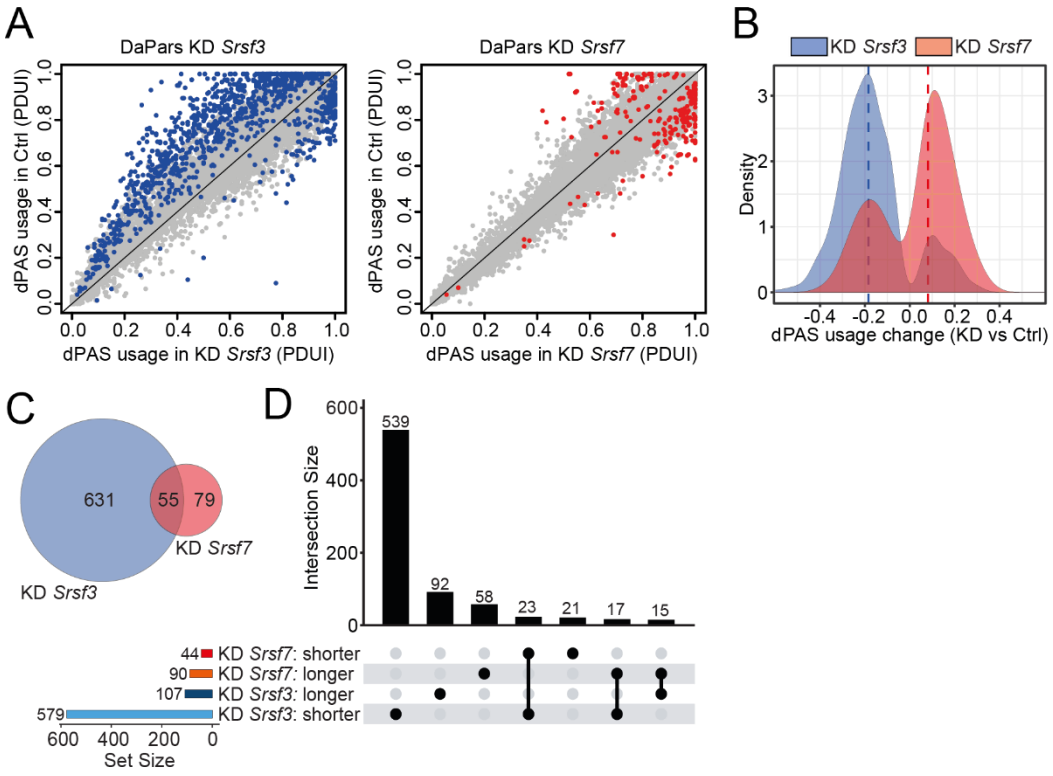


Figure 21: Analysis of changes in APA after KD of Srsf3 and Srsf7 from RNA-Seq data using DaPARS. A) *Scatterplot showing the differential analysis of dPAS usage upon KD of Srsf3 (left) or Srsf7 (right). PDUI values of Ctrl was plotted against KD conditions. Significant events (p.adj-value < 0.1) are highlighted in blue and red, respectively.* ***B)*** *Density plot representing global trends in 3' UTR length changes.* ***C)*** *Venn Diagram showing the overlap of significantly affected transcripts.* ***D)*** *Offset Plot intersecting targets differently affected by SRSF3 or SRSF7 to identify potential antagonistic regulated transcripts.*

To confirm these trends the RNA-Seq data were re-analyzed using MISO (Katz et al., 2010), which is limited to 1187 annotated tandem-3' UTR events. As seen with the DaPARS analysis, KD of Srsf3 affected more targets than KD of Srsf7 (23% vs 7.5%) and the same trend for 3' UTR shortening or extension, respectively, was observed (**Figure 22**A, **SuppTable 3**). Moreover, the percentage of overlapping transcripts was similar to the previous study (41% vs. 52%) (**Figure 22**B). Subsequently, SRSF3 APA targets were investigated for gene ontology (GO) term enrichment using DAVID (Huang et al., 2009a, 2009b). Interestingly, targets were enriched for the term 'RNA processing', 'protein localization' and 'cell cycle', suggesting that those processes are regulated by 3' UTR length via SRSF3 (**Figure 23**).

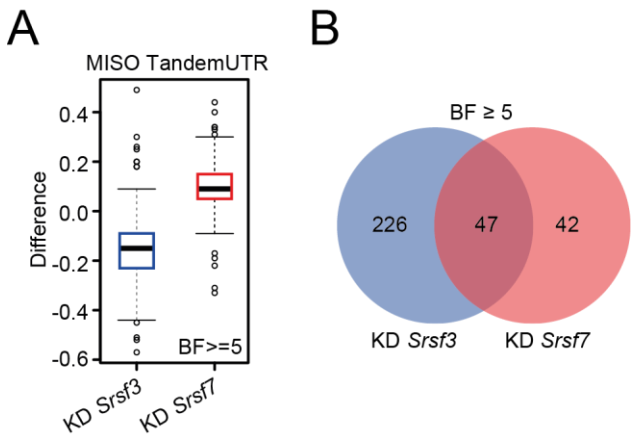


Figure 22: Analysis of Tandem 3'UTR changes after KD of Srsf3 and Srsf7 by MISO. A) *Boxplot summarizing Tandem 3'UTR length changes of significantly affected transcripts (Bayes-factor ≥ 5)* ***B)*** *Venn diagram showing the overlap of co-regulated transcripts.*

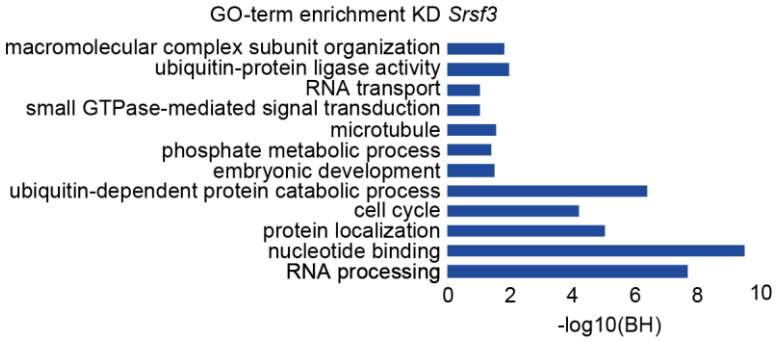


Figure 23: Enriched gene ontology terms of Srsf3 regulated APA targets. *Sorted by the negative decadic logarithm of the Benjamini-Hochberg (BH) value.*

Five antagonistically regulated transcripts (*Ddx21*, *Anp32e*, *Rab11a*, *Hspa4* and *Pphln1*) were chosen from the DaPARS datasets for validation by rapid amplification of cDNA ends (3'RACE)-PCR using gene-specific forward primers (**Figure 24**). The experiment was run in biological triplicates and quantified after agarose gel electrophoresis. For all five targets, KD of *Srsf3* significantly reduced the usage of the dPAS (orange), increasing usage of the pPAS (red) and thereby increasing the expression of transcripts with short 3' UTRs. Similar to the DaPARS and MISO analyses, KD of *Srsf7* had a milder impact, slightly increasing the levels of extended 3' UTRs in the selected transcripts.

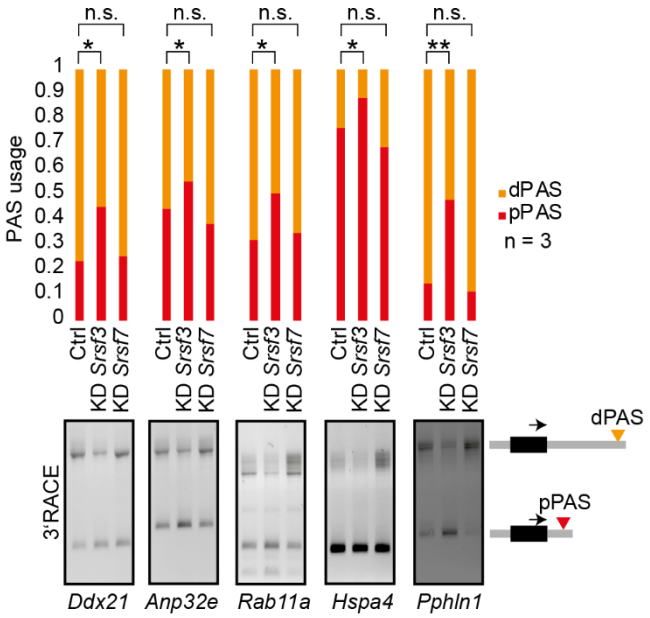


Figure 24: Validation of APA changes of selected targets after KD of Srsf3 and Srsf7 by 3'RACE-PCR. *3' UTR isoforms were amplified using gene specific forward primers, while the reverse primer bound to a common anchor-sequence introduced during RT. The PCR samples were separated on 1.5% agarose gels. The respective long and short 3' UTR isoforms, derived from usage of the proximal (red) or distal (orange) PAS, were quantified and the PAS ratio was calculated. n = 2, significance tested by Student's t-test (* = p<0.05).*

In conclusion, global transcriptome analysis using different algorithms identified all APA targets of SRSF3 and SRSF7 and confirmed their effect on 3' UTR length in opposite directions. SRSF3 emerged as a more potent regulator of APA promoting the expression of transcripts with longer 3' UTRs using the distal PAS. SRSF7 affected fewer targets and promotes pPAS usage and the expression of

transcripts with shorter 3' UTRs. Surprisingly, the number of antagonistically regulated transcripts was low, indicating that SRSF3 and SRSF7 act mostly on distinct transcripts.

5.3 SRSF3 and SRSF7 may compete for binding to proximal PAS

To investigate a potential direct involvement of SRSF3 and SRSF7 in the regulation of APA we next investigated the binding patterns of SRSF3 and SRSF7 around used and regulated PASs using previously published P19 iCLIP data sets from SRSF3 and SRSF7 (Müller-McNicoll et al., 2016).

Meta-gene analysis revealed that for transcripts bearing only a single PAS (n = 4796) iCLIP signals of both SRSF3 and SRSF7 were highly enriched approximately 75 nt upstream of the sPAS with slightly more binding of SRSF3, suggesting a direct involvement in CPA (**Figure 25**A). This picture changed for transcripts that contain more than one PAS within the same 3' UTR and undergo APA (**Figure 25**B). SRSF3 and SRSF7 showed preferential binding at pPASs with a strong enrichment 75 nt upstream of pPASs and a smaller enrichment directly downstream. Interestingly, SRSF7 binding exceeds SRSF3 at pPASs by ca. 20%, whereas the binding patterns of the two SR proteins are very similar around dPASs.

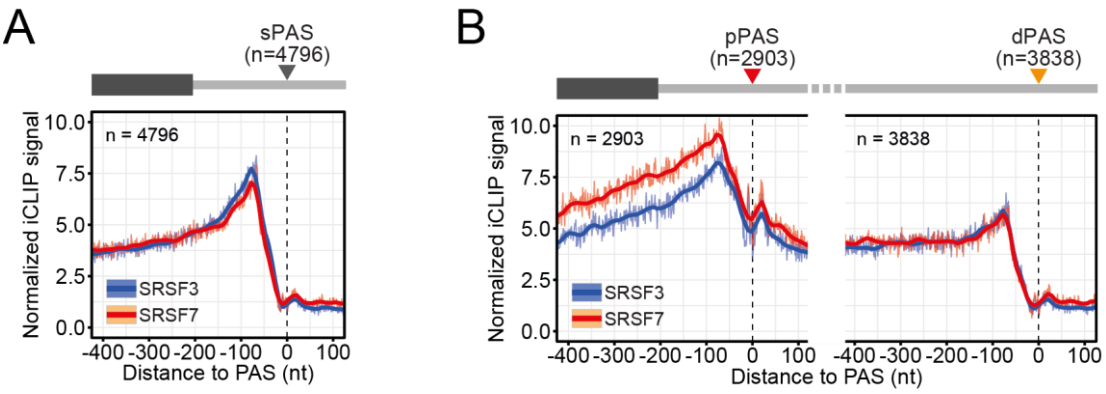


Figure 25: Binding pattern of SRSF3 and SRSF7 around sPAS, pPAS and dPAS. A) *Binding profiles of significant cross-link sites of SRSF3 and SRSF7 around sPAS.* ***B)*** *Binding profiles of significant cross-link sites of SRSF3 and SRSF7 around pPAS and dPAS in transcripts undergoing APA.*

To test whether SRSF3 and SRSF7 bind more to PAS in transcripts that are affected by their depletion, we integrated the iCLIP data with changes in 3' UTR length derived from the DaPARS analysis and with PAS coordinates derived from the MACE-Seq data (**Figure 26**A). Due to the limited number of SRSF7-depending targets identified by DaPARS (134 vs. 686; **Figure 21**) this analysis was restricted to SRSF3-depletion targets only. Indeed, binding of both SR proteins is highly enriched at SRSF3-sensitive pPASs in comparison to the same number of unaffected, random pPASs (**Figure 26**B). This enrichment was not observed at dPASs, indicating that pPASs are the main regulatory site where SRSF3 and SRSF7 affect APA. Furthermore, the higher enrichment of SRSF7 at SRSF3-responsive pPAS suggests that both proteins may compete for binding at the same pPASs to decide on the fate of the respective 3' UTRs.

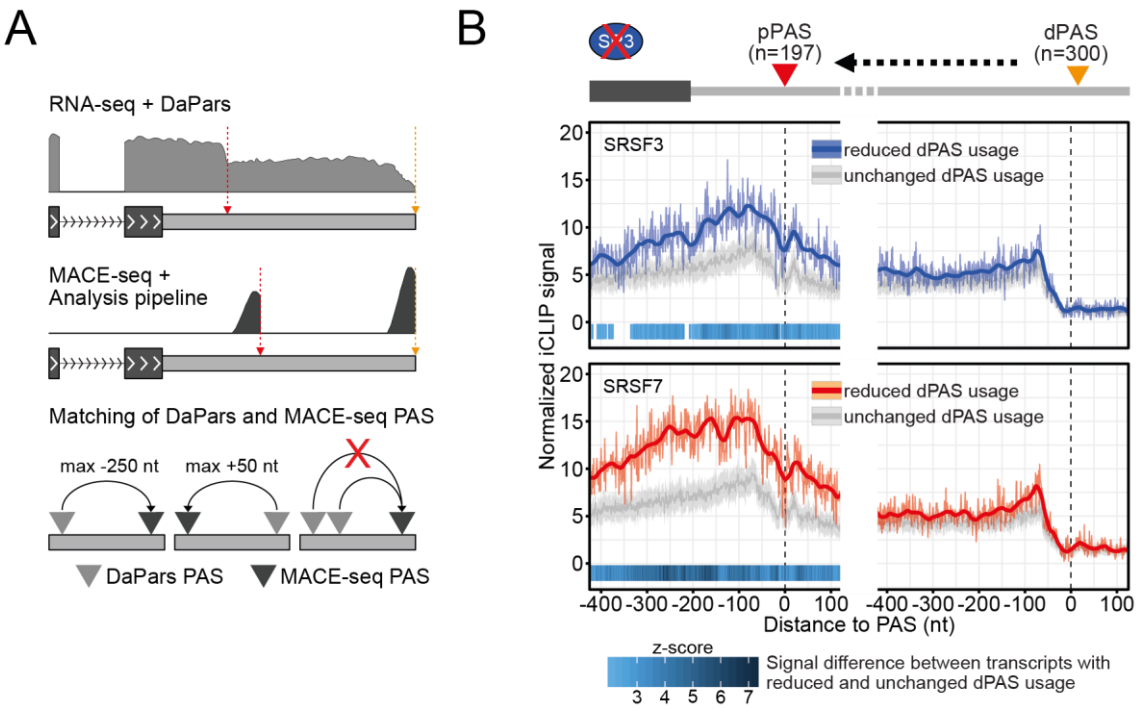


Figure 26: Binding of SRSF3 and SRSF7 to SRSF3-responsive PAS. A) *Scheme to match PAS identified by DaPARS and MACE-Seq.* ***B)*** *Binding profiles of significant cross-link sites of SRSF3 and SRSF7 at pPASs and dPASs in SRSF3-responsive transcripts, compared to unaffected, random PASs.*

5.4 Regulation of the pPAS is splicing-independent and concentration-dependent

So far, we showed that depletion of SRSF3 and SRSF7 alters the expression of hundreds of transcript isoforms that vary in their 3' UTR lengths, and that both SR proteins bind upstream of PAS with a particular enrichment at SRSF3-regulated pPAS. Yet, the underlying mechanism and its connection to splicing are unknown.

SR proteins are essential splicing factors that normally bind to splicing enhancer sequences in exons. Furthermore, more than 20 years ago SRSF3 was identified as a terminal-exon-splicing regulator thereby regulating CDS-APA (Lou et al., 1998). Hence, it was important to investigate whether the effect of SRSF3 and SRSF7 particularly on 3' UTR-APA is interconnected with their splicing functions. For this, three validated APA targets (**Figure 24**) were selected for a reporter gene study: *Ddx21*, *Anp32e*, *Rab11a*. All three genes contain at least two PASs within their 3' UTRs, whose usage changed upon KD of *Srsf3* (increased pPAS usage) and *Srsf7* (increased dPAS usage) (**Figure 27**). Moreover, both SR proteins bind upstream of their pPASs, according to the iCLIP data.

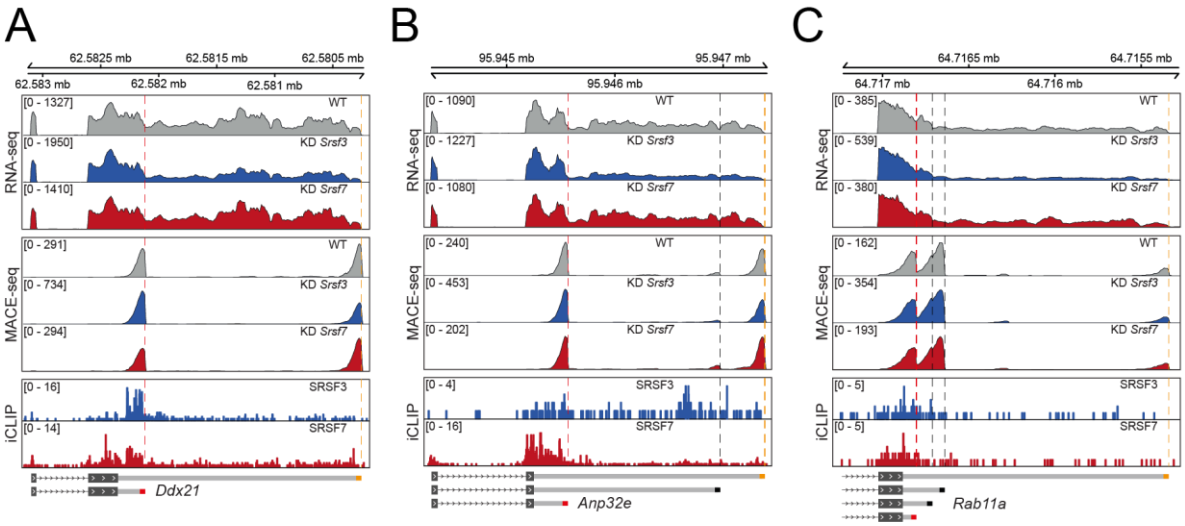


Figure 27: Browsershots of Ddx21 (A), Anp32e (B) and Rab11a (C). *Top: RNA-Seq reads in control condition (WT) and after KD of Srsf3 (blue) or Srsf7 (red). Middle: MACE-Seq reads, indicating pPAS and dPAS in the same conditions as Top. Bottom: Significant iCLIP crosslink events of SRSF3 (blue) and SRSF7 (red).*

The full-length 3' UTRs of the three candidate genes (*Ddx21*, *Anp32e* and *Rab11a*) including 150 nt of the downstream sequence were fused downstream of two intron-less reporter genes, *mCherry* and *firefly luciferase* (Luc), using Gibson Assembly cloning. After transfection of the respective reporter plasmids into P19 wt cells, the expression of endogenous and transgenic 3' UTR-APA isoforms were assessed by semiquantitative 3'RACE-PCR (**Figure 28**A&B). Strikingly, both reporter gene constructs produced the same number and sizes of 3' UTR-APA transcript isoforms as the endogenous genes, while the control reporter genes (*Luc-Ctrl* and *mCherry-Ctrl*) gave rise to only one transcript isoform containing a SV40 PAS. This suggests that APA-regulation of these genes acts independently of introns and splicing.

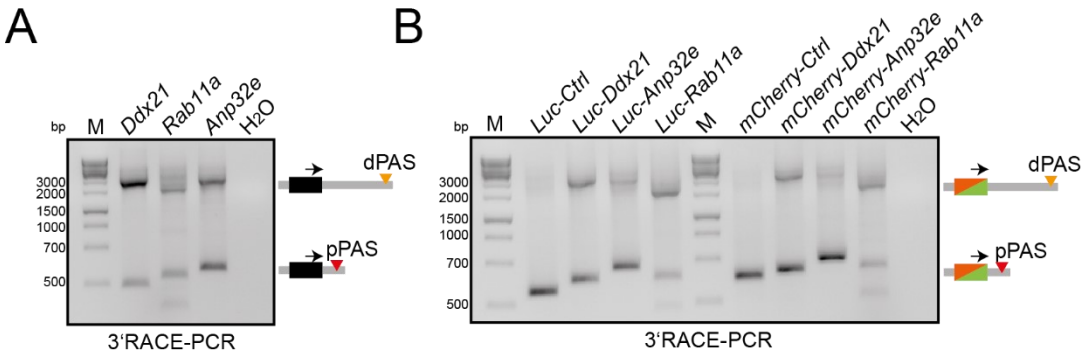


Figure 28: Validation of the existence of two PASs within the 3' UTR of chosen target genes and the respective reporter genes using 3'RACE-PCR. A) *Endogenous transcript isoforms with different 3' UTR lengths were amplified by 3'RACE-PCR using gene specific forward primers.* ***B)*** *The expression of the two expected 3' UTR isoforms of both reporter genes was validated using 3'RACE-PCR and reporter gene-specific forward primers after transfection into P19 wt cells. M = marker (O'GeneRuler 1 kb Plus).*

We next tested whether SRSF3 and SRSF7 affect APA of the *Luc* and *mCherry* reporter transcripts in the same way as the corresponding endogenous transcripts. For this, reporter plasmids were co-transfected with esiRNAs to knock down *Srsf3* and *Srsf7,* respectively, in P19 wt cells. Depletion of both SR proteins was validated by Western blot and the expression of 3' UTR-APA isoforms were analyzed by 3'RACE-PCR (**Figure 29**). Remarkably, depletion of SRSF3 and SRSF7 reproduced the changes in pPAS usage found in the endogenous transcripts. Decreased levels of SRSF3 resulted in shortening of 3' UTRs, while *Srsf7* KD had the opposite effect. These results further confirmed that changes in pPAS usage through altered SRSF3 and SRSF7 levels is independent of splicing.

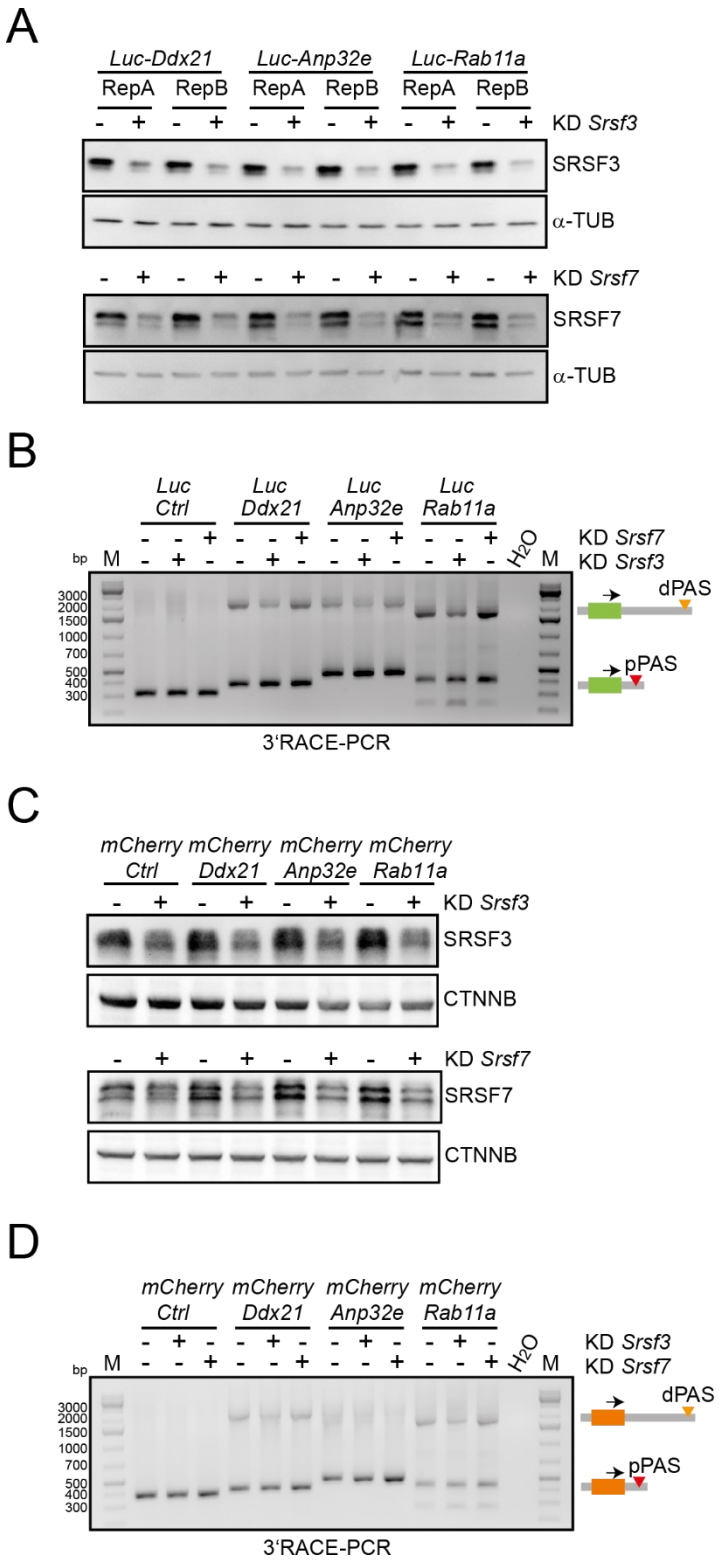


Figure 29: KD of Srsf3 and Srsf7 affects pPAS usage of Luc and mCherry reporter constructs. A&C) *Western blot validation of KD efficiency after transfection of esiRNAs targeting SRSF3 and SRSF7, respectively. α-Tubulin or CTNNB were used as loading controls.* ***B&D)*** *3'RACE-PCR analysis of 3' UTR changes in Luc and mCherry reporter transcripts. Unmodified reporter transcripts were included as controls. M = marker (O'GeneRuler 1 kb Plus)*

To test the concentration-dependency of pPASs usage regulation and a potential competition between both SR proteins, P19 wt cells were co-transfected with plasmids expressing the *Luc-Ddx21* reporter gene and GFP-tagged SRSF3 and SRSF7 in increasing amounts. The levels of the respective SR proteins were determined by Western blot and the ratio of the *Luc-Ddx21* 3' UTR isoforms was quantified using semi-quantitative 3'RACE-PCR using a *Luc* specific forward primer (**Figure 30**A&B). Indeed, increasing amounts of SRSF7 enhanced the production of the shorter 3' UTR isoform and simultaneously decreased the longer isoform. But surprisingly, overexpression of SRSF3 had no effect on pPAS usage. These results reveal that the regulation of pPAS usage by SRSF7 and SRSF3 is splicing-independent and concentration-dependent, but in a distinct manner. While depletion of SRSF3 and overexpression of SRSF7 showed strong effects, the *vice versa* scenarios showed much weaker effects.

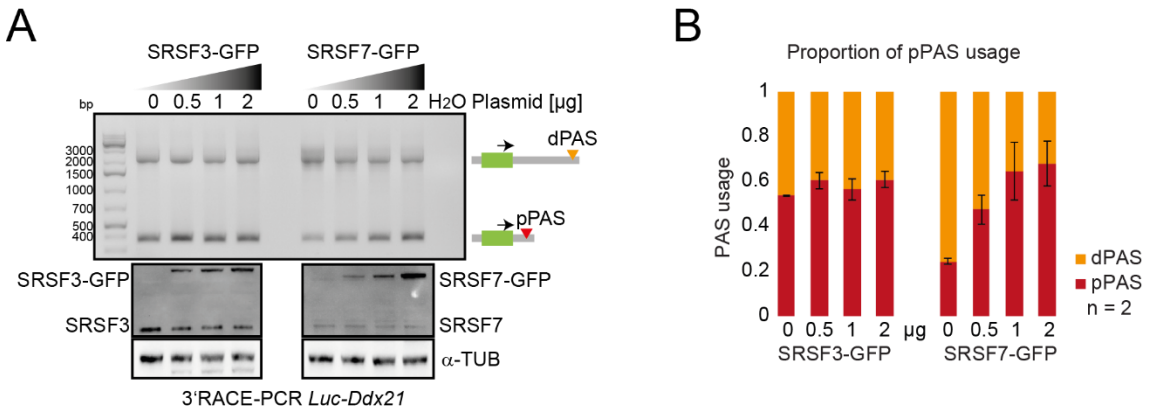


Figure 30: Expression of increasing amounts of SRSF7 increases pPAS usage. A) *Increasing amounts of SRSF3-GFP or SRSF7-GFP were co-transfected with the Luc-Ddx21 reporter gene. Expression of SRSF3 and SRSF7 was validated by Western blot, detection of α-Tubulin was used as loading control. Changes in the levels of 3' UTR-APA isoforms were quantified by 3'RACE-PCR. M = marker (O'GeneRuler 1 kb Plus)* ***B)*** *Proportion of PAS usage from A). The pPAS is colored in red and the dPAS in orange (n = 2).*

iCLIP meta-analyses at the beginning of this section showed that SRSF3 and SRSF7 binding is enriched upstream of SRSF3-regulated pPAS. To test whether binding of SRSF3 and SRSF7 to these sites is required for the regulation of pPAS usage in opposite directions, we mutated all SRSF3 binding motifs (CNYC) in a region 110 nt upstream of the *Ddx21* pPAS to SRSF7 (allSRSF7) motifs (GAY) and vice versa (allSRSF3) (**Figure 31**A). As controls, single point mutations were introduced in the proximal poly(A) signal motif to increase its strength (AGTAAA → AAUAAA) or to weaken the motif (AGTAAA → AGTAAG) (**Figure 31**A). These mutant constructs were transfected into P19 wt cells and 3' UTR-APA was assessed by 3'RACE-PCRs as mentioned before. Interestingly, altering the strength of the proximal poly(A) signal motif completely abrogated 3' UTR-APA of *Ddx21*. A stronger proximal poly(A) signal motif produced only the short 3' UTR isoform, while weakening the proximal poly(A) signal motif terminated its usage leading to exclusive expression of the long 3' UTR isoform (**Figure 31**B&C). Remarkably, increasing the number of SRSF7 binding sites (allSRSF7) increased pPAS usage, while an

increase in SRSF3 binding sites had again no effect on pPAS usage (**Figure 31**B&C). It seems rather strange that the replacement of all SRSF7 motifs by SRSF3 motifs had no visible effect on pPAS usage, but this might be explained by the fact that SRSF7 is able to bind the SRSF3-specific binding motif CNYC with its RRM-portion alone (Cavaloc et al., 1999; Königs et al., 2020). This has been identified previously by SELEX and iCLIP studies. The not evident alteration of pPAS usage might relate to the decreased affinity of the SRSF7-RRM towards this certain motif.

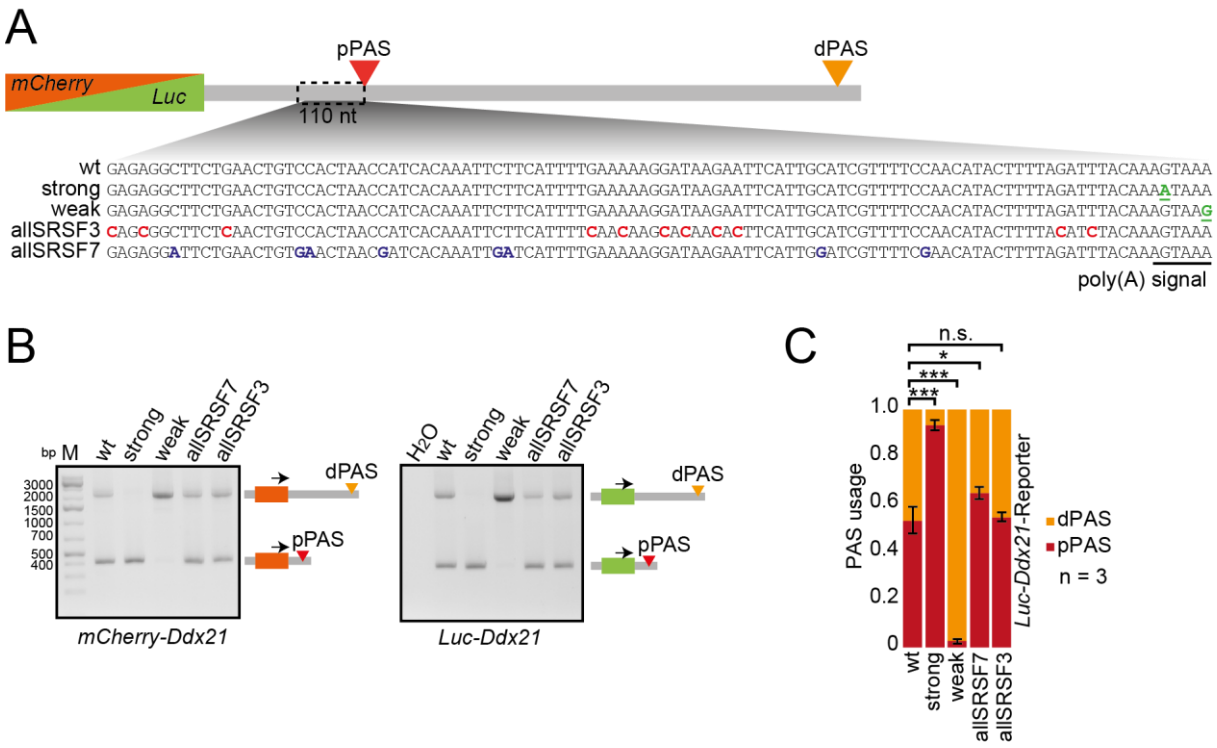


Figure 31: Increasing the binding potential of SRSF7 upstream of the pPAS increases its usage. A) *Scheme of the mutants derived from the Luc and mCherry-Ddx21 reporter genes. SRSF3 binding motifs (CNYC) were changed into SRSF7 binding motifs (GAY), and vice versa, in a region 110 nt upstream of the pPAS. In addition, the pPAS-associated hexameric motif was mutated. Mutated nucleotides are highlighted by color and printed boldly.* ***B)*** *Impact of mutations in the mCherry and Luc-Ddx21 reporter genes determined by 3'RACE-PCR. M = marker (O'GeneRuler 1kb Plus)* ***C)*** *Quantification of changes in pPAS usage in the Luc mutants (n = 3; Student's t-test: * = p < 0.01, ** = p < 0.01, *** = p < 0.005).*

In conclusion, the results presented here from experiments with two distinct intron-less reporter genes suggest that the regulation of pPAS usage by SRSF3 and SRSF7 is independent of splicing. Depletion, overexpression, and mutation experiments revealed that APA regulation depends on the levels and the binding potential of both SR proteins, whereby an intermediate strength of the pPAS is critical. Depletion of SRSF3 and decreasing its binding potential as well as increasing SRSF7 levels and binding potential showed the biggest impact on pPAS usage. This indicates that access of SRSF3 and SRSF7 to the pPAS region is critical for its usage in opposite directions and suggests that the pPAS might be enhanced by SRSF7 binding and blocked by SRSF3 binding, and a competition between

both SR proteins fine-tunes pPAS usage. Yet, it remains unclear why an increase in SRSF3 levels and binding showed no effect on APA and how these two SR proteins regulate pPAS usage in detail.

5.5 SRSF3 and SRSF7 interact differently with cleavage and polyadenylation factors

Earlier results showed enriched binding of SRSF3 and SRSF7 upstream of pPASs and that usage of these PASs depends on binding and expression levels of both SR proteins. However, SRSF3 and SRSF7 showed opposing effects on pPAS usage, raising the question how this is achieved. One possibility would be that SRSF3 and SRSF7 interact differently with CPA factors. To identify CPA factors that interact with SRSF3, GFP-labelled SRSF3 was stably expressed at endogenous levels in murine P19 cells (Müller-McNicoll et al., 2016), SRSF3-GFP containing RNPs were stringently purified without RNase treatment and subjected to quantitative mass spectrometry (performed by the Christian Münch Lab, IBC, Universität Frankfurt) (Klann et al., 2020). As a negative control, P19 cells expressing GFP fused to a nuclear localization sequence (GFP-NLS) were used (Änkö et al., 2012). 832 SRSF3-RNP-interactors were identified, 364 of those were significantly enriched in comparison to the GFP-NLS control (**SuppTable 4**). Within those several key CPA-factors were found, e.g., Fip1, CPSF3 and WDR33 from the CPSF complex), CPSF5 from the CFIm complex and PABPN1, while other CPA factors like CPSF1 and CPSF4 (CPSF complex) and CPSF6 (CFIm) were depleted (**Figure 32**).

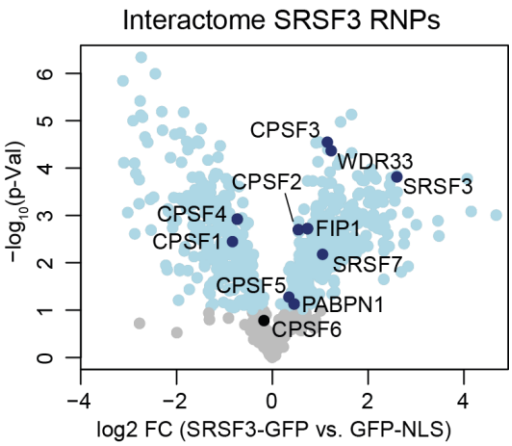


Figure 32: CPA factors that are present in SRSF3-containing RNPs. *Volcano plot of proteins co-purified after pull-down of GFP-tagged SRSF3. Significant hits are highlighted in light blue. CPA factors are indicated.*

SR proteins usually mediate protein-protein interactions via their RS-domains or SR-like domains (Brais et al., 1998; Graveley, 2000; Long & Caceres, 2009). Since SRSF3-GFP containing RNPs were purified without RNase treatment, it is possible that co-purified proteins interact indirectly with SRSF3 via binding to the same transcripts. To identify direct SRSF3 interaction partners the polypeptide sequences of the interacting CPA factors were analyzed for the presence of RS-like domains. Only two of them, CPSF6 and FIP1, contain RS or RS-like domains, enriched in arginine-aspartate/ glutamate repeats at the C-terminus (**Figure 33**). We selected those two

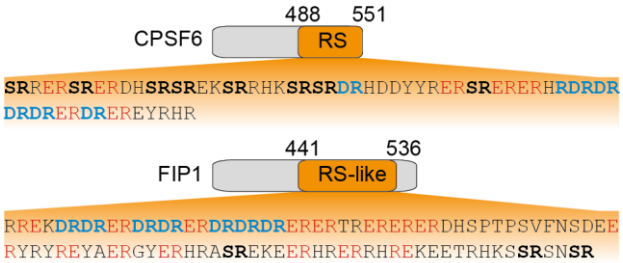


Figure 33: Sequence of the RS/RS-like domains in CPSF6 and FIP1, respectively. *RS dipeptides are indicated in bold, RD and RE dipeptides are highlighted in blue and red, respectively. The numbers above the domain indicate the beginning and the end of the respective domains.*

proteins for a detailed biochemical investigation, because FIP1 was previously described to facilitate pPAS usage, whereas CPSF6, the big subunit of CFIm, was shown to enhance dPAS usage by recruiting FIP1 via its RS-(like) domain (Lackford et al., 2014; Zhu et al., 2018). We also included CPSF5 because it forms a very stable ternary complex with CPSF6, the CFIm complex (**Figure 34**A). P19 cell lines stably expressing GFP-tagged FIP1, CPSF6 and CPSF5 from bacterial artificial chromosomes were generated and tested for correct expression and subcellular localization of the GFP-tagged proteins (**Figure 34**A-E). CPSF5-GFP and GFP-FIP1 expression induced auto-regulation of the endogenous protein by an unknown mechanism, while this was not noticeable for CPSF6-GFP expression (**Figure 34**B-D). Moreover, all GFP-tagged isoforms located within the nucleus, as expected (**Figure 34**E).

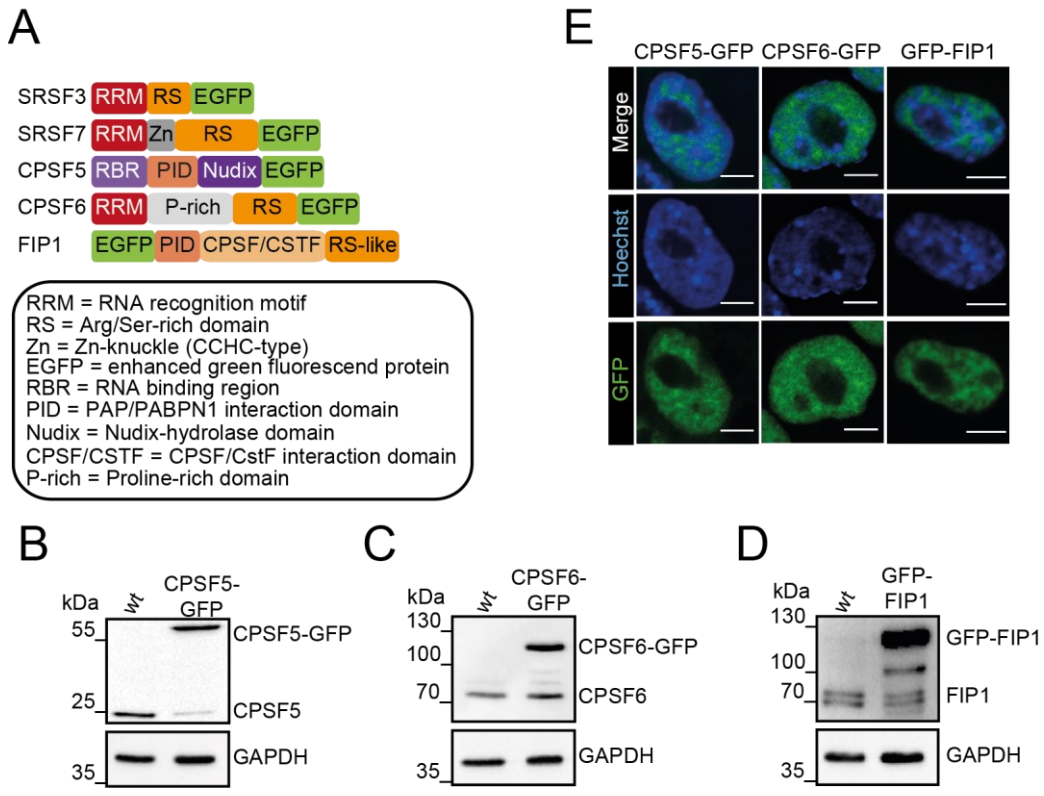


Figure 34: P19 cell lines expressing GFP-tagged isoforms of CPSF5 and CPSF6 (CFIm) as well as FIP1 from bacterial artificial chromosomes (BACs). A) *Scheme of the domain structure of GFP-tagged SRSF3, SRSF7, CPSF5, CPSF6 and FIP1 used in this study.* ***B-D)*** *Western blot analysis to compare expression of the endogenous and transgenic proteins (CPSF5, CPSF6, FIP1) in the stable cell lines. GAPDH was probed as a loading control.* ***E)*** *Confocal fluorescence microscopy to validate the sub-cellular localization of GFP-tagged proteins. HOECHST staining was used to label the nucleus. Scale bars = 5 µm*

5.5.1 SRSF3 interaction with the CPA factors is RNA-mediated, while their interaction with SRSF7 is RNA-independent

The stable cell lines were used to test whether SRSF3 and SRSF7 interact with FIP1, CPSF5 and CPSF6 and whether this interaction is dependent on RNA. For this, co-immunoprecipitations were performed with or without RNase treatment.

Unfortunately, the CPSF6-GFP cell line was not suitable for pull-downs using an α-GFP antibody (data not shown). Therefore, it was substituted by the CPSF5-GFP cell line since both proteins form the highly stable ternary CFIm complex.

First, GFP-tagged CPSF5 and FIP1 were used for Co-IPs and probed for endogenous SR proteins and cleavage factors (**Figure 35**A&B). These experiments confirmed the formation of a stable CFIm complex with stoichiometric interactions between CPSF5 and CPSF6 (both alternative splicing-derived isoforms of 68 kDa and 72 kDa, marked by * and **, respectively), while the interactions between the CPSF5 and FIP1 were rather weak and RNA-dependent. Both cleavage factors interact with SRSF3 and SRSF7 to a similar degree, but surprisingly, the interaction of CPSF5 and FIP1 with SRSF3 was abolished after RNase treatment. In contrast SRSF7 remained a strong interactor of both cleavage factors in the absence of RNA (**Figure 35**&B). All interactions were validated in reverse with Co-IPs using GFP-tagged SRSF3 and SRSF7 as bait probed for endogenous cleavage factors (**Figure 35**C&D).

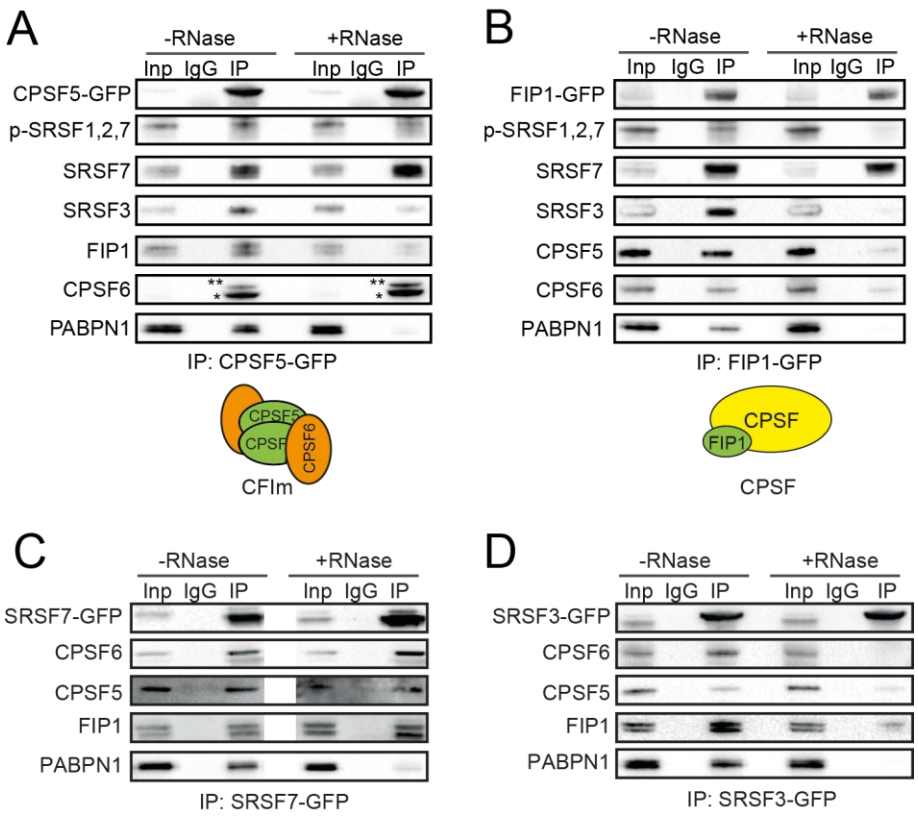


Figure 35: SRSF3 and SRSF7 interact differently with cleavage factors CPSF5 and FIP1, depending on RNA and phosphorylation states of the RS domain. A&B) *Co-immunoprecipitations using GFP-tagged CPSF5 and FIP1 as baits with and without RNase treatment. An antibody detecting phosphorylated SR proteins (mAb1O4) was used to probe the phosphorylation state of co-immunoprecipitated SRSF7. An antibody detecting PABPN1 was used to control for RNase treatment efficiency. * = 68 kDa CPSF6 isoform, ** = 72 kDa CPSF& isoform* ***C&D)*** *Reverse co-immunoprecipitations using GFP-tagged SRSF3 and SRSF7 as baits with and without RNase treatment. An antibody detecting PABPN1 was used to control for RNase treatment efficiency.*

As the activity of SR proteins is regulated via phosphorylation of the serine residues within the RS domain, an antibody detecting exclusively phosphorylated SR proteins (mAb1O4) was used. This showed that in the absence of RNA, CPSF5 preferentially interacts with phosphorylated SRSF7 (p-SRSF1,2,7), while FIP1 seems to prefer hypophosphorylated SRSF7, as the signal vanished in the +RNase sample (**Figure 35**A&B). However, this result must be treated carefully as the mAb1O4 antibody cannot discriminate between phosphorylated SRSF1, SRSF2 and SRSF7.

In conclusion, these data suggest a mechanism how SRSF3 and SRSF7 might regulate pPAS usage. SRSF7 showed strong RNA-independent interactions with both cleavage factors, but especially with FIP1, which could facilitate the usage of bound pPAS. SRSF3 does not interact directly with the cleavage factors, supporting the idea that it might actively block pPAS usage when it binds in close proximity. In addition, the phosphorylation state of SRSF7 seems to discriminate its interaction with CFIm or FIP1 and might reflect active and inactive assemblies.

5.5.2 The interaction between SRSF7 and FIP1 is CFIm-independent

Our data suggest that SRSF7 interacts RNA-independently with CFIm and with FIP1. Since CFIm was shown to recruit FIP1 via an interaction between their respective RS/RS-like domains (Zhu et al., 2018), it was important to test whether the interaction between SRSF7 and FIP1 is bridged by CFIm. For this, Co-IPs using GFP-FIP1 as a bait were performed after KD of *Cpsf6* or *Cpsf5*. Remarkably, KD of *Cpsf6* drastically reduced the protein levels of CPSF5. Although the KD of *Cpsf5* was much weaker we still observed a slight reduction of CPSF6 protein levels. This suggests that both proteins stabilize each other within the CFIm complex (**Figure 36**A&C). Interestingly, the depletion of either CFIm subunit did not impair the interaction between SRSF7 and FIP1, suggesting that SRSF7 directly recruits FIP1, which might lead to the enhanced usage of SRSF7 bound pPASs (**Figure 36**B&D).

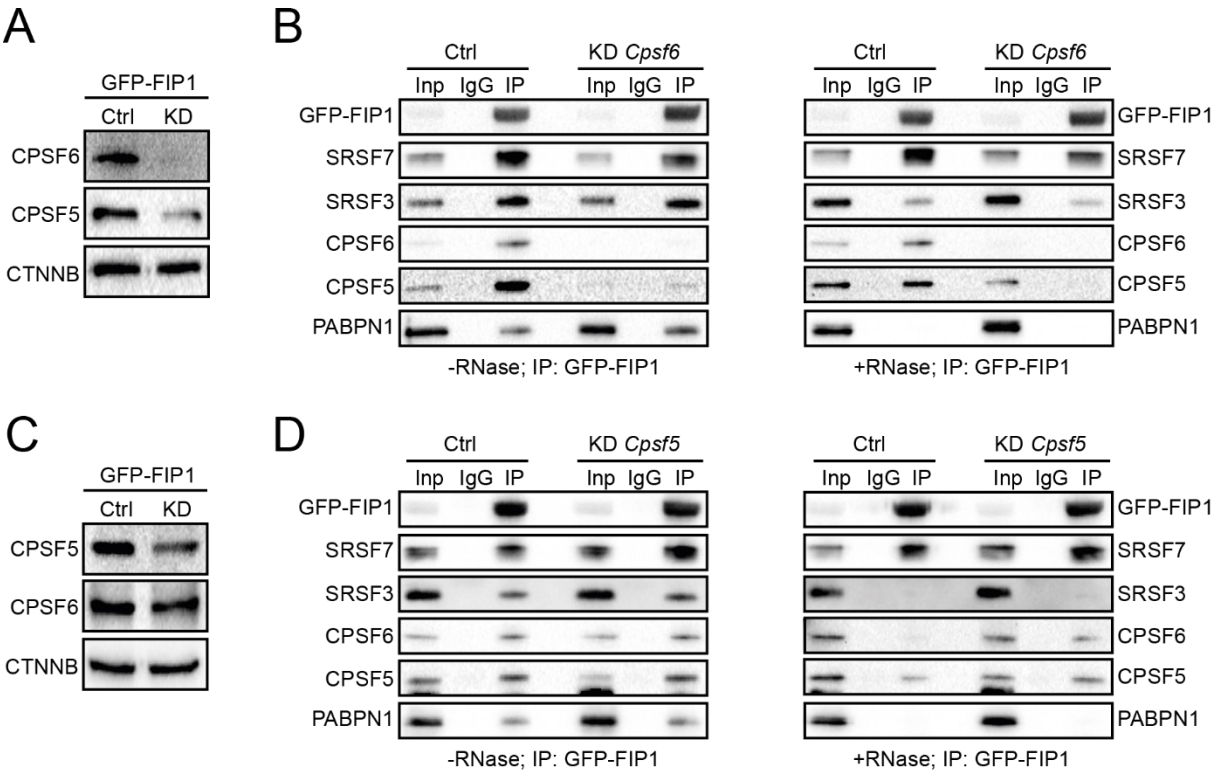


Figure 36: Depletion of CFIm subunits does not affect RNA-independent interaction of FIP1 and SRSF7. A) *Western blot to validate the knockdown efficiency after transfection of Cpsf6 specific esiRNAs. The expression of the core-subunit of the CFIm, CPSF5, was analyzed as well. CTNNB was detected as loading control.* ***B)*** *Co-immunoprecipitations using GFP-tagged cleavage factor FIP1 as bait with and without RNase treatment after depletion of Cpsf6. An antibody detecting PABPN1 was used to control the RNase treatment.* ***C)*** *Western blot to validate the knockdown efficiency after transfection of Cpsf5-specific esiRNAs. The expression of the large subunit of the CFIm, CPSF6, was validated as well. CTNNB was detected as loading control.* ***D)*** *Co-immunoprecipitations using GFP-tagged cleavage factor FIP1 as bait with and without RNase treatment after depletion of Cpsf5. An antibody detecting PABPN1 was used to control the RNase treatment.*

5.5.3 SRSF7 recruits FIP1 via its RS-domain in a phosphorylation-dependent manner

Having shown that FIP1 interacts with hypophosphorylated SRSF7 independent of RNA and the CFIm, we next investigated whether this interaction is mediated by the RS domain of SRSF7 and assessed the role of RS domain phosphorylation. For this we fused the RS domains of SRSF7 and SRSF3 to the tetracycline repressor protein (TetR-RS7, TetR-RS3; **Figure 37**A). In addition, phosphomimetic variants of the SRSF7 RS domain were generated by replacing all serine residues with alanine (A) or with aspartate (D) residues to mimic a non-phosphorylated (TetR-RA7) and a completely phosphorylated RS domain (TetR-RD7), respectively (**SuppTable 5**) (**Figure 37**A). These constructs were transiently transfected into P19 cells and correct expression of the fusion proteins was validated by Western blot (**Figure 37**B). Phosphorylation of the fusion proteins was assessed by shrimp alkaline phosphatase (SAP) treatment (**Figure 37**C). As expected, TetR-RS7 and TetR-RS3 were phosphatase sensitive with a reduction of their molecular weight after SAP treatment, while the variants TetR-RA and -RD7 variants showed no size shift.

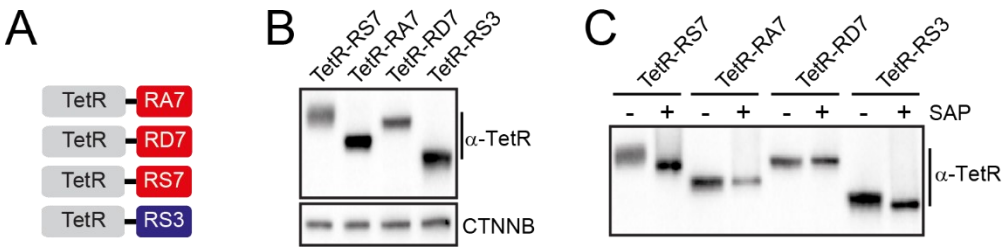


Figure 37: Generation and validation of tetracycline repressor (TetR) protein-RS domain fusion constructs including phosphomimetic variants. A) *Scheme of the tetracycline repressor protein-derived reporter constructs. To mimic non-phosphorylated and completely phosphorylated RS domains, all serine residues (S) within the RS domain were switched to aspartate (D) or alanine (A) residues.* ***B)*** *Validation of the expression of the fusion proteins after transient transfection into P19 cells via Western blot using an antibody against TetR. CTNNB was detected as loading control.* ***C)*** *Validation of the phosphatase-(in)sensitivity of the generated fusion proteins via shrimp-alkaline phosphatase (SAP) treatment and Western blot.*

The TetR constructs were transiently expressed in P19 GFP-FIP1 and CPSF5-GFP cell lines, and Co-IPs were performed with and without RNase treatment. Interestingly, FIP1 interacted visibly with TetR-RS7 representing the endogenous RS domain and with the non-phosphorylated mimic Tet-RA7 (**Figure 38**A). However, supporting the previous data, FIP1 did not interact with the completely phosphorylated mimic TetR-RD7. The RS domain of SRSF3 was included as a control and showed no interactions with FIP1, also corroborating our previous observations. Surprisingly, when the same Co-IP experiments were performed using CPSF5-GFP as bait no interaction between CPSF5 and any of the RS domain variants of SRSF7 were observed (**Figure 38**B).

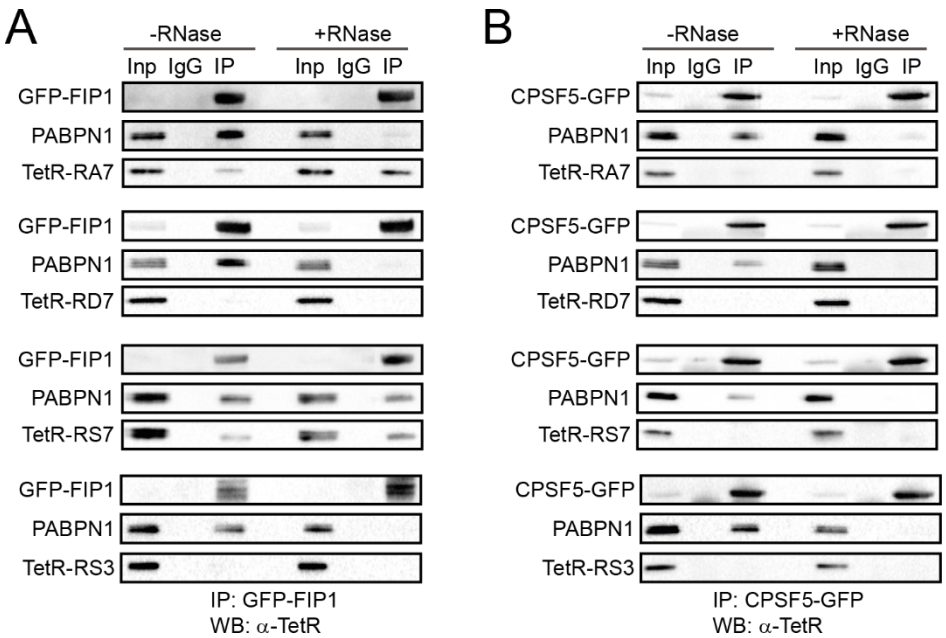


Figure 38: A hypophosphorylated RS domain of SRSF7 is sufficient to mediate interaction with FIP1. A) *Co-immunoprecipitations using GFP-tagged FIP1 as bait with and without RNase treatment after transient transfection of TetR-fusion constructs. An antibody detecting PABPN1 was used to control the RNase treatment.* ***B)*** *Co-immunoprecipitations using GFP-tagged CPSF5 as bait with and without RNase treatment after transient transfection of TetR-fusion constructs. An antibody detecting PABPN1 was used to control the RNase treatment.*

These results suggest that a hypophosphorylated or even non-phosphorylated RS domain of SRSF7 is sufficient to mediate the interaction with FIP1. In contrast, the RS domain alone is not sufficient to to establish the interaction between SRSF7 and CFIm.

To confirm that additional domains of SRSF7 are necessary to mediate the interaction with CFIm and address the role of RS domain phosphorylation, we fused the full-length coding sequences of SRSF7 and SRSF3 downstream of an mCherry reporter gene. As before, phosphomimetic variants were generated by replacing the endogenous RS domain with RA and RD domains, respectively (**Figure 39**A). After transfection into GFP-FIP1 P19 cells, the correct expression and sub-cellular localization of mCherry-fusion proteins was validated by Western blot and confocal fluorescence microscopy (**Figure 39**B&C). mCherry-SRSF7[RS] and the completely phosphorylated mCherry-SRSF7[RD] variant co-localized well with FIP1 in the nucleus. But unexpectedly, mCherry-SRSF7[RA] fusion proteins were found exclusively within the nucleoli of transfected cells and were completely absent in the rest of the nucleus (**Figure 39**C). mCherry-SRSF3[RA] was still visible in the nucleoplasm, but highly concentrated in nucleoli as well. Due to their aberrant localization, the non-phosphorylated mimics were excluded from subsequent experiments.

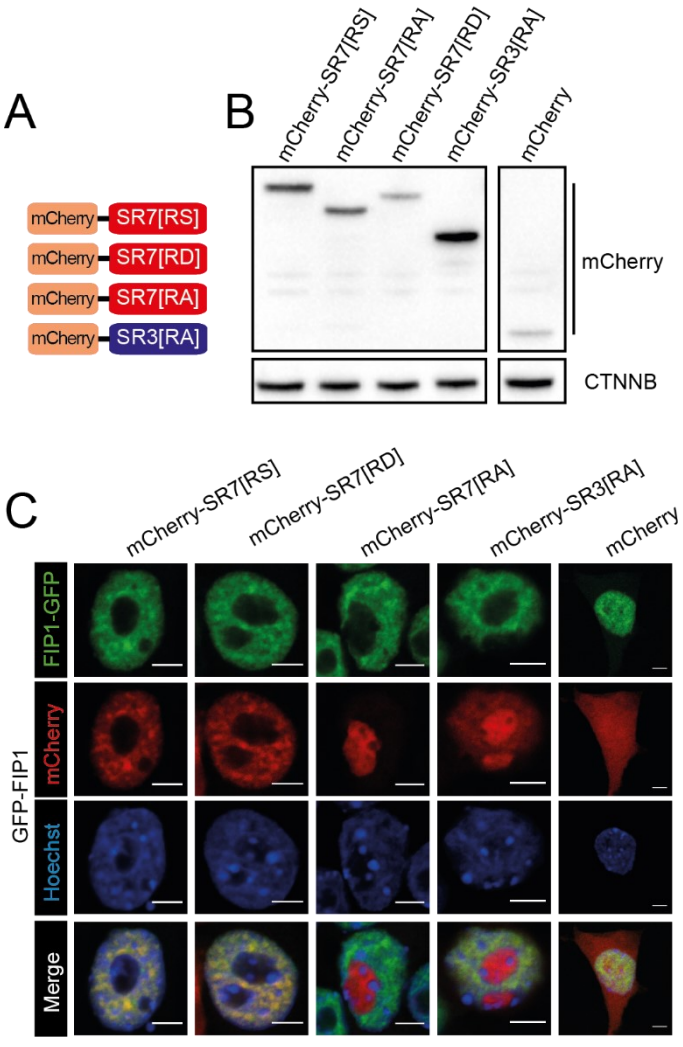


Figure 39: Generation and validation of mCherry-tagged phosphomimetic isoforms of SRSF7 and SRSF3. A) *Scheme of the mCherry-SR protein reporter gene constructs.* ***B)*** *Western blot to validate expression of transiently transfected reporter genes. CTNNB was detected as loading control.* ***C)*** *Validation of sub-nuclear localization of the mCherry reporter proteins in the GFP-FIP1 P19 cell line. HOECHST staining was used to label the nucleus. An empty mCherry vector was transfected as control. Scale bars = 5 µm.*

Co-IPs using mCherry-SRSF7[RS/RD]) constructs showed, that both SRSF7 variants established interactions with CPSF5 (**Figure 40**). This supports the earlier conclusion that additional domains of SRSF7 that were absent in the TetR-RS fusion proteins, are necessary for the interaction of SRSF7 with CFIm. Furthermore, the interaction between mCherry-SRSF7[RD] and CPSF5 was stronger compared to the native RS variant, again suggesting that CPSF5 prefers the interaction with completely phosphorylated RS-domains (**Figure 40**A). FIP1 also interacted with full-length SRSF7 variants. But as seen before, FIP1 prefers a hypo- or non-phosphorylated RS domain, as its interaction decreased when the mCherry-SRSF7[RD] construct was used. These results suggest that the recruitment of FIP1 is regulated by the phosphorylation state of the RS domain of SRSF7.

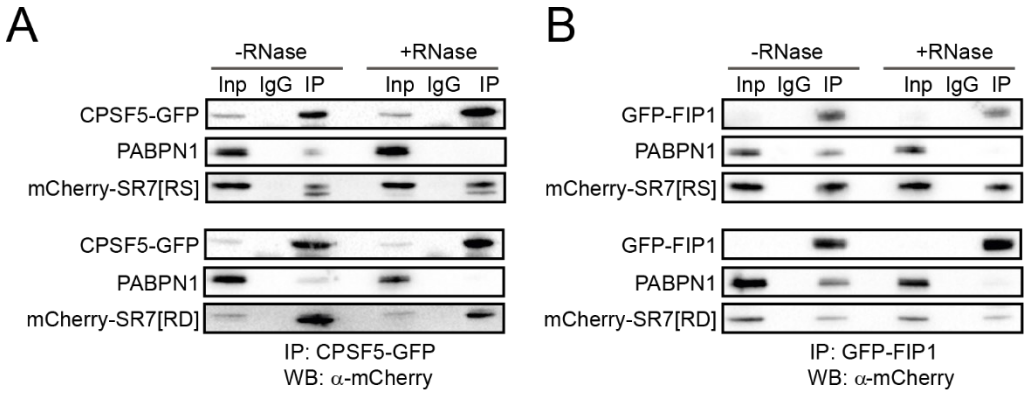


Figure 40: CPSF5 and FIP1 prefer different phosphorylation states of SRSF7. A) *Co-immunoprecipitations using GFP-tagged CPSF5 as bait with and without RNase treatment after transient transfection of mCherry-reporter constructs expressing the native [RS] or a fully phosphorylated [RD] SRSF7 mimic. An antibody detecting PABPN1 was used to control the RNase treatment.* ***B)*** *Co-immunoprecipitations using GFP-tagged cleavage factor FIP1 as bait with and without RNase treatment after transient transfection of mCherry-reporter constructs expressing the native [RS] or a fully phosphorylated [RD] SRSF7 mimic. An antibody detecting PABPN1 was used to control the RNase treatment.*

In conclusion, these experiments show that only SRSF7 interacts independently of RNA with FIP1 and CPSF5, while the interaction with SRSF3 is RNA-mediated. The interaction between SRSF7 and FIP1 is not bridged by the CFIm, suggesting a direct interaction. The RS domain of SRSF7 is sufficient to mediate the interaction with FIP1 while this domain was not sufficient to recruit CFIm, represented by CPSF5. FIP1 and CPSF5 prefer different phosphorylation states of the RS domain of SRSF7 for interaction. This suggests that the three proteins are not in the same complex or they are recruited in a specific order that involves dephosphorylation of SRSF7. However, the results obtained with the non-

phosphorylated RS-mimic [RA] need to be interpreted carefully as the mCherry-SRSF7[RA] fusion proteins were aberrantly enriched within the nucleolus.

5.5.4 Two distinct domains within SRSF7 promote interaction with CFIm and FIP1

Our data support a model whereby SRSF7 actively enhances pPAS usage by recruiting the cleavage and polyadenylation machinery via FIP1, while SRSF3 blocks the pPAS and competes with SRSF7 for binding. Why are SRSF3 and SRSF7 are behaving so differently? Phylogenetic clustering of all 12 members of the SR protein family revealed that SRSF3 and SRSF7 are evolutionary very closely related (**Figure 13**) (Busch & Hertel, 2012). However, both proteins differ from each other in the length of their respective RS domains (SRSF3: 50 aa; SRSF7: 116 aa) and SRSF7 contains a CCHC-type Zn-knuckle, which is unique among the core SR proteins (Cavaloc et al., 1994). To investigate the similarities and differences between both SR proteins in more detail, their polypeptide sequences were aligned using ClustalO (**Figure 9**A). However, both proteins differ from each other in the length of their respective RS domain and SRSF7 contains a CCHC-type Zn-knuckle, which is unique among the core SR proteins. To investigate the similarities and differences between both SR proteins in more detail, their polypeptide sequences were aligned using ClustalO (**Figure 41**). Interestingly, the RRM of SRSF3 and SRSF7 is highly similar, sharing 80% of the residues. However, the following linker is much shorter for SRSF7 compared to SRSF3. Moreover, an unusual region interspersed with hydrophobic residues was found within the RS domain of SRSF7, which is not present in SRSF3 (**Figure 41**). Hence, both proteins diverge by three distinct features, the Zn-knuckle, the linker, and the hydrophobic stretch.

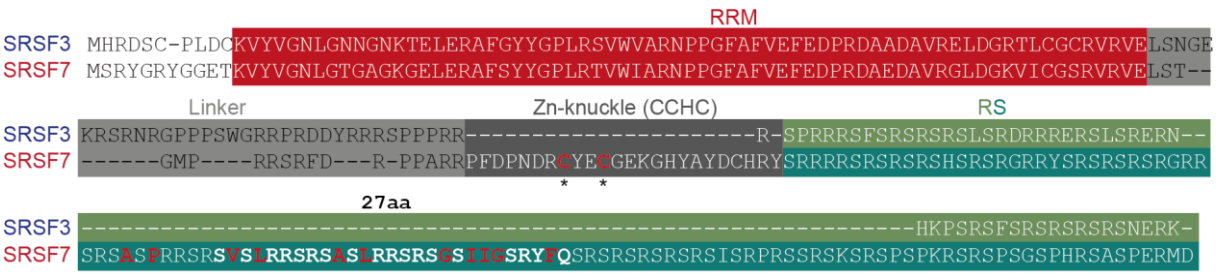


Figure 41: SRSF3 and SRSF7 are closely related but show distinct differences on the polypeptide sequence level. *Polypeptide sequence alignment of murine SRSF3 and SRSF7. Both SR proteins share a highly similar RRM (red) at the N-terminus and a RS domain at the C-terminus (green). In difference SRSF7 contains a shorter linker region (light grey) and a CCHC-type zinc-knuckle (dark grey), which is unique among the SR protein family. In addition, the RS domain of SRSF7 is longer than the one of SRSF3 and it is divided by unusual stretch containing hydrophobic residues (highlighted in red). Asterisks show two cysteine residues (C) which were mutated to alanine to inactivate the Zn-knuckle. Moreover, a region of 27 residues (bold), containing most of the hydrophobic residues, was removed to generate a Δ27aa mutant.*

To investigate whether the Zn-knuckle or the hydrophobic stretch are necessary for the FIP1 and CFIm interaction we generated two P19 cell lines stably expressing GFP-tagged SRSF7 either lacking the hydrophobic stretch (Δ27aa) or containing an inactivated Zn-knuckle (mutZn), where two

essential cysteine residues were mutated into alanine (**Figure 41**, **Figure 42**A). Expression and phosphorylation of the mutant proteins was validated by Western blot and SAP treatment, respectively (**Figure 42**B&C). Moreover, the correct sub-cellular localization was assessed by confocal fluorescence microscopy (**Figure 42**D).

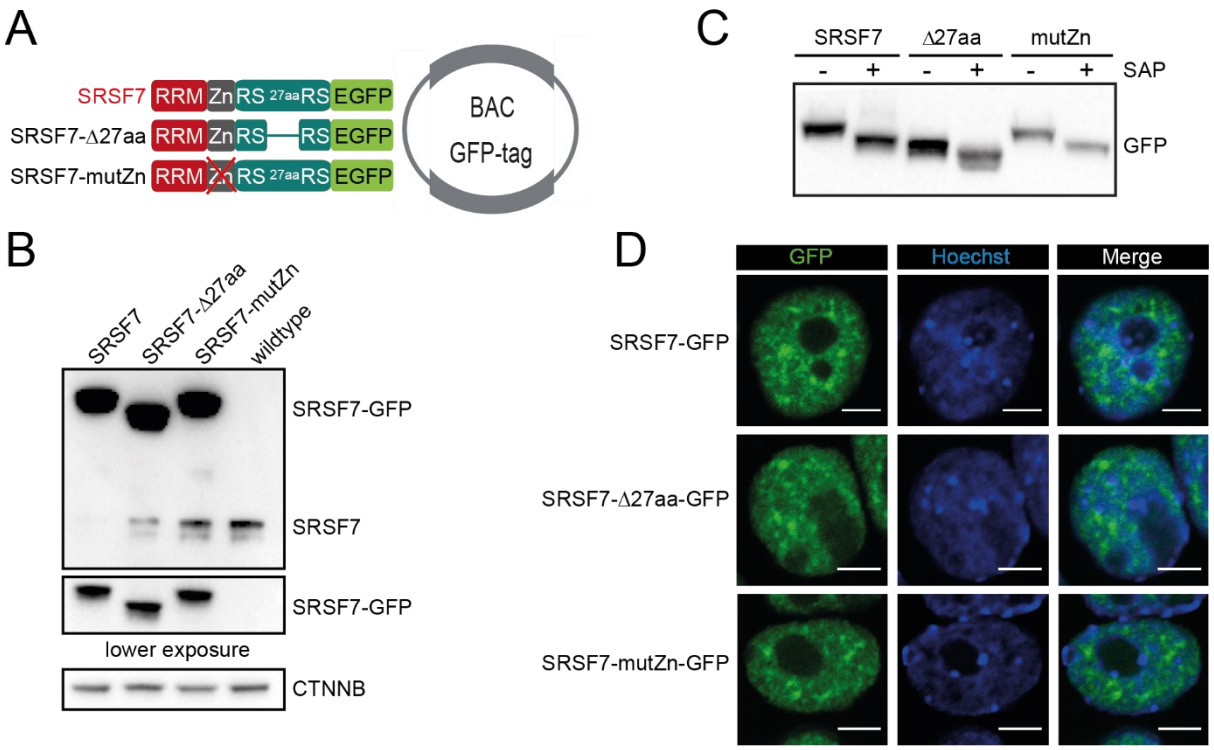


Figure 42: Generation of SRSF7 mutant cell lines lacking either the hydrophobic stretch (Δ27aa) or carrying an inactive Zn-knuckle domain (mutZn). *Schemes of the domains of SRSF7 and the two generated GFP-tagged mutants. These mutants were generated by BAC-recombineering to generate P19 cell lines stably expressing the constructs at physiological levels.* ***B)*** *Validation of expression of the generated mutants by Western blot. A lower exposure of the GFP-tagged mutants was included for better recognition. CTNNB was detected as loading control.* ***C)*** *Validation of phosphatase-sensitivity by Western blot after SAP treatment.* ***D)*** *Validation of correct nuclear localization using confocal fluorescence microscopy. HOECHST staining was used to label the nucleus. Scale bars = 5 µm.*

Subsequently, these cell lines were used to perform Co-IPs using the GFP-tag as the bait and probing for the respective cleavage factors with and without RNase treatment (**Figure 43**). Both SRSF7 mutants showed a decreased RNA-independent interaction with CPSF5 and FIP1 when present RNA was digested with RNase. In the same context, the mutated Zn-knuckle isoform also interacted less with CPSF6. Decreased interactions between the Δ27aa-mutant and the CPA factors were not observed when RNase was absent during the experiments indicating that the mutation did not affect the RNA-binding potential of SRSF7 and that this SRSF7 isoforms still bound to the same target RNAs as the probed CPA factors enabling co-purification. Inactivating the Zn-knuckle of SRSF7 seemed to decrease binding of SRSF7 to target RNAs shared with the CFIm as the amount of co-purified CPSF5 and CPSF6 decreased, compared to wildtype SRSF7 and SRSF7-Δ27aa. Notably the amount of co-purified PABPN1 (control) also decreased, further supporting this hypothesis bearing in mind that Zn-knuckle deficient

SRSF7 still can bind RNA via the CYNC-motif using the RRM-domain (Cavaloc et al., 1999; Königs et al., 2020). Interestingly, this was not observed to the same degree in the combination with FIP1. These results suggest that both SRSF7-specific features are necessary for the RNA-independent interaction with FIP1 and CFIm.

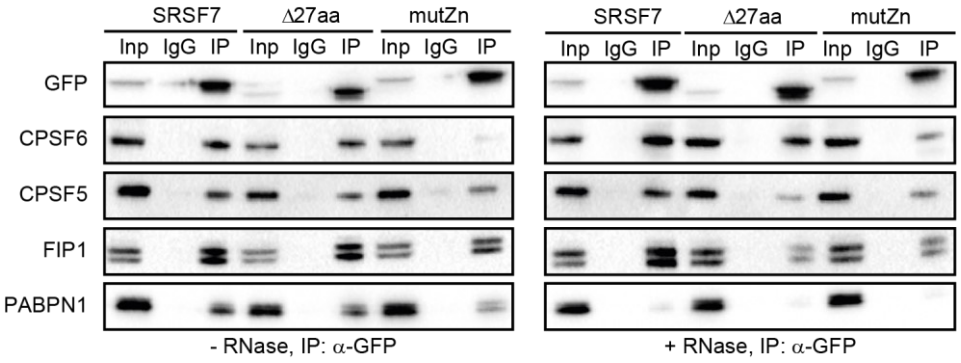


Figure 43: Mutation of both SRSF7-specific features decreases RNA-independent interaction with CPSF5, CPSF6 and FIP1. *Co-Immunoprecipitations using the GFP-tagged mutants as baits were performed with and without RNase treatment. PABPN1 was detected to validate the RNase treatment.*

To further assess whether these domains are also sufficient to establish these interactions, we generated chimeric constructs based on GFP-tagged SRSF3 containing either the Zn-knuckle (SRSF3-ZnF) or the hydrophobic stretch (SRSF3-27aa) or both motifs together (SRSF3-ZnF+27aa) (**Figure 44**A). These chimeric constructs were transiently transfected into P19 cells and their expression and correct phosphorylation was validated by Western blot and SAP treatment (**Figure 44**B&C). Furthermore, their correct sub-cellular localization was validated by confocal fluorescence microscopy (**Figure 44**D).

Interestingly, chimeric proteins were generally much lower expressed than SRSF3- and SRSF7-GFP. Nevertheless, Co-IPs showed an increased interaction between SRSF3 containing the hydrophobic stretch (SRSF3-27aa) and all tested cleavage factors, when compared to the input and wild type SRSF3 (**Figure 45**).

Unexpectedly, insertion of the Zn-knuckle domain alone did not improve interactions between SRSF3 and the cleavage factors. The Zn-knuckle also abolished the increased interaction obtained with the hydrophobic stretch in the double-chimeric construct (SRSF3-ZnF+27aa), suggesting that its insertion might cause steric hindrance and/or that the RNA preference of this mutant is changed as the Zn-knuckle has been shown to favor binding to GAYGAY-motifs with high affinity in wildtype SRSF7 (Cavaloc et al., 1999; Müller-McNicoll et al., 2016; Königs et al., 2020).

In conclusion, these experiments illustrate that the Zn-knuckle domain and the 27-residue hydrophobic stretch within the RS domain of SRSF7 are necessary to strengthen the interaction

between SRSF7 and the CFIm and FIP1. This shows that small differences in the protein architecture of SR proteins can have opposite effects on APA and pPAS activation.

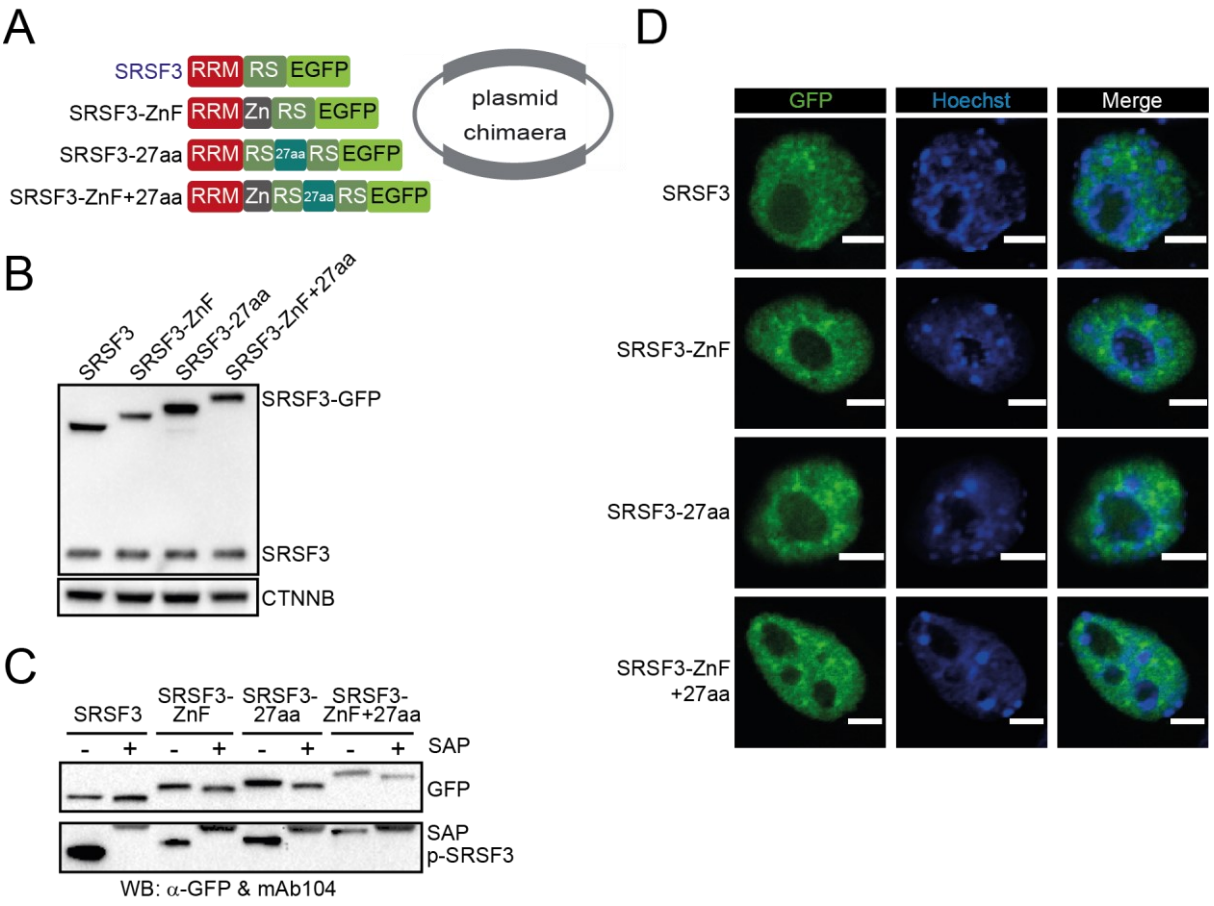


Figure 44: Generation of SRSF3 chimeras containing the Zn-knuckle domain and the hydrophobic stretch derived from SRSF7. A) *Schemes of the GFP-tagged SRSF3 derived chimeras containing the Zn-knuckle (ZnF) or the hydrophobic stretch (27aa), or both, from SRSF7.* ***B)*** *Validation of expression of the generated chimeras by Western blot after transient transfection. CTNNB was detected as loading control.* ***C)*** *Validation of phosphatase-sensitivity by Western blot after SAP treatment. As a control the membrane was probed with an antibody recognizing phosphorylated SR proteins (mAb1O4). This Ab cross-reacted with the SAP enzyme.* ***D)*** *Validation of correct nuclear localization using confocal fluorescence microscopy. HOECHST staining was used to label the nucleus. Scale bars = 5 µm.*

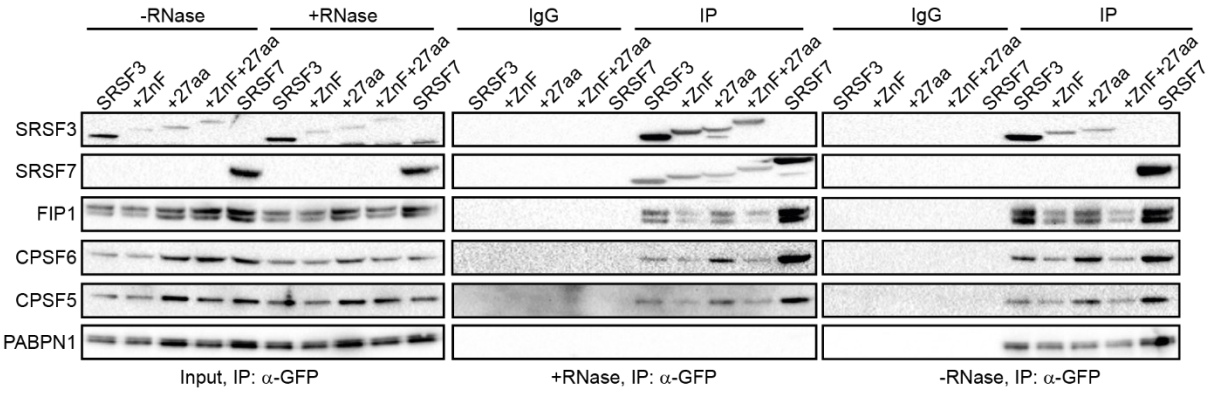


Figure 45: Insertion of the hydrophobic stretch increases interaction between SRSF3 and the cleavage factors CPSF5, CPSF6 and FIP1. *Co-Immunoprecipitations using the GFP-tagged chimeras as baits were performed with and without RNase treatment. GFP-tagged SRSF3 and SRSF7 were included as controls. PABPN1 was detected to validate the RNase treatment.*

5.6 SRSF3 promotes dPAS usage by preserving high levels of CPSF6

Interestingly, depletion of SRSF3 affected five times more APA targets than the depletion of SRSF7 (**Figure 21**). Yet, our experiments showed that only SRSF7 interacts RNA-independently with the cleavage and polyadenylation machinery to activate pPASs. Both SRSF3 and SRSF7 show preferential binding around pPASs, and SRSF3 might regulate APA by blocking the binding of SRSF7 and recruitment of FIP1. Depletion of SRSF3 would increase the binding potential of SRSF7 to pPASs and hence increase their activation. However, we did not observe the opposite effect, i.e., more pPAS inhibition, when we increase SRSF3 levels and binding potential. This suggests that SRSF3 might regulate APA by an additional mechanism.

To test whether the levels of transcripts encoding CPA factors were affected by SRSF3 depletion, we reanalyzed the RNA-Seq data for differential gene expression using DESeq2. Interestingly, the levels of *Cpsf6* transcripts decreased more than 2-fold after KD of *Srsf3*, while no other cleavage factor transcript was affected (**Figure 46**A, **SuppTable 8**). KD of *Srsf7* had no effect on the expression level of any cleavage factor in the RNA-Seq data. qRT-PCR confirmed the decrease of *Cpsf6* transcripts (3-fold) after SRSF3 depletion and showed only a minor decrease after KD of *Srsf7* (**Figure 46**B).

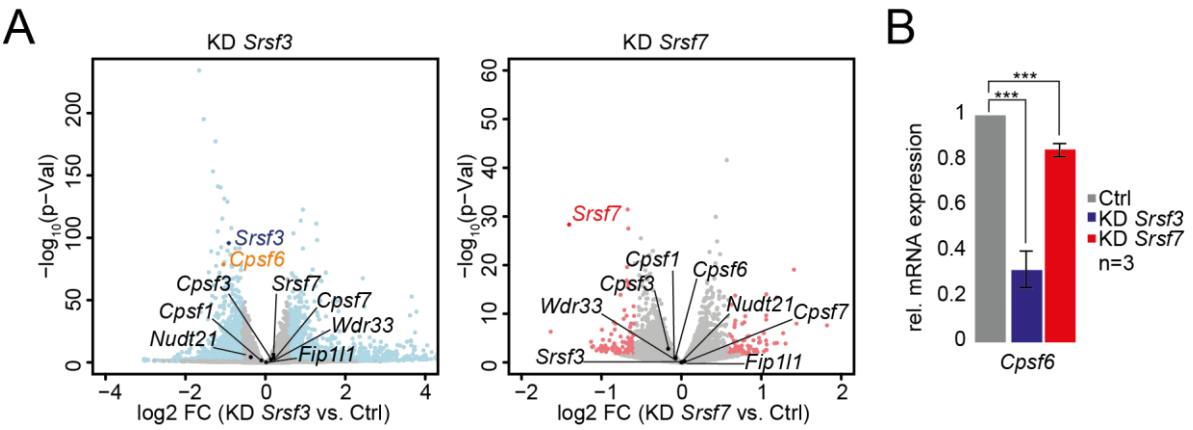


Figure 46: KD of Srsf3 significantly reduces mRNA levels of Cpsf6. A) *Analysis of differential gene expression using gene counts and DESeq2 showed a similar reduction of CPSF6 and Srsf3 mRNA after KD of Srsf3 (left), while depletion of Srsf7 did not affect the expression of any cleavage factor (right). Significant changes (p.adj-value < 0.1) are highlighted in blue or red, respectively.* ***B)*** *Validation of reduced Cpsf6 levels after KD of Srsf3 and Srsf7 by qRT-PCR. n = 3, *** = p-value < 0.005.*

To test whether the decrease in *Cpsf6* mRNA levels after KD of *Srsf3* is due to splicing changes, we inspected the splicing pattern of *Cpsf6* using RNA-Seq data from nuclear fractions after KD of *Srsf3* and *Srsf7* (**Figure 47**A). Usually, *Cpsf6* is constitutively spliced including all 10 exons. However, depletion of SRSF3 accumulates an alternative *Cpsf6* isoform with skipped Exon6, joining Exon5 and Exon7, here called *Cpsf6ΔEx6* (**Figure 47**B). This alternative splicing event generates a premature stop codon (PTC) within Exon7, 87 nt upstream of the next exon-exon junction, making this transcript isoform a potential target for nonsense mediated decay (NMD). In addition, KD of *Srsf3* or *Srsf7* gave rise to another alternative transcript isoform including a small alternative exon of 111 nt between Exon5 and Exon6 (**Figure 46**A). Inclusion of this small exon (X) into the full-length *Cpsf6* transcript produces a larger 72 kDa isoform of CPSF6 (CFIm-72) (Neve et al., 2017). Expression of the different splice isoforms was validated by RT-PCR using primers binding in Exon5 and Exon7 (**Figure 47**C). KD of *Srsf3* produced several shorter transcript isoforms with skipped Exon6. The levels of these spliced isoforms were drastically reduced compared to the control. This was not seen after KD of *Srsf7*, but interestingly the levels of the transcript isoform encoding CFIm-72 were visibly increased. Analysis of SRSF3 iCLIP data (Müller-McNicoll et al., 2016) revealed a much stronger binding of SRSF3 to Exon6 compared to other exons (**Figure 47**D). These data suggest that SRSF3 binds preferentially to Exon6 and facilitates its inclusion into mature *Cpsf6* mRNAs and thereby maintains high *Cpsf6* transcript levels.

Interestingly, we also found other alternative isoforms of *Cpsf6* during these studies. Experiments to identify circular transcripts within P19 cells, conducted by Di Liddo et al., 2019, identified a circular isoform of *Cpsf6* (*circCpsf6*), which is generated via back-splicing connecting the 3'end of Exon9 to the 5'end of Exon2 (**Figure 47**E). The presence of this circular isoform and the corresponding linear transcript lacking the entire coding region connecting Exons1 and 10 (*Cpsf6Ex1-10*) were validated by RT-PCR using specific primers in the affected exons (**Figure 47**F). Moreover, the identity of these transcripts was verified by Sanger sequencing (**Figure 47**G). KD of *Srsf3* slightly decreased the levels of the circular isoform but did not affect the levels of *Cpsf6Ex1-10*.

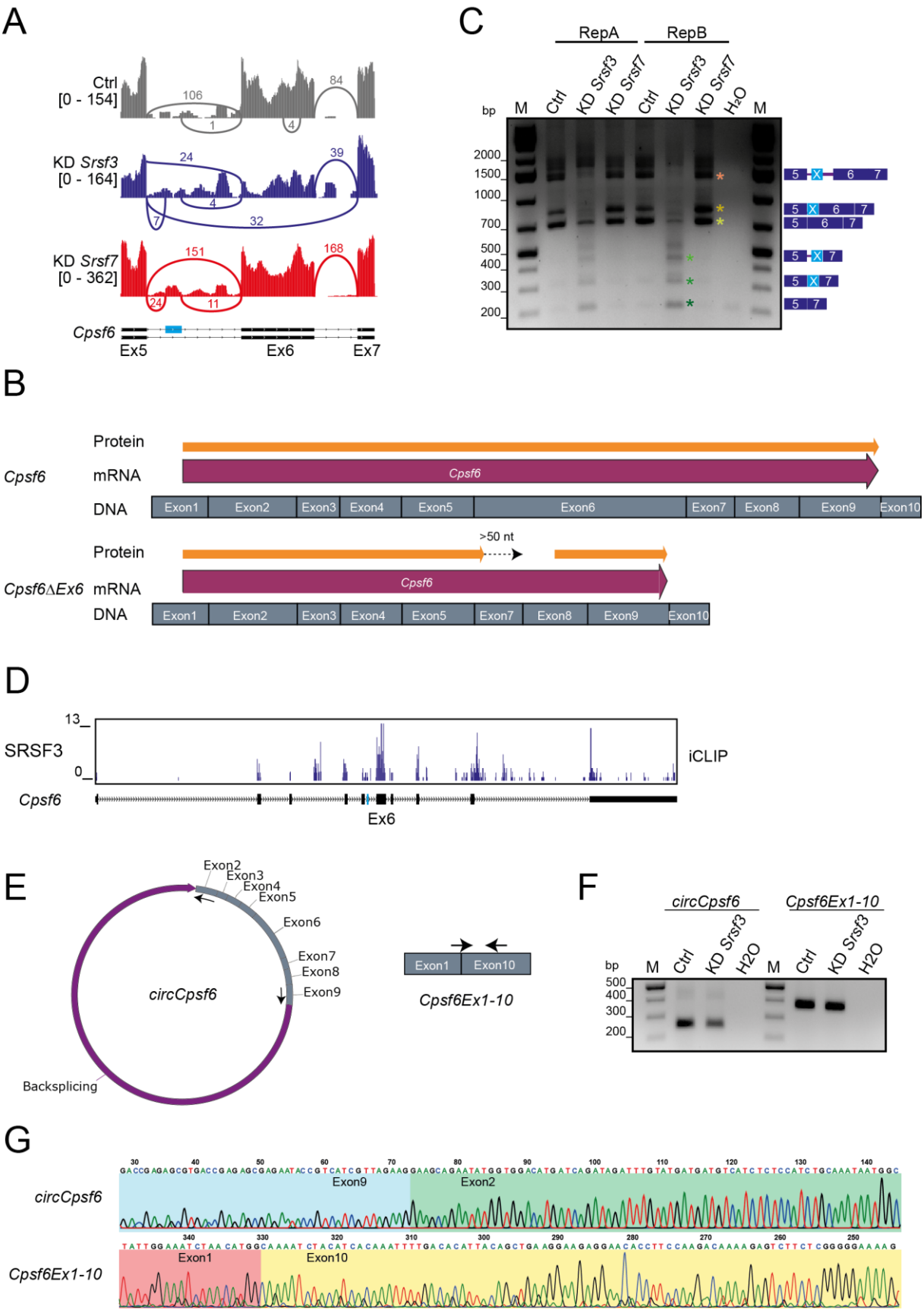


Figure 47: KD of Srsf3 leads to unproductive splicing of Cpsf6. A) *Sashimi plot of RNA-Seq reads at the region between Exon5 to Exon7 of Cpsf6 after KD of Srsf3 or Srsf7. KD of Srsf3 led to skipping of Exon6, while KD of Srsf7 increased insertion of a small exon between Exon5 and Exon6.* ***B)*** *Schemes of the transcripts and correlating proteins derived from Cpsf6 or Cpsf6ΔEx6 transcripts. Skipping of Exon6 creates a PTC in Exon7 and generates a potential NMD target.* ***C)*** *Validation of alternative*

splicing events of Cpsf6 after KD of Srsf3 and Srsf7 by PCR using primers in Exon5 and Exon7. Schemes of the potential alternative isoforms are depicted on the right side matching the size of the respective amplicons. M = marker (O'GeneRuler 1 kb Plus) ***D)*** *Browsershot of SRSF3 binding to Cpsf6 based on iCLIP data. SRSF3 is especially enriched around Exon6.* ***E)*** *Schemes of circCPSF6 and Cpsf6Ex1-10 isoforms. Arrows indicate specific primer for PCR amplification of both isoforms.* ***F)*** *Validation of a circular transcript from Cpsf6 ligating Exon9 and Exon2 as well as the corresponding linear splicing product Exon1-10 by PCR. Primers are specific for each isoform and are indicated in* ***E)****. M = marker (O'GeneRuler 1 kb Plus)* ***G)*** *Validation of the circular and linear transcript junctions by Sanger sequencing.*

Reduction of *Cpsf6* mRNA levels by SRSF3 depletion and unproductive splicing resulted in a 2-fold decrease of CPSF6 protein levels (**Figure 48**). Interestingly, reduction of CPSF6 protein resulted in the co-depletion of CPSF5 similar to what was observed with the *Cpsf6* KD. In contrast, KD of *Srsf7* did not affect expression of either subunit of the CFIm.

In conclusion, Binding of SRSF3 to Exon6 facilitates the inclusion of this exon into the full-length *Cpsf6* transcript, while KD of *Srsf3* results in unproductive splicing and skipping of this exon. This generates an unstable transcript isoform, which is potentially degraded by NMD decreasing CPSF6 and CPSF5 protein levels.

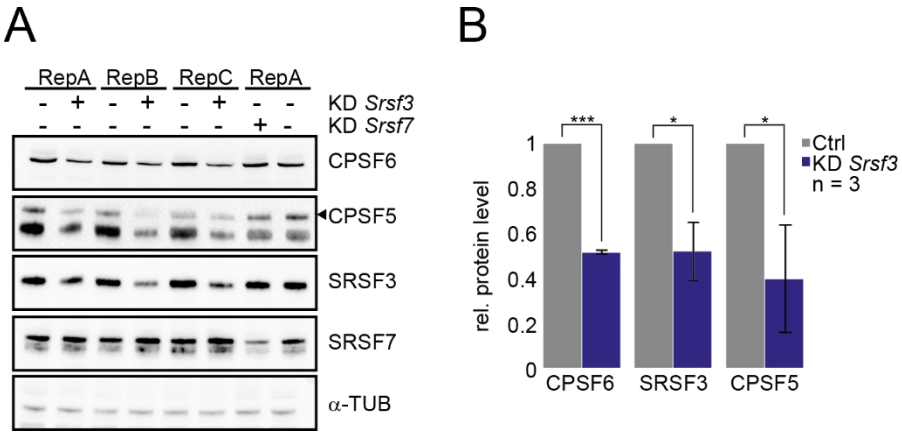


Figure 48: KD of Srsf3 decreases expression of CPSF6 and CPSF5. A) *CPSF6 and CPSF5 levels decreased after KD of Srsf3 as validated by Western blot. Srsf7 KD was used as control. α-TUB was detected as loading control.* ***B)*** *Quantification of A). n = 3, * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.005.*

It was shown that CFIm binds preferentially to dPASs and enhances its usage by the recruitment of FIP1 and the remaining CPA machinery (Zhu et al., 2018). Consequently, depletion of either subunit of the CFIm, *Cpsf5* or *Cpsf6*, in human cells resulted in a global shortening of 3' UTRs (Gruber et al., 2012). To test whether this is also the case in murine P19 cells *Cpsf6* was depleted and total RNA was subjected to RNA-Seq (**Figure 49**, **Table 29**). Sequencing delivered similar number of reads (~ 52 million) for all samples. Between 82-85% of these reads were uniquely mapped to the murine genome (mm10).

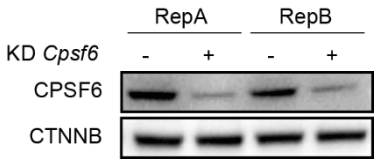


Figure 49: Validation of Cpsf6 KD by RNAi using specific esiRNAs.

Table 29: Statistics of RNA-Seq reads after KD of Cpsf6.

Sample	Replicate	Number of reads	Number of uniquely mapped reads	% uniquely mapped reads
KD *Cpsf6*	A	52723083	43433025	82.38
	B	53083364	44068154	83.02
Ctrl	A	54486515	46091913	84.59
	B	52935779	45217244	85.42

PCA analysis revealed that both KD and control samples clustered well together, while the conditions were well separated, indicating that the two biological replicates show a high degree of reproducibility and suggesting that KD of *Cpsf6* has a strong impact on gene expression (**Figure 50**).

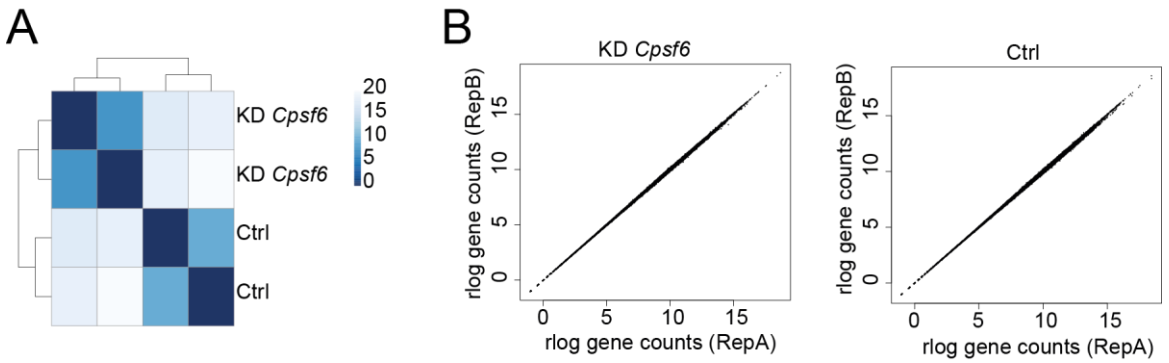


Figure 50: Sample distance and reproducibility between biological replicates of RNA-Seq samples upon KD of Cpsf6. A) *Heatmap of the Euclidean samples distance after rlog transformation for the Cpsf6 KD versus the control condition including both replicates.* ***B)*** *Scatter plot of gene counts (rlog transformed) for both replicates of the respective conditions.*

The RNA-Seq data were analyzed with DaPARS to quantify changes in 3' UTR lengths upon depletion of *Cpsf6*. Similar to human cell lines (Martin et al., 2012; Gruber et al., 2012), KD of *Cpsf6* led to a massive preference of pPAS usage and resulted in the global shortening of 3' UTRs in murine P19 cells (**Figure 51**A&B, S**uppTable 9**).

Since depletion of SRSF3 also leads to a global shortening of 3' UTRs, we next tested whether SRSF3 and CPSF6 regulate the same targets. For this we compared the DaPARS analyses after KD of *Cpsf6* and *Srsf3*. There was a very good overlap between CPSF6 and SRSF3 targets (n=500) (**Figure 51**C) indicating that 73% of all SRSF3-sensitive APA targets also reacted on *Cpsf6* KD. Moreover, most of these targets showed reduced dPAS usage (n=448) (**Figure 51**D&E), while the other possible combinations showed only little overlap. Three validated targets that showed shortened 3' UTRs upon *Srsf3* KD, also showed shortened 3' UTRs upon *Cpsf6 KD* determined by 3'RACE (**Figure 51**F).

In conclusion, these data show that SRSF3 regulate global APA changes by two different mechanisms. On the one hand, it might promote dPAS usage by directly blocking the pPAS by competing with SRSF7 for binding upstream. On the other hand, SRSF3 controls the expression levels of the dPAS enhancer CFIm by alternative splicing of the *Cpsf6* mRNA, affecting APA indirectly.

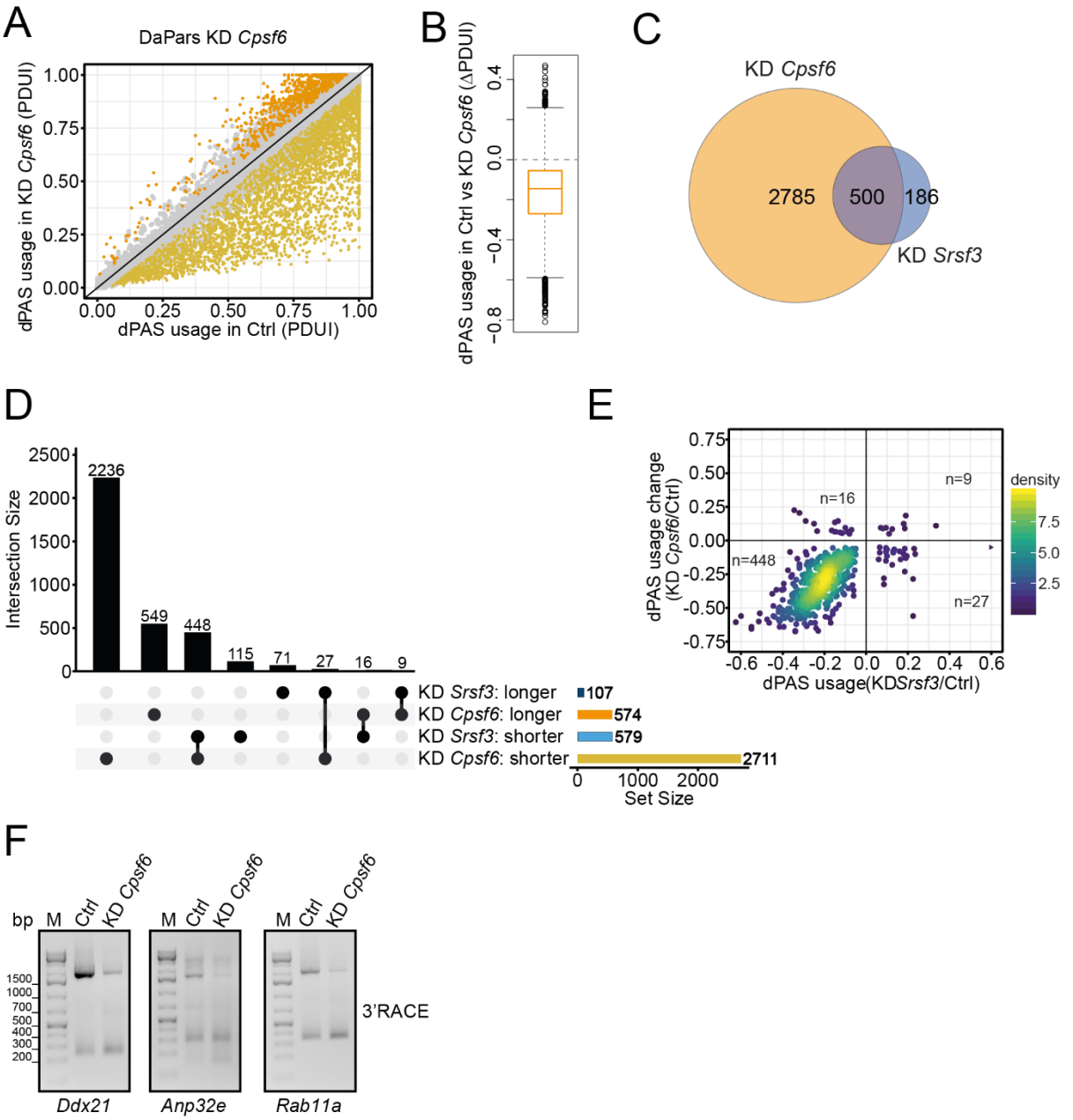


Figure 51: Depletion of CPSF6 and* SRSF3 *leads to global shortening of 3' UTRs. A) *Scatterplot showing that KD of Cpsf6 shortened the 3' UTRs of many transcripts compared to control conditions, while much fewer 3' UTRs were extended. Significant events (p.adj < 0.1) are highlighted in yellow.* ***B)*** *Boxplot showing that KD of Cpsf6 causes a global trend of 3' UTR shortening.* ***C)*** *Venn diagram showing the overlap of transcripts affected by KD of Cpsf6 or Srsf3.* ***D)*** *OffSetPlot of transcripts affected by depletion of Cpsf6 or Srsf3.* ***E)*** *Density map of dPAS usage upon KD of Cpsf6 or Srsf3 compared to control conditions* ***F)*** *3'RACE-PCR shows that three validated SRSF3 targets show 3' UTR shortening upon KD of Cpsf6. M = marker (O'GeneRuler 1 kb Plus).*

5.7 SRSF3-sensitive pPAS usage might be inhibited by CFIm binding

Previous studies established that CFIm enhances dPAS usage by preferentially binding to UGUA motifs at those sites and recruiting FIP1 (Zhu et al., 2018; Gruber & Zavolan, 2019). Yet, it stayed unclear how pPASs become activated globally, when expression of CFIm is limited, given that they are generally weaker than dPAS. Moreover, SRSF3-regulated pPAS seem to be particularly sensitive to low CFIm levels. To test whether these pPAS have features we first compared the prevalence and distribution of CFIm binding motifs (UGUA) around all mapped sPASs, pPASs and dPASs from our MACE-Seq data and compared the patterns to CPSF6- and SRSF3-sensitive PAS. Consistent with previous studies, UGUA motifs were enriched upstream of sPASs and dPASs (**Figure 52**A). Their presence was increased upstream of CPSF6-sensitive dPAS (**Figure 52**B), and they were most enriched upstream of SRSF3-sensitive dPAS, suggesting that their dPAS usage depends on CFIm enhancement (**Figure 52**C). Remarkably, the distribution of CFIm binding sites was very different for pPASs. At these sites, the UGUA motif was found upstream and downstream of the pPAS. Remarkably, this two-sided pattern at the pPAS was even more prominent in SRSF3-sensitive pPAS with an equal enrichment up- and downstream of the pPAS (**Figure 52**C).

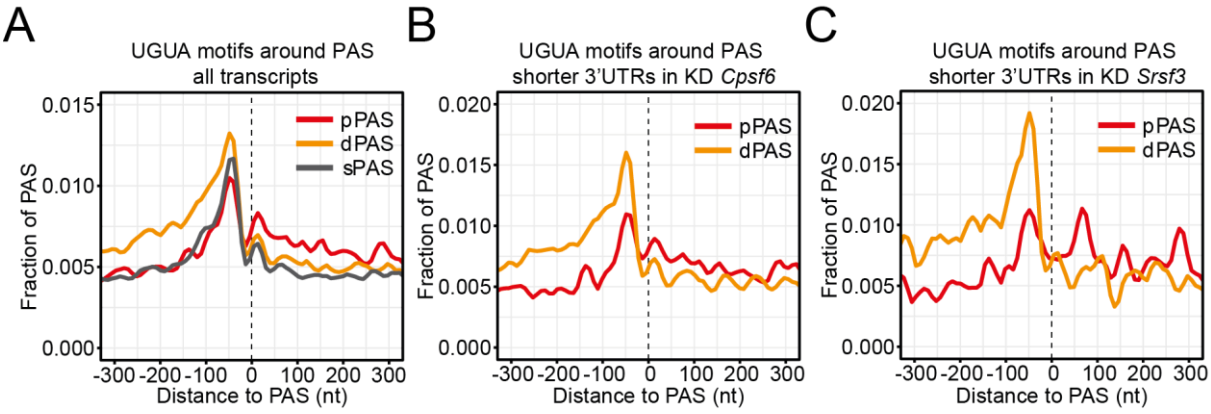


Figure 52: CFIm binding motif UGUA is enriched upstream of sPAS and dPAS but has a bimodal distribution around the pPAS. A) *Enrichment of UGUA motifs up- and downstream of pPASs, dPASs and sPASs.* ***B)*** *UGUA motif enrichment up- and downstream of pPASs and dPASs in CPSF6-sensitive transcripts.* ***C)*** *UGUA motif enrichment up- and downstream of pPASs and dPASs in SRSF3-sensitive transcripts.*

To validate whether these paired UGUA motifs originate from the same transcript we analyzed the frequency of dual-UGUA motifs at PAS (**Figure 53**A). Indeed, the occurrence of dual UGUA motifs was highest in SRSF3-sensitive transcripts enclosing the dPAS and the pPAS. Interestingly, especially SRSF3 APA targets contain UGUA pairs on one side of the respective PAS. At the pPAS UGUA pairs were enriched downstream, while at the dPAS UGUA pairs were enriched upstream. The occurrence of these dual CFIm binding motifs is especially interesting, as the CFIm heterotetramer can bind to two UGUA motifs via the two CPSF5 subunits, looping the intervening RNA around the neighboring CPSF6 subunit (Yang et al., 2011a; Yang et al., 2011b). We next analyzed the distances between two neighboring

UGUA motifs (**Figure 53**B), which revealed that the distance was much larger when the PAS was enclosed (~80 nt) compared to paired motifs up- or downstream of a PAS (~20 nt). These data could explain why SRSF3-regulated APA targets are more sensitive to limiting CFIm levels. The CFIm seem to have a dual function in regulating APA decisions either blocking pPAS usage by looping and hiding the pPAS and/or DSEs, or activating dPAS usage by binding to UGUA pairs upstream (Zhu et al., 2018).

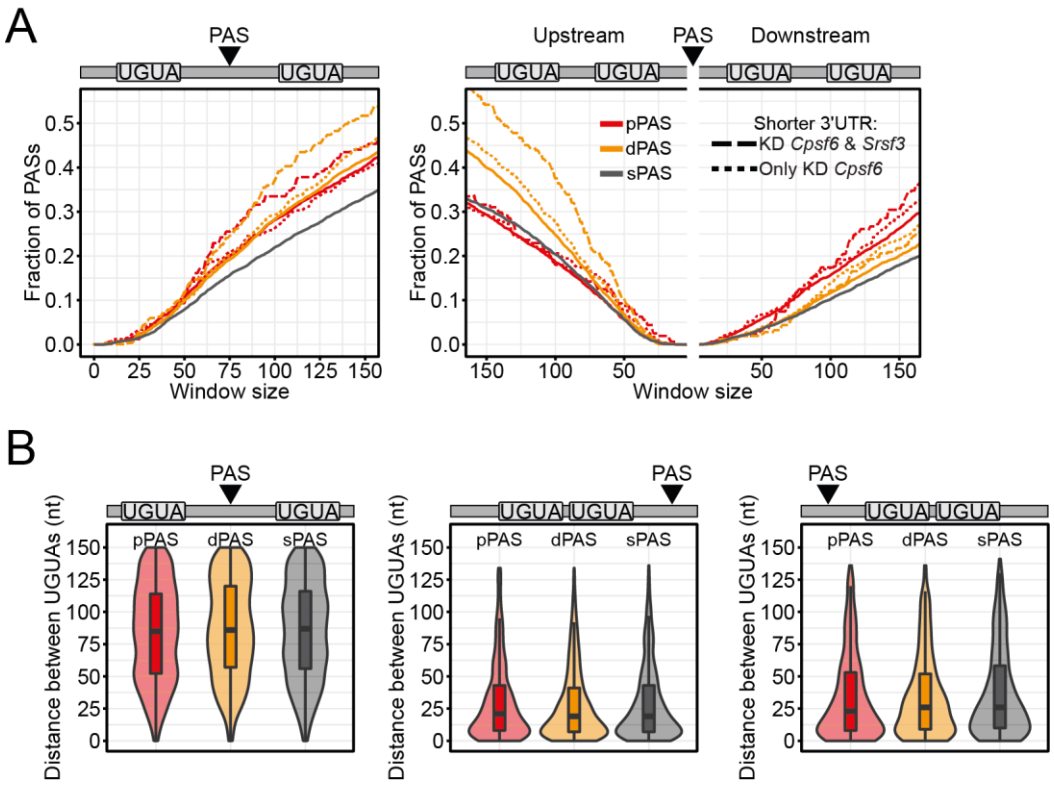


Figure 53: Distribution of UGUA motifs and distances between two adjacent UGUA motifs around sPASs, pPASs and dPASs. ***A)*** *Cumulative fraction of dual UGUA motifs enclosing PAS and preceding or following PAS in sPASs, pPASs and dPASs.* ***B)*** *Violin plots representing the distance in nt between UGUA pairs enclosing (left), preceding (middle) or following (right) PASs.*

To answer this question and verify whether these dual UGUA motifs are indeed recognized by CFIm and whether recruitment of FIP1 takes place, we performed iCLIP of CPSF5-GFP and GFP-FIP1 in P19 cells (6 replicates), to identify the direct binding sites of CFIm and FIP1 (**Figure 54**A-D). To increase the number of true binding sites all replicates were pooled before peak calling. This delivered 1,851,266 unique crosslink events which were matched to 1,659,525 unique positions for CPSF5 and 3,759,237 unique crosslink events which were matched to 3,242,512 unique positions for FIP1. As expected, binding of CPSF5 and FIP1 was especially enriched in the 3' UTR and ncRNA (**Figure 54**E). Pentamer enrichment analysis around X-links (windows of 30 nt in between -30 nt to -5 nt upstream of 5 nt to 30 nt downstream of the respective x-link) revealed the expected binding motifs of both CPA factors, UGUA for CPSF5 (**Figure 54**F) and UG-rich sequences for FIP1 (**Figure 54**G).

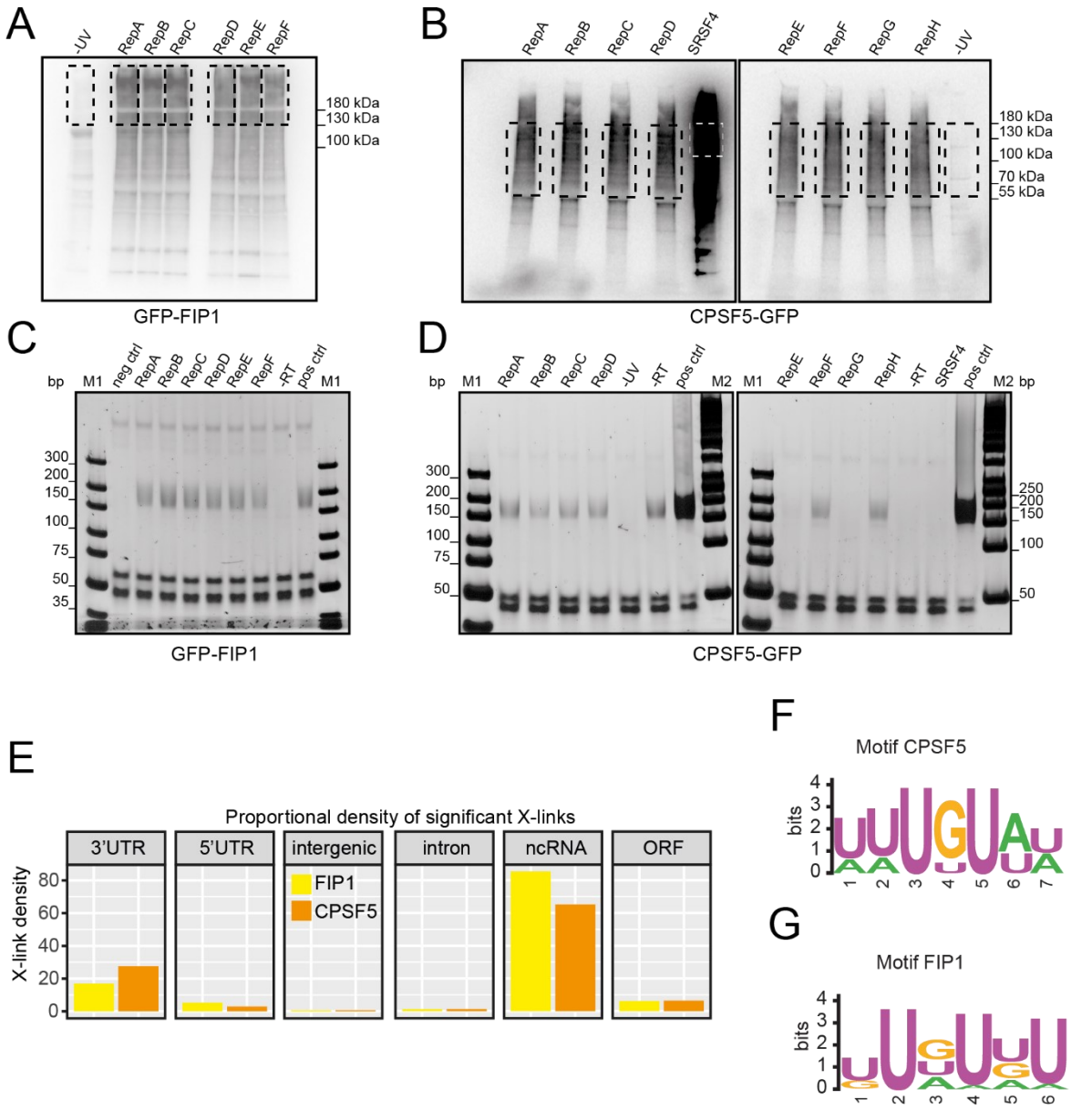


Figure 54: Generation of iCLIP libraries from GFP-FIP1 and CPSF5-GFP. A&B) *Autoradiographs of iCLIP experiments using α-GFP antibody to pull down GFP-FIP1 (**A**) and CPSF5-GFP (**B**) after UV-crosslinking. Crosslinked RNA was labelled radioactively using ^{32}P. Non-crosslinked (-UV) samples served as negative control, while SRSF4-GFP was used as a positive control. Black squares indicate the regions cut to prepare the sequencing libraries.* ***C&D)*** *Amplified libraries of GFP-FIP1 (**C**) and CPSF5-GFP (**D**), separated on TBE-6% Urea gels. Controls lacking reverse transcriptase (-RT), UV-crosslinking (-UV) were included as well as positive controls (pos. ctrl.) from previous successful library amplifications. M1 = marker1 (O'GeneRuler ULR); M2 = marker2 (O'GeneRuler 50 bp)* ***E)*** *Proportion plots showing the density enrichment of significant crosslink-sites in different transcript regions.* ***F&G)*** *Pentamer motif enrichment at iCLIP crosslink sites of CPSF5 (**F**) and FIP1 (**G**).*

Next, we integrated the identified binding sites of CPSF5 and FIP1 with our previously mapped PASs in P19 cells. Binding of both cleavage factors was enriched between 40 - 80 nucleotides upstream of the mapped PAS (**Figure 55**A&B). Similar to the comparable enrichment of UGUA motifs upstream of sPASs, pPASs and dPASs (**Figure 52**A), CPSF5 binds with similar frequencies at all three PAS types. Interestingly, at pPASs CPSF5 showed binding peaks up- and downstream, enclosing the PAS, which was not seen at the sPAS or dPAS, while the binding peak upstream of dPAS was widened, indicating

binding of potentially two CPSF5 proteins (**Figure 55**B). FIP1 always binds up- and downstream of PASs with an increased enrichment slightly downstream of the CPSF5 peak. Interestingly, binding of FIP1 was highest around the pPAS with nearly equal levels before and after the PAS. In contrast, at the sPAS and dPAS overall levels of FIP1 binding were reduced and the peak downstream of the PAS was much smaller than the respective upstream peak, similar to the distribution pattern of dual UGUA motifs shown previously.

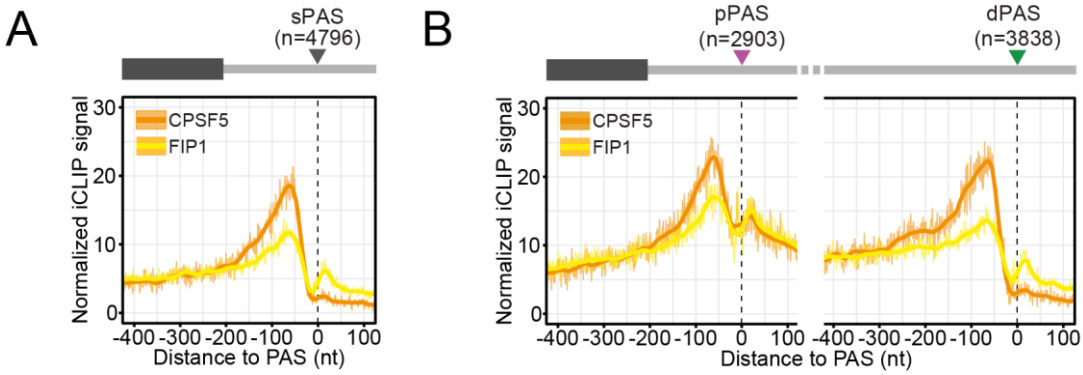


Figure 55: CPSF5 and FIP1 binding enriches upstream of PASs. A&B) *Metaprofiles of normalized iCLIP signals of CPSF5 and FIP1 up- and downstream of sPASs (**A**) or pPASs and dPASs (**B**).*

Notably, binding of CPSF5 increased significantly in SRSF3-sensitive transcripts compared to random transcripts (**Figure 56**A). Especially binding downstream of the pPAS and upstream of the dPAS increased strongly compared to random transcripts. A similar tendency was observed for FIP1, where the up- and downstream peaks around the pPAS display equal intensities. CPSF6-regulated transcripts only showed a slightly significant increase in CPSF5 binding upstream of dPAS (**Figure 56**B). This suggests that pPAS usage in SRSF3-sensitive transcripts is regulated via CFIm binding and FIP1 recruitment. However, not all SRSF3 targets were also affected by CPSF6 depletion. 184 transcripts, similar to the number of SRSF7-sensitive targets, were SRSF3-only targets, again supporting that there might be two distinct mechanisms how SRSF3 regulates APA by promoting dPAS usage. Indeed, SRSF3-only APA targets showed less CPSF5 binding in their 3' UTRs (16.6%), compared to common SRSF3-CPSF6 targets (37%).

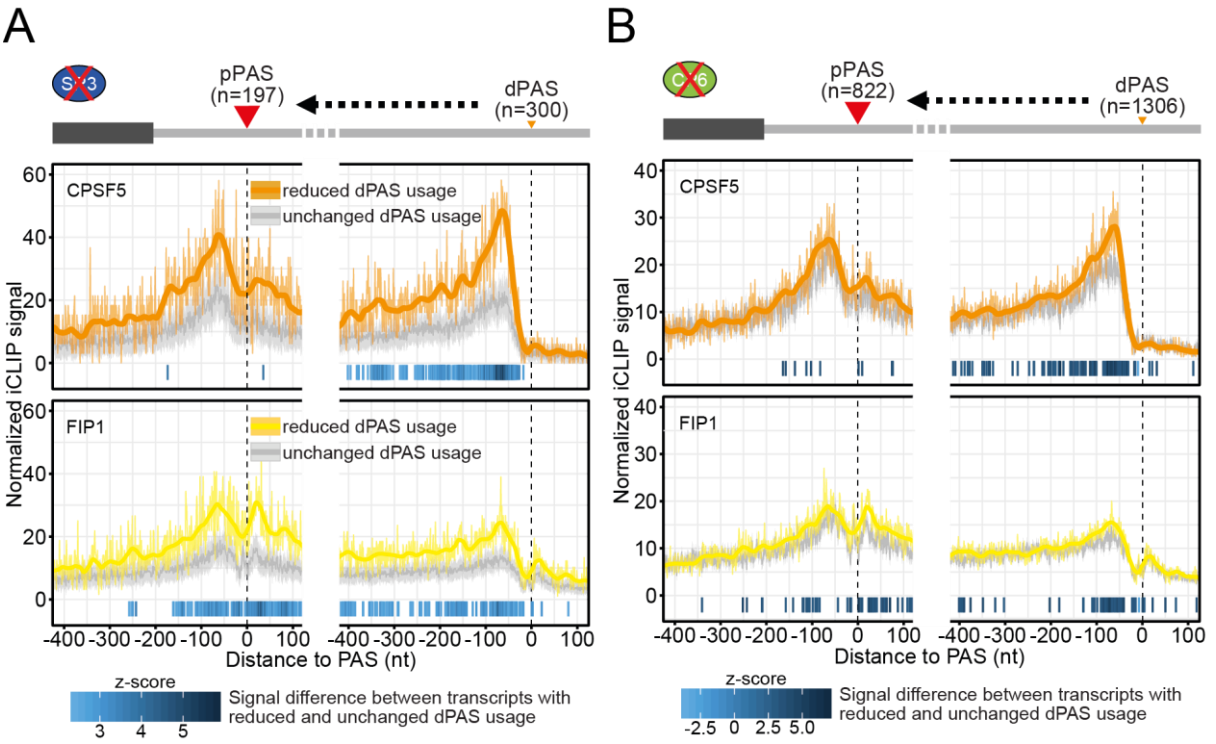


Figure 56: Binding of CPSF5 and FIP1 at pPASs and dPASs is higher in SRSF3-sensitive transcripts. A&B) *Metaprofiles of normalized iCLIP signals of CPSF5 and FIP1 up- and downstream of pPASs and dPASs in SRSF3-sensitive transcripts (**A**) or CPSF6-sensitive transcripts (**B**) compared to random PAS (grey). Significant binding differences are calculated and shown by a z-score.*

Considering that pPASs in SRSF3-sensitive transcripts are suppressed under conditions where iCLIP was performed, cleavage might be prevented by CPSF5 binding and unproductive recruitment of FIP1. To confirm this, transcripts with short and long 3' UTRs in control conditions were separately investigated for their CPSF5 and FIP1 binding patterns (**Figure 57**A). Indeed, CPSF5 and FIP1 were found to bind to pPAS in long 3' UTRs, with FIP1 displaying the previously seen dual binding up- and downstream of the pPAS, although the pPAS is not used. This suggests that a combination of different mechanisms regulates APA. Whereas enhancement of dPAS usage might be sufficient for most transcripts, for certain transcripts, like SRSF3-sensitive targets, additional mechanisms appear to be necessary to suppress pPAS usage. This might be necessary, when pPASs are too strong. To investigate this, we compared the abundance of poly(A) signal hexamer variants at pPAS and dPAS in all transcripts and those getting shorter 3' UTRs by KD of *Cpsf6* or by both *Cpsf6* and *Srsf3* (**Figure 57**B). Indeed, in all transcripts and the CPSF6-only-transcripts subset only ~60% of the pPAS were accompanied by the two strongest cleavage hexamers (AAUAAA and AUUAAA), while this proportion increased to ~75% in SRSF3-sensitive transcripts, indicating that their pPAS are especially strong and might require additional inhibition. In comparison the dPAS showed generally higher ratios of the strong hexamer

motifs as shown previously in **Figure 18**D. Notably, the overall strength of the dPAS was also increased in SRSF3-sensitive transcripts, but to a lesser degree than the pPASs.

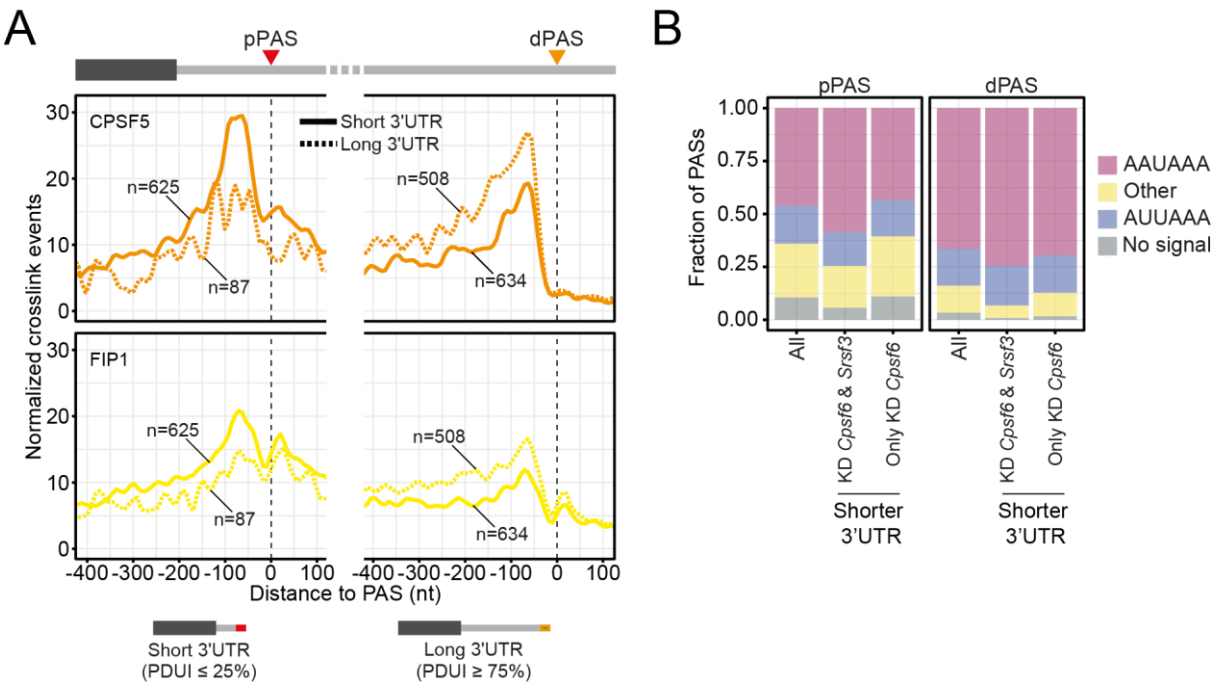


Figure 57: CPSF5 and FIP1 are recruited to pPASs of short and long 3' UTRs. A) *Metaprofiles of normalized iCLIP binding sites of CPSF5 and FIP1 in a window up- and downstream of pPASs and dPASs of all transcripts with short or long 3' UTRs (PDUI ≤ 0.25 or ≥ 0.75, respectively).* ***B)*** *Proportion of poly(A) signal motifs at pPASs (left) or dPASs (right) comparing SRSF3-sensitive and CPSF6-sensitive 3' UTRs with all 3' UTRs.*

Taken together, these results corroborate the hypothesis that the CFIm can regulate APA in different ways in a position-dependent fashion. On the one hand, CFIm might actively inhibit pPAS usage. Either the pPAS is enclosed by CPSF5 binding to two UGUA motifs and looping out central elements of the cleavage site, or by binding to two downstream UGUA motifs thereby hiding DSEs. On the other hand, CFIm enhances dPAS usage by binding to two UGUA motifs upstream of the dPAS. In both cases CFIm can recruit FIP1 and potentially the remaining cleavage machinery. But in the case of unused pPASs FIP1 recruitment must be unfavorable or inactive preventing the cleavage reaction. This bimodal CFIm-functions might be necessary for the expression of long 3' UTRs in SRSF3-sensitive transcripts, as they harbor particularly strong pPAS compared to CPSF6-sensitive targets, which did not exhibit such a bimodal response to CFIm levels.

5.8 SRSF7 and FIP1 levels decrease during neuronal differentiation resulting in global 3’ UTR extension

Finally, we became interested in investigating whether the SR protein-dependent regulation of 3’ UTR-APA is relevant during cellular differentiation. It is known that proliferating and transformed cells preferentially express transcript isoforms with shorter 3’ UTRs compared to differentiated cells. Therefore, the pluripotent P19 cell line used in this study was perfectly suited to analyze changes in APA during differentiation, as they can be easily differentiated into neuronal cells using retinoic acid (Nakayama et al., 2014). Successful differentiation was validated morphological by microscopy (**Figure 58**A). After 8 days of differentiation the cells had developed multiple neurites representing the expected neuronal morphology. In addition, the expression of pluripotency (OCT4) and neuronal markers (SYNAPSIN and NESTIN) was validated by Western blot (**Figure 58**B). Expression of the neuronal marker SYNAPSIN and NESTIN increased and the expression of the pluripotency marker OCT4 decreased.

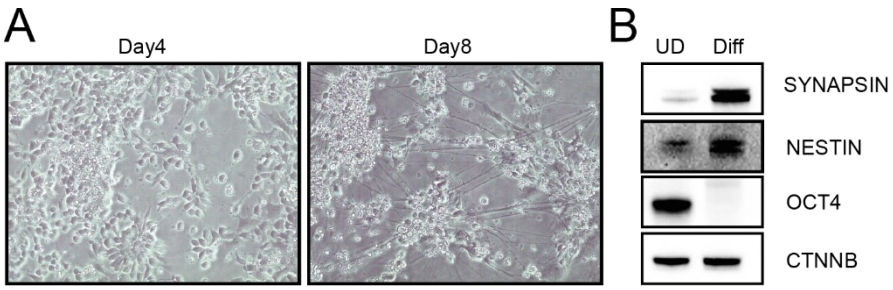


Figure 58: Differentiation of P19 wt cells into neuronal cells using retinoic acid. A) *Morphological changes of P19 cells during differentiation, documented on days 4 and 8 after induction.* ***B)*** *Validation of the pluripotency marker OCT4 and the neuronal markers SYNAPSIN and NESTIN in undifferentiated and differentiated P19 cells by Western blot. CTNNB was probed as loading control.*

Subsequently, total RNA from differentiated and control samples (3 replicates) was subjected to RNA-sequencing. Sequence-reads were quality-filtered and mapped against the murine genome (mm10) as described before. Undifferentiated and differentiated samples separated well from each other as validated by PCA analysis (**Figure 59**A). Changes in the length of 3’ UTRs were quantified using DaPARS as before. As expected, differentiation resulted in 3’ UTR lengthening of many transcripts, while only a few shortened (**Figure 59**B, **SuppTable 10**). This trend was also obvious on the global scale (**Figure 59Figure 58**C).

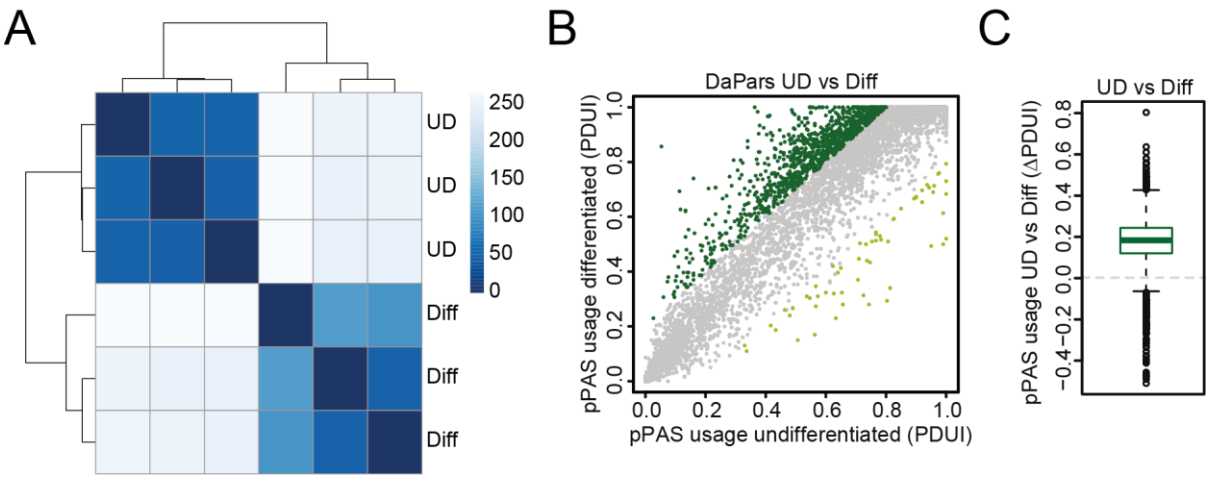


Figure 59: 3' UTRs globally elongated after differentiation. A) *Heatmap of the Euclidean samples distance after rlog transformation for the differentiated versus the control condition including both triplicates.* ***B)*** *Scatterplot showing transcripts that were 3' UTR extended upon differentiation (dark green) or shortened (light green). Significant events (padj < 0.1) are highlighted.* ***C)*** *Boxplot revealing that differentiation leads to a global extension of 3' UTR length.*

To gain more insight into the regulation of 3' UTR length upon differentiation the expression of SRSF3, SRSF7, CPSF6 and FIP1 was analyzed by Western blot in undifferentiated and differentiated cells (**Figure 60**A). Interestingly, expression of SRSF7 and FIP1 was strongly reduced upon differentiation (by ~3-fold), while expression of SRSF3 and CPSF6 did not change. This suggests that the extensions of 3' UTRs during differentiation is accompanied by decreasing levels of the pPAS activating factors SRSF7 and FIP1. This finding supports the previous characterization of FIP1 being an important pluripotency factor important for stem cell renewal, where depletion of *Fip1* resulted in global 3' UTR lengthening (Lackford et al., 2014). To further confirm this, mRNA expression was analyzed using the previously generated RNA-Seq data analyzed with DESeq2 (**Figure 60**B, **SuppTable 11**). Interestingly, *Srsf7* and *Fip1l1* (encoding FIP1) expression levels were less decreased than their protein levels (1.65-fold vs. 3-fold and 1.33-fold vs. 3-fold) in differentiated cells. This suggests that the low protein levels are also due to decreased translation or protein stability. A comparison between targets with elongated 3' UTRs after depletion of SRSF7 or differentiation (when SRSF7 levels were reduced) showed a good overlap (**Figure 60**C&D). Finally, binding of SRSF7 at pPASs and dPASs of transcripts being elongated during differentiation was analyzed (**Figure 60**E). SRSF7 was significantly enriched upstream of pPASs in those transcripts, supporting the pPAS activating function of SRSF7. SRSF3 binding was only slightly enriched at those pPAS and at the dPASs both proteins were equally enriched.

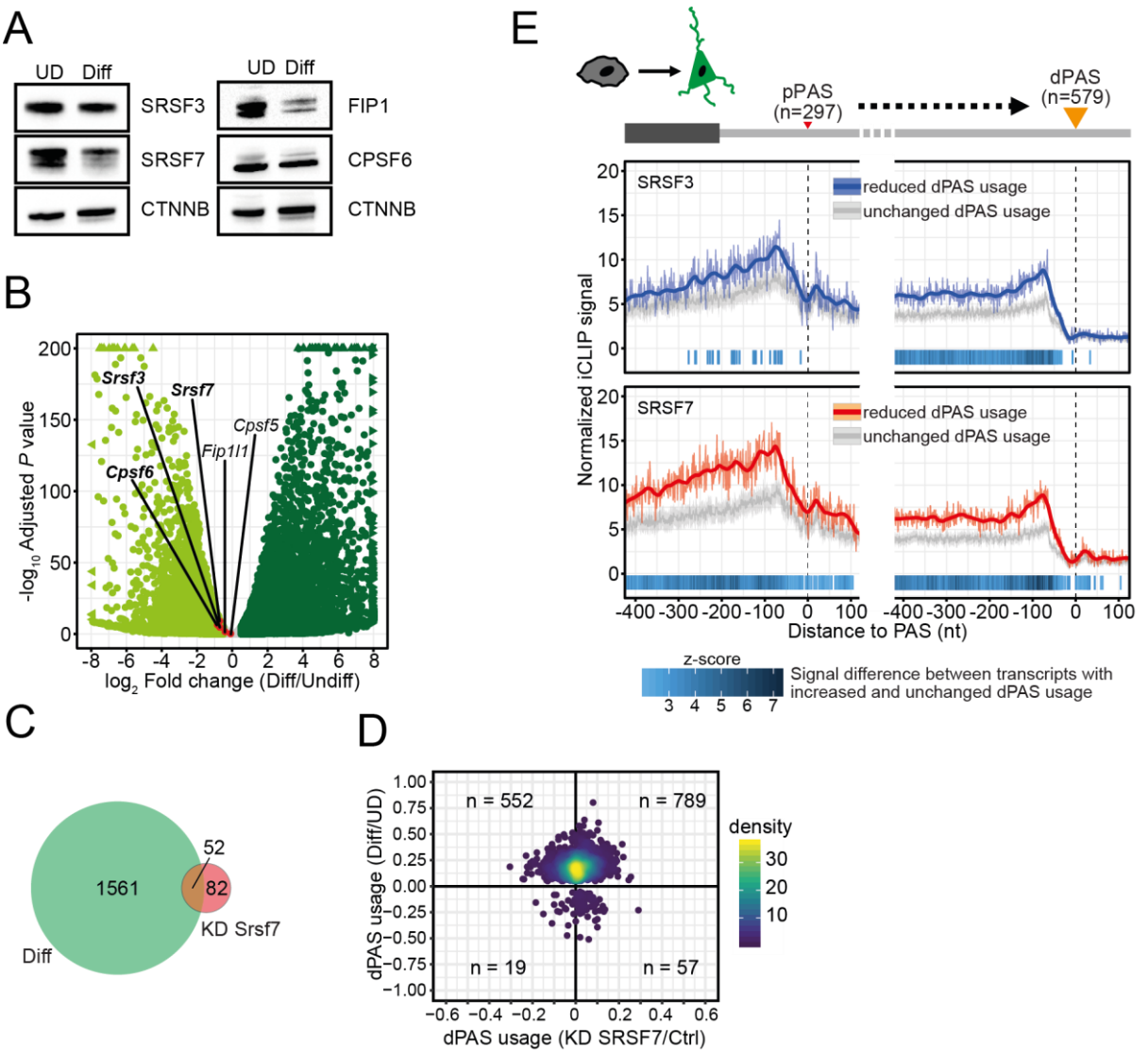


Figure 60: Expression of SRSF7 and FIP1 decreases in neuronal differentiated P19 cells. A) *Western blot analysis of expression of SRSF3, SRSF7, CPSF5 and CPSF6 in undifferentiated and differentiated P19 cells. CTNNB was detected as loading control.* ***B)*** *Volcano plot of differential gene expression after differentiation of P19 cells. Analyzed by DESeq2. Significant genes are highlighted in light green and dark green and Srsf3, Srsf7, Cpsf5, Cpsf6 and Fip1 are indicated.* ***C)*** *Venn diagram showing the overlap of genes changing their 3' UTR length after differentiation or depletion of Srsf7.* ***D)*** *Scatterplot comparing dPAS usage in the differentiated dataset and the Srsf7-depletion dataset.* ***E)*** *Metaprofiles of normalized iCLIP binding sites of SRSF3 and SRSF7 in windows up- and downstream of pPASs and dPASs with elongated 3' UTRs upon differentiation compared to unchanged dPASs.*

In conclusion, these results suggest that SRSF7 and FIP1 regulate pPAS usage in a concentration-dependent manner during differentiation. In the pluripotent state both proteins are highly expressed and likely enhance pPAS activation, but during differentiation the expression levels of both two factors drop and concomitantly usage of the stronger dPASs increases leading to the expression of transcripts with long 3' UTRs.

6 Discussion

The experimental data and findings of this thesis expand the functional range of the essential SR protein family of splicing factors by a significant effect on the regulation of alternative polyadenylation within 3' UTRs. Especially two core SR proteins, SRSF3 and SRSF7, were found to affect 3'UTR-APA globally by modulating proximal polyadenylation site (pPAS) usage, supporting previously published findings (Müller-McNicoll et al., 2016). Moreover, distinct regulation mechanisms were discovered highlighting a direct role of SRSF7 and SRSF3 in activating or blocking pPAS usage, respectively, while SRSF3 additionally modulates 3'UTR-APA indirectly by controlling the expression levels of the dPAS cleavage enhancer CFIm.

6.1 The first poly(A)-tome of pluripotent P19 mouse cells

Until today many studies reported the identification, mapping, and analysis of PASs usage in the murine model organism utilizing various 3'end sequencing strategies, such as 3'READS/3'READS(+), PAPERCLIP, PolyA-Seq, A-Seq and 3P-Seq (Jan et al., 2011; Derti et al., 2012; Gruber et al., 2012; Martin et al., 2012; Hoque et al., 2013; Batra et al., 2014; Gruber et al., 2014; Nam et al., 2014; Li et al., 2015; Hwang et al., 2016; Zheng et al., 2016; Hwang et al., 2017; Jereb et al., 2018). These studies included samples from different cell-lines and different primary cells originating from various tissues. Since 2015 the polyAsite atlas project began to collect the information from those studies to combine them to build a powerful resource data base for poly(A) site-related research (Gruber et al., 2016). Recently, the polyAsite atlas was updated, now including datasets from 178 sequencing libraries (Mus musculus, v.2.0 GRCm38.96, 20.04.2020) combining nearly 1.2 billion reads (Herrmann et al., 2020). Surprisingly, no dataset generated from pluripotent P19 cells has been included.

Murine, pluripotent P19 cells were derived from embryonic teratocarcinoma cells (McBurney & Rogers, 1982) and they can be differentiated into neuronal (Jones-Villeneuve et al., 1983; Nakayama et al., 2014) or mesodermal and endodermal cell types (McBurney et al., 1982). We decided to generate the first poly(A)-tome dataset of P19 cells using MACE-Seq as a sequencing strategy (Müller et al., 2014; Zawada et al., 2014), as we were convinced that PAS usage in these particular cells would be very different compared to testis or muscle cells or other terminally differentiated cells. Even more, we also would expect PAS usage in the P19 cells to differ from other pluripotent cells, such as mouse embryonic stem cells (mESCs), due to their origin as embryonal carcinoma cells derived from teratocarcinomas. We chose MACE-Seq as the sequencing method of choice as it offers the return of robust data on a costs and material saving platform. The biggest advantage is the individual barcoding of each transcript to eliminate PCR duplication biases after library preparation and sequencing.

Unfortunately, after duplicate filtering, most of the reads were lost indicating a library overamplification. To increase the depth of the poly(A)-tome the sequencing reads of all samples were merged and PASs were assigned using a customized pipeline. In total we identified 15,866 PASs with high confidence, which is a much lower number compared to other PAS-profiling reports based on in-house developed 3'UTR-sequencing methods (Derti et al., 2012; Martin et al., 2012; Hoque et al., 2013; Lianoglou et al., 2013), which identified 127,014; 31,906; 70,608 and 80,371 PASs, respectively, and which can be attributed to the low sequencing depth of our MACE-Seq datasets. Currently, mapping of 3'-end sequencing-libraries-derived reads is less effective compared to total RNA-seq, due to technical limitations in read-length/read-quality and read-through of homopolymers, such as the polyA-tail. For example, in the study presented here, on average 55% of the MACE-Seq reads mapped uniquely, while around 82% of total RNA-Seq reads mapped uniquely. In consequence, the exclusion of reads during quality control, mapping and downstream filtering might have led to the loss of potential PASs, as the respective read counts were below the stringent threshold. An increase in the number of 3'end reads (sequencing depth) will certainly expand the poly(A)-tome of P19 cells.

We began by evaluating basic parameters and principal features of all mapped PASs. The vast majority of all PASs were found in protein coding genes (99.6 %; 15,805/15,866) within the 3'UTRs (86,7%; 13,706/15,866), similar to previous reports in murine and human cell-lines (Lianoglou et al., 2013; Hoque et al., 2013). A comparison of the PASs coordinates obtained from MACE-Seq with PASs coordinates annotated by GENCODE showed a good overlap with around 60% of the PAS being in very close proximity. Considering that GENCODE represents a summary of all types of gene features merging data derived from computational analyses, manual annotations, experimental validations, and further evidence-based features, the high overlap supports the usefulness of our MACE-Seq dataset for an extensive analysis of the P19 poly(A)-tome. In addition, evaluation of a well-defined set of potential poly(A) signal motifs (Gruber et al., 2016) revealed an enrichment of the strongest known PAS-motif 'AAUAAA' followed by 'AUUAAA', while all remaining PAS-motifs were found with lower frequency. In summary, the ratio of the different motifs identified in our study resembled the findings from murine data recently published (Hoque et al., 2013; Gruber et al., 2016). 55% of protein-coding genes exhibited only one PAS, while the remaining genes contained at least two PASs, being subject of alternative polyadenylation (APA). Interestingly, these proportions vary from the numbers presented by Derti et al., 2012; Hoque et al., 2013. Those studies had identified that 70% of all transcripts in human cells and up to 80% of all transcripts in murine cells have the potential to undergo 3'UTR-APA and use two or more PASs per transcript. The discrepancy to our finding might be explained by the fact that we used only one cell type in our study (highly proliferating P19 cells), while the other studies compared several differentiated cell-lines and tissues. Similarly to our finding, Gruber et al. identified an enrichment of genes with only one PAS (65%) compared to genes with two, three or four and more

PAS (20%; 9% and 6%, respectively) in proliferating T-cells (Gruber et al., 2014). In the case of the study by Hoque et al., the investigated PASs were extracted from various (commercial) murine cell-line mixes, murine cell-lines and murine tissues by 3'READS and they found on average 4 PASs per gene. Nevertheless, it must be pointed out that, similar to our data, most mRNA genes contained only a single PAS and the number of genes containing additional PASs decreased with the number of PASs mapped. Another explanation could be that highly proliferating and cancer cells shift towards the expression of shorter 3'UTRs by activation of the pPASs, which increases the transcript stability (Sandberg et al., 2008; Mayr & Bartel, 2009). In contrast, differentiation leads to the lengthening of the majority of 3'UTRs (Ji et al., 2009; Chen et al., 2018; Jereb et al., 2018). With respect to the poly(A)-tome of P19 cells, the potentially higher activity of pPASs and expression of transcripts with short 3'UTRs might artificially increase the number of genes that use only one PAS. This is because additional PASs located downstream of the pPASs were no longer present in the transcript and thus could not be picked up during sequencing and mapping, respectively.

Sorting PASs by their number and relative position to the end of the coding sequence into sPAS, pPAS, dPAS and oPAS, we found that sPASs are especially strong being highly enriched in the 'AAUAAA' and 'AUUAAA' PAS-motifs. Interestingly, with the presence of more than one PAS per transcript the frequency of these two strongest PAS-motifs decreases and alternative sequences prevail, whereby the pPASs are always the weakest PAS and the PAS-motif strength increases towards the dPASs. This finding suggests that the pPAS is subject to extensive APA regulation, while the dPASs are used rather by default as they should be recognized easier by the main cleavage-factor CPSF (Schönemann et al., 2014; Chan et al., 2014; Clerici et al., 2017; Clerici et al., 2018). Our results are well in line with similar analyses made in previous studies, which described similar proportional variances in PAS-motif compositions at sPASs, pPASs and dPASs (Martin et al., 2012; Hoque et al., 2013).

In conclusion, we present for the first time the poly(A)-tome of pluripotent P19 cells. In comparison to data from other murine cell-lines and tissues we found that the composition and ratios of PAS-motifs have been highly similar in general as well as in respect to the different PAS types (sPASs, pPASs and dPASs) in transcripts enabling 3'UTR-APA. Interestingly, pluripotent P19 cells express more transcripts with only one PAS than differentiated cells. This might be due to the dominant expression of transcripts with short 3'UTRs, compared to differentiated cells, where 3'UTRs are elongated using the dominant dPAS. Even though our dataset was very useful for our downstream analyses, it must be acknowledged that in comparison to other murine cell-lines and tissues, our MACE-Seq dataset from P19 cells has a limited information content, due to library overamplification and low sequencing depth. It is likely that some PASs have been excluded during the analysis pipeline because of too low read coverage.

6.2 SRSF3 and SRSF7 have opposite effects on pPAS usage

An earlier publication by our group (Müller-McNicoll et al., 2016) suggested that members of the SR protein family of splicing factors also regulate APA. The effect of each individual SR protein on APA was very variable, but two SR proteins - SRSF3 and SRSF7 - stood out as the most potential APA regulators. These findings were supported by publications, which have been describing examples of APA regulations of both proteins, albeit in viral RNAs or connected to alternative splicing (Lou et al., 1998; Valente et al., 2009; Shen et al., 2019). Yet, none of those studies had aimed to comprehensively analyze the APA regulatory functions of SRSF3 and SRSF7 and the underlying molecular mechanism(s). Hence, no complete set of targets of SRSF3 and SRSF7-regulated transcripts had been available nor had the interactions and the molecular principles between SRSF3, SRSF7 and the CPA machinery been determined in detail.

Furthermore, our initial data suggested that both SR proteins regulate APA in opposite directions or even antagonistically, with depletion of SRSF3 leading to 3'UTR shortening of transcripts while depletion of SRSF7 resulted in 3'UTR elongation. To investigate the underlying mechanism of this antagonistic APA regulation, we first obtained the full catalogue of 3'UTR-APA changes affected by SRSF3 and SRSF7. Our MACE-Seq dataset did not have enough sequencing depth for a differential PAS usage determination, therefore we decided to analyze RNA-Seq data with the DaPARS algorithm (Masamha et al., 2014; Xia et al., 2014) to identify APA events that change after depletion of SRSF3 and SRSF7, respectively. The advantage of DaPARS is that it can predict pPAS *de-novo* and does not rely on a previously defined and static dataset or transcript annotations. In comparison to the 2016 dataset, we improved the knockdown (KD) of SRSF3 and SRSF7 and used better downstream analysis tools more suitable for 3'UTR-APA analyses. The DaPARS analysis supported our previous findings, as *Srsf3* KD increased pPAS usage (shortening of 3'UTRs), while *Srsf7* KD decreased pPAS usage (extension of 3'UTRs). Depletion of SRSF3 resulted in many more targets compared to depletion of SRSF7 (686 vs. 138 events). This might be due to certain redundancies between SR protein family members. For example, SRSF1 and SRSF6 have also been suggested to enhance pPAS usage in previous studies with viral RNAs (McNally & McNally, 1996; Maciolek & McNally, 2007; Hudson et al., 2016). Both proteins preferentially bind to purine rich motifs like SRSF7 (Müller-McNicoll et al., 2016), and thus could compensate for the loss of SRSF7. Especially SRSF6 might be a potent substitute for SRSF7 as the knockdown of *Srsf6* KD also showed decreased pPAS usage, but only in 50 transcripts, as shown in our study from 2016. To clarify this and potentially increase the number of SRSF7 targets a co-depletion of SRSF7 and SRSF6/SRSF1 might be an option, simultaneously elucidating further details of the redundancy. Alternatively, a more rapid depletion method, like the TRIM21-system (Clift et al., 2017),

might be useful to identify actual SRSF7-regulated 3'UTR-APA targets before SRSF6 or SRSF1 can step in and revert the effect that the loss of SRSF7 has on targeted 3'UTRs.

Comparing the DaPARS-identified events we found a decent overlap of SRSF7 and SRSF3 targets (55/138 hits). However, most dual-targets (23 hits) were shortened upon depletion of both SR proteins and only 15 targets were regulated antagonistically. This suggests that both SR proteins share a subset of targets, but the majority of targets is specific for each protein. Analysis with MISO (Katz et al., 2010) validated the DaPARS results despite its limitations for 3'UTR analysis. Comparison of individual and shared targets showed that around 50% of the SRSF7 targets (47/89) overlapped with SRSF3 targets, even though much fewer SRSF3-sensitive events were identified by MISO compared to DaPARS. This is due to the very limited number of known APA events available with MISO (1187 tandem UTR events). Nevertheless, these two analyses confirmed our previous findings from 2016 with a new dataset and motivated us to further analyze the underlying mechanism on how SRSF3 and SRSF7 might regulate 3'UTR-APA.

Transcripts that were affected in their 3'UTR length by depletion of SRSF3 were enriched in the categories: 'nucleotide binding', 'RNA processing', 'ubiquitine-dependent protein catabolic process', 'protein localization' and 'cell cycle'. Enrichment of the first two categories suggests that SRSF3 regulates the activity of other RNA binding and RNA processing genes via their 3'UTR length, expanding its impact on fine-tuning of gene expression. Furthermore, 'protein localization' represents a biological process, which has increasingly been linked to APA in recent years (Martin & Ephrussi, 2009; Berkovits & Mayr, 2015; Mayr, 2018). Finally the connection towards cell cycle regulation fits well with the recent report that SRSF3 contributes to cellular senescence (Shen et al., 2019). In that study depletion of SRSF3 resulted in expression of senescence-associated phenotypes. Interestingly, the authors speculated that the underlying mechanism might be connected to increased pPAS usage upon depletion of SRSF3. These findings underline the potential importance of SRSF3 in RNA processing besides its function as a potent splicing enhancer. Unfortunately, the number of targets identified after depletion of SRSF7 was too low to perform a conclusive GO-enrichment analysis.

We used these datasets and analyses to identify candidate transcripts for validation and mechanistic experiments to explore the mechanism(s) of SRSF3 and SRSF7-mediated 3'UTR-APA. We picked five genes with two annotated PASs within their 3'UTR, which predicted to separate well from another on agarose gels (*Ddx21, Anp32e, Rab11a, Hspa4* and *Pphln1*) derived from the set of antagonistically regulated transcripts and validated them by 3'RACE-PCR. This simple, specific, cost-effective and semi-quantitative method allows to determine and compare the ratios of transcript isoforms with different 3'UTR length. Polyadenylated transcripts are reverse transcribed without bias or gene-specificity by an anchored oligod(T)-primer, ensuring that reverse transcription starts at the

beginning of the poly(A)-tail. The oligod(T)-primer contains in addition a binding platform for the reverse primer in the following PCR amplification. The forward primers must be designed to be highly specific for the gene of interest. The drawbacks of this method are related to the final PCR amplification and the separation and visualization on agarose gels. Various cycle numbers have to be tested initially to ensure that amplification is linear and has not entered the stationary phase. Furthermore, it must be acknowledged that PCR is biased towards amplification of short transcripts over long transcripts, which is particularly of interest in this analysis setup. Finally, detection and separation limits of the agarose gels need to be evaluated to fit the expected transcript sizes. Using this method, we confirmed that depletion of SRSF3 increased usage of the dPAS while depletion of SRSF7 increase pPAS usage. The effect of depletion of SRSF7 was much weaker than depletion of SRSF3. Four out of the five tested genes had a strong preference for dPAS usage in the control, while *Hspa4* showed the opposite, suggesting that this gene is highly active in P19 cells, while the remaining four genes are more tightly regulated post-transcriptionally.

These results demonstrated that 3'RACE-PCRs are a well-suited method to analyze and validate transcripts with various 3'UTR lengths especially for abundant transcripts. Alternatively, Northern blots are widely used to detect and quantify APA-derived transcript isoforms. While detection limits and quantification are favored in Northern blotting, the method itself is more labor intensive and complicated, hence also error prone and needs to be optimized carefully.

Overall, we provide evidence that SRSF3 and SRSF7 regulate 3'UTR-APA in opposite directions and rather target distinct transcripts, but also operate antagonistically on a small subset of targets that we validated by 3'RACE-PCR. SRSF3 has a stronger effect on 3'UTR-APA regulation than SRSF7, which might be due to a partial compensation by other SR proteins with similar binding specificity, e.g., SRSF6 or SRSF1.

6.3 SRSF3 and SRSF7 preferentially regulate pPASs and may compete for binding

Cleavage and polyadenylation (CPA) are facilitated and orchestrated by a core of four multimeric cleavage factors which bind in close proximity up- and downstream of the PAS recognizing specific up- and downstream sequence elements (Proudfoot, 1989; Birse et al., 1998; Yonaha & Proudfoot, 2000; Shi et al., 2009). In consequence, aberrant expression of CPA factors affects the outcome of APA as summarized in **Table 1** and by Li et al., 2015. For example, depletion of the CFIm subunits CPSF5 or CPSF6 results in 3' UTR shortening (Gruber et al., 2012; Martin et al., 2012; Li et al., 2015). Depletion of CSTF subunits also results in 3'UTR elongation, while FIP1 overexpression promotes the expression of short 3'UTRs (Yao et al., 2013; Li et al., 2015). Furthermore, CPA-independent RBPs, such as NOVA2 and FUS, have been reported to affect APA specifically in a position-dependent manner (Licatalosi et al., 2008; Masuda et al., 2015).

To understand the potential mechanism of SRSF3 and SRSF7-mediated APA regulation we determined the binding sites of both SR proteins around PASs using published iCLIP datasets (Müller-McNicoll et al., 2016). We found that in genes with a single PAS binding of both, SRSF3 and SRSF7 was strongly enriched around -60 nt upstream of the PAS. In genes with several PASs we identified enrichment of SRSF3 and SRSF7 binding within a similar window upstream of pPAS and dPAS, but with a clear preference for the pPAS. Interestingly, SRSF7 pPASs binding clearly surpasses binding of SRSF3, while dPASs and sPAS binding was similar for both proteins. The positioning of both SR proteins is particularly compelling as their binding window overlaps well with the binding positions of key CPA factors identified by PAR-CLIP (Martin et al., 2012). In this publication, the authors showed that CPSF-subunits were mainly enriched in two windows between -75 nt to -60 nt and -30 nt to -25 nt upstream of PAS. In between these two windows subunits of CFIm were enriched at around -50 nt upstream of the PAS. Hence, we found that SRSF3 and SRSF7 bind in distance to PAS that suggests a direct and physical involvement in the regulation of APA. Furthermore, we provided evidence that pPASs might be preferentially regulated by both SR proteins. Considering the effects of SRSF3 and SRSF7 depletions on the length of targeted 3'UTRs, our data suggests that SRSF3 inactivates bound pPASs, while SRSF7 activates pPAS usage upon binding. Moreover, SRSF3 and SRSF7 binding is enriched upstream of pPASs in SRSF3-regulated genes, when compared to the same number of random unaffected genes. Interestingly, at SRSF3-sensitive pPAS SRSF7 binding was even stronger than SRSF3 binding, suggesting some degree of binding competition to activate or inactivate the pPAS, respectively. Unfortunately, the same analysis could not be conducted for SRSF7-regulated targets as their number was too low.

Taken together, iCLIP revealed enriched binding of both proteins at the pPASs, in general and in SRSF3-regulated genes, suggesting that pPASs are the hotspot of SRSF3 and SRSF7-mediated 3'UTR-APA regulation. Furthermore, as both SR proteins bind in a window upstream of the pPASs that is typically

occupied by other core CPA factors, our data suggest that SRSF3 might inactivate/block pPAS usage, while SRSF7 might activate/increase pPASs usage. Finally, in a subset of targets that are antagonistically regulated (17 events) by SRSF3 and SRSF7, both SR proteins seem to compete for binding upstream of the pPAS to suppress or engage CPA. This subset is most likely underestimated in this study as the effect upon depletion of SRSF7 might be overseen due to the compensation by SRSF1 and SRSF6.

6.4 SRSF3 and SRSF7 regulate 3' UTR-APA independent of splicing, but in a concentration-dependent manner

SRSF3 and SRSF7 are members of the conserved family of SR proteins, which are essential splicing enhancers (Zahler et al., 1992; Zahler et al., 1993) known to bind to specific ESEs (Zhu et al., 2001; Conti et al., 2013). Besides their function in the regulation of alternative splicing, SR proteins were also shown to be involved in post-splicing events, such as APA and nuclear export (Müller-McNicoll et al., 2016; Botti et al., 2017). For APA, SRSF3 was previously described to regulate CDS-APA of the human calcitonin/calcitonin gene-related peptide gene by enhancing alternative terminal exon definition (Lou et al., 1998), thereby connecting AS and APA. But in terms of 3'UTR-APA it remained unclear whether the modulation of pPAS usage by SRSF3 and SRSF7 was connected to or dependent on splicing.

To address this question, we performed experiments with intron-less reporter genes. We selected three 3'UTRs from the previously validated set of target genes: *Ddx21*, *Anp32e* and *Rab11a*, and cloned the full-length 3'UTRs behind *mCherry* and firefly *luciferase*, replacing the present very strong viral SV40 PAS. Transient transfection followed by 3'RACE revealed the presence of 3'UTR-APA isoforms similar in size and number to endogenous APA isoforms. This suggested that in general, not considering the engagement of SR proteins, APA functions mostly independent of splicing. However, the ratios between long and short 3'UTR-isoforms were slightly shifted in reporter transcripts compared to the endogenous transcripts. This suggests that in addition to the 3'UTR sequence-composition the presence of 5'UTRs and coding sequences also influence PAS usage. Importantly, depletion of either of the two SR proteins had the same effect on pPAS usage in reporter transcripts and endogenous transcripts. Depletion of SRSF3 increased the levels of reporter transcripts with short 3'UTRs and depletion of SRSF7 increased reporter transcripts with long 3'UTRs, independent of the coding sequence of *mCherry* or *luciferase*. Hence, also the modulation of 3'UTR-APA by SRSF3 and SRSF7 is independent of splicing or introns. SR proteins are normally dephosphorylated during splicing, which allows splicing surveillance and the recruitment of the nuclear export factor NXF1 (Mermoud et al., 1994; Misteli et al., 1998; Shi et al., 2006; Ma et al., 2010; Zhou & Fu, 2013; Müller-McNicoll et al.,

2016; Wegener & Müller-McNicoll, 2019). What does the splicing-independent 3'UTR APA mean for the phosphorylation cycle of SR proteins and their connection to RNA surveillance and nuclear export of correctly spliced and matured RNAs? One possibility is that SR proteins undergo a similar phosphorylation and dephosphorylation cycle during CPA as during splicing. Both are processes that occur in a large complex and need to be orchestrated precisely, starting from the binding to the target RNA sequence, recruitment of the remaining CPA machinery, catalysis followed by disassembly of the CPA machinery, and the initiation of nuclear export. This hypothesis is supported by the fact that the phosphatase PP1, which dephosphorylates SR proteins during splicing, was also found in purified 3'end processing complexes (Shi et al., 2009; Aubol et al., 2017). Depletion of PP1 and the related phosphatase PP2A did not abrogate cleavage of 3'ends but interfered with the synthesis of the poly(A)-tail highlighting the importance of de-phosphorylation at the end of the CPA process, similar to splicing. Therefore, it might be interesting to investigate the phosphorylation of SR proteins during 3'UTR-APA in more detail in future experiments to further compare the relationships between the presumably related processes of AS and APA.

To test the concentration-dependency of APA regulation by SRSF3 and SRSF7, we co-transfected the *Luc-Ddx21* reporter plasmids with increasing amounts of a vector expressing GFP-tagged SRSF3 or SRSF7. Increasing amounts of SRSF7 strongly increased pPAS usage resulting in increased expression of *Ddx21* isoforms with short 3'UTRs, indicating that 3'UTR-APA is regulated by SRSF7 in a concentration-dependent manner. Increased amounts of SRSF7 might allow SRSF7 to outcompete other RBPs, such as SRSF3, for binding upstream of pPAS, and activate the respective site. The phenomenon of strongly increased pPAS usage observed with overexpression of SRSF7 again suggests that SRSF1 or SRSF6 might substitute for SRSF7 when its expression is decreased, reducing the observable effect on 3'UTR-APA, but they cannot outcompete overexpressed SRSF7.

In comparison, overexpression of SRSF3 had no effect on alternative polyadenylation of the *Ddx21* reporter. This was surprisingly, as we had seen strong changes on pPAS usage with decreasing levels of SRSF3. One explanation might be that overexpression of SRSF3 was much less efficient, compared to overexpression of SRSF7. SRSF3 showed a strong autoregulation. Increasing the expression of GFP-tagged SRSF3 resulted in a reduction of endogenous SRSF3. SRSF3 maintains a constant protein expression through alternative splicing and inclusion of a poison cassette exon (PCE), followed by nonsense-mediated decay (NMD) of this isoform (Jumaa & Nielsen, 2000). This mechanism has been suggested for the auto-regulation of several other SR proteins as well (Lareau et al., 2007; Risso et al., 2012; Müller-McNicoll et al., 2019). Interestingly, increasing amounts of GFP-tagged SRSF7 did not affect the expression of endogenous SRSF7 as strongly as shown for SRSF3, although SRSF7 also was suggested to be affected by PCE-mediated autoregulation (Lareau et al., 2007; Pervouchine et al.,

2019). The non-apparent autoregulation of SRSF7 might be related to the overall low detection-level of endogenous SRSF7. In general, the mediocre or transient over-expression of SRSF7 was shown to lead to inclusion of a PCE and subsequent reduction of SRSF7 levels (Könige et al., 2020). Extended overexpression of SRSF7 results in the protection of bicistronic *SRSF7-PCE* transcripts from decay leading to the expression of two partial proteins derived from these Split-ORFs. The N-terminal polypeptide half (SRSF7_RRM) subsequently outcompetes full-length SRSF7 preventing inclusion of the PCE. Instead SRSF7_RRM promotes retention of the flanking introns and the formation of nuclear bodies at the SRSF7 transcription site.

Previously, it was shown that SR proteins compete with each other for binding sites to modulate splicing (Pandit et al., 2013; Zhou & Fu, 2013). Here we suggested that increased expression of SRSF7 might favor pPAS activation by more efficient competition for binding sites upstream of the pPAS. To test this, we modified the intron-less *Ddx21* reporter constructs by mutating all potential *bona-fide* SRSF3 binding sites upstream of the pPAS into SRSF7 motifs and *vice versa*. Converting all SRSF3 binding motifs into SRSF7 binding motifs increased pPAS usage. Shifting binding towards SRSF3 by converting all SRSF7 binding motifs into SRSF3 binding motifs slightly decreased pPAS usage. This effect became more evident in follow-up experiments using the *mCherry*-constructs (Schwich et al., in-press). Furthermore, insertion of a UGUA motif upstream of the CSE, which endogenously is depleted of UGUA motifs, strongly increased pPAS usage. In addition, all SRSF3 and SRSF7 binding motifs were removed in that construct. Interestingly, this single UGUA motif also strongly increased pPAS usage. Taken together, this suggests that SRSF3 and SRSF7 modulate pPAS usage in opposite directions in the absence of UGUA at mediocre CSEs by direct competition in a concentration- and binding-dependent manner.

Proximal pPAS often harbor PAS-motifs that deviate from the *consensus* poly(A) signal motif ‘AAUAAA’, which decreases their usage (Beaudoing et al., 2000; Gruber et al., 2016). In line with this, the pPAS-associated sequence motif of *Ddx21* was ‘AGUAAA’ representing a weaker poly(A) signal motif. We mutated the *Ddx21* pPAS poly(A) signal motif to ‘AAUAAA’ to strengthen the pPAS and to ‘AGUAAG’ to weaken it. Surprisingly, these single point-mutations had a very pronounced effect on pPAS usage. Strengthening the poly(A) signal motif resulted in the expression of only the short transcript isoform, while weakening it totally abolished CPA a the pPAS leading to the exclusive expression of the long transcript isoform. These results support the common model where pPAS motifs tend to deviate from the ‘AAUAAA’ motif to enable a fine-tuned regulation of transcript isoform expression. They also suggest that weaker pPAS can be enhanced, activated, or repressed by auxiliary *trans*-acting factors like SRSF3 and SRSF7. In contrast, the dPAS functions as the default cleavage site displaying a stronger poly(A) signal motif.

In conclusion, the SRSF3 and SRSF7 mediated-regulation of 3'UTR-APA is independent of prior splicing reactions but dependent on their protein concentration and binding strength. Furthermore, we show that activation of the pPAS is prone to be regulated by direct competition between SRSF3 and SRSF7 most likely upstream of intermediate CSEs lacking UGUA motifs. This suggests that the regulation of 3'UTR-APA is an additional separate function of SR proteins during maturation of mRNAs. Furthermore, we demonstrate the relevance of poly(A) signal motif and minor deviations from its *bona-fide* motif enabling the regulation of 3'UTR-APA facilitated by pPASs usage, potentially depending on auxiliary proteins.

6.5 SRSF7 interacts with CFIm and FIP1 factors independent of RNA

Various studies already elevated distinct crosstalk and interplay between different major mRNA-related processes, such as transcription, capping, (alternative) splicing and 3'end processing (reviewed in Proudfoot et al., 2002). Certainly, with the evolution of exon-based splicing the connection between splicing and cleavage factors gained importance to successfully define the terminal exon and release the nascent transcript by cleavage and polyadenylation (Niwa et al., 1990; Berget, 1995). To identify interactions between SRSF3 and SRSF7 and CPA factors in murine P19 cells we used quantitative mass-spectrometry on purified SRSF3-GFP-containing RNPs and identified several cleavage factors enriching with SRSF3, such as CPSF5, CPSF3, WDR33, FIP1 and CPSF2. These findings are well in line with previous proteomic-studies investigating the compositions of the spliceosome and the 3'end processing machinery, which also identified cleavage factors being present in the spliceosome and splicing factors being present within the 3'end processing complexes, respectively (Zhou et al., 2002; Rappsilber et al., 2002; Shi et al., 2009). Evaluating these potential interactions between SRSF3 and other proteins it must be considered that the samples were prepared without RNase treatment. Hence, we cannot distinguish between indirect (RNA-dependent) and direct (RNA-independent) protein-protein interactions. Unfortunately, no such interactome analysis was available for SRSF7 in P19 cells to further dissect similarities and differences towards the SRSF3 interactome on a proteome wide scale.

We scanned all major subunits of the CPA factors for RS and RS-like domains, since SR proteins interact with other proteins mostly via their RS-domains (Wu & Maniatis, 1993; Kohtz et al., 1994; Graveley, 2000; Long & Caceres, 2009). Surprisingly, only two CPA-factors contain such domains: the CFIm-subunit CPSF6 and the CPSF-subunit FIP1. It has been reported earlier by Zhu et al., 2018 that the RS/RS-like domains of CPSF6 and FIP1 are capable to mediate protein-protein interactions, similar to RS proteins and that CPSF6 recruits FIP1 to activate dPASs. Although FIP1 was only mildly enriched in the SRSF3-interactome and CPSF6 was not found, we focused on those two CPA factors since the presence of RS or RS-like domains should enable protein-protein interactions with SRSF3 and SRSF7.

Generating P19 cells lines stably expressing GFP-tagged CFIm subunits CPSF5 and CPSF6, and FIP1 for Co-IPs, we found that expression of CPSF5-GFP resulted in a strong decrease in the levels of endogenous CPSF5, indicating an autoregulatory mechanism of CPSF5. Autoregulation of CPSF5 had been suggested once when the function of *Nudt21* (CPSF5) in the regulation of APA during spermatogenesis was studied (Sartini et al., 2008). Interestingly, auto-regulation was not observed for CPSF6 and FIP1. Although the levels of endogenous FIP1 decreased slightly, both transgenic proteins were over-expressed, despite being expressed from integrated BACs, which should in principle enable (post)-transcriptional regulation and result in expression at nearly endogenous levels. Yet, depending on the genomic location and number of copies of the BAC integrating to the host's genome, the expression of the transgenic isoforms may increase. In addition, the overexpression of FIP1 gave rise to a smaller FIP1 isoform, which might be derived from an alternatively spliced *Fip1l1* isoform. As this isoform was detectable with the antibody that recognizes the N-terminal part of FIP1, alternative splicing seem to affect the C-terminal region of the protein, which would match with the annotated FIP1 isoform 4 (identifier: Q6UN15-4) in the uniport database (The UniProt Consortium, 2019). However, we did not observe any obvious phenotype or negative physiological effects on the cells.

Analogues to localization studies by Dettwiler et al., 2004 and Cardinale et al., 2007, we found that CPSF5- and CPSF6-GFP dispersed in the nucleoplasm. In difference to these two studies, we did not find CPSF6 in paraspeckles, as P19 cells are pluripotent and do not contain paraspeckles (unpublished observation). FIP1 was also dispersed in the nucleoplasm suggesting that the slight overexpression of CPSF6 and FIP1 did not interfere with their subcellular localization.

Surprisingly, GFP-tagged CPSF6 was not efficiently pulled down by the α-GFP antibody used in Co-IPs, suggesting that the C-terminal GFP-tag might interfere with antibody-binding. Since CPSF6 and CPSF5 form the very stable ternary complex of CFIm (Kim et al., 2010; Yang et al., 2011a), we could use GFP-tagged CPSF5 in the CoIP-experiments. Indeed, CPSF6 was co-purified in stoichiometric ratios in CPSF5-GFP pulldowns after RNase treatment. In addition to the main isoform of CPSF6 (CFIm-68), we also co-purified the slightly larger alternative spliced isoform CFIm-72. Interestingly, co-purification of CFIm-68 and CFIm-72 increased after RNA depletion suggesting that RNA-bound CFIm might contain rather CPSF5/CPSF7 than CPSF5/CPSF6, but this remains to be tested. All investigated cleavage and splicing proteins were co-purified with one another in the various pull-down combinations in the presence of RNA, suggesting that all these factors bind to common target RNAs. However, after RNA was digested using RNase, the interaction of SRSF3 with CPSF5, CPSF6 and FIP1 vanished, while interactions with SRSF7 remained at equal levels. Our finding that SRSF3 does not interact RNA-independently with CFIm subunits is in contrast to a study published by Dettwiler et al., 2004. In this study, the authors performed GST-pulldown assays using the C-terminal RS-like domain of CPSF6 as

the bait and SR proteins SRSF3, SRSF7 and SRSF10 being expressed from cDNA transfections. All three SR proteins were identified to interact with the bait, suggesting that SRSF3 could in principle interact with CPSF6 *in vitro*, but it does not do so in P19 cells, as our Co-IP data only support the RNA-independent interaction between CFIm and SRSF7 but not with SRSF3.This experiment also included GST-labelled CPSF5 as bait and no interaction with SRSF3, SRSF7 or SRSF10 was detected, confirming that the interaction with CFIm occurs via CPSF6. Besides the CFIm, SRSF7 also interacts RNA-independently with FIP1, which is recruited to dPASs via CFIm (Zhu et al., 2018). In contrast, we did not detect RNA-independent interactions between CPSF5 (CFIm) and FIP1 in P19 cells. One reason for this is the transient interaction of the CPA complex and the fact that P19 cells preferentially use pPAS, which limits the chance of CFIm to bind to the dPAS and recruit FIP1. Finally, we found that FIP1 preferentially co-purifies de-phosphorylated SRSF7, while CPSF5 rather interacts with phosphorylated SRSF7. Unfortunately, SRSF1 and SRSF2 run at the same size as SRSF7 and the mAb1O4 antibody, which detects all phosphorylated SR-proteins cannot discriminate them. Therefore, this finding must be treated with caution.

In conclusion, we determined that SRSF7 interacts independent of RNA with subunits CPSF5 and CPSF6 of the CFIm and with FIP1 a subunit of the CPSF complex (**Figure 61**A). These interactions most likely are established between the dephosphorylated SRSF7 proteins as analyzed using a phosphorylation-sensitive antibody. In contrast, SRSF3 was found to not interact RNA-independently with these cleavage factors. Nevertheless, we found that both SR proteins and CFIm and FIP1 bind to the same RNA targets. Together, this suggests that SRSF3 might inhibit pPAS usage by binding in close proximity of the pPAS and blocking recruitment of CPA factors to USEs, while SRSF7 actively support recruitment of CPA factors towards the USEs increasing pPAS usage (**Figure 61**B).

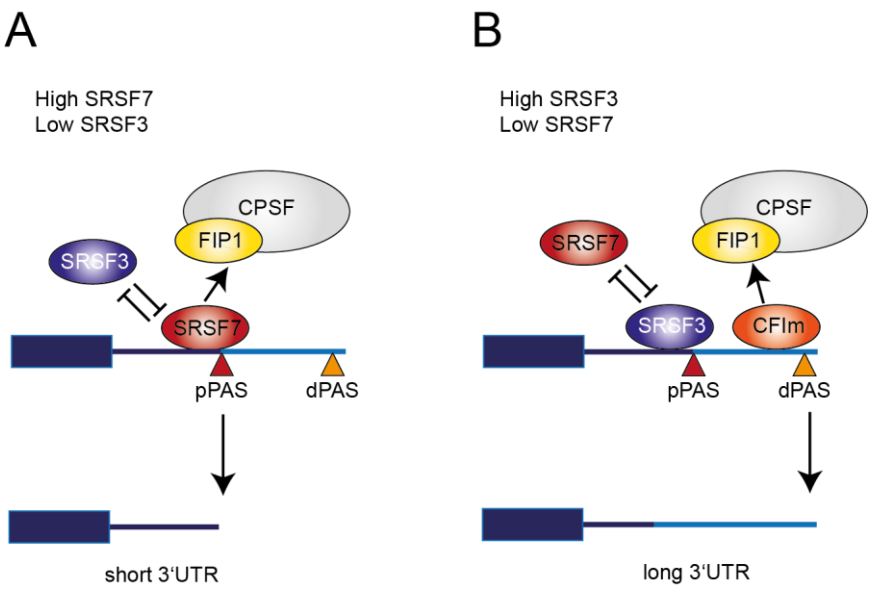


Figure 61: Model of the competitive mechanisms how SRSF3 and SRSF7 might regulate 3'UTR-APA at the pPAS. A) *High levels of SRSF7 enable successful competition with SRSF3 binding upstream of the pPASs followed by FIP1 recruitment and cleavage at the pPAS leading to the expression of short 3' UTRs.* ***B)*** *High levels of SRSF3 outcompete SRSF7 for binding at pPASs and block SRSF7 mediated recruitment of FIP1. Furthermore, high level of SRSF3 ensure high level of CFIm enhancing CPA-machinery recruitment towards the dPAS and cleavage leading to the expression of transcripts with long 3' UTRs.*

6.5.1 SRSF7 and FIP1 are co-purified independent of CFIm and RNA

It has been reported before that FIP1 is recruited to dPASs by CFIm via the RS and RS-like domains present in both cleavage factors (Zhu et al., 2018). To test whether the interaction between SRSF7 and FIP1 is mediated by CFIm we performed Co-IPs after depletion of both CFIm subunits. Depletion of CPSF6 was very effective and decreased CPSF5 levels, suggesting that the binding of both proteins stabilize each other. Depletion of CFIm did not affect the RNA-independent interaction between SRSF7 and FIP1 suggesting that these two proteins interact independent of CFIm. This indicates that SRSF7 might recruit FIP1 to the pPAS similar to the recruitment of FIP1 to dPAS by CFIm.

KD of *Nudt21* was less effective than *Cpsf6* KD, which might be due to the esiRNAs used. Depletion of CPSF5 did also not affect RNA-independent co-purification of SRSF7 with FIP1, supporting the notion that SRSF7 and FIP1 interact independent of CFIm and further suggest that SRSF7 might mirror the functions of CFIm in CPA-machinery assemblage at pPASs.

It was shown that the recruitment of FIP1 via CFIm to dPASs is mediated via their RS/RS-like domains (Zhu et al., 2018). Similarly, the splicing factor U2AF65 recruits CFIm in the alternative configuration (CPSF5 and CPSF7) via its RS-domain (Millevoi et al., 2006). SR proteins also mediate protein-protein interactions via their RS-domains (Wu & Maniatis, 1993; Kohtz et al., 1994; Graveley, 2000; Long & Caceres, 2009; Zhu et al., 2018). To further investigate the SRSF7-FIP1 interaction we fused the RS domain of SRSF7 with the unrelated TetR-protein. As our Co-IP experiments suggested

that the interaction might depend on the phosphorylation level of SRSF7, we included phosphomimetic variants representing hyper-phosphorylated (RD7) and hypo-phosphorylated (RA7) SRSF7-RS-domains. Co-IPs with these TetR-constructs confirmed this suggestion because we identified an RNA-independent interaction between GFP-FIP1 and TetR-RA7 and TetR-RS7, while TetR-RD7 was not co-purified after RNA digestion. Moreover, the interaction seems to occur preferentially with the hypo-phosphorylated SRSF7-RA variant, compared to the endogenous SRSF7-RS-domain, which might be only partially phosphorylated. This finding deviates from the phosphorylation-cycle of SR proteins during splicing. Here partially phosphorylated SR proteins are stored in nuclear speckles (Cáceres et al., 1997; Galganski et al., 2017) and upon hyper-phosphorylation by CLK1 they leave nuclear speckles, bind to their RNA targets and recruit the spliceosome via RS-domain-mediated interactions. SR proteins are de-phosphorylated during the splicing reaction, prior to the recruitment of export factors, such as NXF1, that mediate nuclear export of matured transcripts (Mermoud et al., 1994; Misteli et al., 1998; Shi et al., 2006; Ma et al., 2010; Müller-McNicoll et al., 2016). In contrast to splicing, the interaction with FIP1 seems to be rather established in a partially de-phosphorylated state of SRSF7 and is stabilized by hypo-phosphorylation. We cannot rule out that dephosphorylation and recruitment of CPA factors take place during the CPA reaction, as the phosphatase PP1A was found to be present in the purified holo-CPA machinery (Shi et al., 2009). Strikingly, the SRSF7-RS-domain alone is not sufficient to establish an interaction with CFIm, represented by CPSF5-GFP, suggesting that additional regions of SRSF7 are necessary to mediate the RNA-independent interaction with CFIm. Taken together this suggests that in general SRSF7 can bind to RNA and proteins simultaneously via its RRM and Zn-knuckle on the one hand and the RS-domain on the other hand. Yet, this does not seem to be the case while trying to establish an interaction between SRSF7 and CFIm. Here the simultaneous interaction between SRSF7, CFIm and RNA might not be possible as more than the SRSF7-RS-domain is required to facilitate the interaction towards CFIm. Yet it needs to be investigated which additional domains, either the Zn-knuckle, the RRM or both domains might be necessary to recruit CFIm. This contrasts with the interaction between SRSF7 and FIP1, which is established via the RS-domain itself.

To be able to test the phosphorylation-dependency of CFIm-SRSF7 interaction, we fused the full coding sequence of SRSF7 and the respective RA/RD-phosphomimetics to mCherry. Localization analyses in GFP-FIP1 P19 cells revealed that the SRSF7-RS and the hyper-phosphorylated mimic SRSF7-RD were dispersed in the nucleoplasm and co-localized well with GFP-tagged FIP1. Surprisingly, mCherry-SRSF7[RA] was detected solely within the nucleolus of P19 cells. The nucleolus is a sub-compartment of the nucleus, which is the main site of ribosomal RNA biogenesis, maturation and assembly of ribosomal subunits (Scheer & Hock, 1999; Raska et al., 2006). Short stretches of basic residues (6 to 10 aa, enriched in arginine and lysine) were defined as NLSs, but several of these sequences also enable the nucleolar targeting of the respective protein (Dang & Lee, 1989). It remained

unclear what distinguishes an NLS from a nucleolar localization sequence. A recent study reported that an enrichment of arginine residues (≥6 Rs) is sufficient for nucleolar localization while an enrichment in lysins did not result in significant nucleolar localization (Martin et al., 2015). It was suggested that a certain concentration of positive charges represents an isoelectric threshold for nucleolus enrichment, which compensates the negative charge of the concentrated rRNA. This theory might explain the high concentration of mCherry-SRSF7[RA] in the nucleolus as the high number of positively charged arginine residues (41 Rs) are neither compensated by phosphorylated serines nor by negatively charged aspartate residues, enabling to surpass the threshold for nucleolar localization. The threshold theory is further supported by the fact that SRS3[RA]-mCherry partially localized to the nucleoplasm showing less enrichment in the nucleolus. This might be due to the much shorter RS domain (51 aa, 19 Rs) of SRSF3, which in comparison to SRSF7 results in a lower isoelectric point. Due to the observed aberrant localization of SRSF7-RA we decided to exclude this phosphomimetic in subsequent experiments.

Co-IP experiments using CPSF5-GFP as the bait showed that full-length SRSF7-RS-mCherry co-purified with CPSF5 independent of RNA, supporting our previous suggestion that additional domains beside the RS domain of SRSF7 are necessary to mediate the interaction with CPSF5. Interestingly, mCherry-SRSF7[RD] showed an increased interaction with CPSF5, suggesting that this interaction is mediated between the hyper-phosphorylated SR protein and CFIm. This would argue that this interaction is established in close context to newly synthesized mRNA as hyper-phosphorylated SR proteins are recruited to nascent transcript for subsequent initiation of (alternative) splicing (Zhou & Fu, 2013). As according to our data SRSF7 cannot establish simultaneous interactions between CFIm and RNA we speculate that this might be a mechanism how CFIm could abrogate recruitment of SRSF7 towards AS sites and pPASs and thereby suppresses pPAS usage. Yet, additional experiments would be necessary to support this hypothesis. In contrast, Co-IPs using GFP-FIP1 as bait fortified our conclusion that the interaction between SRSF7 and FIP1 is rather mediated by a semi- to hypo-phosphorylated state, as the interaction decreased between FIP1 and mCherry-SRSF7[RD]. This suggests that the interaction is either established while SRSF7 is stored in a semi-phosphorylated state in nuclear speckles or alternatively during- or post-splicing, when SRSF7 is de-phosphorylated. These differences in the phosphorylation states of SRSF7 necessary to establish interactions towards the CFIm and FIP1 are unexpected and might indicate a spatial and temporal separation of the specific interactions between SRSF7 and CFIm and FIP1. In addition, the different phosphorylation-dependent interactions are unexpected as both CPA factors contain an RS/RS-like domain. However, a comparison of the residual composition of these two domains (**Figure 33**) reveals two distinct differences: i) the RS-domain of CPSF6 is shorter than the RS-like domain of FIP1 (63 aa vs. 95 aa) and ii) the RS-domain of CPSF6 contains proportional more RS di-peptides (36%) than the RS-like domain of FIP1 (12%), while FIP1 contains more RE/RD di-peptides (60% and 28% in FIP1 vs. 32% and 32% in CPSF6). This suggests

that the basic charges of these domains might have a function in discriminating the shown distinct interactions with differently phosphorylated SRSF7.

Together these experiments suggest that the RNA-independent interactions established between SRSF7 and CFIm and FIP1 depend on the phosphorylation status of SRSF7. Moreover, the interaction between SRSF7 and the CFIm require more than the RS domain of SRSF7, while this domain alone is sufficient to interact with FIP1.

6.5.2 The Zn-knuckle and hydrophobic-stretch of SRSF7 are necessary to establish RNA-independent interactions with CFIm

We raised the question what the differences between SRSF3 and SRSF7 are, especially as these two SR proteins are evolutionary very closely related to each other (Busch & Hertel, 2012), and which distinct regions of SRSF7 are necessary to mediate the RNA-independent interaction with CFIm.

SRSF7 stands out from all other SR proteins by containing a CCHC-type zinc-knuckle in between the RRM and the RS-domain. Zn-knuckles and -fingers are very common protein domains mainly characterized for nucleotide-binding. In line with this, the Zn-knuckle of SRSF7 was shown to contribute to its RNA-binding specificity (GAYGAY) (Cavaloc et al., 1999; Müller-McNicoll et al., 2016; Königs et al., 2020). In addition, CCHC-type Zn-knuckles have been associated with protein-protein interactions (Fox et al., 1998; Fox et al., 1999; MacKay & Crossley, 1998; Matthews et al., 2000). A detailed comparison of the protein sequences of SRSF3 and SRSF7 disclosed that the RRMs of both proteins are highly similar (sharing 80% of the residues) (Busch & Hertel, 2012), while their RS-domains are very different: i) the RS-domain of SRSF3 is much shorter (51 vs. 116 aa), however with a similar content of RS-dipeptides (21 % vs. 20 %); ii) the center of the SRSF7-RS-domain contains a stretch of 27 aa enriched in highly positively charged arginine and hydrophobic residues, such as alanine, leucine, isoleucine, glycine and valine (16/27 aa). Usually, such a hydrophobic region would be expected to be buried inside the core of a protein, but it is located within the floppy and unstructured RS-domain, suggesting that it might facilitate protein-protein interactions via arginine residues, which can establish salt-bridges with aromatic and charged sidechains or alternatively via the hydrophobic residues. Moreover we have shown that this region is subject to alternative splicing (Königs et al., 2020). To elucidate the contribution of these domains/regions to the RNA-independent interaction between SRSF7, CFIm and FIP1 we generated P19 cell-lines stably expressing mutant GFP-tagged SRSF7 lacking the 27aa stretch (Δ27aa) or containing an inactive CCHC-Zn-knuckle (mutZn) from a BAC background. The two SRSF7 mutants were expressed at the same levels as the native GFP-tagged SRSF7, while the auto-regulatory capacities were reduced especially in the mutZn-isoform. This evasion from autoregulation has been

attributed to the loss of sequence-specific binding of SRSF7 within its PCE and increased intronic binding at CNYC-motifs mediated via the RRM portion of the protein (Müller-McNicoll et al., 2016; Königs et al., 2020). Apart from this, the mutated SRSF7 isoforms showed a similar behavior in terms of phosphorylation states and sub-cellular localization. Co-IPs revealed that deletion of the 27aa stretch decreased the RNA-independent interaction between SRSF7 and FIP1 as well as to CPSF5, suggesting that the arginine-rich and hydrophobic-stretch within the RS-domain of SRSF7 enhances the recruitment of FIP1 and CPSF5. The interaction with CPSF6 was less affected by the 27aa deletion, suggesting that the interaction with CPSF6 is established via other domains of SRSF7. This was surprising, as the CFIm consists of CPSF5 and CPSF6 subunits in stoichiometric ratios. We would have expected to co-purify also CPSF5 if CPSF6 interacts with SRSF7 independent of the hydrophobic stretch. This could be an indication that predominantly CFIm complexes consisting of CPSF5 and CPSF7 were affected by this mutation, although this remains to be analyzed as no suitable antibody detecting CPSF7 has been available during this study. Another possibility would be that SRSF7 interferes with the composition of CFIm and replaces the RNA-binding subunit CPSF5. However, there are no reports available that such an alternative CFIm complex might exist and we were not able to detect SRSF7 in native purifications of the CFIm complex (unpublished data). Inactivation of the CCHC-Zn-knuckle also led to reduced RNA-independent interactions with CPSF5, CPSF6 and FIP1, suggesting that this Zn-knuckle has a dual RNA- and protein-binding function in SRSF7 in relation to the CFIm complex. In case of FIP1 we have shown that the RS-domain of SRSF7 is sufficient to interact with FIP1. Nevertheless, the reduced interaction upon the Zn-knuckle mutation might indicate that this interaction is mediated directly at the RNA requiring a distinct confirmation established with the contribution of the Zn-knuckle e.g., in proximity of pPASs. This notion is further supported by the fact that the Zn-knuckle mutation reduced the RNA-dependent co-purification of SRSF7 with PABPN1, CPSF5 and CPSF6, while FIP1 still co-purified well, suggesting that SRSF7 and FIP1 still bind to the same transcripts, but cannot initiate the cleavage and polyadenylation reactions as the bound transcripts might not be correctly processed, as the co-purification of PABPN1, which stabilizes and catalyzes the synthesis of the poly(A)-tail upon cleavage, was reduced indicating the absence of a poly(A)-tail. These results underline the importance of these SRSF7-specific domains enabling the sequence-specific recruitment of FIP1.

To test whether these domains were not only necessary to establish these interactions, but whether they might be sufficient, we generated SRSF3 chimaeras containing the SRSF7 Zn-knuckle, the 27aa hydrophobic-stretch or both domains together inserted at the identical relative positions compared to SRSF7. Insertion of these domains did not affect expression, auto-regulation, phosphorylation, or sub-cellular localization of SRSF3. Interestingly, the 27aa hydrophobic stretch alone visibly increased the RNA-independent interaction with CPSF5, CPSF6 and FIP1, but not the Zn-knuckle or both domains together. One reason might be that SRSF3 is the smallest SR protein in the

family and adopts a very compact conformation, which is disturbed by the insertion of a Zn-knuckle. In SRSF7 the RRM is followed by a 35aa long linker region upstream of the CCHC-Zn-knuckle and the hydrophobic stretch is incorporated in the center of a much longer RS-domain. The linker-region in SRSF3 measures only 19 aa and the RS-domain is much shorter, suggesting that the adjacent sequences around these two distinct domains of SRSF7 is important to enable protein-protein interactions (**Figure 27**, **Figure 30**, **Figure 31**). The limited space within SRSF3 might lead to steric hindrances blocking these interactions. Nevertheless, in a follow-up experiment the SRSF3 double chimaera containing the Zn-knuckle and 27aa hydrophobic stretch was capable to mimic the activation of pPAS usage in SRSF7-sensitive transcripts, similar to OE of SRSF7, while the SRSF3-27aa-chimaera showed no affect upon transient overexpression (Schwich et al., in-press). This suggests that the 27aa hydrophobic stretch is sufficient to establish the interaction towards FIP1, but additional sequence-specific binding mediated via the Zn-knuckle is necessary to enhance pPAS usage.

In conclusion, we determined that in P19 cells only SRSF7, but not SRSF3, can establish RNA-independent interactions with CFIm and FIP1 depending on the phosphorylation status of the RS-domain. This finding deviates to the previous finding from Dettwiler et al., 2004, where in vitro GST-pulldowns were performed rather using individual domains of CPSF6 than using the native protein. We identified two distinct domains within SRSF7, the CCHC-Zn-knuckle, and a 27aa hydrophobic stretch, which are crucial for SRSF7 to interact independent of RNA with CFIm and FIP1. Of these two, the Zn-knuckle-domain was necessary to interact with CPSF5, CPSF6 and FIP1, while CPSF6 can interact with SRSF7 in the absence of the 27aa hydrophobic stretch, suggesting that the interactions are mediated by a combination of the different domains. Additionally, the Zn-knuckle seems to contribute to sequence-specific binding necessary for the interaction between SRSF7 and FIP1. Furthermore, experiments with the SRSF3 chimaeras revealed that the surrounding linker and extended RS-domains of SRSF7 might be important to reproduce the Co-IP results. Yet, the double chimeric construct was functional and enhanced pPAS usage like native SRSF7. These interactions appear highly interesting to investigate further in the future to determine the molecular principles of these interactions.

6.6 SRSF3 promotes dPAS usage by controlling CFIm expression levels

Shen et al. suggested that SRSF3 negatively regulates pPAS usage by blocking the recruitment of the CPA machinery to pPASs. In line with this suggestion, we here and in the recently accepted publication Schwich et al., in-press present evidence and confirm SRSF3 binds close to the pPAS. Furthermore, we show that SRSF3 is outcompeted by SRSF7, does not interact with FIP1 or CFIm and finally that reduced levels of SRFS3 enhance usage of the pPAS. However, the competition on enhancement or reduction of pPAS usage could not be the only mechanism how SRSF3 influences

3'UTR-APA as depletion of SRSF3 affected eight times more individual 3'UTR-APA events compared to depletion of SRSF7 (631 vs. 78 events). While this discrepancy could partially lead back to SRSF1- and SRSF6-dependent compensation of the KD of the *Srsf7*, we could not exclude that several SRSF3-specific targets might be indirectly regulated. Hence, we were wondering if there are alternative mechanisms how SRSF3 might affect PAS choice. Many RNA-binding proteins, such as SR and hnRNP proteins, have been found to cross-regulate other RBPs via alternative splicing in combination with NMD (Lareau et al., 2007; Long & Caceres, 2009; Rossbach et al., 2009; Sun et al., 2010; McGlincy et al., 2010). Therefore, we analyzed differential expression of mRNAs encoding CPA factors after depletion of SRSF3 or SRSF7 using our RNA-Seq datasets. Interestingly, the expression of *Cpsf6* was markedly decreased upon knockdown of *Srsf3*, but not upon knockdown of *Srsf7*. This result was confirmed by qPCR and Western blot, suggesting that SRSF3 cross-regulates expression of the CFIm subunit CPSF6 by a specific mechanism as none of the other CPA-factors was affected. Interestingly, CPSF5 levels were also decreased, suggesting that CPSF6 might stabilize the core CFIm protein and that decreased levels of CPSF6 might disrupt assembly of the CFIm, as we did not find any evidence that depletion of SRSF3 affects *Nudt21* (Gruber et al., 2012; Kim et al., 2010). Different to examples of splicing factors auto- or cross-regulating expression of their own or related transcripts upon overexpression, expression of SRSF3 and CPSF6 are regulated in the same direction, suggesting a feed-forward loop (Müller-McNicoll et al., 2019). With respect to the potential mechanism(s) how SRSF3 affects PAS usage and 3'UTR-APA, this adds an additional layer beyond the interactions at the protein-protein level.

Analyzing the splicing pattern of *Cpsf6* upon *Srsf3* KD we identified two alternative splicing events. First, upon depletion of SRSF3 a small region/alternative exon (111 nt) of Intron5 was spliced into the mature transcript, second, the region from Intron5 until Exon7 were excluded, connecting Exon5 to Exon7 in the mature transcript. These findings were confirmed on splicing gels showing that depletion of SRSF3 resulted in the expression of several smaller transcript isoforms, not present in control conditions, which were identified as different isoforms that skip Exon6 and include either the alternative exon or the intron in between. Coherently, expression of full-length transcripts was strongly decreased. In line with the notion that SR proteins facilitate inclusion of bound exons (Shen & Mattox, 2012; Fu & Ares, 2014) we confirmed strong binding of SRSF3 to Exon6 of *Cpsf6* by iCLIP, suggesting that particularly the inclusion of this exon depends on SRSF3 expression levels. Hence, depletion of SRSF3 would lead to exclusion of Exon6 and expression of the detected shorter transcripts at the expense of full-length *Cpsf6*. *In silico* analysis of the smaller *Cpsf6*-isoforms directly connecting Exon5 to Exon7 revealed the formation of a new termination codon within Exon7 more than 50 nt upstream of the next adjacent exon-exon-junction. This positioning might turn the termination codon into a PTC, which should target these transcripts for NMD (Nagy & Maquat, 1998; Thermann et al., 1998; Zhang

et al., 1998). Alternatively, the alternatively spliced isoform could be translated into a truncated CPSF6 isoform containing 241 aa and an estimated molecular weight of 25 kDa. Yet, no such truncated isoform was detected by Western blot using the CPSF6-specific antibody which binds in between residues 1 to 50 (Exon1 and Exon2) and therefore should be capable of detecting this potential isoform (data not shown). On the other hand, the levels of the Exon6-skipped transcript isoforms did not increase after treatment with cycloheximide (CHX) to block NMD (data not shown), suggesting that those are no NMD-targets and that the decrease of the main isoform is simply related to increased expression of alternative isoforms, consequently decreasing expression of CPSF6. Yet, it remains to be verified whether these alternative *Cpsf6* isoform are true NMD-targets, e.g., by co-depletion of SRSF3 and UPF1 as an alternative to block NMD, followed by splicing gels.

Interestingly, depletion of SRSF7 caused the increased inclusion of the alternative 111 nt exon but did not affect the skipping of Exon6. Splicing gels showed an increased expression of larger transcript isoforms likely containing the alternative 111 nt exon, which would be translated into the 72 kDa CPSF6 isoform (Rüegsegger et al., 1996; Venkataraman et al., 2005; Neve et al., 2017). Inclusion of this alternative exon affects a proline-rich domain separating the N-terminal RRM and C-terminal RS-like domains of CPSF6. This domain has been speculated to enhance CFIm formation and sequence-specific binding of CFIm in accordance with the function of a proline-rich domain in the splicing enhancer U2AF (Kielkopf et al., 2001; Dettwiler et al., 2004). Yet, Dettwiler et al. did not find evidence that the proline-rich domain enhances the interaction between CPSF5 and CPSF6, although they succeeded to generate a construct containing the combination of the proline-rich and RS-domain. In addition, the proline-rich domain has been discussed as a potential nuclear retention signal (NuRS) while being compared to the first identified NuRS found in hnRNPC1 (Nakielny & Dreyfuss, 1996; Dettwiler et al., 2004). The NuRS of hnRNPC1 is enriched in prolines, pervaded with clusters of basic residues and potential phosphorylation and glycosylation sites. A basic analysis of the residual composition of the proline-rich domain of CPSF6 revealed a lack of basic residues as well as distinct residues required for phosphorylation or glycosylation, suggesting that this domain is no NuRS. The additional residues, introduced by the alternative exon, are mostly non-polar amino acids (35%; leucine, isoleucine, and proline) increasing this non-polar domain without adding substantial residues necessary for the NuRS-related function. Therefore, the function of this alternative isoform is unknown, but the increased inclusion of the 111 nt alternative exon upon depletion of SRSF7 might compensate for the loss of SRSF7 in pPAS activation assuming that this isoform has positive effects on CFIm activity. Alternatively, the 72 kDa isoform of CPSF6 might have negative effects on CPA and therefore its expression is controlled by the presence of SRSF7. These questions need to be addressed in future studies.

A recent study Di Liddo et al., 2019 identified a circular *Cpsf6* transcript during a genome-wide search in HeLa cells, which is generated by back-splicing of Exon9 to Exon1. circRNAs were known since the 1970s but have been re-discovered recently (Hansen, 2018; Wilusz, 2018; Xiao et al., 2020). Their functions vary from serving as sponges binding various miRNAs or RBPs or representing a scaffold for protein folding, protein translation and facilitating transcription and splicing (Kristensen et al., 2018). Comprehensibly, splicing factors have been implicated in the generation of circRNAs as a consequence of constitutive or alternative splicing (Ashwal-Fluss et al., 2014; Starke et al., 2015). We confirmed the expression of *circCpsf6* in P19 cells by PCR amplification of the unusual connection between the 3'end of Exon9 and the 5'end of Exon2 and Sanger sequencing. We also detected the respective linear byproduct of exon skipping (Exon1-Exon10). Interestingly, the expression of *circCpsf6* seems to depend on SRSF3, as its levels decreased after depletion of SRSF3, while expression of the linear transcript containing only Exon1 and Exon10 did not change. This suggests that SRSF3 might be involved in the generation of *circCpsf6*. Nevertheless, the identity of the circular transcript needs to be validated in detail including proof for RNase I/RNase R resistance and absence of a poly(A)-tail. While the true function of this circular transcript remains unclear, we can speculate that SRSF3-promoted expression of *circCpsf6* affects 3'UTR-APA by suppressing pPAS enhancement via SRSF7 by sponging SRSF7 molecules and preventing those from binding upstream of pPASs leading to increased expression of transcripts with long 3'UTRs (**Figure 62**A). This hypothesis is supported by iCLIP data of SRSF7 (Müller-McNicoll et al., 2016) showing binding of SRSF7 and the enriched presence of GAY-motifs across all exons of *Cpsf6* (data not shown). Depletion of SRSF3 would abrogate expression of native CPSF6 and *circCpsf6* neglecting the sponging effect, resulting in the expression transcripts with short 3'UTRs (**Figure 62**B). This could be a second indirect-regulatory mechanism how SRSF3 regulate 3'UTR-APA via *Cpsf6*. Future studies may further elucidate the function of this circular transcript.

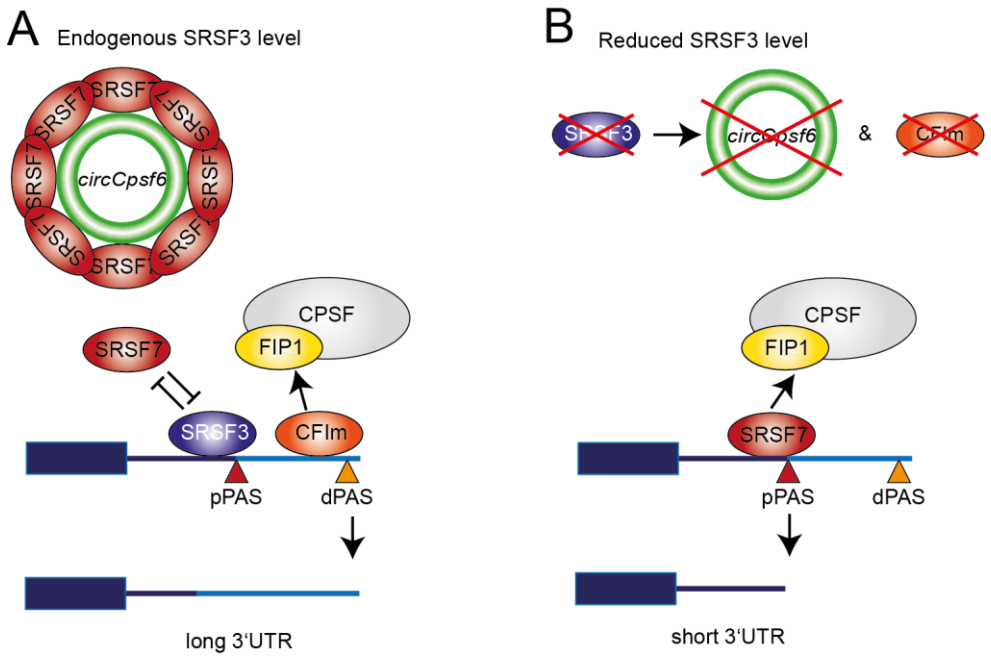


Figure 62: Model how the SRSF3-dependent expression of circCpsf6 and CFIm could affect 3'UTR-APA. A) *Expression of SRSF3 promotes expression of circCpsf6 (green circle) that might sponge SRSF7 molecules and prevent those from enhancing pPAS usage leading to the expression of transcripts with long 3'UTRs. In addition, expression of SRSF3 promotes constant expression of CFIm facilitating dPAS activation.* ***B)*** *KD of Srsf3 co-depletes circCpsf6 and CFIm allowing SRSF7 to enhance pPAS usage resulting in the expression of transcripts with short 3'UTRs.*

Depletion of CPSF6 had been shown genome-wide to shorten 3'UTRs as UGUA-containing dPAS are not activated in the absence of CFIm (Gruber et al., 2012; Martin et al., 2012; Li et al., 2015; Zhu et al., 2018). Depletion of SRSF3 also results in increased pPAS usage and expression of short-3'UTR transcripts of distinct genes. Our co-depletion of CPSF6 and CPSF5 upon *Srsf3* KD suggest that SRSF3 regulates 3'UTR-APA also indirectly by promoting the expression of the CFIm. Hence, depletion of SRSF3 and CPSF6 should affect the same transcripts. Indeed, comparison of SRSF3 and CPSF6 targets revealed a good overlap (500/686 events) with a decent overlap in transcripts that are shortened in both depletion experiments (448/686 events).

In conclusion, we determined that high levels of SRSF3 enhance expression of the transcript encoding the 68 kDa isoform of CPSF6. Decreased levels of SRSF3 result in unproductive splicing generating much shorter and potentially unstable NMD targets lacking the highly SRSF3-bound Exon6. Decreased expression of the 68 kDa isoform of CPSF6 also decreases CPSF5 levels thereby disturbing CFIm assembly and consequently dPASs activation. Comparison between SRSF3 and CPSF6 3'UTR-APA targets revealed a large overlap suggesting that SRSF3 regulates 3'UTR-APA by two independent mechanisms: i) binding to 3'UTR binding sites upstream of pPASs and blocking the recruitment of the CPA machinery, ii) promoting expression of CPSF6 and thereby stabilizing CFIm to drive CPA at the dPAS. The latter seems to be the main mechanism how SRSF3 affect 3'UTR-APA. Finally, an initiall AS analysis exposed that depletion of SRSF7 increases expression of the alternative transcript encoding

the 72 kDa isoform of CPSF6. This might be an interesting starting point to analyze this protein in more detail and determine its function in CPA and mRNA processing.

6.7 CFIm binding at SRSF3-sensitive pPASs may inhibit their usage

CFIm is known to predominately activate dPAS usage by recognizing two UGUA motifs via its CPSF5 subunits prior the recruitment of FIP1 and the remaining CPA machinery via its CPSF6 subunit (Hu et al., 2005; Venkataraman et al., 2005; Zhu et al., 2018; Gruber & Zavolan, 2019). Thus, enhanced pPAS usage upon depletion of SRSF3 might be due to reduced CPSF6 levels and therefore reduced activation of dPASs. To investigate this possibility, we compared the occurrence of UGUA motifs around pPASs and dPASs in SRSF3 and CPSF6 targets. Most UGUA motifs were enriched upstream of PASs in a distance of ~ -50 nt, fitting well with previously reported distances of -50 to -39 nt that were most efficient for CPA (Zhu et al., 2018). CFIm-specific binding motifs were increased in a window up to -300 nt upstream of dPASs confirming that the dPASs functions as the default poly(A)-site through enhanced CFIm binding and the presence of strong PAS motifs (Beaudoing et al., 2000). Interestingly, the proportion of UGUA motifs increased upstream of dPASs in CPSF6- and SRSF3-sensitive transcripts, showing the highest peak in SRSF3-sensitive targets, compared to UGUA motifs upstream of pPASs, which remained at equal levels across the different datasets. However, we observed only slightly more UGUA motifs upstream of dPASs compared to sPASs and pPASs, suggesting that CFIm binding not only enhances dPAS usage as suggested. Another interesting finding from this analysis was that pPASs showed a pronounced second peak of UGUA motif enrichment downstream of the pPAS. Interestingly, this downstream peak was massively enhanced in SRSF3-sensitive transcripts, with similar proportions of UGUA motifs on either side of these pPASs. As the assembled CFIm contains two CPSF5-subunits, which can contact RNA directly at their UGUA-motifs it has been postulated that the CFIm recognizes two UGUA motifs looping the RNA-sequence in between around the CPSF6/7 subunits (Yang et al., 2010; Yang et al., 2011a). This has also been proposed as a potential mechanism how CFIm skips upstream and potentially weaker pPASs. The surprising bipartite enrichment of UGUA motifs directly up- and downstream of SRSF3-sensitive pPASs (-40 nt; 60 nt), supports this model and could represent a new mechanism of how SRSF3-sensitive pPASs are regulated via CFIm expression.

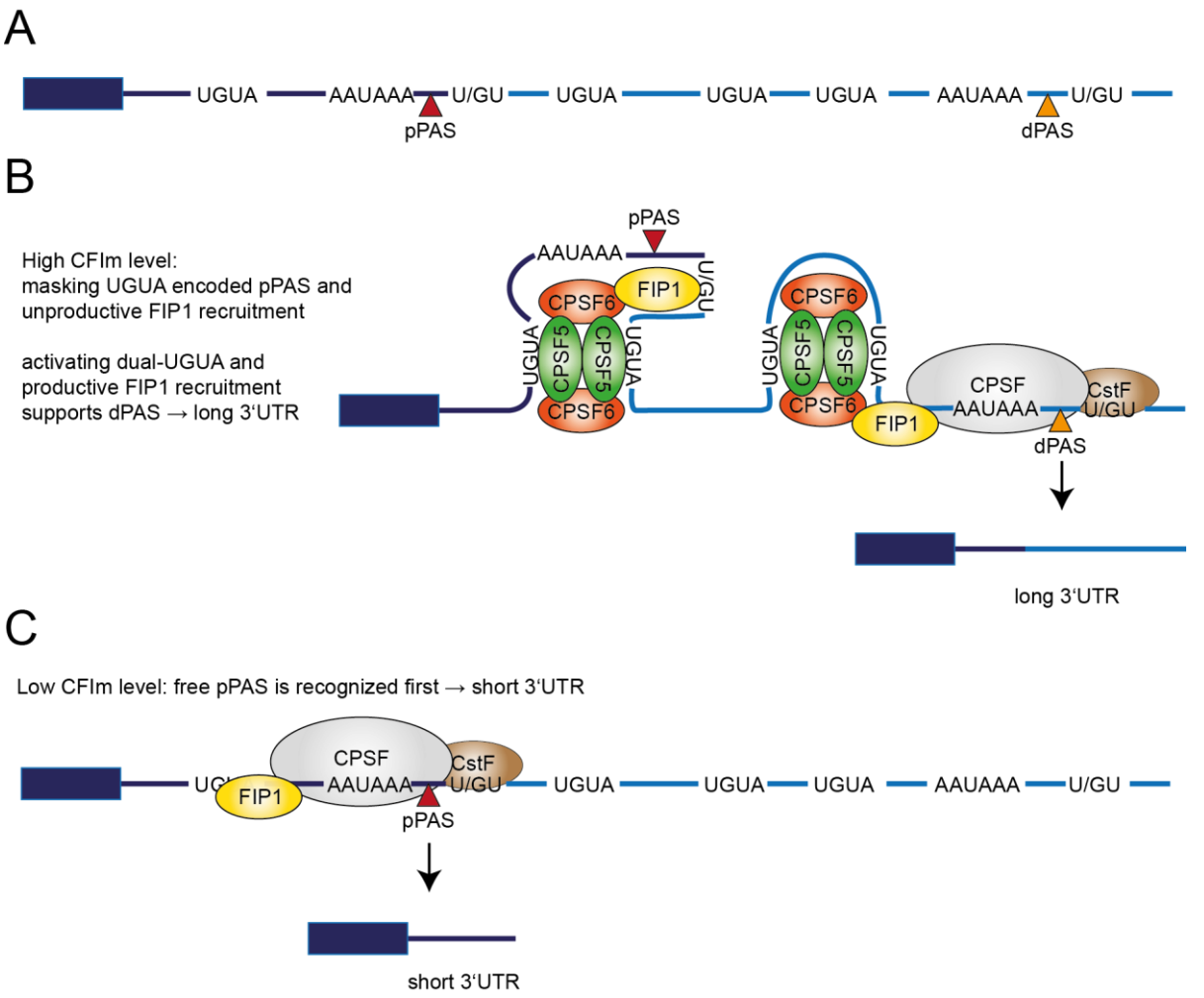


Figure 63: Model how bipartite and dual UGUA-motifs might inhibit or activate CPA at proximal and distal PAS, respectively. A) *Scheme of a transcript containing two strong (AAUAAA) PAS, showing distinct positions of UGUA-motifs.* ***B)*** *High levels of CFIm (presented as its subunits CPSF5 [green] and CPSF6 [orange]) result in CFIm binding to bipartite UGUA-motifs immediately up- and downstream of the pPAS looping out the CSE and/or DSEs leading to unproductive FIP1 (yellow) recruitment. Thereby cleavage is inhibited. Binding of CFIm to dual UGUA-motifs immediately upstream of the dPAS result in functional recruitment of FIP1 and the remaining CPA machinery (represented by CPSF [grey] and CstF [brown]) resulting in cleavage and the expression of transcripts with long 3'UTRs.* ***C)*** *Depletion of SRSF3 results in decreased expression of CFIm and circCpsf6, unmasking the pPAS enabling assembly of the CPA machinery and cleavage resulting in the expression of transcripts with short 3'UTRs.*

To confirm the presence of dual UGUA motifs around pPASs we repeated the motif search and specified the queries to distinguish between single and UGUA pairs. Interestingly, we found that bipartite UGUA motifs were mostly enriched at both the dPASs and pPASs of CPSF6- and SRSF3-sensitive transcripts compared to transcripts that are only CPSF6-sensitive. This suggests that SRSF3-sensitive pPASs and dPASs could be skipped by looping the intervening RNA. When we looked at dual UGUA motifs being either up- or downstream of PASs the picture was more differentiated between pPASs and dPASs. Dual UGUA motifs upstream of PAS were enriched at dPASs, especially in transcripts with shortened 3'UTRs upon CPSF6 and SRSF3 depletion, while dual motifs downstream of PAS were enriched at pPASs, again showing the strongest enrichment in CPSF6/SRSF3-sensitive transcripts. The

strong enrichment of dual UGUA motifs upstream of dPASs corresponds well with the finding from Zhu et al., 2018, showing that dual-UGUA motifs upstream of a PAS increase its usage compared to a single UGUA motif. Accordingly, the enriched presence of dual UGUA motifs downstream of pPASs might facilitate skipping of those PASs, either by looping out DSEs or by disturbing CstF-binding to DSE at the pPASs (Nunes et al., 2010). Interestingly, the distances between two UGUA motifs enclosing a PAS were much larger (~80 nt), compared to those located on either side of the respective PAS (~25 nt). It has been reported before that the distance between two subsequent UGUA motifs is flexible, but that a minimal distance of 21 nt is required to enable binding of both CPSF5-subunits of the CFIm (Yang et al., 2011a). The same study showed that slightly increasing the distance benefitted the recruitment of CFIm towards the model-RNA-molecules. This suggests that especially the dual UGUA motifs up- and downstream of a PAS display a nearly perfect platform for CFIm recruitment. The increased distance found in between bipartite UGUA motifs might inactivate the respective PAS by looping out the poly(A)-signal located in between, yet this potential mechanism needs to be investigated by additional experiments.

Having identified distinct locations enriched for UGUA motifs at the different PAS-types, we wanted to confirm that those sites are indeed bound by CFIm. For this, we prepared iCLIP-libraries from GFP-tagged CPSF5 and FIP1 derived from our stable P19 cell-lines. We had to pool six replicates each to obtain sufficient crosslink events for downstream analyses. The low number of crosslink events of individual samples might be explained by the fact that each transcript contains only one active CPA site where CPSF5 and FIP1 can bind, in comparison to SR proteins which bind to every exon in mature transcripts and thus generate many more crosslink events. Moreover, binding of CPSF5 and FIP1 to transcript is only transient, while SR proteins remain bound to mature transcript, which strongly enhances their crosslink potential. Nevertheless, we identified sufficient significant crosslink sites for CPSF5 and FIP1. Interestingly, both proteins were most enriched in non-coding RNAs, followed by 3'UTRs, while being almost completely depleted from 5'UTRs, introns, and intergenic regions. The strong enrichment with ncRNAs suggests a potential storage of CPSF5 and FIP1 in sub-nuclear compartments, such as nuclear speckles or paraspeckles (not present in P19 cells), which are constructed by the ncRNAs *Malat1* and *Neat1*, respectively (Hutchinson et al., 2007; Clemson et al., 2009; Sasaki et al., 2009; Sunwoo et al., 2009; Chen & Carmichael, 2009; Tripathi et al., 2010; Spector & Lamond, 2011). Interestingly, paraspeckles are constructed by the long *Neat1_2* isoform, which has been shown to be expressed via APA upon hnRNPK-mediated arrest of the CFIm at the upstream alternative PAS of the short *Neat1_1* isoform (Naganuma et al., 2012). Potentially, high expression of CFIm and enriched binding at the alternative PAS result in preferential expression of *Neat1_1* might be related to the absence of paraspeckles in P19 cells. Alternatively, it has been shown that the snoRNA SNORD50A interacts with FIP1 and inhibits subsequent PAS processing, suggesting that the enrichment

with ncRNAs might also have additional mechanistical functions (Huang et al., 2017). To verify this hypothesis the identified ncRNAs need to be characterized more in depth. Motif enrichment analysis around CPSF5 and FIP1 crosslink sites confirmed the well-established UGUA-motif and UG-rich sequences for CPSF5 and FIP1, respectively (Kaufmann et al., 2004; Venkataraman et al., 2005). Although, the FIP1-associated binding motif was highly enriched in uracil, its core resembles the binding motif of CPSF5 (UGU[U/G/A]) suggesting that crosslinks might also stem from CPSF5 interacting with FIP1.

To be able to investigate the binding patterns of CPSF5 and FIP1 in SRSF3 and CPSF6 targets we integrated the PAS-coordinates derived from MACE-Seq with the crosslink sites derived from CPSF5-GFP and GFP-FIP1 iCLIP. At sPASs CPSF5 binding peaked at around 50 nt upstream of the respective PAS and binding diminished completely at and downstream of the sPAS. FIP1 peaked at the same location but was less enriched. The distance between CPSF5 binding and the downstream PAS agreed well with numbers obtained in PAR-CLIP experiments with CPSF5 and CPSF6 by Martin et al., 2012 and Zhu et al., 2018. However, in the study of Martin et al. FIP1 binding sites were located about 25 nt upstream of the PAS in the middle between CFIm binding sites and the PAS. This deviation might be due to the number and kind of PASs included into this analysis. We included all sPAS identified in our dataset (n = 4,796), while Martin et al. selected the 3,000 most abundant cleavage events. Another iCLIP analysis of FIP1 conducted by Lackford et al., 2014 declared the FIP1 binding site to be around -25 nt upstream of the PAS, but their crosslink profile showed a broad distribution of FIP1 binding within a window of -100 nt to -25 nt, arguing that the binding site of FIP1 is rather variable and caused by the rather degenerated/unspecific U-rich region downstream of PAS. Interestingly, in addition to the main peak upstream of sPASs our FIP1 iCLIP dataset showed a second binding peak in a window between 0 to 25 nt downstream of the sPASs. A similar pattern can be assumed from the binding profile presented in Martin et al., 2012. As CPSF5 binding was depleted in this region it can be assumed that crosslinks of FIP1 downstream of the sPASs might be independent of the CFIm and represent sequence specific binding since the downstream sequences are generally enriched in uracil and guanosine to facilitate recruitment of the necessary CstF-complex (Zarudnaya et al., 2003; Tian et al., 2005; Nunes et al., 2010). When we switched the focus to transcripts with multiple PASs and separated pPASs from dPASs, we found a similar distribution of CPSF5 and FIP1 crosslink sites at dPASs, supporting the assumption that dPASs resemble the default CPA site of transcripts and behave similar as sPASs. In contrast, the pattern at pPASs diverged suggesting that those sites are the regulatory hotspot of 3'UTR-APA. While CPSF5 enrichment upstream of the pPASs resembled the enrichment pattern at sPASs and dPASs, FIP1 binding was increased compared to the dPASs, suggesting that FIP1 recruitment mirrors the enhanced activation of pPAS in P19 cells. Compellingly, CPSF5 and FIP1 binding was similar upstream and downstream of pPASs confirming that certain bipartite UGUA motifs are

bound by the CFIm. We saw no obvious change in the binding-patterns and -levels of CPSF5 and FIP1 around pPASs and dPASs after depletion of CPSF6. However, transcripts with reduced dPAS usage after SRSF3 depletion exposed an increased binding of both CPA-factors compared to non-affected transcripts. This supports a model whereby SRSF3-regulated targets are especially dependent on CFIm-mediated recruitment of FIP1.

When we analyzed the binding patterns of CPSF5 and FIP1 at pPASs and dPASs separated by transcripts with a short 3'UTR (ΔPDUI ≤ 25 %) or a long 3'UTR (ΔPDUI ≥ 75 %) under control conditions, we found that CPSF5 and FIP1 bind to pPASs in transcripts with long 3'UTRs, suggesting that they do not activate the pPAS. One possibility could be an unproductive recruitment of CPSF5 and/or FIP1 disturbing CPA at those pPASs. This hypothesis is supported by a recent publication by Wang et al., 2020 in which it was shown that ICP27 interacted with FIP1 to assemble an inactive CPA complex upon HSV-1 infection. Similarly, the U1 snRNP complex was shown to inhibit intronic PAS activation (telescripting) although CFIm and FIP1 were found to bind at those suppressed PAS (So et al., 2019). Interestingly, those CFIm complexes were enriched in CPSF7, which has been shown to rather inhibit PAS usage compared to its paralogue CPSF6 (Gruber et al., 2012; Martin et al., 2012). In addition, Millevoi et al., 2006 showed that the RNA binding of CPSF7 deviated from CPSF6 and does not interact with SRSF7. Finally, depletion of CPSF7 did not affect 3'UTR-APA (Rüegsegger et al., 1998; Li et al., 2015; So et al., 2019). Hence, it might be possible that an alternative CFIm composition containing CPSF5 and CPSF7 might be recruited towards suppressed pPAS resulting in a dead-assembly of the remaining CPA machinery.

In conclusion, we identified a mechanism how 3'UTR-APA is regulated via CFIm dependent on the expression of SRSF3. SRSF3 promotes high expression levels of CPSF6 which in consequence stabilizes expression of CPSF5. This leads to a predominant usage of dPASs and the expression of transcripts with long 3'UTRs. CPSF5-specific dual UGUA motifs are distributed differently at pPASs and dPASs. While dPASs contain dual UGUA motifs upstream that promote their activation, pPASs contain more dual UGUA motifs downstream of the PASs. These pPAS could be inactivated through an unproductive recruitment of the CPA machinery. Additionally, we identified UGUA motifs enclosing the pPAS, which might loop-out poly(A) signals to render those pPAS inactive. iCLIP experiments showed that only a small fraction of transcripts might be regulated by bipartite CFIm binding at pPASs, while the majority of CPSF5 binds upstream of pPASs. On the other hand, FIP1 binds upstream and downstream of certain transcripts, which could be due to the U/GU-rich sequences within the DSE region. Nevertheless, it will be interesting to further study which function FIP1 might have binding downstream of PASs. Finally, we found initial evidence that SRSF3-sensitive targets might be regulated by unproductive recruitment of CPSF5 and FIP1, indicating that several similar mechanisms networking

with each other are in place to tightly regulate APA at strong pPASs within the 3'UTR. All these results define a distinct mechanism how SRSF3 might be able to render 3'UTR-APA events indirectly in difference to its sibling SRSF7, where we identified a direct and RNA-independent interaction with FIP1 and the CFIm to activate pPAS usage.

6.8 dPAS usage is promoted in neuronal differentiated P19 cells, while SRSF7 and FIP1 expression decreased

Various studies, including analysis of individual genes and more genome-wide analysis utilizing microarrays or RNA-Seq, had shown that neuronal cells express transcripts with elongated 3'UTRs in the model organisms *Drosophila melanogaster*, *Mus musculus* and in humans (Pelka et al., 2005; Zhang et al., 2005b; Costessi et al., 2006; Smibert et al., 2012; Miura et al., 2013). 3'UTR-APA and the associated length of distinct transcript isoforms in neurons have been associated with the sub-cellular localization of the respective transcripts. An et al., 2008 provided evidence that the short 3'UTR isoform of *Bdnf* located to the cell body, while the long 3'UTR isoform was found in the dendrites. However, recent genome-wide analyses challenged that finding, stating that the length of 3'UTRs does not generally influence localization of the transcripts to certain compartments within neurons (Tushev et al., 2018; Middleton et al., 2019; Ciolli Mattioli et al., 2019). It seems that no general statement for certain localizations or functions can be made for neuronal transcripts with long 3'UTRs compared to their short alternative isoforms. However, it seems that the respective isoforms derived from the same gene express rather distinct functions either by localized translation e.g., establishing distinct protein-protein interactions, or independent of translation mediating *trans*-acting functions e.g., as miRNA sponges or coordinating AS & APA (Valluy et al., 2015; Kishi et al., 2019; Zhang et al., 2019; Bae & Miura, 2020). Interestingly, several RBPs and especially splicing factors, such as *D*. *melanogaster* ELAV, its human orthologue ELAVL, NOVA2 and FUS, have been indicated to regulate the expression of 3'UTR-elongated neuronal transcripts (Licatalosi et al., 2008; Hilgers et al., 2012; Masuda et al., 2015; Grassi et al., 2018).

Hence, we wanted to investigate the potential engagement of SRSF3 and SRSF7 in the regulation of 3'UTR-APA after differentiation. Therefore, we differentiated the P19 cells into neuronal cells as described by Nakayama et al., 2014. The protocol worked well and resulted in successful differentiation after 8 days, as controlled by the expression of the commonly known pluripotency marker Oct4 and the differentiation markers Synapsin and Nestin. Neuronal differentiation of P19 cells resulted in global 3'UTR-length extension. Interestingly, SRSF7, especially its de-phosphorylated isoform, and FIP1 were decreased after differentiation. The decrease of SRSF7 expression could be induced by increased

inclusion of the PCE via AS. Yet, in a recent report Leclair et al. extensively investigated PCE-related expression of SR proteins during differentiation and found that inclusion of the PCE of SRSF7 is only slightly increased (~25%) during neuronal differentiation, arguing for an unknown alternative mechanism (Leclair et al., 2020). In this publication it was also shown that the PCE of SRSF3 was more prone to be included upon neuronal differentiation (~50%) which would imply preferred degradation of SRSF3 in neuronal cells. We could not confirm this finding in our experiments, as SRSF3 expression decreased to a minor extent compared to the extensive decrease of SRSF7 expression. Furthermore, a study on juvenescence focusing on SRSF7 reported that SRSF7 was highly expressed in juvenile cells of the cerebral cortex (Kadota et al., 2020). SRSF7 expression decreased during senescence of these cells and upon increased cell densities, suggesting that the decrease of SRSF7 during differentiation of P19 cells might be connected to senescence or cell densities. Interestingly, Kadota et al. found a direct connection between SRSF7-expression and the retinoic-acid-dependent differentiation potential of Neuro2a cells. Depletion of SRSF7 neglected the ability of Neuro2a cells to differentiate into neuronal cells upon retinoic acid-treatment, suggesting that SRSF7 might be a neuronal differentiation factor.

SRSF3 and CFIm levels did not change significantly upon differentiation enhancing expression of transcripts with long 3'UTRs via fail-safe usage of the dPASs, supporting the hypothesis that pPASs represent the hotspot of 3'UTR-APA regulation. Consequently, we speculate that expression of potential pPAS enhancers should be decreased in differentiated cells. Preliminarily we checked the differential expression using our differentiation DESeq2 dataset regarding cleavage factors, which were shown to cause 3'UTR lengthening upon depletion (Li et al., 2015). Indeed, expression of *Fip1l1*, *Pcf11*, *Sympk* (encoding for SYMPLEKIN) and several transcripts encoding for CstF-subunits was significantly reduced upon differentiation (data not shown). Yet, we only validated that the reduced expression of *Fip1l1* resulted in decreased expression of FIP1. Expression of the other cleavage factors remains to be validated on the protein level. Coherently, Li et al. reported that depletion of FIP1 or PCF11 affected 3'UTR-lengthening as strong as differentiation (Log_2[Lengthening/Shortening] values of 1.5; 1.4 and 1.8, respectively). FIP1 was defined as a pluripotency factor, whose depletion succeeded in the elongation in 3'UTRs (Lackford et al., 2014). As we showed in this thesis that SRSF7 can recruit FIP1 towards pPASs, this suggests that the higher expression of transcripts with short 3'UTRs during pluripotency may be connected to this SRSF7-mediated recruitment of FIP1. Differentiation and resulting reduced expression of SRSF7 and FIP1 may abrogate efficient cleavage and polyadenylation at pPASs leading to expression of long 3'UTRs using the default dPASs.

A comparison of targets identified by DaPARS after differentiation and depletion of SRSF7 revealed a defined overlap (52/134 SRSF7 targets), indicating that a distinct subset of transcripts depends on SRSF7 for their 3'UTR length. This subset might be highly specific to SRSF7 as the direct

depletion of FIP1 was shown to affect 374 genes in embryonal stem cells, elongating the 3'UTR of 311 of those transcripts by enhanced usage of the dPAS (Lackford et al., 2014). Subsequently, using differential expression analysis via DeSeq2, we determined that mRNA expression of SRSF7 and FIP1 were only decreased slightly, not matching the levels of decrease observed in the decrease of the protein levels, suggesting translational or post-translational regulation of SRSF7 and FIP1. Those underlying mechanisms will require additional research to clarify.

Finally, we monitored binding of SRF3 and SRSF7 at pPASs and dPASs of transcripts with elongated 3'UTRs upon differentiation. Compared to unchanged transcripts binding of SRSF3 and SRSF7 was slightly increased upstream of the respective PASs. Persuasively, SRSF7 binding upstream of affected pPASs was even more increased, supporting a potential 3'UTR-APA regulation of specific transcripts in proliferating P19 cells and during differentiation into neuronal cells.

In conclusion, we provide evidence that SRSF7 might be involved in the regulated expression of transcripts containing short 3'UTRs in proliferating cells, such as P19 cells. Reduction of SRSF7 during neuronal differentiation, together with co-depletion of the pluripotency factor FIP1, inactivates pPAS usage resulting in transcripts with long 3'UTRs. In this context, various other neuronal cell types, such as astrocytes, microglia and oligodendrocytes were found to not show this 3'UTR-elongation phenotype, suggesting that 3'UTR-APA has to be regulated differently (Wang et al., 2014b; Xia et al., 2014; Ye et al., 2018). It might be interesting to analyze if SRSF7 is involved. In addition, depletion of the cleavage factor CPSF5 had been connected to the development of glioblastomas, hence mis-regulation of SRSF7 might also cause neuronal diseases (Masamha et al., 2014).

7 Conclusion and outlook

In the light of these findings, SR proteins cannot be reduced solely to their essential functions in constitutive and alternative splicing to modulate gene expression. Over the last three decades individual SR proteins were described on single events to have additional functions during the life cycle of mRNAs including alternative polyadenylation (Lou et al., 1998; Alt et al., 1980). The development of genome-wide analysis to investigate changes in gene expression and direct interactions between DNA, RNA and proteins enabled more comprehensive analysis of SR protein functions. Here, extending the findings from our previous publication (Müller-McNicoll et al., 2016), SRSF3 and SRSF7 were now shown and described in detail to affect the outcome of 3' UTR APA in opposite directions in distinct target genes in a splicing-independent mode starting a new chapter of SR protein-affected gene expressions apart of constitutive and alternative splicing.

Interestingly, even if SRSF3 and SRSF7 developed highly co-evolutionary and share basic characteristics, e.g., both SR proteins shuttle well between the nucleus and the cytoplasm and both affect 3'UTR-APA. Here, for the first time, we described that SRSF3 and SRSF7 have distinct mode of action and outcome on regulating 3'UTR-APA. We identified SRSF7 to directly promote activation of pPASs, increasing the expression of short 3' UTR-transcripts, by RNA-independent recruitment of FIP1. Yet, the number of identified targets was much lower compared to SRSF3. This might be due to certain redundancies within the SR protein family. SRSF1 and SRSF6 also recognize purine-rich motifs, similar to SRSF7 (Müller-McNicoll et al., 2016). Therefore, these two SR proteins might substitute upon knockdown of *Srsf7*. To verify this hypothesis, I suggest co-depletion experiments targeting SRSF7/SRSF1, SRSF7/SRSF6 or SRSF7/SRSF1/SRSF6, followed by total RNA-Seq and DaPARS analysis to see if the number of targets increases and if the direction of 3' UTR-changes stays the same. As SR proteins are essential for cellular survival and knockouts are nearly impossible, the co-depletion might be difficult or inefficient. An alternative experiment could be to target SRSF7 for rapid degradation, e.g., by the TRIM-21 system as described by Clift et al., 2017. That way there might be a time frame in which SRSF7 levels are decreased and neither SRSF1 nor SRSF6 could have the time to substitute in the 3'end processing. Furthermore, it might be interesting to determine if SRSF1 and SRSF6 are capable to recruit early factors of the CPA machinery, such as FIP1 or CFIm to facilitate the activation of pPASs. Hence, Co-IPs with these two SR proteins might be helpful experiments to clarify the function of SRSF7 in 3' UTR-APA.

The subcellular and subnuclear localization, as well as the activity of SR proteins in constitutive and alternative splicing are tightly connected to differential phosphorylation states of the RS-domains, which has been summarized in the phosphorylation cycle of SR proteins (Zhou & Fu, 2013; Wegener &

Müller-McNicoll, 2019). In this study, we have shown that the SRSF7-dependent regulation of 3'UTR-APA is independent of splicing; hence it seems a valid question to ask whether the phosphorylation during CPA matches the cycle during splicing or if there are differences. Our results identified that SRSF7 interacts directly with FIP1 in a potentially semi-phosphorylated state (tested by using the native RS-domain), yet the interaction seemed to be strengthened using RA-phosphomimetics representing the completely de-phosphorylated SRSF7. But the experiments involving RA-mimetics must be considered carefully, due to the potential charge-dependent enrichment to nucleoli. Yet, together with the finding that FIP1 did not interact with the RD-mimetics representing hyperphosphorylated SRSF7 this suggests a deviating phosphorylation cycle compared to splicing. Nevertheless, as the SR protein-specific phosphatase PP1 was found in purified CPA complexes (Shi et al., 2009; Aubol et al., 2017) the regulation and localization of SR proteins might not differ too much between splicing and CPA. In conclusion I propose the following sequence of events: First SRSF3 and SRSF7 get hyper-phosphorylated within the nucleus by CLK1/4 prior recruitment to their target pPAS at nascent mRNAs. At the pPAS both SR proteins might be partially de-phosphorylated via a yet unknown mechanism, potentially by the presence of the partially preassembled CPA machinery including PP1. The partial de-phosphorylation enables SRSF7 to recruit FIP1, subsequently recruiting and/or stabilizing the remaining CPA machinery and consequently increasing the contact between SRSF7 and PP1 resulting in hypo-phosphorylation of SRSF7 during cleavage, like the de-phosphorylation during splicing. This would be a crucial event as only the hypo-phosphorylated SRSF7 was shown to recruit NXF1 to facilitate nuclear export of the bound transcript, in this case a transcript with a short 3'UTR (Müller-McNicoll et al., 2016). In contrast, SRSF3 might as well be partially de-phosphorylated but cannot recruit FIP1 and the CPA machinery. Therefore, it seems unlikely that SRSF3 becomes hypo-phosphorylated at the pPAS neglecting NXF1 recruitment. Nevertheless, SRSF3 also binds upstream of the dPAS and could be de-phosphorylated by the CFIm-recruited CPA machinery, subsequently allowing NXF1 recruitment and nuclear export of transcripts with long 3'UTRs. Additional experiments could help to clarify the phosphorylation-dependent mechanism of SRSF-regulated 3'UTR-APA. I suggest monitoring the interaction between SRSF7 and FIP1 after depletion of either the CLK1/4 kinases phosphorylating RS-domains, or depletion of PP1 and PP2A to block de-phosphorylation of RS-domains. The subsequent effect on APA could be easily controlled by 3'RACE-PCR either of endogenous transcripts or by using the reporter-constructs generated for the purpose of this thesis.

High-throughput methods enabling genome-wide mapping and analysis of m^6A methylation within transcripts identified terminal exons and 3' UTRs as hot-spots of those reversible nucleotide modifications (Batista et al., 2014; Ke et al., 2015; Molinie et al., 2016; Bartosovic et al., 2017). With those findings m^6A has been associated with potential functions in the regulation of alternative splicing and APA. Recently, a genome wide approach identified m^6A methylation enriched at non-canonical

PAS, considering a direct regulation of those cleavage sites by the methylation (Zhang et al., 2020). Interestingly, when we compared the canonical m^{6}A motif 'RR**A**CH' (with R = G/A, H = A/C/U and the central A being the methylatable adenosine) we found a potential overlap with the SRSF7 motif 'GAYGAY' (with Y = C/T). We speculate, that the methylation writer proteins, such as METTL3 and METTL14 might also recognize the GAY motif to add the m^{6}A modification to the SRSF7 motif, thereby potentially inhibiting recruitment of SRSF7 to the respective site. Strikingly, an initial manual research in m^{6}A databases using few SRSF7 targets, such as *Ddx21*, revealed positive methylation events in SRSF7 binding motifs upstream of pPASs, confirming our hypothesis (data not shown). To analyze this on the full scale it would be important to first extract the m^{6}A-tome of P19 cells, e.g., by using miCLIP (Linder et al., 2015) or m^{6}A-CLIP (Ke et al., 2015), to confirm that these modifications are enriched within the 3' UTR of P19 cells and to identify potential overlaps with the SRSF7 motif. In addition, I propose to repeat the SRSF7 iCLIP experiment upon depletion of the m^{6}A-writing proteins, such as METTL3/METTL14, as I would hypothesize that this might result in increased enrichment of SRSF7 upstream of pPASs. Subsequently, the effect of m^{6}A methylation on SRSF7 recruitment and the effect on distinct 3' UTRs, such as *Ddx21* should be tested e.g., by 3'RACE-PCR-analysis. Finally, I would propose a collaboration with the FRET-experts at the AK Hengesbach of the University Frankfurt to establish FRET-experiments to test if m^{6}A-methylation upstream of pPAS specifically inhibits SRSF7 binding. Therefore, artificially synthesized RNA-oligonucleotides based on the sequence upstream of SRSF7-specific which either contain or lack pre-methylated adenosines could be coupled with a suitable FRET-acceptor, such as mCherry. I assume that flushing with EGFP-tagged SRSF7 as a FRET-donor should enable efficient FRET, resulting in the detection of mCherry-emitted fluorescence, only if SRSF7 can bind to the respective target in the absence of the m^{6}A-modification. In the presence of those m^{6}A modifications no energy-transfer and therefore no fluorescence emerging from mCherry should be detectable.

In contrast to SRSF7, we found that SRSF3 was unable to interact RNA-independently with either the CFIm or FIP1, yet the depletion of SRSF3 affected much more targets than the depletion of SRSF7. Hence, we proposed that SRSF3 actively blocks the pPAS competing with SRSF7 for binding at distinct shared pPASs. We began by dissecting the differences between both closely related SR proteins and identified the CCHC-type Zinc-knuckle and a 27 aa hydrophobic stretch within the RS-domain of SRSF7 as major differences between SRSF3 and SRSF7. Experiments including various SRSF7 mutants and SRSF3 chimaeras showed that these two domains are necessary within SRSF7 to facilitate RNA-independent interaction towards FIP1, while these domains alone were not sufficient to enable SRSF3 to behave likewise. We concluded that these domains need to be embedded in rather flexible surroundings, which are not present in the rather concise SRSF3 protein. Nevertheless, simultaneous inclusion of these two domains into SRSF3 resulted in increased pPAS usage resembling endogenous

SRSF7. To further analyze this and define the sufficient minimal setup of domains to enable RNA-independent interactions between SRSF3 and FIP1 the existing chimaera constructs could be extended with artificial flexible linkers of various lengths to verify if those domains are sufficient to recruit FIP1.

Moreover, we identified that SRSF3 mainly affects 3'UTR-APA by promoting the expression of CPSF6, the large subunit of CFIm, a potent dPAS activator. Expression of SRSF3 facilitates inclusion of Exon 6 into the *Cpsf6* pre-mRNA which subsequently is translated into native CPSF6. Depletion of SRSF3 results in the skipping of Exon 6, which is normally massively bound by SRSF3, generating a potential NMD-targeted transcript and subsequent reduced CPSF6 levels. While differential gene expression analysis of RNA-Seq data suggested a reduction of the respective mRNA, it remains to be verified that the emerging alternatively spliced transcript is routed for NMD. Therefore, I propose experiments in P19 cells co-depleting SRSF3 and the NMD-associated co-factor UPF1 to inhibit NMD and monitoring the levels of the alternative Δ*Ex6Cpsf6* transcript (Kim & Maquat, 2019). Alternatively, micro-RNA miR-128 could be expressed which also was shown to inhibit NMD by reducing levels of NMD-related key-factors (Bruno et al., 2011).

Additional to the connotation of CFIm depletion leading to decreased dPAS usage (Li et al., 2015) we identified another potential mechanism how CFIm binding around the pPAS might inhibit the usage of those PASs. Therefore, two UGUA motifs seem to be necessary that either enclose the poly(A) signal or are located closely downstream of it. This could facilitate looping of the short enclosed sequence around the CFIm molecule, similar to the proposed long-distance model by Yang et al., 2011a and Martin et al., 2012, or mask certain CSEs and/or DSEs inhibiting recruitment of CstF. To further address this, artificial UGUA motifs could be introduced in the existing reporter constructs at positions and distances as determined by our iCLIP experiments presented here, followed by 3'RACE-PCR analysis.

To our surprise we found CPSF5 and FIP1 strongly enriched with ncRNAs, surpassing binding to 3' UTRs by far, after performing iCLIP. Certain ncRNAs, such as *NEAT1* and *MALAT1*, are important central structures of nuclear storage compartments of RBPs like paraspeckles and nuclear speckles, respectively. Therefore, we speculated that this might be an indication for the association of CPSF5 and FIP1 with one of those compartments. Yet, nothing is known if these two cleavage factors may be stored to certain speckles and if so, which kind of speckle it might be. Interestingly, a recent publication by Jang et al., 2019 provided evidence that CPSF6 shuttles between the nucleus and cytoplasm and that re-import into the nucleus is mediated via phosphorylation of it RS-like domain by TPNO3 similarly to SR proteins. Moreover, CPSF6 has been identified to localize to paraspeckles and nuclear speckles (Dettwiler et al., 2004), suggesting that those compartments might also function as a storage for CPSF5 and FIP1. To further investigate this, I would suggest first to analyze the ncRNAs recognized by CPSF5 and FIP1 to identify potential enrichment for certain lncRNAs, such as *MALAT1*. Subsequently, I would

propose to follow up with co-localization experiments using specific nuclear speckle marker, such as SC35 and other identified lncRNAs which might sponge CFIm subunits or FIP1.

Getting back to our initial publication from 2016, in which we determined the mechanism how SR proteins function as important mediators of nuclear export of matured transcript via the interaction with the nuclear export co-factor NXF1 (Müller-McNicoll et al., 2016). In this paper, we also identified SR protein (especially SRSF3 and SRSF7) as potential APA regulators on a global scale, expanding their function from consecutive and alternative splicing to 3'end processing and nuclear surveillance of pre-mRNA maturation. With this thesis, we extended the insights into the molecular mechanisms how SRSF3 and SRSF7 affect 3' UTR-APA by either denying or activating cleavage at pPASs, respectively, resulting in the expression of transcripts with short or long 3' UTRs. Recruitment of NXF1 to those transcripts, either by SRSF3 at inhibited pPASs, or by SRSF7 at activated pPASs might lead to the nuclear export of nearly identical transcripts only differing in the length of their 3' UTR. As recent genome-wide studies suggested that the differing lengths of 3' UTRs might be associated with localized translation of distinct transcript (Tushev et al., 2018; Middleton et al., 2019; Ciolli Mattioli et al., 2019), it might be very interesting to follow the final destination and fate of certain transcripts. In line with this hypothesis, Middleton et al. identified an enrichment of SRSF3 binding motifs in dendritic-localized transcript isoforms containing long 3'UTRs. To achieve this, I would suggest beginning with RNA fluorescence *in situ* hybridization (RNA-FISH) of the respective transcript isoforms. A general challenge will be to distinguish the short isoforms, as this particular sequence of the 3' UTR will be shared with the alternative long isoform. This might be achieved by designing the FISH probe to overlap into the poly(A) tail of the short 3' UTR isoform. Furthermore, I would propose a local cooperation with the Heilemann and Schuman labs from the University Frankfurt, to establish DNA-PAINT (Points Accumulation for Imaging in Nanoscale Topography) in the analysis of subcellular localization of 3' UTR-APA-derived transcripts and their local translation to overcome potential localization issues due to the refraction limits of conventional confocal microscopy (Jungmann et al., 2014; Jungmann et al., 2016; Schnitzbauer et al., 2017). This might be particularly interesting in neuronal differentiated P19 cells as this kind of compartmentalization of transcripts has been extensively studied in neuronal cells (reviewed in Glock et al., 2017).

Finally, in this thesis we dissected a new function of SR proteins in the regulation of 3'UTR-APA, expanding the post-transcriptional functions of SR proteins beyond constitutive and alternative splicing. We merely focused on the function of SRSF3 and SRSF7 related to 3'UTR-APA in the course of this thesis. But we have to take into consideration that SR proteins are mainly (alternative) splicing regulators and therefore, should affect CDS-APA to a certain degree. For example, SRSF3 had been identified early on as a regulator of CDS-APA (Lou et al., 1998). In consequence it might be worthwhile

to reexamine the underlying 3’end-sequencing data focusing to identify CDA-APA events upon knockdown of *Srsf3* and *Srsf7*.

8 References

Abu-Baker, A., and Rouleau, G.A. (2007). Oculopharyngeal muscular dystrophy: recent advances in the understanding of the molecular pathogenic mechanisms and treatment strategies. Biochimica et biophysica acta *1772*, 173–185.

Agarwala, S.D., Blitzblau, H.G., Hochwagen, A., and Fink, G.R. (2012). RNA methylation by the MIS complex regulates a cell fate decision in yeast. PLoS genetics *8*, e1002732.

Alt, F.W., Bothwell, A.L., Knapp, M., Siden, E., Mather, E., Koshland, M., and Baltimore, D. (1980). Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3' ends. Cell *20*, 293–301.

An, J.J., Gharami, K., Liao, G.-Y., Woo, N.H., Lau, A.G., Vanevski, F., Torre, E.R., Jones, K.R., Feng, Y., Lu, B., and Xu, B. (2008). Distinct role of long 3' UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons. Cell *134*, 175–187.

Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. Genome research *22*, 2008–2017.

Anderson, P., and Kedersha, N. (2009). RNA granules: post-transcriptional and epigenetic modulators of gene expression. Nature reviews. Molecular cell biology *10*, 430–436.

Änkö, M.-L., Müller-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 biology *13*, R17.

Arefeen, A., Liu, J., Xiao, X., and Jiang, T. (2018). TAPAS: tool for alternative polyadenylation site analysis. Bioinformatics (Oxford, England) *34*, 2521–2529.

Arif, A., Jia, J., Moodt, R.A., DiCorleto, P.E., and Fox, P.L. (2011). Phosphorylation of glutamyl-prolyl tRNA synthetase by cyclin-dependent kinase 5 dictates transcript-selective translational control. Proceedings of the National Academy of Sciences of the United States of America *108*, 1415–1420.

Ashwal-Fluss, R., Meyer, M., Pamudurti, N.R., Ivanov, A., Bartok, O., Hanan, M., Evantal, N., Memczak, S., Rajewsky, N., and Kadener, S. (2014). circRNA biogenesis competes with pre-mRNA splicing. Molecular cell *56*, 55–66.

Aubol, B.E., Hailey, K.L., Fattet, L., Jennings, P.A., and Adams, J.A. (2017). Redirecting SR Protein Nuclear Trafficking through an Allosteric Platform. Journal of molecular biology *429*, 2178–2191.

Awasthi, S., and Alwine, J.C. (2003). Association of polyadenylation cleavage factor I with U1 snRNP. RNA (New York, N.Y.) *9*, 1400–1409.

Bae, B., and Miura, P. (2020). Emerging Roles for 3' UTRs in Neurons. International journal of molecular sciences *21*.

Balcerak, A., Trebinska-Stryjewska, A., Konopinski, R., Wakula, M., and Grzybowska, E.A. (2019). RNA-protein interactions: disorder, moonlighting and junk contribute to eukaryotic complexity. Open biology *9*, 190096.

Banerjee, S., and Barraud, P. (2014). Functions of double-stranded RNA-binding domains in nucleocytoplasmic transport. RNA biology *11*, 1226–1232.

Bartosovic, M., Molares, H.C., Gregorova, P., Hrossova, D., Kudla, G., and Vanacova, S. (2017). N6-methyladenosine demethylase FTO targets pre-mRNAs and regulates alternative splicing and 3'-end processing. Nucleic acids research *45*, 11356–11370.

Batista, P.J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D.M., Lujan, E., Haddad, B., Daneshvar, K., Carter, A.C., Flynn, R.A., Zhou, C., Lim, K.-S., Dedon, P., Wernig, M., Mullen, A.C., Xing, Y., Giallourakis, C.C., and Chang, H.Y. (2014). m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell stem cell *15*, 707–719.

Batra, R., Charizanis, K., Manchanda, M., Mohan, A., Li, M., Finn, D.J., Goodwin, M., Zhang, C., Sobczak, K., Thornton, C.A., and Swanson, M.S. (2014). Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease. Molecular cell *56*, 311–322.

Beadle, G.W., and Tatum, E.L. (1941). Genetic Control of Biochemical Reactions in Neurospora. Proceedings of the National Academy of Sciences of the United States of America *27*, 499–506.

Beaudoing, E., Freier, S., Wyatt, J.R., Claverie, J.M., and Gautheret, D. (2000). Patterns of variant polyadenylation signal usage in human genes. Genome research *10*, 1001–1010.

Beckmann, B.M., Castello, A., and Medenbach, J. (2016). The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions. Pflugers Archiv European journal of physiology *468*, 1029–1040.

Bedard, K.M., Daijogo, S., and Semler, B.L. (2007). A nucleo-cytoplasmic SR protein functions in viral IRES-mediated translation initiation. The EMBO journal *26*, 459–467.

Ben-Dov, C., Hartmann, B., Lundgren, J., and Valcárcel, J. (2008). Genome-wide analysis of alternative pre-mRNA splicing. The Journal of biological chemistry *283*, 1229–1233.

Berg, M.G., Singh, L.N., Younis, I., Liu, Q., Pinto, A.M., Kaida, D., Zhang, Z., Cho, S., Sherrill-Mix, S., Wan, L., and Dreyfuss, G. (2012). U1 snRNP determines mRNA length and regulates isoform expression. Cell *150*, 53–64.

Berget, S.M. (1995). Exon recognition in vertebrate splicing. The Journal of biological chemistry *270*, 2411–2414.

Berglund, J.A., Abovich, N., and Rosbash, M. (1998). A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. Genes & development *12*, 858–867.

Berkovits, B.D., and Mayr, C. (2015). Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. Nature *522*, 363–367.

Birse, C.E., Minvielle-Sebastia, L., Lee, B.A., Keller, W., and Proudfoot, N.J. (1998). Coupling termination of transcription to messenger RNA maturation in yeast. Science (New York, N.Y.) *280*, 298–301.

Boeynaems, S., Alberti, S., Fawzi, N.L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., van den Bosch, L., Tompa, P., and Fuxreiter, M. (2018). Protein Phase Separation: A New Phase in Cell Biology. Trends in cell biology *28*, 420–435.

Bokar, J.A., Shambaugh, M.E., Polayes, D., Matera, A.G., and Rottman, F.M. (1997). Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. RNA *3*, 1233–1247.

Botti, V., McNicoll, F., Steiner, M.C., Richter, F.M., Solovyeva, A., Wegener, M., Schwich, O.D., Poser, I., Zarnack, K., Wittig, I., Neugebauer, K.M., and Müller-McNicoll, M. (2017). Cellular differentiation state modulates the mRNA export activity of SR proteins. The Journal of cell biology *216*, 1993–2009.

Boulle, K. de, Verkerk, A.J., Reyniers, E., Vits, L., Hendrickx, J., van Roy, B., van den Bos, F., Graaff, E. de, Oostra, B.A., and Willems, P.J. (1993). A point mutation in the FMR-1 gene associated with fragile X mental retardation. Nature genetics *3*, 31–35.

Boutz, P.L., Bhutkar, A., and Sharp, P.A. (2015). Detained introns are a novel, widespread class of post-transcriptionally spliced introns. Genes & development *29*, 63–80.

Bradley, T., Cook, M.E., and Blanchette, M. (2015). SR proteins control a complex network of RNA-processing events. RNA (New York, N.Y.) *21*, 75–92.

Brais, B., Bouchard, J.P., Xie, Y.G., Rochefort, D.L., Chrétien, N., Tomé, F.M., Lafrenière, R.G., Rommens, J.M., Uyama, E., Nohira, O., Blumen, S., Korczyn, A.D., Heutink, P., Mathieu, J., Duranceau, A., Codère, F., Fardeau, M., Rouleau, G.A., and Korcyn, A.D. (1998). Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. Nature genetics *18*, 164–167.

Bressan, G.C., Moraes, E.C., Manfiolli, A.O., Kuniyoshi, T.M., Passos, D.O., Gomes, M.D., and Kobarg, J. (2009). Arginine methylation analysis of the splicing-associated SR protein SFRS9/SRP30C. Cellular & molecular biology letters *14*, 657–669.

Brown, K.M., and Gilmartin, G.M. (2003). A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. Molecular cell *12*, 1467–1476.

Brown, R.S. (2005). Zinc finger proteins: getting a grip on RNA. Current opinion in structural biology *15*, 94–98.

Bruno, I.G., Karam, R., Huang, L., Bhardwaj, A., Lou, C.H., Shum, E.Y., Song, H.-W., Corbett, M.A., Gifford, W.D., Gecz, J., Pfaff, S.L., and Wilkinson, M.F. (2011). Identification of a microRNA that activates gene expression by repressing nonsense-mediated RNA decay. Molecular cell *42*, 500–510.

Buchan, J.R. (2014). mRNP granules. Assembly, function, and connections with disease. RNA biology *11*, 1019–1030.

Busch, A., and Hertel, K.J. (2012). Evolution of SR protein and hnRNP splicing regulatory factors. Wiley interdisciplinary reviews. RNA *3*, 1–12.

Cáceres, J.F., and Kornblihtt, A.R. (2002). Alternative splicing: multiple control mechanisms and involvement in human disease. Trends in genetics TIG *18*, 186–193.

Cáceres, J.F., Misteli, T., Screaton, G.R., Spector, D.L., and Krainer, A.R. (1997). Role of the Modular Domains of SR Proteins in Subnuclear Localization and Alternative Splicing Specificity. The Journal of Cell Biology *138*, 225–238.

Camper, S.A., Albers, R.J., Coward, J.K., and Rottman, F.M. (1984). Effect of undermethylation on mRNA cytoplasmic appearance and half-life. Molecular and cellular biology *4*, 538–543.

Cardinale, S., Cisterna, B., Bonetti, P., Aringhieri, C., Biggiogera, M., and Barabino, S.M.L. (2007). Subnuclear localization and dynamics of the Pre-mRNA 3' end processing factor mammalian cleavage factor I 68-kDa subunit. Molecular biology of the cell *18*, 1282–1292.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., Krijgsveld, J., and Hentze, M.W. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell *149*, 1393–1406.

Castello, A., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.-M., Foehr, S., Curk, T., Krijgsveld, J., and Hentze, M.W. (2016). Comprehensive Identification of RNA-Binding Domains in Human Cells. Molecular cell *63*, 696–710.

Castello, A., Fischer, B., Hentze, M.W., and Preiss, T. (2013). RNA-binding proteins in Mendelian disease. Trends in genetics TIG *29*, 318–327.

Cavaloc, Y., Bourgeois, C.F., Kister, L., and Stévenin, J. (1999). The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. RNA *5*, 468–483.

Cavaloc, Y., Popielarz, M., Fuchs, J.P., Gattoni, R., and Stévenin, J. (1994). Characterization and cloning of the human splicing factor 9G8: a novel 35 kDa factor of the serine/arginine protein family. The EMBO journal *13*, 2639–2649.

Chan, S.L., Huppertz, I., Yao, C., Weng, L., Moresco, J.J., Yates, J.R., Ule, J., Manley, J.L., and Shi, Y. (2014). CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. Genes & development *28*, 2370–2380.

Chang, H., Lim, J., Ha, M., and Kim, V.N. (2014). TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. Molecular cell *53*, 1044–1052.

Cheah, M.T., Wachter, A., Sudarsan, N., and Breaker, R.R. (2007). Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. Nature *447*, 497–500.

Chen, C.-Y., Chen, S.-T., Juan, H.-F., and Huang, H.-C. (2012). Lengthening of 3'UTR increases with morphological complexity in animal evolution. Bioinformatics (Oxford, England) *28*, 3178–3181.

Chen, L.-L., and Carmichael, G.G. (2009). Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. Molecular cell *35*, 467–478.

Chen, M., Lyu, G., Han, M., Nie, H., Shen, T., Chen, W., Niu, Y., Song, Y., Li, X., Li, H., Chen, X., Wang, Z., Xia, Z., Li, W., Tian, X.-L., Ding, C., Gu, J., Zheng, Y., Liu, X., Hu, J., Wei, G., Tao, W., and Ni, T. (2018). 3' UTR lengthening as a novel mechanism in regulating cellular senescence. Genome research.

Chen, S., Wang, R., Zheng, D., Zhang, H., Chang, X., Wang, K., Li, W., Fan, J., Tian, B., and Cheng, H. (2019). The mRNA Export Receptor NXF1 Coordinates Transcriptional Dynamics, Alternative Polyadenylation, and mRNA Export. Molecular cell *74*, 118-131.e7.

Chen, W., Jia, Q., Song, Y., Fu, H., Wei, G., and Ni, T. (2017). Alternative Polyadenylation: Methods, Findings, and Impacts. Genomics, proteomics & bioinformatics *15*, 287–300.

Cheng, H., Dufu, K., Lee, C.-S., Hsu, J.L., Dias, A., and Reed, R. (2006a). Human mRNA export machinery recruited to the 5' end of mRNA. Cell *127*, 1389–1400.

Cheng, Y., Miura, R.M., and Tian, B. (2006b). Prediction of mRNA polyadenylation sites by support vector machine. Bioinformatics (Oxford, England) *22*, 2320–2325.

Chiou, N.-T., Shankarling, G., and Lynch, K.W. (2013). hnRNP L and hnRNP A1 induce extended U1 snRNA interactions with an exon to repress spliceosome assembly. Molecular cell *49*, 972–982.

Chu, Y., Elrod, N., Wang, C., Li, L., Chen, T., Routh, A., Xia, Z., Li, W., Wagner, E.J., and Ji, P. (2019). Nudt21 regulates the alternative polyadenylation of Pak1 and is predictive in the prognosis of glioblastoma patients. Oncogene.

Ciolli Mattioli, C., Rom, A., Franke, V., Imami, K., Arrey, G., Terne, M., Woehler, A., Akalin, A., Ulitsky, I., and Chekulaeva, M. (2019). Alternative 3' UTRs direct localization of functionally diverse protein isoforms in neuronal compartments. Nucleic acids research *47*, 2560–2573.

Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A., and Lawrence, J.B. (2009). An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Molecular cell *33*, 717–726.

Clerici, M., Faini, M., Aebersold, R., and Jinek, M. (2017). Structural insights into the assembly and polyA signal recognition mechanism of the human CPSF complex. eLife *6*.

Clerici, M., Faini, M., Muckenfuss, L.M., Aebersold, R., and Jinek, M. (2018). Structural basis of AAUAAA polyadenylation signal recognition by the human CPSF complex. Nature structural & molecular biology *25*, 135–138.

Cléry, A., Blatter, M., and Allain, F.H.-T. (2008). RNA recognition motifs: boring? Not quite. Current opinion in structural biology *18*, 290–298.

Clift, D., McEwan, W.A., Labzin, L.I., Konieczny, V., Mogessie, B., James, L.C., and Schuh, M. (2017). A Method for the Acute and Rapid Degradation of Endogenous Proteins. Cell *171*, 1692-1706.e18.

Conti, L. de, Baralle, M., and Buratti, E. (2013). Exon and intron definition in pre-mRNA splicing. Wiley interdisciplinary reviews. RNA *4*, 49–60.

Corden, J.L. (1990). Tails of RNA polymerase II. Trends in biochemical sciences *15*, 383–387.

Costessi, L., Devescovi, G., Baralle, F.E., and Muro, A.F. (2006). Brain-specific promoter and polyadenylation sites of the beta-adducin pre-mRNA generate an unusually long 3'-UTR. Nucleic acids research *34*, 243–253.

Cougot, N., Babajko, S., and Séraphin, B. (2004). Cytoplasmic foci are sites of mRNA decay in human cells. The Journal of Cell Biology *165*, 31–40.

Cozzarelli, N.R., Gerrard, S.P., Schlissel, M., Brown, D.D., and Bogenhagen, D.F. (1983). Purified RNA polymerase III accurately and efficiently terminates transcription of 5S RNA genes. Cell *34*, 829–835.

Curk, T., Ule, J., Rot, G., Gorup, C., and Konig, J. (2019). iCount: protein-RNA interaction iCLIP data analysis (in preparation), 2019.

Dang, C.V., and Lee, W.M. (1989). Nuclear and nucleolar targeting sequences of c-erb-A, c-myb, N-myc, p53, HSP70, and HIV tat proteins. The Journal of biological chemistry *264*, 18019–18023.

Davidovic, L., Jaglin, X.H., Lepagnol-Bestel, A.-M., Tremblay, S., Simonneau, M., Bardoni, B., and Khandjian, E.W. (2007). The fragile X mental retardation protein is a molecular adaptor between the neurospecific KIF3C kinesin and dendritic RNA granules. Human molecular genetics *16*, 3047–3058.

Davis, M.J., Hanson, K.A., Clark, F., Fink, J.L., Zhang, F., Kasukawa, T., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., and Teasdale, R.D. (2006). Differential use of signal peptides and membrane domains is a common occurrence in the protein output of transcriptional units. PLoS genetics *2*, e46.

DeLigio, J.T., Stevens, S.C., Nazario-Munoz, G.S., MacKnight, H.P., Doe, K.K., Chalfant, C.E., and Park, M.A. (2019). SERINE/ARGININE-RICH SPLICING FACTOR 3 MODULATES THE ALTERNATIVE SPLICING OF CYTOPLASMIC POLYADENYLATION ELEMENT BINDING PROTEIN 2. Molecular cancer research MCR.

Derti, A., Garrett-Engele, P., Macisaac, K.D., Stevens, R.C., Sriram, S., Chen, R., Rohl, C.A., Johnson, J.M., and Babak, T. (2012). A quantitative atlas of polyadenylation in five mammals. Genome research *22*, 1173–1183.

Desterro, J., Bak-Gordon, P., and Carmo-Fonseca, M. (2019). Targeting mRNA processing as an anticancer strategy. Nature reviews. Drug discovery.

Dettwiler, S., Aringhieri, C., Cardinale, S., Keller, W., and Barabino, S.M.L. (2004). Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein-protein interactions, and subcellular localization. The Journal of biological chemistry *279*, 35788–35797.

Devany, E., Park, J.Y., Murphy, M.R., Zakusilo, G., Baquero, J., Zhang, X., Hoque, M., Tian, B., and Kleiman, F.E. (2016). Intronic cleavage and polyadenylation regulates gene expression during DNA damage response through U1 snRNA. Cell discovery *2*, 16013.

Di Giammartino, D.C., Li, W., Ogami, K., Yashinskie, J.J., Hoque, M., Tian, B., and Manley, J.L. (2014). RBBP6 isoforms regulate the human polyadenylation machinery and modulate expression of mRNAs with AU-rich 3' UTRs. Genes & development *28*, 2248–2260.

Di Liddo, A., Oliveira Freitas Machado, C. de, Fischer, S., Ebersberger, S., Heumüller, A.W., Weigand, J.E., Müller-McNicoll, M., and Zarnack, K. (2019). A combined computational pipeline to detect circular RNAs in human cancer cells under hypoxic stress. Journal of molecular cell biology *11*, 829–844.

Dictenberg, J.B., Swanger, S.A., Antar, L.N., Singer, R.H., and Bassell, G.J. (2008). A direct role for FMRP in activity-dependent dendritic mRNA transport links filopodial-spine morphogenesis to fragile X syndrome. Developmental cell *14*, 926–939.

Ding, L., Paszkowski-Rogacz, M., Nitzsche, A., Slabicki, M.M., Heninger, A.-K., Vries, I. de, Kittler, R., Junqueira, M., Shevchenko, A., Schulz, H., Hubner, N., Doss, M.X., Sachinidis, A., Hescheler, J., Iacone, R., Anastassiadis, K., Stewart, A.F., Pisabarro, M.T., Caldarelli, A., Poser, I., Theis, M., and Buchholz, F. (2009). A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. Cell stem cell *4*, 403–415.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G.K., Khatun, J., Williams, B.A., Zaleski, C., Rozowsky, J., Röder, M., Kokocinski, F., Abdelhamid, R.F., Alioto, T., Antoshechkin, I., Baer, M.T., Bar, N.S., Batut, P., Bell, K., Bell, I., Chakrabortty, S., Chen, X., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Falconnet, E., Fastuca, M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M.J., Gao, H., Gonzalez, D., Gordon, A., Gunawardena, H., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Luo, O.J., Park, E., Persaud, K., Preall, J.B., Ribeca, P., Risk, B., Robyr, D., Sammeth, M., Schaffer, L., See, L.-H., Shahab, A., Skancke, J., Suzuki, A.M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Wrobel, J., Yu, Y., Ruan, X., Hayashizaki, Y., Harrow, J., Gerstein, M., Hubbard, T., Reymond, A., Antonarakis, S.E., Hannon, G., Giddings, M.C., Ruan, Y., Wold, B., Carninci, P., Guigó, R., and Gingeras, T.R. (2012). Landscape of transcription in human cells. Nature *489*, 101–108.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England) *29*, 15–21.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., Sorek, R., and Rechavi, G. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201–206.

Dreyfus, M., and Régnier, P. (2002). The Poly(A) Tail of mRNAs. Cell *111*, 611–613.

Du, H., Zhao, Y., He, J., Zhang, Y., Xi, H., Liu, M., Ma, J., and Wu, L. (2016). YTHDF2 destabilizes m(6)A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. Nature communications *7*, 12626.

Du, L., and Warren, S.L. (1997). A functional interaction between the carboxy-terminal domain of RNA polymerase II and pre-mRNA splicing. The Journal of Cell Biology *136*, 5–18.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., Moor, B. de, Brazma, A., and Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics (Oxford, England) *21*, 3439–3440.

Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nature protocols *4*, 1184–1191.

Edmond, V., Moysan, E., Khochbin, S., Matthias, P., Brambilla, C., Brambilla, E., Gazzeri, S., and Eymin, B. (2011). Acetylation and phosphorylation of SRSF2 control cell fate decision in response to cisplatin. The EMBO journal *30*, 510–523.

Egloff, S., O'Reilly, D., and Murphy, S. (2008). Expression of human snRNA genes from beginning to end. Biochemical Society transactions *36*, 590–594.

Eldridge, A.G., Li, Y., Sharp, P.A., and Blencowe, B.J. (1999). The SRm160/300 splicing coactivator is required for exon-enhancer function. Proceedings of the National Academy of Sciences of the United States of America *96*, 6125–6130.

Elkon, R., Drost, J., van Haaften, G., Jenal, M., Schrier, M., Oude Vrielink, J.A.F., and Agami, R. (2012). E2F mediates enhanced alternative polyadenylation in proliferation. Genome Biology *13*, R59.

Escudero-Paunetto, L., Li, L., Hernandez, F.P., and Sandri-Goldin, R.M. (2010). SR proteins SRp20 and 9G8 contribute to efficient export of herpes simplex virus 1 mRNAs. Virology *401*, 155–164.

Fan, J., Kuai, B., Wu, G., Wu, X., Chi, B., Wang, L., Wang, K., Shi, Z., Zhang, H., Chen, S., He, Z., Wang, S., Zhou, Z., Li, G., and Cheng, H. (2017). Exosome cofactor hMTR4 competes with export adaptor ALYREF to ensure balanced nuclear RNA pools for degradation and export. The EMBO journal *36*, 2870–2886.

Finger, L.D., Johansson, C., Rinaldi, B., Bouvet, P., and Feigon, J. (2004). Contributions of the RNA-binding and linker domains and RNA structure to the specificity and affinity of the nucleolin RBD12/NRE interaction. Biochemistry *43*, 6937–6947.

Flavell, S.W., Kim, T.-K., Gray, J.M., Harmin, D.A., Hemberg, M., Hong, E.J., Markenscoff-Papadimitriou, E., Bear, D.M., and Greenberg, M.E. (2008). Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. Neuron *60*, 1022–1038.

Fong, N., and Bentley, D.L. (2001). Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. Genes & development *15*, 1783–1795.

Fox, A.H., Kowalski, K., King, G.F., MacKay, J.P., and Crossley, M. (1998). Key residues characteristic of GATA N-fingers are recognized by FOG. The Journal of biological chemistry *273*, 33595–33603.

Fox, A.H., Liew, C., Holmes, M., Kowalski, K., Mackay, J., and Crossley, M. (1999). Transcriptional cofactors of the FOG family interact with GATA proteins by means of multiple zinc fingers. The EMBO journal *18*, 2812–2822.

Fox-Walsh, K., Davis-Turak, J., Zhou, Y., Li, H., and Fu, X.-D. (2011). A multiplex RNA-seq strategy to profile poly(A+) RNA: application to analysis of transcription response and 3' end formation. Genomics *98*, 266–271.

Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I.T., García Girón, C., Gonzalez, J.M., Grego, T., Hardy, M., Hourlier, T., Hunt, T., Izuogu, O.G., Lagarde, J., Martin, F.J., Martínez, L., Mohanan, S., Muir, P., Navarro, F.C.P., Parker, A., Pei, B., Pozo, F., Ruffier, M., Schmitt, B.M., Stapleton, E., Suner, M.-M., Sycheva, I., Uszczynska-Ratajczak, B., Xu, J., Yates, A., Zerbino, D., Zhang, Y., Aken, B., Choudhary, J.S., Gerstein, M., Guigó, R., Hubbard, T.J.P., Kellis, M., Paten, B., Reymond, A., Tress, M.L., and Flicek, P. (2019). GENCODE reference annotation for the human and mouse genomes. Nucleic acids research *47*, D766-D773.

Fu, X.-D., and Ares, M. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. Nature reviews. Genetics *15*, 689–701.

Fu, Y., and Wang, Y. (2018). SRSF7 knockdown promotes apoptosis of colon and lung cancer cells. Oncology letters *15*, 5545–5552.

Fustin, J.-M., Doi, M., Yamaguchi, Y., Hida, H., Nishimura, S., Yoshida, M., Isagawa, T., Morioka, M.S., Kakeya, H., Manabe, I., and Okamura, H. (2013). RNA-methylation-dependent RNA processing controls the speed of the circadian clock. Cell *155*, 793–806.

Galganski, L., Urbanek, M.O., and Krzyzosiak, W.J. (2017). Nuclear speckles: molecular organization, biological function and role in disease. Nucleic acids research *45*, 10350–10368.

Galiana-Arnoux, D., Lejeune, F., Gesnel, M.-C., Stevenin, J., Breathnach, R., and Del Gatto-Konczak, F. (2003). The CD44 alternative v9 exon contains a splicing enhancer responsive to the SR proteins 9G8, ASF/SF2, and SRp20. The Journal of biological chemistry *278*, 32943–32953.

Garneau, N.L., Wilusz, J., and Wilusz, C.J. (2007). The highways and byways of mRNA decay. Nature reviews. Molecular cell biology *8*, 113–126.

Gautheret, D., Poirot, O., Lopez, F., Audic, S., and Claverie, J.M. (1998). Alternate polyadenylation in human mRNAs: a large-scale analysis by EST clustering. Genome research *8*, 524–530.

Gennarino, V.A., Alcott, C.E., Chen, C.-A., Chaudhury, A., Gillentine, M.A., Rosenfeld, J.A., Parikh, S., Wheless, J.W., Roeder, E.R., Horovitz, D.D.G., Roney, E.K., Smith, J.L., Cheung, S.W., Li, W., Neilson, J.R., Schaaf, C.P., and Zoghbi, H.Y. (2015). NUDT21-spanning CNVs lead to neuropsychiatric disease and altered MeCP2 abundance via alternative polyadenylation. eLife *4*.

Gentleman, R., Carey, V., Huber, W., and Hahne, F. (2019). genefilter: methods for filtering genes from high-throughput experiments. R package version 1.66.0, 2019.

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. Nature reviews. Genetics *15*, 829–845.

Glock, C., Heumüller, M., and Schuman, E.M. (2017). mRNA transport & local translation in neurons. Current opinion in neurobiology *45*, 169–177.

Glover-Cutter, K., Kim, S., Espinosa, J., and Bentley, D.L. (2008). RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. Nature structural & molecular biology *15*, 71–78.

Gong, C., and Maquat, L.E. (2011). lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. Nature *470*, 284–288.

Grassi, E., Mariella, E., Lembo, A., Molineris, I., and Provero, P. (2016). Roar: detecting alternative polyadenylation with standard mRNA sequencing libraries. BMC bioinformatics *17*, 423.

Grassi, E., Santoro, R., Umbach, A., Grosso, A., Oliviero, S., Neri, F., Conti, L., Ala, U., Provero, P., DiCunto, F., and Merlo, G.R. (2018). Choice of Alternative Polyadenylation Sites, Mediated by the RNA-Binding Protein Elavl3, Plays a Role in Differentiation of Inhibitory Neuronal Progenitors. Frontiers in cellular neuroscience *12*, 518.

Graveley, B.R. (2000). Sorting out the complexity of SR protein functions. RNA *6*, 1197–1211.

Gruber, A.J., Schmidt, R., Gruber, A.R., Martin, G., Ghosh, S., Belmadani, M., Keller, W., and Zavolan, M. (2016). A comprehensive analysis of 3' end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. Genome research *26*, 1145–1159.

Gruber, A.J., and Zavolan, M. (2019). Alternative cleavage and polyadenylation in health and disease. Nature reviews. Genetics.

Gruber, A.R., Martin, G., Keller, W., and Zavolan, M. (2012). Cleavage factor Im is a key regulator of 3' UTR length. RNA biology *9*, 1405–1412.

Gruber, A.R., Martin, G., Müller, P., Schmidt, A., Gruber, A.J., Gumienny, R., Mittal, N., Jayachandran, R., Pieters, J., Keller, W., van Nimwegen, E., and Zavolan, M. (2014). Global 3' UTR shortening has a limited effect on protein abundance in proliferating T cells. Nature communications *5*, 5465.

Grummt, I. (1998). Regulation of Mammalian Ribosomal Gene Transcription by RNA Polymerase I. In Progress in Nucleic Acid Research and Molecular Biology - Volume 62, K. Moldave, ed. (Academic Press), pp. 109–154.

Grüter, P., Tabernero, C., Kobbe, C. von, Schmitt, C., Saavedra, C., Bachi, A., Wilm, M., Felber, B.K., and Izaurralde, E. (1998). TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. Molecular Cell *1*, 649–659.

Gunderson, S.I., Polycarpou-Schwarz, M., and Mattaj, I.W. (1998). U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. Molecular Cell *1*, 255–264.

Ha, K.C.H., Blencowe, B.J., and Morris, Q. (2018). QAPA: a new method for the systematic analysis of alternative polyadenylation from RNA-seq data. Genome biology *19*, 45.

Hansen, T.B. (2018). Improved circRNA Identification by Combining Prediction Algorithms. Frontiers in cell and developmental biology *6*, 20.

Hargous, Y., Hautbergue, G.M., Tintaru, A.M., Skrisovska, L., Golovanov, A.P., Stevenin, J., Lian, L.-Y., Wilson, S.A., and Allain, F.H.-T. (2006). Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. The EMBO journal *25*, 5126–5137.

Herrmann, C.J., Schmidt, R., Kanitz, A., Artimo, P., Gruber, A.J., and Zavolan, M. (2020). PolyASite 2.0: a consolidated atlas of polyadenylation sites from 3' end sequencing. Nucleic acids research *48*, D174-D179.

Hilgers, V., Lemke, S.B., and Levine, M. (2012). ELAV mediates 3' UTR extension in the Drosophila nervous system. Genes & development *26*, 2259–2264.

Hill, C.H., Boreikaitė, V., Kumar, A., Casañal, A., Kubík, P., Degliesposti, G., Maslen, S., Mariani, A., Loeffelholz, O. von, Girbig, M., Skehel, M., and Passmore, L.A. (2019). Activation of the Endonuclease that Defines mRNA 3' Ends Requires Incorporation into an 8-Subunit Core Cleavage and Polyadenylation Factor Complex. Molecular cell.

Hirose, Y., and Manley, J.L. (1998). RNA polymerase II is an essential mRNA polyadenylation factor. Nature *395*, 93–96.

Hogg, J.R., and Goff, S.P. (2010). Upf1 senses 3'UTR length to potentiate mRNA decay. Cell *143*, 379–389.

Hoque, M., Ji, Z., Zheng, D., Luo, W., Li, W., You, B., Park, J.Y., Yehia, G., and Tian, B. (2013). Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. Nature methods *10*, 133–139.

Horiuchi, K., Kawamura, T., Iwanari, H., Ohashi, R., Naito, M., Kodama, T., and Hamakubo, T. (2013). Identification of Wilms' tumor 1-associating protein complex and its role in alternative splicing and the cell cycle. The Journal of biological chemistry *288*, 33292–33302.

Howard, J.M., and Sanford, J.R. (2015). The RNAissance family: SR proteins as multifaceted regulators of gene expression. Wiley interdisciplinary reviews. RNA *6*, 93–110.

Hu, G., Kim, J., Xu, Q., Leng, Y., Orkin, S.H., and Elledge, S.J. (2009). A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. Genes & development *23*, 837–848.

Hu, J., Lutz, C.S., Wilusz, J., and Tian, B. (2005). Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. RNA (New York, N.Y.) *11*, 1485–1493.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids research *37*, 1–13.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols *4*, 44–57.

Huang, H., Zhang, J., Harvey, S.E., Hu, X., and Cheng, C. (2017). RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF. Genes & development *31*, 2296–2309.

Huang, Y., Gattoni, R., Stévenin, J., and Steitz, J.A. (2003). SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. Molecular cell *11*, 837–843.

Huang, Y., and Steitz, J.A. (2001). Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. Molecular cell *7*, 899–905.

Huang, Y., Yario, T.A., and Steitz, J.A. (2004). A molecular link between SR protein dephosphorylation and mRNA export. Proceedings of the National Academy of Sciences of the United States of America *101*, 9666–9670.

Hudson, S.W., McNally, L.M., and McNally, M.T. (2016). Evidence that a threshold of serine/arginine-rich (SR) proteins recruits CFIm to promote rous sarcoma virus mRNA 3' end formation. Virology *498*, 181–191.

Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L.E., Sibley, C.R., Sugimoto, Y., Tajnik, M., König, J., and Ule, J. (2014). iCLIP: protein-RNA interactions at nucleotide resolution. Methods (San Diego, Calif.) *65*, 274–287.

Hutchinson, J.N., Ensminger, A.W., Clemson, C.M., Lynch, C.R., Lawrence, J.B., and Chess, A. (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. BMC genomics *8*, 39.

Hwang, H.-W., Park, C.Y., Goodarzi, H., Fak, J.J., Mele, A., Moore, M.J., Saito, Y., and Darnell, R.B. (2016). PAPERCLIP Identifies MicroRNA Targets and a Role of CstF64/64tau in Promoting Non-canonical poly(A) Site Usage. Cell Reports *15*, 423–435.

Hwang, H.-W., Saito, Y., Park, C.Y., Blachère, N.E., Tajima, Y., Fak, J.J., Zucker-Scharff, I., and Darnell, R.B. (2017). cTag-PAPERCLIP Reveals Alternative Polyadenylation Promotes Cell-Type Specific Protein Diversity and Shifts Araf Isoforms with Microglia Activation. Neuron *95*, 1334-1349.e5.

Hwang, J., and Maquat, L.E. (2011). Nonsense-mediated mRNA decay (NMD) in animal embryogenesis: to die or not to die, that is the question. Current opinion in genetics & development *21*, 422–430.

Ip, J.Y., Schmidt, D., Pan, Q., Ramani, A.K., Fraser, A.G., Odom, D.T., and Blencowe, B.J. (2011). Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. Genome research *21*, 390–401.

Jan, C.H., Friedman, R.C., Ruby, J.G., and Bartel, D.P. (2011). Formation, regulation and evolution of Caenorhabditis elegans 3'UTRs. Nature *469*, 97–101.

Jang, S., Cook, N.J., Pye, V.E., Bedwell, G.J., Dudek, A.M., Singh, P.K., Cherepanov, P., and Engelman, A.N. (2019). Differential role for phosphorylation in alternative polyadenylation function versus nuclear import of SR-like protein CPSF6. Nucleic acids research.

Jenal, M., Elkon, R., Loayza-Puch, F., van Haaften, G., Kühn, U., Menzies, F.M., Oude Vrielink, J.A.F., Bos, A.J., Drost, J., Rooijers, K., Rubinsztein, D.C., and Agami, R. (2012). The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. Cell *149*, 538–553.

Jereb, S., Hwang, H.-W., van Otterloo, E., Govek, E.-E., Fak, J.J., Yuan, Y., Hatten, M.E., and Darnell, R.B. (2018). Differential 3' Processing of Specific Transcripts Expands Regulatory and Protein Diversity Across Neuronal Cell Types. eLife *7*.

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.

Ji, Z., Lee, J.Y., Pan, Z., Jiang, B., and Tian, B. (2009). Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. Proceedings of the National Academy of Sciences of the United States of America *106*, 7028–7033.

Ji, Z., Luo, W., Li, W., Hoque, M., Pan, Z., Zhao, Y., and Tian, B. (2011). Transcriptional activity regulates alternative cleavage and polyadenylation. Molecular systems biology *7*, 534.

Ji, Z., and Tian, B. (2009). Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. PloS one *4*, e8419.

Jia, R., Ajiro, M., Yu, L., McCoy, P., and Zheng, Z.-M. (2019). Oncogenic splicing factor SRSF3 regulates ILF3 alternative splicing to promote cancer cell proliferation and transformation. RNA (New York, N.Y.).

Jia, R., Liu, X., Tao, M., Kruhlak, M., Guo, M., Meyers, C., Baker, C.C., and Zheng, Z.-M. (2009). Control of the papillomavirus early-to-late switch by differentially expressed SRp20. Journal of virology *83*, 167–180.

Jiang, L., Huang, J., Higgs, B.W., Hu, Z., Xiao, Z., Yao, X., Conley, S., Zhong, H., Liu, Z., Brohawn, P., Shen, D., Wu, S., Ge, X., Jiang, Y., Zhao, Y., Lou, Y., Morehouse, C., Zhu, W., Sebastian, Y., Czapiga, M., Oganesyan, V., Fu, H., Niu, Y., Zhang, W., Streicher, K., Tice, D., Zhao, H., Zhu, M., Xu, L., Herbst, R., Su, X., Gu, Y., Li, S., Huang, L., Gu, J., Han, B., Jallal, B., Shen, H., and Yao, Y. (2016). Genomic Landscape Survey Identifies SRSF1 as a Key Oncodriver in Small Cell Lung Cancer. PLoS genetics *12*, e1005895.

Jin, Y., Yang, Y., and Zhang, P. (2011). New insights into RNA secondary structure in the alternative splicing of pre-mRNAs. RNA biology *8*, 450–457.

Jones-Villeneuve, E.M., Rudnicki, M.A., Harris, J.F., and McBurney, M.W. (1983). Retinoic acid-induced neural differentiation of embryonal carcinoma cells. Molecular and cellular biology *3*, 2271–2279.

Jorda, J., Xue, B., Uversky, V.N., and Kajava, A.V. (2010). Protein tandem repeats - the more perfect, the less structured. The FEBS journal *277*, 2673–2682.

Jumaa, H., and Nielsen, P.J. (2000). Regulation of SRp20 exon 4 splicing. Biochimica et biophysica acta *1494*, 137–143.

Jungmann, R., Avendaño, M.S., Dai, M., Woehrstein, J.B., Agasti, S.S., Feiger, Z., Rodal, A., and Yin, P. (2016). Quantitative super-resolution imaging with qPAINT. Nature methods *13*, 439–442.

Jungmann, R., Avendaño, M.S., Woehrstein, J.B., Dai, M., Shih, W.M., and Yin, P. (2014). Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. Nature methods *11*, 313–318.

Kadota, Y., Jam, F.A., Yukiue, H., Terakado, I., Morimune, T., Tano, A., Tanaka, Y., Akahane, S., Fukumura, M., Tooyama, I., and Mori, M. (2020). Srsf7 Establishes the Juvenile Transcriptome through Age-Dependent Alternative Splicing in Mice. iScience *23*, 101242.

Kahles, A., Lehmann, K.-V., Toussaint, N.C., Hüser, M., Stark, S.G., Sachsenberg, T., Stegle, O., Kohlbacher, O., Sander, C., and Rätsch, G. (2018). Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. Cancer cell *34*, 211-224.e6.

Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., and Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature *468*, 664–668.

Kamieniarz-Gdula, K., Gdula, M.R., Panser, K., Nojima, T., Monks, J., Wiśniewski, J.R., Riepsaame, J., Brockdorff, N., Pauli, A., and Proudfoot, N.J. (2019). Selective Roles of Vertebrate PCF11 in Premature and Full-Length Transcript Termination. Molecular cell.

Karni, R., Stanchina, E. de, Lowe, S.W., Sinha, R., Mu, D., and Krainer, A.R. (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. Nature structural & molecular biology *14*, 185–193.

Kasowitz, S.D., Ma, J., Anderson, S.J., Leu, N.A., Xu, Y., Gregory, B.D., Schultz, R.M., and Wang, P.J. (2018). Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. PLoS genetics *14*, e1007412.

Katahira, J., Okuzaki, D., Inoue, H., Yoneda, Y., Maehara, K., and Ohkawa, Y. (2013). Human TREX component Thoc5 affects alternative polyadenylation site choice by recruiting mammalian cleavage factor I. Nucleic acids research *41*, 7060–7072.

Kataoka, N., Bachorik, J.L., and Dreyfuss, G. (1999). Transportin-SR, a nuclear import receptor for SR proteins. The Journal of Cell Biology *145*, 1145–1152.

Katz, Y., Wang, E. T., Stilterra, J., Schwartz, S., Wong, B., Thorvaldsdóttir, H., Robinson, J. T., Mesirov, J. P., Airoldi, E. M., & Burge, C. B. (2014). Sashimi plots: Quantitative visualization of alternative isoform expression from RNA-seq data.

Katz, Y., Wang, E.T., Airoldi, E.M., and Burge, C.B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nature methods *7*, 1009–1015.

Kaufmann, I., Martin, G., Friedlein, A., Langen, H., and Keller, W. (2004). Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. The EMBO journal *23*, 616–626.

Ke, S., Alemu, E.A., Mertens, C., Gantman, E.C., Fak, J.J., Mele, A., Haripal, B., Zucker-Scharff, I., Moore, M.J., Park, C.Y., Vågbø, C.B., Kusśnierczyk, A., Klungland, A., Darnell, J.E., and Darnell, R.B. (2015). A majority of m6A residues are in the last exons, allowing the potential for 3' UTR regulation. Genes & development *29*, 2037–2053.

Keller, W., Bienroth, S., Lang, K.M., and Christofori, G. (1991). Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA. The EMBO journal *10*, 4241–4249.

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 research *12*, 996–1006.

Kielkopf, C.L., Rodionova, N.A., Green, M.R., and Burley, S.K. (2001). A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. Cell *106*, 595–605.

Kiledjian, M., and Dreyfuss, G. (1992). Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. The EMBO journal *11*, 2655–2664.

Kim, E., Du, L., Bregman, D.B., and Warren, S.L. (1997). Splicing factors associate with hyperphosphorylated RNA polymerase II in the absence of pre-mRNA. The Journal of cell biology *136*, 19–28.

Kim, J., Park, R.Y., Chen, J.-K., Jeong, S., and Ohn, T. (2014). Splicing factor SRSF3 represses the translation of programmed cell death 4 mRNA by associating with the 5'-UTR region. Cell death and differentiation *21*, 481–490.

Kim, S., Yamamoto, J., Chen, Y., Aida, M., Wada, T., Handa, H., and Yamaguchi, Y. (2010). Evidence that cleavage factor Im is a heterotetrameric protein complex controlling alternative polyadenylation. Genes to cells devoted to molecular & cellular mechanisms *15*, 1003–1013.

Kim, Y.K., and Maquat, L.E. (2019). UPFront and center in RNA decay: UPF1 in nonsense-mediated mRNA decay and beyond. RNA *25*, 407–422.

Kishi, J.Y., Lapan, S.W., Beliveau, B.J., West, E.R., Zhu, A., Sasaki, H.M., Saka, S.K., Wang, Y., Cepko, C.L., and Yin, P. (2019). SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. Nature methods *16*, 533–544.

Kittler, R., Heninger, A.-K., Franke, K., Habermann, B., and Buchholz, F. (2005). Production of endoribonuclease-prepared short interfering RNAs for gene silencing in mammalian cells. Nature methods *2*, 779–784.

Klann, K., Tascher, G., and Münch, C. (2020). Functional Translatome Proteomics Reveal Converging and Dose-Dependent Regulation by mTORC1 and eIF2α. Molecular cell *77*, 913-925.e4.

Klerk, E. de, Venema, A., Anvar, S.Y., Goeman, J.J., Hu, O., Trollet, C., Dickson, G., den Dunnen, J.T., van der Maarel, S.M., Raz, V., and Hoen, P.A.C. 't (2012). Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. Nucleic acids research *40*, 9089–9101.

Kohtz, J.D., Jamison, S.F., Will, C.L., Zuo, P., Lührmann, R., Garcia-Blanco, M.A., and Manley, J.L. (1994). Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. Nature *368*, 119–124.

König, 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. Nature structural & molecular biology *17*, 909–915.

Königs, V., Machado Freitas, C.O., Arnold, B.L., Blümel, N., Solovyeva, A., Ruiz De Los Mozos, I., Wittig, I., McNicoll, F., Schulz, M.H., and Müller-McNicoll, M. (2019). SR proteins maintain protein homeostasis through translation of separate protein domains, generation of intron-retained arcRNAs and formation of nuclear bodies.

Königs, V., Oliveira Freitas Machado, C. de, Arnold, B., Blümel, N., Solovyeva, A., Löbbert, S., Schafranek, M., Ruiz de Los Mozos, I., Wittig, I., McNicoll, F., Schulz, M.H., and Müller-McNicoll, M. (2020). SRSF7 maintains its homeostasis through the expression of Split-ORFs and nuclear body assembly. Nature structural & molecular biology *27*, 260–273.

Kota, V., Sommer, G., Hazard, E.S., Hardiman, G., Twiss, J.L., and Heise, T. (2018). SUMO Modification of the RNA-Binding Protein La Regulates Cell Proliferation and STAT3 Protein Stability. Molecular and cellular biology *38*.

Krämer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. Annual review of biochemistry *65*, 367–409.

Kristensen, L.S., Hansen, T.B., Venø, M.T., and Kjems, J. (2018). Circular RNAs in cancer: opportunities and challenges in the field. Oncogene *37*, 555–565.

Kuhn, A., and Grummt, I. (1989). 3'-end formation of mouse pre-rRNA involves both transcription termination and a specific processing reaction. Genes & development *3*, 224–231.

Kühn, U., Gündel, M., Knoth, A., Kerwitz, Y., Rüdel, S., and Wahle, E. (2009). Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor. The Journal of biological chemistry *284*, 22803–22814.

Kühn, U., and Wahle, E. (2004). Structure and function of poly(A) binding proteins. Biochimica et biophysica acta *1678*, 67–84.

Kurosaki, T., Popp, M.W., and Maquat, L.E. (2019). Publisher Correction: Quality and quantity control of gene expression by nonsense-mediated mRNA decay. Nature reviews. Molecular cell biology *20*, 384.

Kyburz, A., Friedlein, A., Langen, H., and Keller, W. (2006). Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing. Molecular Cell *23*, 195–205.

La Mata, M. de, Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. Molecular cell *12*, 525–532.

La Mata, M. de, and Kornblihtt, A.R. (2006). RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. Nature structural & molecular biology *13*, 973–980.

Lackford, B., Yao, C., Charles, G.M., Weng, L., Zheng, X., Choi, E.-A., Xie, X., Wan, J., Xing, Y., Freudenberg, J.M., Yang, P., Jothi, R., Hu, G., and Shi, Y. (2014). Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. The EMBO journal *33*, 878–889.

Lai, M.C., Lin, R.I., Huang, S.Y., Tsai, C.W., and Tarn, W.Y. (2000). A human importin-beta family protein, transportin-SR2, interacts with the phosphorylated RS domain of SR proteins. The Journal of biological chemistry *275*, 7950–7957.

Lai, M.C., Lin, R.I., and Tarn, W.Y. (2001). Transportin-SR2 mediates nuclear import of phosphorylated SR proteins. Proceedings of the National Academy of Sciences of the United States of America *98*, 10154–10159.

Lai, M.-C., and Tarn, W.-Y. (2004). Hypophosphorylated ASF/SF2 binds TAP and is present in messenger ribonucleoproteins. The Journal of biological chemistry *279*, 31745–31749.

Lamichhane, R., Daubner, G.M., Thomas-Crusells, J., Auweter, S.D., Manatschal, C., Austin, K.S., Valniuk, O., Allain, F.H.-T., and Rueda, D. (2010). RNA looping by PTB: Evidence using FRET and NMR spectroscopy for a role in splicing repression. Proceedings of the National Academy of Sciences of the United States of America *107*, 4105–4110.

Lang, W.H., and Reeder, R.H. (1995). Transcription termination of RNA polymerase I due to a T-rich element interacting with Reb1p. Proceedings of the National Academy of Sciences of the United States of America *92*, 9781–9785.

Langmead, B. (2010). Aligning short sequencing reads with Bowtie. Current protocols in bioinformatics *Chapter 11*, Unit 11.7.

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.

Lawrence, M., Gentleman, R., and Carey, V. (2009). rtracklayer: an R package for interfacing with genome browsers. Bioinformatics (Oxford, England) *25*, 1841–1842.

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. PLoS computational biology *9*, e1003118.

Leclair, N.K., Brugiolo, M., Urbanski, L., Lawson, S.C., Thakar, K., Yurieva, M., George, J., Hinson, J.T., Cheng, A., Graveley, B.R., and Anczuków, O. (2020). Poison Exon Splicing Regulates a Coordinated Network of SR Protein Expression during Differentiation and Tumorigenesis. Molecular cell *80*, 648-665.e9.

Lee, J.Y., Yeh, I., Park, J.Y., and Tian, B. (2007). PolyA_DB 2: mRNA polyadenylation sites in vertebrate genes. Nucleic acids research *35*, D165-8.

Lee, S.-H., Singh, I., Tisdale, S., Abdel-Wahab, O., Leslie, C.S., and Mayr, C. (2018). Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. Nature *561*, 127–131.

Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H., and Kim, V.N. (2004). MicroRNA genes are transcribed by RNA polymerase II. The EMBO journal *23*, 4051–4060.

Lee, Y., and Rio, D.C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annual review of biochemistry *84*, 291–323.

Legnini, I., Alles, J., Karaiskos, N., Ayoub, S., and Rajewsky, N. (2019). FLAM-seq: full-length mRNA sequencing reveals principles of poly(A) tail length control. Nature methods.

Lemay, J.-F., Lemieux, C., St-André, O., and Bachand, F. (2010). Crossing the borders: poly(A)-binding proteins working on both sides of the fence. RNA biology *7*, 291–295.

Lesbirel, S., Viphakone, N., Parker, M., Parker, J., Heath, C., Sudbery, I., and Wilson, S.A. (2018). The m6A-methylase complex recruits TREX and regulates mRNA export. Scientific reports *8*, 13827.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England) *25*, 2078–2079.

Li, W., You, B., Hoque, M., Zheng, D., Luo, W., Ji, Z., Park, J.Y., Gunderson, S.I., Kalsotra, A., Manley, J.L., and Tian, B. (2015). Systematic profiling of poly(A)+ transcripts modulated by core 3' end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. PLoS genetics *11*, e1005166.

Li, Y., Chen, Z.Y., Wang, W., Baker, C.C., and Krug, R.M. (2001a). The 3'-end-processing factor CPSF is required for the splicing of single-intron pre-mRNAs in vivo. RNA (New York, N.Y.) *7*, 920–931.

Li, Z., Zhang, Y., Ku, L., Wilkinson, K.D., Warren, S.T., and Feng, Y. (2001b). The fragile X mental retardation protein inhibits translation via interacting with mRNA. Nucleic acids research *29*, 2276–2283.

Lianoglou, S., Garg, V., Yang, J.L., Leslie, C.S., and Mayr, C. (2013). Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. Genes & development *27*, 2380–2396.

Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., Darnell, J.C., and Darnell, R.B. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature *456*, 464–469.

Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., and Fu, X.-D. (2008). The splicing factor SC35 has an active role in transcriptional elongation. Nature structural & molecular biology *15*, 819–826.

Lin, S., Xiao, R., Sun, P., Xu, X., and Fu, X.-D. (2005). Dephosphorylation-dependent sorting of SR splicing factors during mRNP maturation. Molecular cell *20*, 413–425.

Linder, B., Grozhik, A.V., Olarerin-George, A.O., Meydan, C., Mason, C.E., and Jaffrey, S.R. (2015). Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. Nature methods *12*, 767–772.

Linder, P., and Jankowsky, E. (2011). From unwinding to clamping - the DEAD box RNA helicase family. Nature reviews. Molecular cell biology *12*, 505–516.

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., Dai, Q., Chen, W., and He, C. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nature chemical biology *10*, 93–95.

Liu, K.J., and Harland, R.M. (2005). Inhibition of neurogenesis by SRp38, a neuroD-regulated RNA-binding protein. Development (Cambridge, England) *132*, 1511–1523.

Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., and Pan, T. (2015). N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. Nature *518*, 560–564.

Liu, N., Zhou, K.I., Parisien, M., Dai, Q., Diatchenko, L., and Pan, T. (2017). N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. Nucleic acids research *45*, 6051–6063.

Liu, Z., and Huang, Y. (2014). Advantages of proteins being disordered. Protein science a publication of the Protein Society *23*, 539–550.

Long, J.C., and Caceres, J.F. (2009). The SR protein family of splicing factors: master regulators of gene expression. The Biochemical journal *417*, 15–27.

Long, Y., Sou, W.H., Yung, K.W.Y., Liu, H., Wan, S.W.C., Li, Q., Zeng, C., Law, C.O.K., Chan, G.H.C., Lau, T.C.K., and Ngo, J.C.K. (2018). Distinct mechanisms govern the phosphorylation of different SR protein splicing factors. The Journal of biological chemistry.

Longman, D., McGarvey, T., McCracken, S., Johnstone, I.L., Blencowe, B.J., and Cáceres, J.F. (2001). Multiple interactions between SRm160 and SR family proteins in enhancer-dependent splicing and development of C. elegans. Current biology CB *11*, 1923–1933.

Lou, H., Neugebauer, K.M., Gagel, R.F., and Berget, S.M. (1998). Regulation of alternative polyadenylation by U1 snRNPs and SRp20. Molecular and cellular biology *18*, 4977–4985.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology *15*, 550.

Lukong, K.E., Chang, K., Khandjian, E.W., and Richard, S. (2008). RNA-binding proteins in human genetic disease. Trends in genetics TIG *24*, 416–425.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nature reviews. Molecular cell biology *8*, 479–490.

Ma, C.-T., Ghosh, G., Fu, X.-D., and Adams, J.A. (2010). Mechanism of Dephosphorylation of the SR Protein ASF/SF2 By Protein Phosphatase 1. Journal of molecular biology *403*, 386–404.

Maciolek, N.L., and McNally, M.T. (2007). Serine/arginine-rich proteins contribute to negative regulator of splicing element-stimulated polyadenylation in rous sarcoma virus. Journal of virology *81*, 11208–11217.

Maciolek, N.L., and McNally, M.T. (2008). Characterization of Rous sarcoma virus polyadenylation site use in vitro. Virology *374*, 468–476.

MacKay, J.P., and Crossley, M. (1998). Zinc fingers are sticking together. Trends in biochemical sciences *23*, 1–4.

Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., and Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic acids research.

Maeder, C., Kutach, A.K., and Guthrie, C. (2008). ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C-terminus of Prp8. Nature structural & molecular biology *16*, 42–48.

Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., and Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. Nature *444*, 953–956.

Manley, J.L., and Krainer, A.R. (2010). A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). Genes & development *24*, 1073–1074.

Martin, G., Gruber, A.R., Keller, W., and Zavolan, M. (2012). Genome-wide analysis of pre-mRNA 3' end processing reveals a decisive role of human cleavage factor I in the regulation of 3' UTR length. Cell reports *1*, 753–763.

Martin, G., Ostareck-Lederer, A., Chari, A., Neuenkirchen, N., Dettwiler, S., Blank, D., Rüegsegger, U., Fischer, U., and Keller, W. (2010). Arginine methylation in subunits of mammalian pre-mRNA cleavage factor I. RNA (New York, N.Y.) *16*, 1646–1659.

Martin, K.C., and Ephrussi, A. (2009). mRNA localization: gene expression in the spatial dimension. Cell *136*, 719–730.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10.

Martin, R.M., Ter-Avetisyan, G., Herce, H.D., Ludwig, A.K., Lättig-Tünnemann, G., and Cardoso, M.C. (2015). Principles of protein targeting to the nucleolus. Nucleus (Austin, Tex.) *6*, 314–325.

Martinson, H.G. (2011). An active role for splicing in 3'-end formation. Wiley interdisciplinary reviews. RNA *2*, 459–470.

Masamha, C.P., Xia, Z., Yang, J., Albrecht, T.R., Li, M., Shyu, A.-B., Li, W., and Wagner, E.J. (2014). CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. Nature *510*, 412–416.

Maslon, M.M., Heras, S.R., Bellora, N., Eyras, E., and Cáceres, J.F. (2014). The translational landscape of the splicing factor SRSF1 and its role in mitosis. eLife, e02028.

Masuda, A., Takeda, J., Okuno, T., Okamoto, T., Ohkawara, B., Ito, M., Ishigaki, S., Sobue, G., and Ohno, K. (2015). Position-specific binding of FUS to nascent RNA regulates mRNA length. Genes & development *29*, 1045–1057.

Mata, J. (2013). Genome-wide mapping of polyadenylation sites in fission yeast reveals widespread alternative polyadenylation. RNA biology *10*, 1407–1414.

Matthews, J.M., Kowalski, K., Liew, C.K., Sharpe, B.K., Fox, A.H., Crossley, M., and MacKay, J.P. (2000). A class of zinc fingers involved in protein-protein interactions biophysical characterization of CCHC fingers from fog and U-shaped. European Journal of Biochemistry *267*, 1030–1038.

Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A.V., Patil, D.P., Linder, B., Pickering, B.F., Vasseur, J.-J., Chen, Q., Gross, S.S., Elemento, O., Debart, F., Kiledjian, M., and Jaffrey, S.R. (2017). Reversible methylation of m6Am in the 5' cap controls mRNA stability. Nature *541*, 371–375.

Mayr, C. (2016). Evolution and Biological Roles of Alternative 3'UTRs. Trends in cell biology *26*, 227–237.

Mayr, C. (2018). Protein complexes assemble as they are being made. Nature *561*, 186–187.

Mayr, C., and Bartel, D.P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell *138*, 673–684.

Mazroui, R., Huot, M.-E., Tremblay, S., Filion, C., Labelle, Y., and Khandjian, E.W. (2002). Trapping of messenger RNA by Fragile X Mental Retardation protein into cytoplasmic granules induces translation repression. Human molecular genetics *11*, 3007–3017.

Mbita, Z., Meyer, M., Skepu, A., Hosie, M., Rees, J., and Dlamini, Z. (2012). De-regulation of the RBBP6 isoform 3/DWNN in human cancers. Molecular and cellular biochemistry *362*, 249–262.

McBurney, M.W., Jones-Villeneuve, E.M., Edwards, M.K., and Anderson, P.J. (1982). Control of muscle and neuronal differentiation in a cultured embryonal carcinoma cell line. Nature *299*, 165–167.

McBurney, M.W., and Rogers, B.J. (1982). Isolation of male embryonal carcinoma cells and their chromosome replication patterns. Developmental Biology *89*, 503–508.

McCracken, S., Lambermon, M., and Blencowe, B.J. (2002). SRm160 splicing coactivator promotes transcript 3'-end cleavage. Molecular and cellular biology *22*, 148–160.

McCracken, S., Longman, D., Johnstone, I.L., Cáceres, J.F., and Blencowe, B.J. (2003). An evolutionarily conserved role for SRm160 in 3'-end processing that functions independently of exon junction complex formation. The Journal of biological chemistry *278*, 44153–44160.

McGlincy, N.J., Tan, L.-Y., Paul, N., Zavolan, M., Lilley, K.S., and Smith, C.W.J. (2010). Expression proteomics of UPF1 knockdown in HeLa cells reveals autoregulation of hnRNP A2/B1 mediated by alternative splicing resulting in nonsense-mediated mRNA decay. BMC genomics *11*, 565.

McGuire, A.M., Pearson, M.D., Neafsey, D.E., and Galagan, J.E. (2008). Cross-kingdom patterns of alternative splicing and splice recognition. Genome Biology *9*, R50.

McNally, L.M., and McNally, M.T. (1996). SR protein splicing factors interact with the Rous sarcoma virus negative regulator of splicing element. Journal of virology *70*, 1163–1172.

McNally, L.M., and McNally, M.T. (1998). An RNA splicing enhancer-like sequence is a component of a splicing inhibitor element from Rous sarcoma virus. Molecular and cellular biology *18*, 3103–3111.

Mermoud, J.E., Cohen, P.T., and Lamond, A.I. (1994). Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. The EMBO journal *13*, 5679–5688.

Meyer, K.D., and Jaffrey, S.R. (2017). Rethinking m6A Readers, Writers, and Erasers. Annual review of cell and developmental biology *33*, 319–342.

Meyer, K.D., Patil, D.P., Zhou, J., Zinoviev, A., Skabkin, M.A., Elemento, O., Pestova, T.V., Qian, S.-B., and Jaffrey, S.R. (2015). 5' UTR m(6)A Promotes Cap-Independent Translation. Cell *163*, 999–1010.

Meyer, S., Temme, C., and Wahle, E. (2004). Messenger RNA turnover in eukaryotes: pathways and enzymes. Critical reviews in biochemistry and molecular biology *39*, 197–216.

Michlewski, G., Sanford, J.R., and Cáceres, J.F. (2008). The splicing factor SF2/ASF regulates translation initiation by enhancing phosphorylation of 4E-BP1. Molecular cell *30*, 179–189.

Middleton, S.A., Eberwine, J., and Kim, J. (2019). Comprehensive catalog of dendritically localized mRNA isoforms from sub-cellular sequencing of single mouse neurons. BMC biology *17*, 5.

Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S.Z., Keller, W., Antoniou, M., and Vagner, S. (2006). An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries. The EMBO journal *25*, 4854–4864.

Millhouse, S., and Manley, J.L. (2005). The C-terminal domain of RNA polymerase II functions as a phosphorylation-dependent splicing activator in a heterologous protein. Molecular and cellular biology *25*, 533–544.

Minvielle-Sebastia, L., Preker, P.J., Wiederkehr, T., Strahm, Y., and Keller, W. (1997). The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3'-end formation. Proceedings of the National Academy of Sciences of the United States of America *94*, 7897–7902.

Misteli, T., Cáceres, J.F., Clement, J.Q., Krainer, A.R., Wilkinson, M.F., and Spector, D.L. (1998). Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. The Journal of Cell Biology *143*, 297–307.

Misteli, T., Cáceres, J.F., and Spector, D.L. (1997). The dynamics of a pre-mRNA splicing factor in living cells. Nature *387*, 523–527.

Miura, P., Shenker, S., Andreu-Agullo, C., Westholm, J.O., and Lai, E.C. (2013). Widespread and extensive lengthening of 3' UTRs in the mammalian brain. Genome research *23*, 812–825.

Mo, S., Ji, X., and Fu, X.-D. (2013). Unique role of SRSF2 in transcription activation and diverse functions of the SR and hnRNP proteins in gene expression regulation. Transcription *4*, 251–259.

Mohan, N., Kumar, V., Kandala, D.T., Kartha, C.C., and Laishram, R.S. (2018). A Splicing-Independent Function of RBM10 Controls Specific 3' UTR Processing to Regulate Cardiac Hypertrophy. Cell Reports *24*, 3539–3553.

Molinie, B., Wang, J., Lim, K.S., Hillebrand, R., Lu, Z.-X., van Wittenberghe, N., Howard, B.D., Daneshvar, K., Mullen, A.C., Dedon, P., Xing, Y., and Giallourakis, C.C. (2016). m(6)A-LAIC-seq reveals the census and complexity of the m(6)A epitranscriptome. Nature methods *13*, 692–698.

Moore, M.J., and Sharp, P.A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. Nature *365*, 364–368.

Müller, S., Rycak, L., Afonso-Grunz, F., Winter, P., Zawada, A.M., Damrath, E., Scheider, J., Schmäh, J., Koch, I., Kahl, G., and Rotter, B. (2014). APADB: a database for alternative polyadenylation and microRNA regulation events. Database the journal of biological databases and curation *2014*.

Müller-McNicoll, M., Botti, V., Jesus Domingues, A.M. de, 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 & development *30*, 553–566.

Müller-McNicoll, M., and Neugebauer, K.M. (2013). How cells get the message: dynamic assembly and function of mRNA-protein complexes. Nature reviews. Genetics *14*, 275–287.

Müller-McNicoll, M., Rossbach, O., Hui, J., and Medenbach, J. (2019). Auto-regulatory feedback by RNA-binding proteins. Journal of molecular cell biology *11*, 930–939.

Mure, F., Corbin, A., Benbahouche, N.E.H., Bertrand, E., Manet, E., and Gruffat, H. (2018). The splicing factor SRSF3 is functionally connected to the nuclear RNA exosome for intronless mRNA decay. Scientific reports *8*, 12901.

Naganuma, T., Nakagawa, S., Tanigawa, A., Sasaki, Y.F., Goshima, N., and Hirose, T. (2012). Alternative 3'-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. The EMBO journal *31*, 4020–4034.

Nagy, E., and Maquat, L.E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends in biochemical sciences *23*, 198–199.

Nakayama, Y., Wada, A., Inoue, R., Terasawa, K., Kimura, I., Nakamura, N., and Kurosaka, A. (2014). A rapid and efficient method for neuronal induction of the P19 embryonic carcinoma cell line. Journal of neuroscience methods *227*, 100–106.

Nakielny, S., and Dreyfuss, G. (1996). The hnRNP C proteins contain a nuclear retention sequence that can override nuclear export signals. The Journal of Cell Biology *134*, 1365–1373.

Nam, J.-W., Rissland, O.S., Koppstein, D., Abreu-Goodger, C., Jan, C.H., Agarwal, V., Yildirim, M.A., Rodriguez, A., and Bartel, D.P. (2014). Global analyses of the effect of different cellular contexts on microRNA targeting. Molecular cell *53*, 1031–1043.

Nasim, F.-U.H., Hutchison, S., Cordeau, M., and Chabot, B. (2002). High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5' splice site selection in support of a common looping out and repression mechanism. RNA *8*, 1078–1089.

Neugebauer, K.M., Stolk, J.A., and Roth, M.B. (1995). A conserved epitope on a subset of SR proteins defines a larger family of Pre-mRNA splicing factors. The Journal of cell biology *129*, 899–908.

Neve, J., Burger, K., Li, W., Hoque, M., Patel, R., Tian, B., Gullerova, M., and Furger, A. (2016). Subcellular RNA profiling links splicing and nuclear DICER1 to alternative cleavage and polyadenylation. Genome research *26*, 24–35.

Neve, J., Patel, R., Wang, Z., Louey, A., and Furger, A.M. (2017). Cleavage and polyadenylation: Ending the message expands gene regulation. RNA biology *14*, 865–890.

Ng, P., Tan, J.J.S., Ooi, H.S., Lee, Y.L., Chiu, K.P., Fullwood, M.J., Srinivasan, K.G., Perbost, C., Du, L., Sung, W.-K., Wei, C.-L., and Ruan, Y. (2006). Multiplex sequencing of paired-end ditags (MS-PET): a strategy for the ultra-high-throughput analysis of transcriptomes and genomes. Nucleic acids research *34*, e84.

Ng, P., Wei, C.-L., Sung, W.-K., Chiu, K.P., Lipovich, L., Ang, C.C., Gupta, S., Shahab, A., Ridwan, A., Wong, C.H., Liu, E.T., and Ruan, Y. (2005). Gene identification signature (GIS) analysis for transcriptome characterization and genome annotation. Nature methods *2*, 105–111.

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. (2007). Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes & development *21*, 708–718.

Ni, T., Yang, Y., Hafez, D., Yang, W., Kiesewetter, K., Wakabayashi, Y., Ohler, U., Peng, W., and Zhu, J. (2013). Distinct polyadenylation landscapes of diverse human tissues revealed by a modified PA-seq strategy. BMC genomics *14*, 615.

Nishikura, K. (2010). Functions and regulation of RNA editing by ADAR deaminases. Annual review of biochemistry *79*, 321–349.

Niwa, M., Rose, S.D., and Berget, S.M. (1990). In vitro polyadenylation is stimulated by the presence of an upstream intron. Genes & development *4*, 1552–1559.

Nunes, N.M., Li, W., Tian, B., and Furger, A. (2010). A functional human Poly(A) site requires only a potent DSE and an A-rich upstream sequence. The EMBO journal *29*, 1523–1536.

Nussbacher, J.K., Tabet, R., Yeo, G.W., and Lagier-Tourenne, C. (2019). Disruption of RNA Metabolism in Neurological Diseases and Emerging Therapeutic Interventions. Neuron *102*, 294–320.

Ogorodnikov, A., Levin, M., Tattikota, S., Tokalov, S., Hoque, M., Scherzinger, D., Marini, F., Poetsch, A., Binder, H., Macher-Göppinger, S., Probst, H.C., Tian, B., Schaefer, M., Lackner, K.J., Westermann, F., and Danckwardt, S. (2018). Transcriptome 3'end organization by PCF11 links alternative polyadenylation to formation and neuronal differentiation of neuroblastoma. Nature communications *9*, 5331.

Ozsolak, F., Kapranov, P., Foissac, S., Kim, S.W., Fishilevich, E., Monaghan, A.P., John, B., and Milos, P.M. (2010). Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. Cell *143*, 1018–1029.

Pan, Z., Zhang, H., Hague, L.K., Lee, J.Y., Lutz, C.S., and Tian, B. (2006). An intronic polyadenylation site in human and mouse CstF-77 genes suggests an evolutionarily conserved regulatory mechanism. Gene *366*, 325–334.

Pandit, S., Zhou, Y., Shiue, L., Coutinho-Mansfield, G., Li, H., Qiu, J., Huang, J., Yeo, G.W., Ares, M., and Fu, X.-D. (2013). Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. Molecular cell *50*, 223–235.

Park, H.J., Ji, P., Kim, S., Xia, Z., Rodriguez, B., Li, L., Su, J., Chen, K., Masamha, C.P., Baillat, D., Fontes-Garfias, C.R., Shyu, A.-B., Neilson, J.R., Wagner, E.J., and Li, W. (2018). 3' UTR shortening represses tumor-suppressor genes in trans by disrupting ceRNA crosstalk. Nature genetics *50*, 783–789.

Park, J.W., Tokheim, C., Shen, S., and Xing, Y. (2013). Identifying differential alternative splicing events from RNA sequencing data using RNASeq-MATS. Methods in molecular biology (Clifton, N.J.) *1038*, 171–179.

Park, S.K., and Jeong, S. (2016). SRSF3 represses the expression of PDCD4 protein by coordinated regulation of alternative splicing, export and translation. Biochemical and biophysical research communications *470*, 431–438.

Parker, R., and Song, H. (2004). The enzymes and control of eukaryotic mRNA turnover. Nature structural & molecular biology *11*, 121–127.

Patil, D.P., Chen, C.-K., Pickering, B.F., Chow, A., Jackson, C., Guttman, M., and Jaffrey, S.R. (2016). m(6)A RNA methylation promotes XIST-mediated transcriptional repression. Nature *537*, 369–373.

Pelka, G.J., Watson, C.M., Christodoulou, J., and Tam, P.P.L. (2005). Distinct expression profiles of Mecp2 transcripts with different lengths of 3'UTR in the brain and visceral organs during mouse development. Genomics *85*, 441–452.

Pendleton, K.E., Chen, B., Liu, K., Hunter, O.V., Xie, Y., Tu, B.P., and Conrad, N.K. (2017). The U6 snRNA m6A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. Cell *169*, 824-835.e14.

Pervouchine, D., Popov, Y., Berry, A., Borsari, B., Frankish, A., and Guigó, R. (2019). Integrative transcriptomic analysis suggests new autoregulatory splicing events coupled with nonsense-mediated mRNA decay. Nucleic acids research *47*, 5293–5306.

Peterson, M.L., and Perry, R.P. (1989). The regulated production of mu m and mu s mRNA is dependent on the relative efficiencies of mu s poly(A) site usage and the c mu 4-to-M1 splice. Molecular and cellular biology *9*, 726–738.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry *25*, 1605–1612.

Ping, X.-L., Sun, B.-F., Wang, L., Xiao, W., Yang, X., Wang, W.-J., Adhikari, S., Shi, Y., Lv, Y., Chen, Y.-S., Zhao, X., Li, A., Yang, Y., Dahal, U., Lou, X.-M., Liu, X., Huang, J., Yuan, W.-P., Zhu, X.-F., Cheng, T., Zhao, Y.-L., Wang, X., Rendtlew Danielsen, J.M., Liu, F., and Yang, Y.-G. (2014). Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. Cell research *24*, 177–189.

Poser, I., Sarov, M., Hutchins, J.R.A., Hériché, J.-K., Toyoda, Y., Pozniakovsky, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A.W., Pelletier, L., Kittler, R., Hua, S., Naumann, R., Augsburg, M., Sykora, M.M., Hofemeister, H., Zhang, Y., Nasmyth, K., White, K.P., Dietzel, S., Mechtler, K., Durbin, R., Stewart, A.F., Peters, J.-M., Buchholz, F., and Hyman, A.A. (2008). BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nature methods *5*, 409–415.

Preker, P.J., Ohnacker, M., Minvielle-Sebastia, L., and Keller, W. (1997). A multisubunit 3' end processing factor from yeast containing poly(A) polymerase and homologues of the subunits of mammalian cleavage and polyadenylation specificity factor. The EMBO journal *16*, 4727–4737.

Proudfoot, N.J. (1989). How RNA polymerase II terminates transcription in higher eukaryotes. Trends in biochemical sciences *14*, 105–110.

Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. Cell *108*, 501–512.

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

Ram, O., and Ast, G. (2007). SR proteins: a foot on the exon before the transition from intron to exon definition. Trends in genetics TIG *23*, 5–7.

Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. Genome research *12*, 1231–1245.

Raska, I., Shaw, P.J., and Cmarko, D. (2006). Structure and function of the nucleolus in the spotlight. Current opinion in cell biology *18*, 325–334.

Ratti, A., and Buratti, E. (2016). Physiological functions and pathobiology of TDP-43 and FUS/TLS proteins. Journal of neurochemistry *138 Suppl 1*, 95–111.

Risso, G., Pelisch, F., Quaglino, A., Pozzi, B., and Srebrow, A. (2012). Regulating the regulators: serine/arginine-rich proteins under scrutiny. IUBMB life *64*, 809–816.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nature biotechnology *29*, 24–26.

Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J., and Dunker, A.K. (2001). Sequence complexity of disordered protein. Proteins: Structure, Function, and Genetics *42*, 38–48.

Rosonina, E., Bakowski, M.A., McCracken, S., and Blencowe, B.J. (2003). Transcriptional activators control splicing and 3'-end cleavage levels. The Journal of biological chemistry *278*, 43034–43040.

Rossbach, O., Hung, L.-H., Schreiner, S., Grishina, I., Heiner, M., Hui, J., and Bindereif, A. (2009). Auto- and cross-regulation of the hnRNP L proteins by alternative splicing. Molecular and cellular biology *29*, 1442–1451.

Rot, G., Wang, Z., Huppertz, I., Modic, M., Lenče, T., Hallegger, M., Haberman, N., Curk, T., Mering, C. von, and Ule, J. (2017). High-Resolution RNA Maps Suggest Common Principles of Splicing and Polyadenylation Regulation by TDP-43. Cell reports *19*, 1056–1067.

Rottman, F., Shatkin, A.J., and Perry, R.P. (1974). Sequences containing methylated nucleotides at the 5' termini of messenger RNAs: possible implications for processing. Cell *3*, 197–199.

Roundtree, I.A., Evans, M.E., Pan, T., and He, C. (2017a). Dynamic RNA Modifications in Gene Expression Regulation. Cell *169*, 1187–1200.

Roundtree, I.A., Luo, G.-Z., Zhang, Z., Wang, X., Zhou, T., Cui, Y., Sha, J., Huang, X., Guerrero, L., Xie, P., He, E., Shen, B., and He, C. (2017b). YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs. eLife *6*.

Rueden, C.T., Schindelin, J., Hiner, M.C., DeZonia, B.E., Walter, A.E., Arena, E.T., and Eliceiri, K.W. (2017). ImageJ2: ImageJ for the next generation of scientific image data. BMC bioinformatics *18*, 529.

Rüegsegger, U., Beyer, K., and Keller, W. (1996). Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. The Journal of biological chemistry *271*, 6107–6113.

Rüegsegger, U., Blank, D., and Keller, W. (1998). Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits. Molecular cell *1*, 243–253.

Ruepp, M.-D., Aringhieri, C., Vivarelli, S., Cardinale, S., Paro, S., Schümperli, D., and Barabino, S.M.L. (2009). Mammalian pre-mRNA 3' end processing factor CF I m 68 functions in mRNA export. Molecular biology of the cell *20*, 5211–5223.

Ruepp, M.-D., Schümperli, D., and Barabino, S.M.L. (2011). mRNA 3' end processing and more--multiple functions of mammalian cleavage factor I-68. Wiley interdisciplinary reviews. RNA *2*, 79–91.

Rueter, S.M., Dawson, T.R., and Emeson, R.B. (1999). Regulation of alternative splicing by RNA editing. Nature *399*, 75–80.

Ryan, K., Calvo, O., and Manley, J.L. (2004). Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. RNA (New York, N.Y.) *10*, 565–573.

Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A., and Burge, C.B. (2008). Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. Science (New York, N.Y.) *320*, 1643–1647.

Sanford, J.R., Gray, N.K., Beckmann, K., and Cáceres, J.F. (2004). A novel role for shuttling SR proteins in mRNA translation. Genes & development *18*, 755–768.

Sanford, J.R., Wang, X., Mort, M., Vanduyn, 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 research *19*, 381–394.

Sapra, A.K., Ankö, 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. Molecular cell *34*, 179–190.

Sartini, B.L., Wang, H., Wang, W., Millette, C.F., and Kilpatrick, D.L. (2008). Pre-messenger RNA cleavage factor I (CFIm): potential role in alternative polyadenylation during spermatogenesis. Biology of reproduction *78*, 472–482.

Sasaki, Y.T.F., Ideue, T., Sano, M., Mituyama, T., and Hirose, T. (2009). MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. Proceedings of the National Academy of Sciences of the United States of America *106*, 2525–2530.

Sashital, D.G., and Butcher, S.E. (2006). Flipping off the riboswitch: RNA structures that control gene expression. ACS chemical biology *1*, 341–345.

Schäfer, P., Tüting, C., Schönemann, L., Kühn, U., Treiber, T., Treiber, N., Ihling, C., Graber, A., Keller, W., Meister, G., Sinz, A., and Wahle, E. (2018). Reconstitution of mammalian Cleavage Factor II involved in 3' processing of mRNA precursors. RNA (New York, N.Y.).

Scheer, U., and Hock, R. (1999). Structure and function of the nucleolus. Current opinion in cell biology *11*, 385–390.

Scheres, S.H., and Nagai, K. (2017). CryoEM structures of spliceosomal complexes reveal the molecular mechanism of pre-mRNA splicing. Current opinion in structural biology *46*, 130–139.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., and Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. Nature methods *9*, 676–682.

Schnitzbauer, J., Strauss, M.T., Schlichthaerle, T., Schueder, F., and Jungmann, R. (2017). Super-resolution microscopy with DNA-PAINT. Nature protocols *12*, 1198–1228.

Schönemann, L., Kühn, U., Martin, G., Schäfer, P., Gruber, A.R., Keller, W., Zavolan, M., and Wahle, E. (2014). Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. Genes & development *28*, 2381–2393.

Schwich, O.D., Blümel, N., Keller, M., Wegener, M., Thonta Setty, S., Brunstein, M.E., Poser, I., Los Mozos, I.R. de, Suess, B., Münch, C., McNicoll, F., Zarnack, K., and Müller-McNicoll, M. (in-press).

SRSF3 and SRSF7 modulate 3'UTR length through suppression or activation of proximal polyadenylation sites and regulation of CFIm levels. Genome Biology.

Shen, E.C., Henry, M.F., Weiss, V.H., Valentini, S.R., Silver, P.A., and Lee, M.S. (1998). Arginine methylation facilitates the nuclear export of hnRNP proteins. Genes & development *12*, 679–691.

Shen, H., and Green, M.R. (2006). RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. Genes & development *20*, 1755–1765.

Shen, H., Zheng, X., Shen, J., Zhang, L., Zhao, R., and Green, M.R. (2008). Distinct activities of the DExD/H-box splicing factor hUAP56 facilitate stepwise assembly of the spliceosome. Genes & development *22*, 1796–1803.

Shen, M., and Mattox, W. (2012). Activation and repression functions of an SR splicing regulator depend on exonic versus intronic-binding position. Nucleic acids research *40*, 428–437.

Shen, T., Li, H., Song, Y., Li, L., Lin, J., Wei, G., and Ni, T. (2019). Alternative polyadenylation dependent function of splicing factor SRSF3 contributes to cellular senescence. Aging.

Shenasa, H., and Hertel, K.J. (2019). Combinatorial regulation of alternative splicing. Biochimica et biophysica acta. Gene regulatory mechanisms.

Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., Frank, J., and Manley, J.L. (2009). Molecular architecture of the human pre-mRNA 3' processing complex. Molecular cell *33*, 365–376.

Shi, Y., Reddy, B., and Manley, J.L. (2006). PP1/PP2A phosphatases are required for the second step of Pre-mRNA splicing and target specific snRNP proteins. Molecular Cell *23*, 819–829.

Shimba, S., Bokar, J.A., Rottman, F., and Reddy, R. (1995). Accurate and efficient N-6-adenosine methylation in spliceosomal U6 small nuclear RNA by HeLa cell extract in vitro. Nucleic acids research *23*, 2421–2426.

Shin, C., and Manley, J.L. (2002). The SR protein SRp38 represses splicing in M phase cells. Cell *111*, 407–417.

Simard, M.J., and Chabot, B. (2002). SRp30c is a repressor of 3' splice site utilization. Molecular and cellular biology *22*, 4001–4010.

Singh, G., Pratt, G., Yeo, G.W., and Moore, M.J. (2015). The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. Annual review of biochemistry *84*, 325–354.

Singh, N.N., Singh, R.N., and Androphy, E.J. (2007). Modulating role of RNA structure in alternative splicing of a critical exon in the spinal muscular atrophy genes. Nucleic acids research *35*, 371–389.

Sinha, R., Allemand, E., Zhang, Z., Karni, R., Myers, M.P., and Krainer, A.R. (2010). Arginine methylation controls the subcellular localization and functions of the oncoprotein splicing factor SF2/ASF. Molecular and cellular biology *30*, 2762–2774.

Śledź, P., and Jinek, M. (2016). Structural insights into the molecular mechanism of the m(6)A writer complex. eLife *5*.

Smibert, P., Miura, P., Westholm, J.O., Shenker, S., May, G., Duff, M.O., Zhang, D., Eads, B.D., Carlson, J., Brown, J.B., Eisman, R.C., Andrews, J., Kaufman, T., Cherbas, P., Celniker, S.E., Graveley, B.R., and Lai, E.C. (2012). Global patterns of tissue-specific alternative polyadenylation in Drosophila. Cell Reports *1*, 277–289.

So, B.R., Di, C., Cai, Z., Venters, C.C., Guo, J., Oh, J.-M., Arai, C., and Dreyfuss, G. (2019). A Complex of U1 snRNP with Cleavage and Polyadenylation Factors Controls Telescripting, Regulating mRNA Transcription in Human Cells. Molecular cell *76*, 590-599.e4.

Solomon, O., Oren, S., Safran, M., Deshet-Unger, N., Akiva, P., Jacob-Hirsch, J., Cesarkas, K., Kabesa, R., Amariglio, N., Unger, R., Rechavi, G., and Eyal, E. (2013). Global regulation of alternative splicing by adenosine deaminase acting on RNA (ADAR). RNA (New York, N.Y.) *19*, 591–604.

Spector, D.L., and Lamond, A.I. (2011). Nuclear speckles. Cold Spring Harbor perspectives in biology *3*.

Staley, J.P., and Guthrie, C. (1999). An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. Molecular cell *3*, 55–64.

Starke, S., Jost, I., Rossbach, O., Schneider, T., Schreiner, S., Hung, L.-H., and Bindereif, A. (2015). Exon circularization requires canonical splice signals. Cell Reports *10*, 103–111.

Strässer, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondón, A.G., Aguilera, A., Struhl, K., Reed, R., and Hurt, E. (2002). TREX is a conserved complex coupling transcription with messenger RNA export. Nature *417*, 304–308.

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. Nature structural & molecular biology *17*, 306–312.

Sunwoo, H., Dinger, M.E., Wilusz, J.E., Amaral, P.P., Mattick, J.S., and Spector, D.L. (2009). MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. Genome research *19*, 347–359.

Surendranath, V., Theis, M., Habermann, B.H., and Buchholz, F. (2013). Designing efficient and specific endoribonuclease-prepared siRNAs. Methods in molecular biology (Clifton, N.J.) *942*, 193–204.

Swanson, C.M., Sherer, N.M., and Malim, M.H. (2010). SRp40 and SRp55 promote the translation of unspliced human immunodeficiency virus type 1 RNA. Journal of virology *84*, 6748–6759.

Swartz, J.E., Bor, Y.-C., Misawa, Y., Rekosh, D., and Hammarskjold, M.-L. (2007). The shuttling SR protein 9G8 plays a role in translation of unspliced mRNA containing a constitutive transport element. The Journal of biological chemistry *282*, 19844–19853.

Szymczyna, B.R., Bowman, J., McCracken, S., Pineda-Lucena, A., Lu, Y., Cox, B., Lambermon, M., Graveley, B.R., Arrowsmith, C.H., and Blencowe, B.J. (2003). Structure and function of the PWI motif: a novel nucleic acid-binding domain that facilitates pre-mRNA processing. Genes & development *17*, 461–475.

Tacke, R., Chen, Y., and Manley, J.L. (1997). Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. Proceedings of the National Academy of Sciences of the United States of America *94*, 1148–1153.

Takagaki, Y., and Manley, J.L. (1997). RNA recognition by the human polyadenylation factor CstF. Molecular and cellular biology *17*, 3907–3914.

Takagaki, Y., and Manley, J.L. (2000). Complex protein interactions within the human polyadenylation machinery identify a novel component. Molecular and cellular biology *20*, 1515–1525.

Takagaki, Y., Seipelt, R.L., Peterson, M.L., and Manley, J.L. (1996). The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. Cell *87*, 941–952.

Tange, T.Ø., Damgaard, C.K., Guth, S., Valcárcel, J., and Kjems, J. (2001). The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. The EMBO journal *20*, 5748–5758.

The UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. Nucleic acids research *47*, D506-D515.

Thermann, R., Neu-Yilik, G., Deters, A., Frede, U., Wehr, K., Hagemeier, C., Hentze, M.W., and Kulozik, A.E. (1998). Binary specification of nonsense codons by splicing and cytoplasmic translation. The EMBO journal *17*, 3484–3494.

Tian, B., Hu, J., Zhang, H., and Lutz, C.S. (2005). A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic acids research *33*, 201–212.

Tian, B., and Manley, J.L. (2017). Alternative polyadenylation of mRNA precursors. Nature reviews. Molecular cell biology *18*, 18–30.

Tian, B., Pan, Z., and Lee, J.Y. (2007). Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. Genome research *17*, 156–165.

Tran, D.D.H., Saran, S., Williamson, A.J.K., Pierce, A., Dittrich-Breiholz, O., Wiehlmann, L., Koch, A., Whetton, A.D., and Tamura, T. (2014). THOC5 controls 3'end-processing of immediate early genes via interaction with polyadenylation specific factor 100 (CPSF100). Nucleic acids research *42*, 12249–12260.

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., Blencowe, B.J., Prasanth, S.G., and Prasanth, K.V. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Molecular cell *39*, 925–938.

Tushev, G., Glock, C., Heumüller, M., Biever, A., Jovanovic, M., and Schuman, E.M. (2018). Alternative 3' UTRs Modify the Localization, Regulatory Potential, Stability, and Plasticity of mRNAs in Neuronal Compartments. Neuron *98*, 495-511.e6.

Vagner, S., Vagner, C., and Mattaj, I.W. (2000). The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. Genes & development *14*, 403–413.

Valente, L., and Nishikura, K. (2005). ADAR Gene Family and A-to-I RNA Editing: Diverse Roles in Posttranscriptional Gene Regulation. In Progress in nucleic acid research and molecular biology: Volume 79, K. Moldave, ed. (Amsterdam: Elsevier Academic Press), pp. 299–338.

Valente, S.T., Gilmartin, G.M., Venkatarama, K., Arriagada, G., and Goff, S.P. (2009). HIV-1 mRNA 3' end processing is distinctively regulated by eIF3f, CDK11, and splice factor 9G8. Molecular cell *36*, 279–289.

Valluy, J., Bicker, S., Aksoy-Aksel, A., Lackinger, M., Sumer, S., Fiore, R., Wüst, T., Seffer, D., Metge, F., Dieterich, C., Wöhr, M., Schwarting, R., and Schratt, G. (2015). A coding-independent function of an alternative Ube3a transcript during neuronal development. Nature neuroscience *18*, 666–673.

Valverde, R., Edwards, L., and Regan, L. (2008). Structure and function of KH domains. The FEBS journal *275*, 2712–2726.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nature reviews. Genetics *10*, 252–263.

Vassileva, M.T., and Matunis, M.J. (2004). SUMO modification of heterogeneous nuclear ribonucleoproteins. Molecular and cellular biology *24*, 3623–3632.

Vasudevan, S., Peltz, S.W., and Wilusz, C.J. (2002). Non-stop decay--a new mRNA surveillance pathway. BioEssays news and reviews in molecular, cellular and developmental biology *24*, 785–788.

Venables, J.P., Klinck, R., Koh, C., Gervais-Bird, J., Bramard, A., Inkel, L., Durand, M., Couture, S., Froehlich, U., Lapointe, E., Lucier, J.-F., Thibault, P., Rancourt, C., Tremblay, K., Prinos, P., Chabot, B., and Elela, S.A. (2009). Cancer-associated regulation of alternative splicing. Nature structural & molecular biology *16*, 670–676.

Venkataraman, K., Brown, K.M., and Gilmartin, G.M. (2005). Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. Genes & development *19*, 1315–1327.

Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., Fu, Y.-H., Kuhl, D.P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F., Eussen, B.E., van Ommen, G.-J.B., Blonden, L.A., Riggins, G.J., Chastain, J.L., Kunst, C.B., Galjaard, H., Thomas Caskey, C., Nelson, D.L., Oostra, B.A., and Warren, S.T. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell *65*, 905–914.

Viphakone, N., Sudbery, I., Griffith, L., Heath, C.G., Sims, D., and Wilson, S.A. (2019). Co-transcriptional Loading of RNA Export Factors Shapes the Human Transcriptome. Molecular cell *75*, 310-323.e8.

Vries, H. de, Rüegsegger, U., Hübner, W., Friedlein, A., Langen, H., and Keller, W. (2000). Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. The EMBO journal *19*, 5895–5904.

Wagner, E.J., and Garcia-Blanco, M.A. (2001). Polypyrimidine Tract Binding Protein Antagonizes Exon Definition. Molecular and cellular biology *21*, 3281–3288.

Wahle, E. (1991). Purification and characterization of a mammalian polyadenylate polymerase involved in the 3' end processing of messenger RNA precursors. The Journal of biological chemistry *266*, 3131–3139.

Wang, P., Doxtader, K.A., and Nam, Y. (2016). Structural Basis for Cooperative Function of Mettl3 and Mettl14 Methyltransferases. Molecular cell *63*, 306–317.

Wang, R., Zheng, D., Wei, L., Ding, Q., and Tian, B. (2019). Regulation of Intronic Polyadenylation by PCF11 Impacts mRNA Expression of Long Genes. Cell reports *26*, 2766-2778.e6.

Wang, W., Wei, Z., and Li, H. (2014a). A change-point model for identifying 3'UTR switching by next-generation RNA sequencing. Bioinformatics (Oxford, England) *30*, 2162–2170.

Wang, W., Wei, Z., and Li, H. (2014b). A change-point model for identifying 3'UTR switching by next-generation RNA sequencing. Bioinformatics (Oxford, England) *30*, 2162–2170.

Wang, X., Hennig, T., Whisnant, A.W., Erhard, F., Prusty, B.K., Friedel, C.C., Forouzmand, E., Hu, W., Erber, L., Chen, Y., Sandri-Goldin, R.M., Dölken, L., and Shi, Y. (2020). Herpes simplex virus blocks host transcription termination via the bimodal activities of ICP27. Nature communications *11*, 293.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., Ren, B., Pan, T., and He, C. (2014c). N6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117–120.

Wang, X., McLachlan, J., Zamore, P.D., and Hall, T.M. (2002). Modular Recognition of RNA by a Human Pumilio-Homology Domain. Cell *110*, 501–512.

Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., and He, C. (2015). N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. Cell *161*, 1388–1399.

Wang, Z., Chatterjee, D., Jeon, H.Y., Akerman, M., Vander Heiden, M.G., Cantley, L.C., and Krainer, A.R. (2012). Exon-centric regulation of pyruvate kinase M alternative splicing via mutually exclusive exons. Journal of molecular cell biology *4*, 79–87.

Wang, Z., Kayikci, M., Briese, M., Zarnack, K., Luscombe, N.M., Rot, G., Zupan, B., Curk, T., and Ule, J. (2010). iCLIP predicts the dual splicing effects of TIA-RNA interactions. PLoS biology *8*, e1000530.

Wang, Z.-L., Li, B., Luo, Y.-X., Lin, Q., Liu, S.-R., Zhang, X.-Q., Zhou, H., Yang, J.-H., and Qu, L.-H. (2018). Comprehensive Genomic Characterization of RNA-Binding Proteins across Human Cancers. Cell reports *22*, 286–298.

Warkocki, Z., Odenwälder, P., Schmitzová, J., Platzmann, F., Stark, H., Urlaub, H., Ficner, R., Fabrizio, P., and Lührmann, R. (2009). Reconstitution of both steps of Saccharomyces cerevisiae splicing with purified spliceosomal components. Nature structural & molecular biology *16*, 1237–1243.

Wegener, M., and Müller-McNicoll, M. (2019). Nuclear retention of mRNAs - quality control, gene regulation and human disease. Seminars in cell & developmental biology *79*, 131–142.

Wilkening, S., Pelechano, V., Järvelin, A.I., Tekkedil, M.M., Anders, S., Benes, V., and Steinmetz, L.M. (2013). An efficient method for genome-wide polyadenylation site mapping and RNA quantification. Nucleic acids research *41*, e65.

Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. Cold Spring Harbor perspectives in biology *3*.

Williams, R.M., Obradovic, Z., Mathura, V., Braun, W., Garner, E.C., Young, J., Takayama, S., Brown, C.J., and Dunker, A.K. (2001). The Protein Non-Folding Problem: Amino Acid Determinants Of Intrinsic Order And Disorder. In Pacific Symposium on Biocomputing 2001, Mauna Lani, Hawaii, 3 - 7 January 2001, R.B Altman, ed. (Singapore u.a.: WORLD SCIENTIFIC), pp. 89–100.

WILLIS, I.M. (1993). RNA polymerase III. Genes, factors and transcriptional specificity. European Journal of Biochemistry *212*, 1–11.

Wilusz, J.E. (2018). A 360° view of circular RNAs: From biogenesis to functions. Wiley interdisciplinary reviews. RNA *9*, e1478.

Wong, J., Garner, B., Halliday, G.M., and Kwok, J.B.J. (2012). Srp20 regulates TrkB pre-mRNA splicing to generate TrkB-Shc transcripts with implications for Alzheimer's disease. Journal of neurochemistry *123*, 159–171.

Wu, H., Sun, S., Tu, K., Gao, Y., Xie, B., Krainer, A.R., and Zhu, J. (2010). A splicing-independent function of SF2/ASF in microRNA processing. Molecular cell *38*, 67–77.

Wu, J.Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell *75*, 1061–1070.

Wu, X., Liu, M., Downie, B., Liang, C., Ji, G., Li, Q.Q., and Hunt, A.G. (2011). Genome-wide landscape of polyadenylation in Arabidopsis provides evidence for extensive alternative polyadenylation. Proceedings of the National Academy of Sciences of the United States of America *108*, 12533–12538.

Xia, Z., Donehower, L.A., Cooper, T.A., Neilson, J.R., Wheeler, D.A., Wagner, E.J., and Li, W. (2014). Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. Nature communications *5*, 5274.

Xiang, Y., Ye, Y., Lou, Y., Yang, Y., Cai, C., Zhang, Z., Mills, T., Chen, N.-Y., Kim, Y., Muge Ozguc, F., Diao, L., Karmouty-Quintana, H., Xia, Y., Kellems, R.E., Chen, Z., Blackburn, M.R., Yoo, S.-H., Shyu, A.-B., Mills, G.B., and Han, L. (2018). Comprehensive Characterization of Alternative Polyadenylation in Human Cancer. Journal of the National Cancer Institute *110*, 379–389.

Xiao, M.-S., Ai, Y., and Wilusz, J.E. (2020). Biogenesis and Functions of Circular RNAs Come into Focus. Trends in cell biology *30*, 226–240.

Xiao, W., Adhikari, S., Dahal, U., Chen, Y.-S., Hao, Y.-J., Sun, B.-F., Sun, H.-Y., Li, A., Ping, X.-L., Lai, W.-Y., Wang, X., Ma, H.-L., Huang, C.-M., Yang, Y., Huang, N., Jiang, G.-B., Wang, H.-L., Zhou, Q., Wang, X.-J., Zhao, Y.-L., and Yang, Y.-G. (2016). Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. Molecular cell *61*, 507–519.

Xuan, J.-J., Sun, W.-J., Lin, P.-H., Zhou, K.-R., Liu, S., Zheng, L.-L., Qu, L.-H., and Yang, J.-H. (2018). RMBase v2.0: deciphering the map of RNA modifications from epitranscriptome sequencing data. Nucleic acids research *46*, D327-D334.

Yadegari, H., Biswas, A., Akhter, M.S., Driesen, J., Ivaskevicius, V., Marquardt, N., and Oldenburg, J. (2016). Intron retention resulting from a silent mutation in the VWF gene that structurally influences the 5' splice site. Blood *128*, 2144–2152.

Yang, Q., Coseno, M., Gilmartin, G.M., and Doublié, S. (2011a). Crystal structure of a human cleavage factor CFI(m)25/CFI(m)68/RNA complex provides an insight into poly(A) site recognition and RNA looping. Structure (London, England 1993) *19*, 368–377.

Yang, Q., Gilmartin, G.M., and Doublié, S. (2010). Structural basis of UGUA recognition by the Nudix protein CFI(m)25 and implications for a regulatory role in mRNA 3' processing. Proceedings of the National Academy of Sciences of the United States of America *107*, 10062–10067.

Yang, Q., Gilmartin, G.M., and Doublié, S. (2011b). The structure of human cleavage factor I(m) hints at functions beyond UGUA-specific RNA binding: a role in alternative polyadenylation and a potential link to 5' capping and splicing. RNA biology *8*, 748–753.

Yang, W., Hsu, P.L., Yang, F., Song, J.-E., and Varani, G. (2018). Reconstitution of the CstF complex unveils a regulatory role for CstF-50 in recognition of 3'-end processing signals. Nucleic acids research *46*, 493–503.

Yao, C., Choi, E.-A., Weng, L., Xie, X., Wan, J., Xing, Y., Moresco, J.J., Tu, P.G., Yates, J.R., and Shi, Y. (2013). Overlapping and distinct functions of CstF64 and CstF64τ in mammalian mRNA 3' processing. RNA (New York, N.Y.) *19*, 1781–1790.

Yao, P., Potdar, A.A., Arif, A., Ray, P.S., Mukhopadhyay, R., Willard, B., Xu, Y., Yan, J., Saidel, G.M., and Fox, P.L. (2012). Coding region polyadenylation generates a truncated tRNA synthetase that counters translation repression. Cell *149*, 88–100.

Ye, C., Long, Y., Ji, G., Li, Q.Q., and Wu, X. (2018). APAtrap: identification and quantification of alternative polyadenylation sites from RNA-seq data. Bioinformatics (Oxford, England) *34*, 1841–1849.

Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.-D., and Gage, F.H. (2009). An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nature structural & molecular biology *16*, 130–137.

Yeo, G.W., van Nostrand, E., Holste, D., Poggio, T., and Burge, C.B. (2005). Identification and analysis of alternative splicing events conserved in human and mouse. Proceedings of the National Academy of Sciences of the United States of America *102*, 2850–2855.

Yin, X., Jiang, X., Wang, J., Qian, S., Liu, F., and Qian, W. (2018). SIRT1 Deacetylates SC35 and Suppresses Its Function in Tau Exon 10 Inclusion. Journal of Alzheimer's disease JAD *61*, 561–570.

Yonaha, M., and Proudfoot, N.J. (1999). Specific transcriptional pausing activates polyadenylation in a coupled in vitro system. Molecular Cell *3*, 593–600.

Yonaha, M., and Proudfoot, N.J. (2000). Transcriptional termination and coupled polyadenylation in vitro. The EMBO journal *19*, 3770–3777.

Yu, M.C., Bachand, F., McBride, A.E., Komili, S., Casolari, J.M., and Silver, P.A. (2004). Arginine methyltransferase affects interactions and recruitment of mRNA processing and export factors. Genes & development *18*, 2024–2035.

Yu, S., Pritchard, M., Kremer, E., Lynch, M., Nancarrow, J., Baker, E., Holman, K., Mulley, J.C., Warren, S.T., and Schlessinger, D. (1991). Fragile X genotype characterized by an unstable region of DNA. Science (New York, N.Y.) *252*, 1179–1181.

Yudin, D., Hanz, S., Yoo, S., Iavnilovitch, E., Willis, D., Gradus, T., Vuppalanchi, D., Segal-Ruder, Y., Ben-Yaakov, K., Hieda, M., Yoneda, Y., Twiss, J.L., and Fainzilber, M. (2008). Localized regulation of axonal RanGTPase controls retrograde injury signaling in peripheral nerve. Neuron *59*, 241–252.

Yun, C.Y., and Fu, X.D. (2000). Conserved SR protein kinase functions in nuclear import and its action is counteracted by arginine methylation in Saccharomyces cerevisiae. The Journal of Cell Biology *150*, 707–718.

Yuryev, A., Patturajan, M., Litingtung, Y., Joshi, R.V., Gentile, C., Gebara, M., and Corden, J.L. (1996). The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. Proceedings of the National Academy of Sciences of the United States of America *93*, 6975–6980.

Zahler, A.M., Lane, W.S., Stolk, J.A., and Roth, M.B. (1992). SR proteins: a conserved family of pre-mRNA splicing factors. Genes & development *6*, 837–847.

Zahler, A.M., Neugebauer, K.M., Lane, W.S., and Roth, M.B. (1993). Distinct functions of SR proteins in alternative pre-mRNA splicing. Science (New York, N.Y.) *260*, 219–222.

Zarudnaya, M.I., Kolomiets, I.M., Potyahaylo, A.L., and Hovorun, D.M. (2003). Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. Nucleic acids research *31*, 1375–1386.

Zawada, A.M., Rogacev, K.S., Müller, S., Rotter, B., Winter, P., Fliser, D., and Heine, G.H. (2014). Massive analysis of cDNA Ends (MACE) and miRNA expression profiling identifies proatherogenic pathways in chronic kidney disease. Epigenetics *9*, 161–172.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., Gil, L., Gordon, L., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O.G., Janacek, S.H., Juettemann, T., To, J.K., Laird, M.R., Lavidas, I., Liu, Z., Loveland, J.E., Maurel, T., McLaren, W., Moore, B., Mudge, J., Murphy, D.N., Newman, V., Nuhn, M., Ogeh, D., Ong, C.K., Parker, A., Patricio, M., Riat, H.S., Schuilenburg, H., Sheppard, D., Sparrow, H., Taylor, K., Thormann, A., Vullo, A., Walts, B., Zadissa, A., Frankish, A., Hunt, S.E., Kostadima, M., Langridge, N., Martin, F.J., Muffato, M., Perry, E., Ruffier, M., Staines, D.M., Trevanion, S.J., Aken, B.L., Cunningham, F., Yates, A., and Flicek, P. (2018). Ensembl 2018. Nucleic acids research *46*, D754-D761.

Zhang, H., Hu, J., Recce, M., and Tian, B. (2005a). PolyA_DB: a database for mammalian mRNA polyadenylation. Nucleic acids research *33*, D116-20.

Zhang, H., Lee, J.Y., and Tian, B. (2005b). Biased alternative polyadenylation in human tissues. Genome Biology *6*, R100.

Zhang, H., Shi, X., Huang, T., Zhao, X., Chen, W., Gu, N., and Zhang, R. (2020). Dynamic landscape and evolution of m6A methylation in human. Nucleic acids research *48*, 6251–6264.

Zhang, J., and Corden, J.L. (1991). Identification of phosphorylation sites in the repetitive carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. The Journal of biological chemistry *266*, 2290–2296.

Zhang, J., Sun, X., Qian, Y., LaDuca, J.P., and Maquat, L.E. (1998). At least one intron is required for the nonsense-mediated decay of triosephosphate isomerase mRNA: a possible link between nuclear splicing and cytoplasmic translation. Molecular and cellular biology *18*, 5272–5283.

Zhang, Z., So, K., Peterson, R., Bauer, M., Ng, H., Zhang, Y., Kim, J.H., Kidd, T., and Miura, P. (2019). Elav-Mediated Exon Skipping and Alternative Polyadenylation of the Dscam1 Gene Are Required for Axon Outgrowth. Cell Reports *27*, 3808-3817.e7.

Zhang, Z., Theler, D., Kaminska, K.H., Hiller, M., La Grange, P. de, Pudimat, R., Rafalska, I., Heinrich, B., Bujnicki, J.M., Allain, F.H.-T., and Stamm, S. (2010). The YTH domain is a novel RNA binding domain. The Journal of biological chemistry *285*, 14701–14710.

Zhao, B.S., Roundtree, I.A., and He, C. (2017). Post-transcriptional gene regulation by mRNA modifications. Nature reviews. Molecular cell biology *18*, 31–42.

Zhao, X., Yang, Y., Sun, B.-F., Shi, Y., Yang, X., Xiao, W., Hao, Y.-J., Ping, X.-L., Chen, Y.-S., Wang, W.-J., Jin, K.-X., Wang, X., Huang, C.-M., Fu, Y., Ge, X.-M., Song, S.-H., Jeong, H.S., Yanagisawa, H., Niu, Y., Jia, G.-F., Wu, W., Tong, W.-M., Okamoto, A., He, C., Rendtlew Danielsen, J.M., Wang, X.-J., and Yang, Y.-G. (2014). FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. Cell research *24*, 1403–1419.

Zheng, D., Liu, X., and Tian, B. (2016). 3'READS+, a sensitive and accurate method for 3' end sequencing of polyadenylated RNA. RNA (New York, N.Y.) *22*, 1631–1639.

Zheng, G., Dahl, J.A., Niu, Y., Fedorcsak, P., Huang, C.-M., Li, C.J., Vågbø, C.B., Shi, Y., Wang, W.-L., Song, S.-H., Lu, Z., Bosmans, R.P.G., Dai, Q., Hao, Y.-J., Yang, X., Zhao, W.-M., Tong, W.-M., Wang, X.-J., Bogdan, F., Furu, K., Fu, Y., Jia, G., Zhao, X., Liu, J., Krokan, H.E., Klungland, A., Yang, Y.-G., and He, C. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Molecular cell *49*, 18–29.

Zhong, S., Li, H., Bodi, Z., Button, J., Vespa, L., Herzog, M., and Fray, R.G. (2008). MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. The Plant cell *20*, 1278–1288.

Zhou, X., Li, R., Michal, J.J., Wu, X.-L., Liu, Z., Zhao, H., Xia, Y., Du, W., Wildung, M.R., Pouchnik, D.J., Harland, R.M., and Jiang, Z. (2016). Accurate Profiling of Gene Expression and Alternative Polyadenylation with Whole Transcriptome Termini Site Sequencing (WTTS-Seq). Genetics *203*, 683–697.

Zhou, Z., and Fu, X.-D. (2013). Regulation of splicing by SR proteins and SR protein-specific kinases. Chromosoma *122*, 191–207.

Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. Nature *419*, 182–185.

Zhou, Z., Luo, M.J., Straesser, K., Katahira, J., Hurt, E., and Reed, R. (2000). The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. Nature *407*, 401–405.

Zhu, J., Mayeda, A., and Krainer, A.R. (2001). Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. Molecular Cell *8*, 1351–1361.

Zhu, Y., Wang, X., Forouzmand, E., Jeong, J., Qiao, F., Sowd, G.A., Engelman, A.N., Xie, X., Hertel, K.J., and Shi, Y. (2018). Molecular Mechanisms for CFIm-Mediated Regulation of mRNA Alternative Polyadenylation. Molecular cell *69*, 62-74.e4.

9 Supplemental

9.1 List of primers

SuppTable 1: Primers used in this thesis.

Primer#	Name	Sequence (5’ -3’)	Species	Concentration (qRT-PCR)
48	qPCR_LAP_rev	TCTTCCAGGCTCGACGAACC		
111	SRSF7_esi_fw	CGTAATACGACTCACTATAGGGATTCGCCTTTGTGGAATTTG		
112	SRSF7_esi_rev	CGTAATACGACTCACTATAGGGCTTGAGCGGGATTGGAAATA		
245	RS_SRSF3_fw	GACCCGGGaggagtcctccacctc		
246	RS_SRSF3_rev	GACCCGGGctatttcctttcatttgacc		
247	RS_SRSF7_fw	GACCCGGGagccgacgaagaagaagc		
248	RS_SRSF7_rev	GACCCGGGtcagtccattctttctggact		
253	RS_SRSF3_S-to-D_fw	GACCCGGGagggaccctccacctc		
254	RS_SRSF3_S-to-D_rev	GACCCGGGctatttcctttcattgtccc		
255	RS_SRSF7_S-to-D_fw	GACCCGGGgaccgacgaagaagagac		
256	RS_SRSF7_S-to-D_rev	GACCCGGGtcagtccattctttctgggtc		
261	RS_SRSF3_S-to-A_fw	GACCCGGGagggcccctccacctc		
262	RS_SRSF3_S-to-A_rev	GACCCGGGctatttcctttcattggccc		
263	RS_SRSF7_S-to-A_fw	GACCCGGGgcccgacgaagaagagcc		
264	RS_SRSF7_S-to-A_rev	GACCCGGGtcagtccattctttctggggc		
502	Laptag fwd	AATCACCCTGGGCATGGACG		
580	qPCR_U6-1_for	GCTCGCTTCGGCAGC	Mouse	1000 mM
581	qPCR_U6-1_rev	AAATATGGAACGCTTCACGAATT	Mouse	1000 mM
706	esi_mCPSF5_fwd (2)	TAATACGACTCACTATAGGGAGATTGAGCTGTATGACAA		
707	esi_mCPSF5 rev	TAATACGACTCACTATAGGGAGAaatacaaataagaaattcactt		
741	esi_mSRSF3_1_fwd	cgtaatacgactcactatagggagagaaatcacaagccgtctc		
742	esi_mSRSF3_1_rev	Cgtaatacgactcactatagggagaggcttgtgttcacagcag		
745	esi_mFip1L1_fwd	TAATACGACTCACTATAGGGAGAgcacagtgacttggcaaagga		
746	esi_mFip1L1_rev	TAATACGACTCACTATAGGGAGActgccaggtgcatcgaggtc		

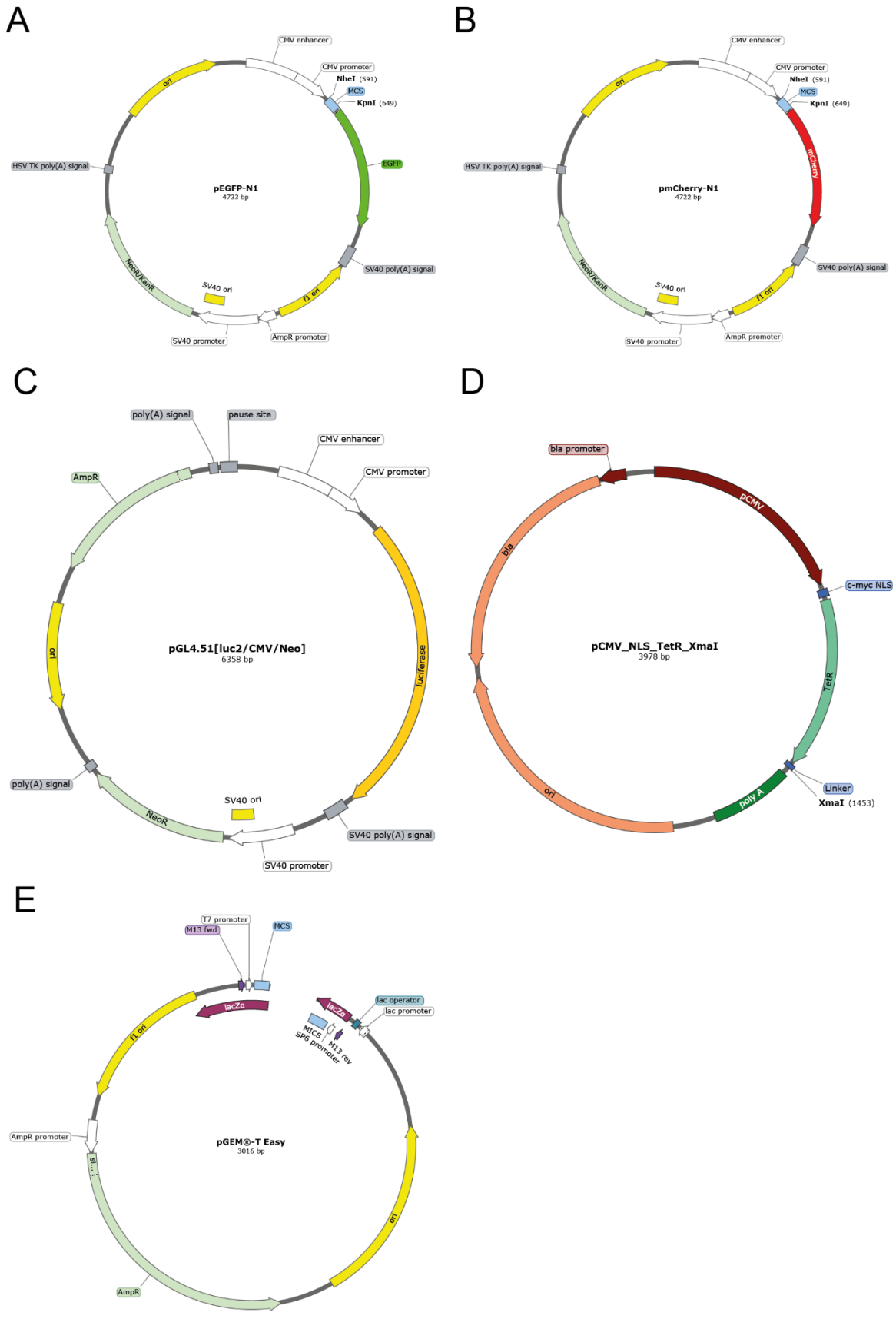
747	esi_Luc_fwd	TAATACGACTCACTATAGGGAGATTCTTCATGCCCGTGTTGGGTGC		
748	esi_Luc_rev	TAATACGACTCACTATAGGGAGAAATCCGGTACTGCCACTACTGTTCATGA		
784	esi_mCPSF6_R2_fwd	TAATACGACTCACTATAGGGAGAgtccagcagtttgcttcttgtgat		
785	esi_mCPSF6_R2_rev	TAATACGACTCACTATAGGGAGAtgctgaacacaagcctttccct		
834	F_Calr	TCAAGTCCGGGACAATCTTTG		
835	R_Calr	TCCTCTTTACGCTTCTTGTCCTC		
851	3'RACE_PCR_rev	CCAGTGAGCAGAGTGACGAGGACTCGAGCTCAAGC		
925	mCherry3' _fwd	GCGCCTACAACGTCAACATCAAG		
931	3'RACE-RT-Anchored	CCAGTGAGCAGAGTGACGAGGACTGAGCTCAAGCTTTTTTTTTTTTTTTTTTV		
958	gBlock_SRSF3_fwd	Atgcatcgtgattcctgtcccttgg		
959	gBlocks_SRSF3-RA_rev	ctatttcctttcattggccctagctcgg		
960	gBLocks_SRSF3-RD_rev	ctatttcctttcattgtccctgtctcggt		
964	gBlocks_SRSF7_fwd	atgtcacgctacgggcgg		
965	gBlocks_SRSF7-RA_rev	Tcagtccattctttctggagctgcg		
966	gBlocks_SRSF7-RD_rev	tcagtccattctttctgggtctgcgt		
994	splicing_CPSF6_Ex4_fwd	TGGTCAGAGTCCTGTTGTAACTCCA		
1129	RT1clip	[Phos]NNAACCNNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC		
1130	RT2clip	[Phos]NNACAANNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC		
1132	RT12clip	[Phos]NNGTGGNNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC		
1134	RT6clip	[Phos]NNCCGGNNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC		
1137	RT9clip	[Phos]NNGCCANNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC		
1138	RT10clip	[Phos]NNGACCNNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC		
1139	CutC4_oligo	GTTCAGGATCCACGACGCTCTTC		
1140	P3Solexa	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT		

1141	P5Solexa	AATGATACGGCGACCACCGAGATCTacactctttccctacacgacgctcttccgatct		
1305	GA_pmCherry_for	AGGCGTAAATTGTAAGCGTTAATATTTTGT		
1306	GA_pmCherry_rev	CTACTTGTACAGCTCGTCCATGCC		
1312	GA_pLuc2_fwd	aaatcgataaggatccgtttgcgtattggg		
1313	GA_pLuc2_rev	ttacacggcgatcttgccgccct		
1328	SRSF3-RA-KpnI_rev	GCGGTACCactttcctttcattggcccctagctcggg		
1329	SRSF3-RD-KpnI_rev	GCGGTACCactttcctttcattgtcctgtctcggtcac		
1332	SRSF7-RA-KpnI_rev	GCGGTACCacgtccattctttctggagctgcgg		
1333	SRSF7-RD-KpnI_rev	GCGGTACCacgtccattctttctgggtctgcgtctctgtg		
1362	GA_Luc_Ddx21_fwd	caagatcgccgtgtaaGTAGAGGCCAGAAGGGACTGTTCC		
1363	GA_Luc_Ddx21_rev	Aaacggatccttatcgatttaggattgcgttcatttttatgacaatttagccaagg		
1364	GA_mCherry_Ddx21_fwd	gacgagctgtacaagtagGTAGAGGCCAGAAGGGACTGTTCCC		
1365	GA_mCherry_Ddx21_rev	Cgcttacaatttacgcctaggattgcgttcatttttatgacaatttagccaagg		
1366	GA_Luc_Anp32e_fwd	caagatcgccgtgtaaACCTCCAGGACCAGGCCAC		
1367	GA_Luc_Anp32e_rev	aaacggatccttatcgattttaccaccactgctgttggagtatgagt		
1368	GA_mCherry_Anp32e_fwd	gacgagctgtacaagtagACCTCCAGGACCAGGCCAC		
1369	GA_mCherry_Anp32e_rev	cgcttacaatttacgccttaccaccactgctgttggagtatgagt		
1382	GA_Luc_Rab11a_fwd	caagatcgccgtgtaaGGCGTCTCTTCCCCTAGAAGGC		
1383	GA_Luc_Rab11a_rev	aaacggatccttatcgatttctctgcagatctaaagtctacctgaactgacagt		
1384	GA_mCherry_Rab11a_fwd	gacgagctgtacaagtagGGCGTCTCTTCCCCTAGAAGGC		
1385	GA_mCherry_Rab11a_rev	Cgcttacaatttacgcctctctgcagatctaaagtctacctgaactgacagt		
1386	PCR_Luc2_PAS-Seq_fwd	aggtgcctaaaggactgaccgg		
1387	Seq_Luc-Ddx21_inUTR_fwd	GTTCCAGTGAAGTCTCCAGACAAGGG		
1388	Seq_Luc-Ddx21_inUTR_rev	ATCATCTTGAACTGGCAAACTCTTTTATCCCA		
1389	Seq_Luc-Rab11a_inUTR_fwd	CATTGTGGAAGTCAGTTTCTAAAATGCCTTAAT		
1390	Seq_Luc-Rab11a_inUTR_rev	GGTACCACTTGAGTATAAATTAACTTCTCACTG		

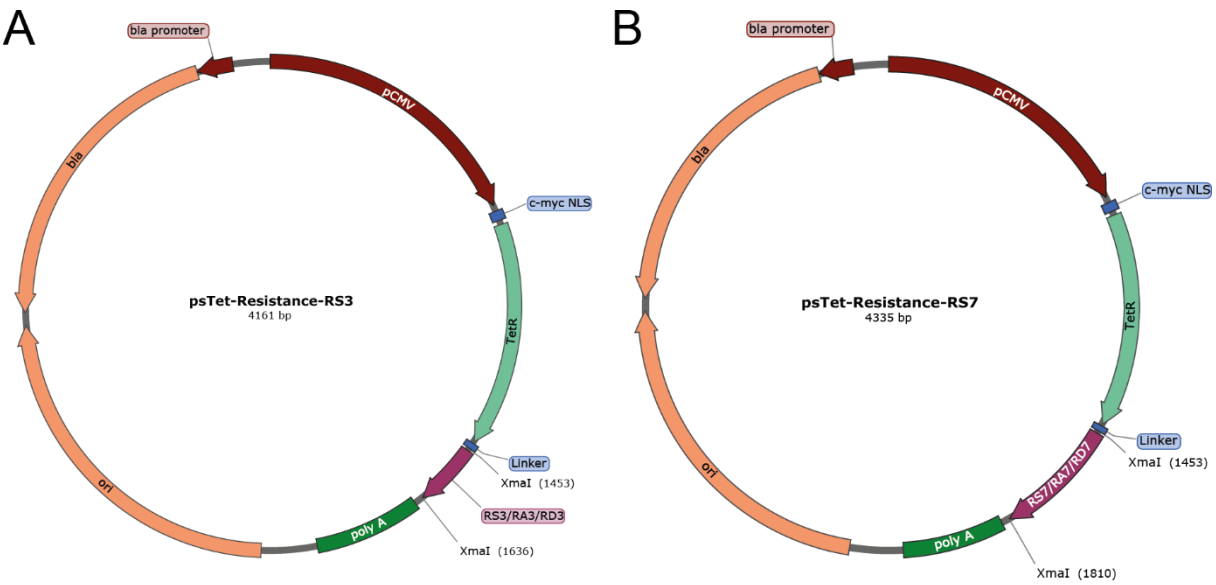
1391	Seq_Luc-Anp32e_inUTR_fwd	TGGCAGTAGTCTTTTGAATGATAAAGCTGG		
1392	Seq_LucAnp32e_inUTR_rev	GCATTTATTTTGAACCTGGCATAGTTTGC		
1397	Ddx21_UTR_Vector_fwd	ACCAACCATGGATCTGCCTATCTTGG		
1398	Ddx21_UTR_Vector_rev	TGCAAGAAAGGGGGCGCTGAG		
1399	Ddx21_strongpPAS_Frag_for	GCCCCCTTTCTTGCAGAGAGGCTTCTGAACTGTCCACTAACCA		
1400	Ddx21_strongpPAS_Frag_rev	GCAGATCCATGGTTGGTTTTATTTTGTAAATCTAAAAGTATGTTGGAAAACGATGCAATGAATTCT		
1601	circ_m+hCPSF6_Ex9_fwd	TATTACAGAGAGAAGCAGAGAACGAGAGAG		
1602	circ_m+hCPSF6_Ex2_rev	GCTCCTTTACCCACATCATCACCAACAG		
1603	qPCR_linCPSF6_fwd	GAAGAGTTCAACCAGGAAGCAGAATA		
1604	qPCR_linCPSF6_rev	CATCTGTTGTCCACCATGTTAGATTTCCAATATA		
1759	qPCR_mCPSF6Ex1Ex10_fwd	ATATTGGAAATCTAACATGGCAAAATCTACATCACAAAT		
1760	qPCR_mCPSF6Ex1Ex10_rev	AAAAACAAGCATTTACTTGACTTTTTTCCAATTGCT		
1797	qPCR_Ddx21_sharedUTR_rev	AGGCAGATCCATGGTTGGTTTTACT	Mouse	500 µM
1925	RACE_Ddx21_fwd	CACCCTGCTGACAAAGCCCGA		
1926	RACE_Rab11a_fwd	GCGTCTCTTCCCCTAGAAGGCTGT		
1927	RACE_Anp32e_fwd	CCTGGGCGCTGGAGAGCGAT		
1954	CPSF6_Exon7_rev	ctcagcttcactcagaggtgttcttgca		
2018	qPCR_Ddx21_dPAS1_fwd	GAGGAGGTGGAGAAGCGTGC	Mouse	500 µM
2019	qPCR_Ddx21_dPAS1_rev	GAGGACCCCACCACTCAAGC	Mouse	500 µM
2101	Ddx21_pPASmutant_Frag_rev	GCAGATCCATGGTTGGTCTTACTTTGTAAATCTAAAAGTATGTTGGAAAACGATGCAATG		
2102	Ddx21_noSR3motif_Frag_for	GCCCCCTTTCTTGCAGAGAGGATTCTGAACTGTGAACTAACGATCACAA		
2103	Ddx21_noSR3motif_Frag_rev	GCAGATCCATGGTTGGTTTTACTTTGTAAATCTAAAAGTATGTTCGAAAACGATCCAATGAA		
2104	Ddx21_noSR7motif_Frag_for	GCCCCCTTTCTTGCACAGCGGCTTCTCAACTGTCCACTAACC		
2105	Ddx21_noSR7motif_Frag_rev	GCAGATCCATGGTTGGTTTTACTTTGTAGATGTAAAAGTATGTTGGAAAACGATGCAAT		

9.2 Supplemental figures

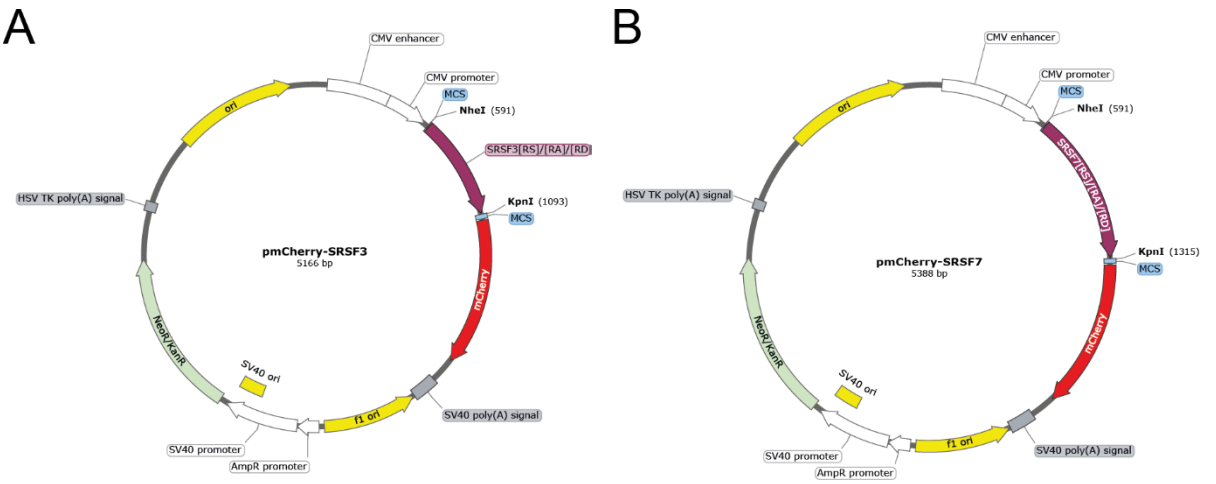
9.2.1 Plasmid maps



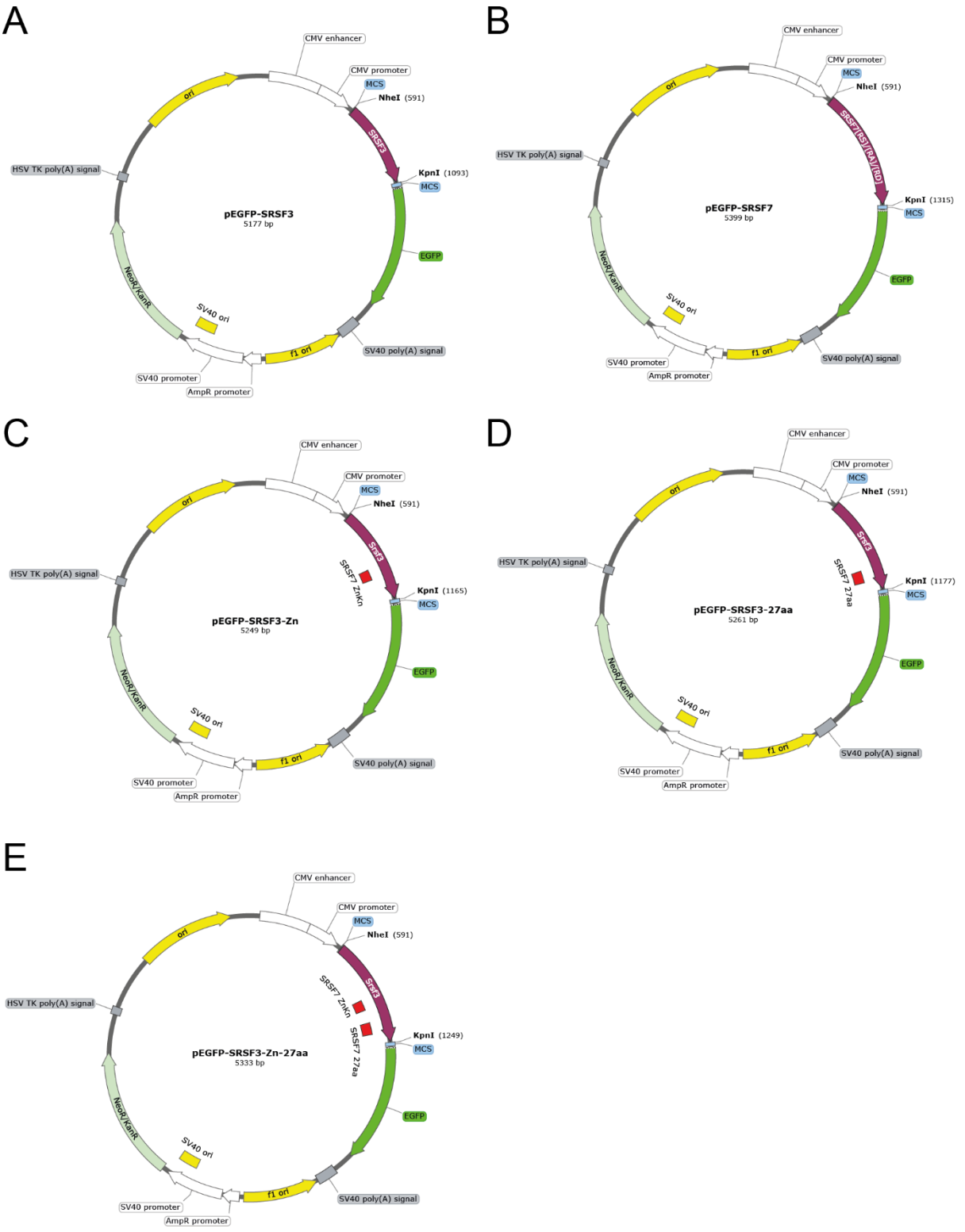
SuppFigure 1: Plasmid maps of basic vectors used for various cloning purposes. A&B) *Maps of pEGFP-N1 and pmCherry-N1 (Clontech). The plasmids contain a kanamycin resistance as selectin marker and were used for restriction enzyme cloning using the NheI and KpnI sites highlighted.* ***C)*** *The firefly luciferase plasmid pGL4.51 by Promega was used to clone 3' UTRs of target genes for validation purposes using Gibson Assembly® cloning. The plasmid carries an ampicillin resistance for selection purposes.* ***D)*** *The single chain (sc) Tet-Repressor (Tet-R) protein plasmid (Suess Lab, TU Darmstadt, Germany) was used to fuse the scTet-R with the RS domains and phosphomimetic RA/RD domains of SRSF3 and SRSF7 for co-immunoprecipitation studies using the XmaI restriction site highlighted. The plasmid contains an ampicillin resistance cassette to be used for selection purposes.* ***E)*** *The linearized pGEM®-T Easy (Promega) was used as a sub-cloning vector to store cloning intermediates with A-overhangs. The plasmid contains an ampicillin resistance and lac-Z cassette for selection purposes, besides M13 primer binding sites enabling straightforward Sanger sequencing-based validation of ligated inserts.*



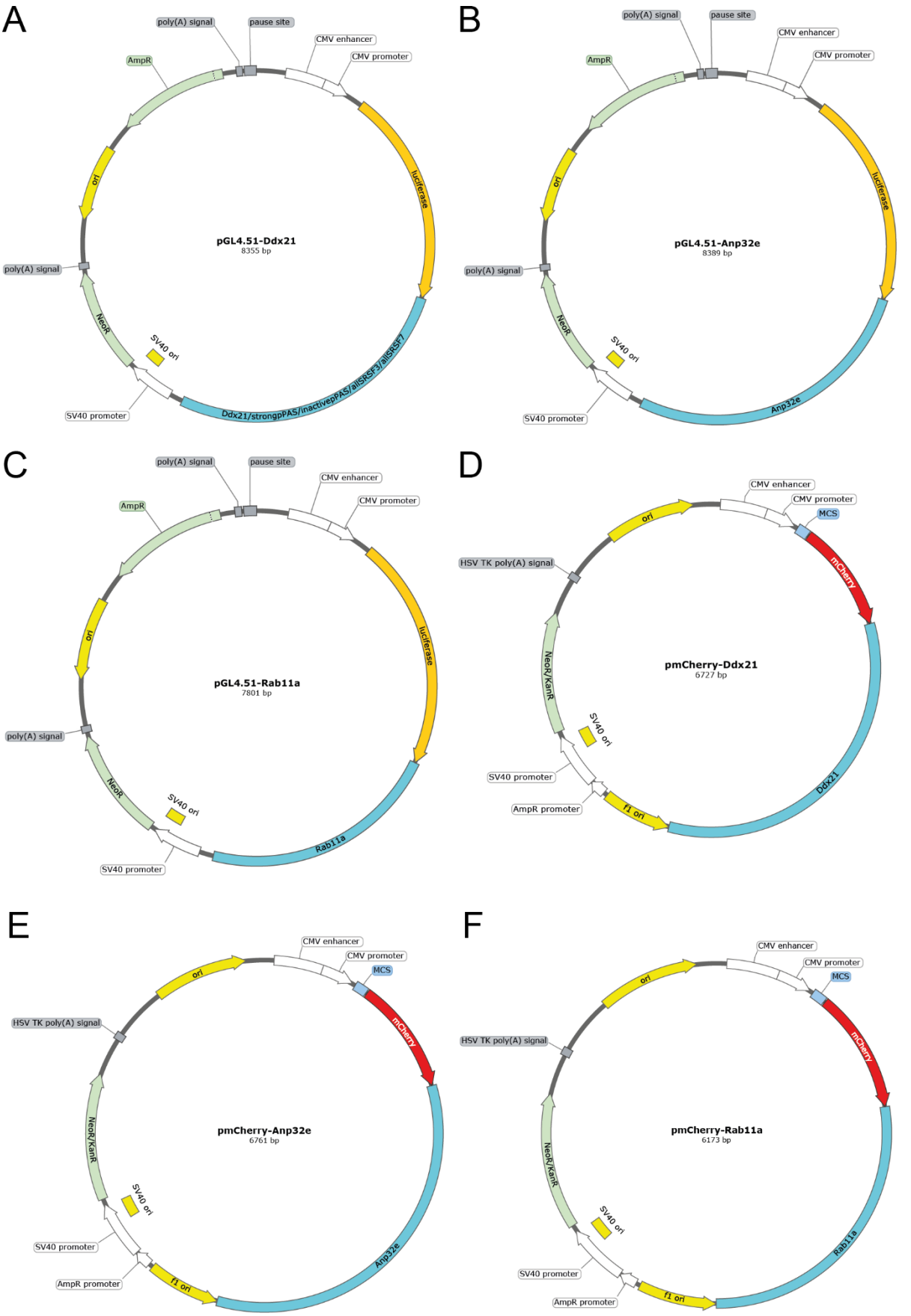
SuppFigure 2: Plasmid maps of psTet-Resistance-RS-Domain-phosphomimetic constructs. A) *Plasmid map of the single-chain Tet-Repressor protein fused with downstream RS, or phosphomimetic RA/RD domains originating from SRSF3.* ***B)*** *Plasmid map of the single-chain Tet-Repressor protein fused with downstream RS, or phosphomimetic RA/RD domains originating from SRSF7.*



SuppFigure 3: Plasmid maps of pmCherry-SRSF3/SRSF7-phosphomimetic constructs. A) *Plasmid map of pmCherry-N1 fused in frame with the complete coding sequence of SRSF3 wt or the phosphomimetics (RA/RD) lacking the stop codon.* ***B)*** *Plasmid map of pmCherry-N1 fused in frame with the complete coding sequence of SRSF7 wt or the phosphomimetics (RA/RD) lacking the stop codon.*



SuppFigure 4: Plasmid maps of pEGFP-SRSF3, pEGFP-SRSF7 and pEGFP-SRSF3-Chimera constructs. A&B) *Plasmid map of the pEGFP-N1 fused in frame with the full-length coding sequences of SRSF3 and SRSF7 lacking the stop codon.* ***C)*** *Plasmid map of pEGFP-N1 fused in frame with the coding sequence of SRSF3 containing the Zn-knuckle sequence originating from SRSF7.* ***D)*** *Plasmid map of pEGFP-N1 fused in frame with the coding sequence of SRSF3 containing the hydrophobic stretch (27aa) originating from SRSF7.* ***E)*** *Plasmid map of pEGFP-N1 fused in frame with the coding sequence of SRSF3 containing the Zn-knuckle and hydrophobic stretch (27aa) originating from SRSF7.*



SuppFigure 5: Plasmid maps of Luciferase-Reporter-3' UTR constructs. *Plasmid map of the firefly luciferase (pGL4.51) plasmid fused with the full length 3' UTR of Ddx21, including 150 nt downstream of the distal polyadenylation site containing downstream sequence elements (DSE). The 3' UTR of Ddx21 was mutated by strengthening or inactivating the proximal*

polyadenylation site and turning all potential SRSF3 binding sites into SRSF7 (and vice versa) in a window of 100nt upstream of the proximal polyadenylation site. ***B and C)*** *Plasmid map of the firefly luciferase (pGL4.51) plasmid fused with the full length 3' UTR of Anp32e or Rab11a, including 150 nt downstream of the distal polyadenylation site containing downstream sequence elements (DSE).* ***D-F)*** *Plasmid map of the pmCherry-N1 plasmid fused with the full length 3' UTR of Ddx21, Anp32e or Rab11a, including 150 nt downstream of the distal polyadenylation site containing downstream sequence elements (DSE).*

9.2.2 Alignments of phosphomimetics

A

```
RS3     AGGAGTCCTCCACCTCGGCGCAGATCCCCAAGAAGGAGAAGCTTTTCCCGAAGCCGGAGC      60
RA3     ----------------------------------------------------------ggGCA      5
RD3     aggGACcctccacctcggcgcagaGACccaagaaggagaGACtttGACcgaGACcggGAC      60
                                                                 **

RS3     AGGTCACTTTCTAGAGATAGGAGAAGAGAAAGGTCTCTGTCTCGTGAGAGAAATCACAAG      120
RA3     aggGCCcttGCTagagataggagaagagaaaggGCCctgGCCcgtgagagaaatcacaag      65
RD3     aggGACcttGACagagataggagaagagaaaggGACctgGACcgtgagagaaatcacaag      120
        ***   ***   ********************   ***   ******************

RS3     CCGTCTCGATCCTTCTCTAGGTCTCGTAGCCGATCTAGGTCAAATGAAAGGAAATAG  177
RA3     ccgGCCcgaGCAttcGCCaggGCCcgtGCCcgaGCTaggGCCaatgaaaggaaatag  122
RD3     ccgGACcgaGACttcGACaggGACcgtGACcgaGACaggGACaatgaaaggaaatag  177
        ***   ***   ***   ***   ***  ****   ***   ***************
```

B

```
RS3     RSPPPRRRSPRRRSFSRSRSRSLSRDRRRERSLSRERNHKPSRSFSRSRSRSRSNERK 58
RD3     RDPPPRRRDPRRRDFDRDRDRDLDRDRRRERDLDRERNHKPDRDFDRDRDRDRDNERK 58
RA3     -------------------ARALARDRRRERALARERNHKPARAFARARARARANERK 39
                          * * ******* * ******* * * * * * * ****
```

SuppFigure 6: Nucleotide and amino acid sequence alignment of phosphomimetics derived from the RS domain of SRSF3. ***A)*** *Nucleotide sequence alignment of the endogenous RS domain of SRSF3 (RS3) and the RD/RA phosphomimetics (RA3/RD3).* ***B)*** *Alignment of the translated sequences from A) for the respective phosphomimetics.*

A

```
RS7     AGCCGACGAAGAAGAAGCAGGTCACGATCTAGATCCCATTCCCGATCCAGGGGAAGGCGA     60
RA7     GCCCGACGAAGAAGAGCAAGGGCCCGAGCTAGAGCCCATGCCCGAGCAAGGGGAAGGCGA     60
RD7     GACCGACGAAGAAGAGACAGGGACCGAGACAGAGACCATGACCGAGACAGGGGAAGGCGA     60
          *************   ***   ***   ***  ****  ****   ************

RS7     TACTCTCGCTCCCGCAGCAGGAGCCGAGGACGGAGGTCAAGATCAGCATCTCCTCGCCGA     120
RA7     TACGCCCGCGCACGCGCTAGGGCCCGAGGACGGAGGGCTAGAGCCGCAGCACCTCGCCGA     120
RD7     TACGACCGCGACCGCGACAGGGACCGAGGACGGAGGGACAGAGACGCAGACCCTCGCCGA     120
        ***   ***   ***   ***  *************   ***   ***   *********

RS7     TCAAGGTCTGTGTCTCTTCGTAGATCAAGATCAGCTTCACTCAGAAGATCTAGGTCTGGT     180
RA7     GCCAGGGCCGTGGCTCTTCGTAGAGCCAGAGCCGCTGCCCTCAGAAGAGCAAGGGCCGGT     180
RD7     GACAGGGACGTGGACCTTCGTAGAGACAGAGACGCTGACCTCAGAAGAGACAGGGACGGT     180
           ***   ***   *********   ***   ***   *********   ***   ***

RS7     TCTATAATAGGATCGAGGTATTTCCAATCCCGCTCAAGGTCGAGATCAAGATCCAGGTCT     240
RA7     GCCATAATAGGAGCTAGGTATTTCCAAGCCCGCGCCAGGGCCAGAGCAAGAGCCAGGGCC     240
RD7     GACATAATAGGAGACAGGTATTTCCAAGACCGCGACAGGGACAGAGACAGAGACAGGGAC     240
           *********   ************  ****   ***   ***   ***  ****

RS7     ATTTCACGACCAAGAAGCAGCCGATCAAAATCCAGATCTCCATCTCCTAAAAGAAGTCGT     300
RA7     ATTGCTCGACCAAGAGCCGCACGAGCCAAAGCCAGAGCTCCAGCCCCTAAAAGAGCCCGT     300
RD7     ATTGACCGACCAAGAGACGACCGAGACAAAGACAGAGACCCAGACCCTAAAAGAGACCGT     300
        ***   *********  *   ***   ***  ****   ***   *********   ***

RS7     TCCCCATCAGGAAGTCCACACAGAAGTGCAAGTCCAGAAAGAATGGACTGA        351
RA7     GCCCCAGCAGGAGCCCCACACAGAGCCGCAGCTCCAGAAAGAATGGACTGA        351
RD7     GACCCAGACGGAGACCCACACAGAGACGCAGACCCAGAAAGAATGGACTGA        351
          ****   ***   *********   ***   ******************
```

B

```
RS7     SRRRRSRSRSRSHSRSRGRRYSRSRSRSRGRRSRSASPRRSRSVSLRRSRSASLRRSRSG     60
RA7     ARRRRARARARAHARARGRRYARARARARGRRARAAAPRRARAVALRRARAAALRRARAG     60
RD7     DRRRRDRDRDRDHDRDRGRRYDRDRDRDRGRRDRDADPRRDRDVDLRRDRDADLRRDRDG     60
         **** * * * * * ***** * * * **** * * *** * * *** * * *** * *

RS7     SIIGSRYFQSRSRSRSRSRSISRPRSSRSKSRSPSPKRSRSPSGSPHRSASPERMD   116
RA7     AIIGARYFQARARARARARAIARPRAARAKARAPAPKRARAPAGAPHRAAAPERMD   116
RD7     DIIGDRYFQDRDRDRDRDRDIDRPRDDRDKDRDPDPKRDRDPDGDPHRDADPERMD   116
         *** **** * * * * * * ***  * * * * *** * * * *** * *****
```

SuppFigure 7: Nucleotide and amino acid sequence alignment of phosphomimetics derived from the RS domain of SRSF7. ***A)*** *Nucleotide sequence alignment of the endogenous RS domain of SRSF7 (RS7) and the RD/RA phosphomimetics (RA7/RD7).* ***B)*** *Alignment of the translated sequences from A) for the respective phosphomimetics.*

A

```
SRSF3-endogenous      atgcatcgtgattcctgtcccttggattgtaaggtttatgtaggtaatcttggaaataat 60
SRSF3-RA              atgcatcgtgattcctgtcccttggattgtaaggtttatgtaggtaatcttggaaataat 60
SRSF3-RD              atgcatcgtgattcctgtcccttggattgtaaggtttatgtaggtaatcttggaaataat 60
                      ************************************************************

SRSF3-endogenous      ggaaacaagactgaattagaacgggcttttggctattatggaccactcagaagtgtgtgg 120
SRSF3-RA              ggaaacaagactgaattagaacgggcttttggctattatggaccactcagaagtgtgtgg 120
SRSF3-RD              ggaaacaagactgaattagaacgggcttttggctattatggaccactcagaagtgtgtgg 120
                      ************************************************************

SRSF3-endogenous      gttgctcgaaaccctcctggctttgctttcgtcgaatttgaggatccccgagatgctgct 180
SRSF3-RA              gttgctcgaaaccctcctggctttgctttcgtcgaatttgaggatccccgagatgctgct 180
SRSF3-RD              gttgctcgaaaccctcctggctttgctttcgtcgaatttgaggatccccgagatgctgct 180
                      ************************************************************

SRSF3-endogenous      gatgctgtccgggaactagatggaagaacactgtgtggctgccgtgtaagagtggaactg 240
SRSF3-RA              gatgctgtccgggaactagatggaagaacactgtgtggctgccgtgtaagagtggaactg 240
SRSF3-RD              gatgctgtccgggaactagatggaagaacactgtgtggctgccgtgtaagagtggaactg 240
                      ************************************************************

SRSF3-endogenous      tcgaatggtgaaaagagaagtcggaatcgtgggccgcctccctcttggggtcgtcgtcct 300
SRSF3-RA              tcgaatggtgaaaagagaagtcggaatcgtgggccgcctccctcttggggtcgtcgtcct 300
SRSF3-RD              tcgaatggtgaaaagagaagtcggaatcgtgggccgcctccctcttggggtcgtcgtcct 300
                      ************************************************************

SRSF3-endogenous      cgagatgattaccgcaggaggagtcctccacctcggcgcagatccccaagaaggagaagc 360
SRSF3-RA              cgagatgattaccgcaggagggcccctccacctcggcgcagagcaccaagaaggagagcc 360
SRSF3-RD              cgagatgattaccgcaggagggaccctccacctcggcgcagagacccaagaaggagagac 360
                      ********************   ******************   ************  *

SRSF3-endogenous      ttttcccgaagccggagcaggtcactttctagagataggagaagagaaaggtctctgtct 420
SRSF3-RA              tttgctcgagcccgggcaagggcccttgctagagataggagaagagaaagggccctggcc 420
SRSF3-RD              tttgaccgagaccgggacagggaccttgacagagataggagaagagaaagggacctggac 420
                      ***   ***  ****   ***   ***   ********************   ***

SRSF3-endogenous      cgtgagagaaatcacaagccgtctcgatccttctctaggtctcgtagccgatctaggtca 480
SRSF3-RA              cgtgagagaaatcacaagccggcccgagcattcgccagggcccgtgcccgagctagggcc 480
SRSF3-RD              cgtgagagaaatcacaagccggaccgagacttcgacagggaccgtgaccgagacagggac 480
                      **********************   ***   ***   ***   ***  ****   ***

SRSF3-endogenous      aatgaaaggaaatag 495
SRSF3-RA              aatgaaaggaaatag 495
SRSF3-RD              aatgaaaggaaatag 495
                      ***************
```

B

```
SRSF3-endogenous      MHRDSCPLDCKVYVGNLGNNGNKTELERAFGYYGPLRSVWVARNPPGFAFVEFEDPRDAA 60
SRSF3-RA              MHRDSCPLDCKVYVGNLGNNGNKTELERAFGYYGPLRSVWVARNPPGFAFVEFEDPRDAA 60
SRSF3-RD              MHRDSCPLDCKVYVGNLGNNGNKTELERAFGYYGPLRSVWVARNPPGFAFVEFEDPRDAA 60
                      ************************************************************

SRSF3-endogenous      DAVRELDGRTLCGCRVRVELSNGEKRSRNRGPPPSWGRRPRDDYRRRSPPPRRRSPRRRS 120
SRSF3-RA              DAVRELDGRTLCGCRVRVELSNGEKRSRNRGPPPSWGRRPRDDYRRRAPPPRRRAPRRRA 120
SRSF3-RD              DAVRELDGRTLCGCRVRVELSNGEKRSRNRGPPPSWGRRPRDDYRRRDPPPRRRDPRRRD 120
                      ********************************************** ****** ****

SRSF3-endogenous      FSRSRSRSLSRDRRRERSLSRERNHKPSRSFSRSRSRSRSNERK* 164
SRSF3-RA              FARARARALARDRRRERALARERNHKPARAFARARARARANERK* 164
SRSF3-RD              FDRDRDRDLDRDRRRERDLDRERNHKPDRDFDRDRDRDRDNERK* 164
                      * * * * * ******* * ******* * * * * * * *****
```

SuppFigure 8: Nucleotide and amino acid sequence alignment of phosphomimetics derived from full length SRSF3. A) *Nucleotide sequence alignment of the endogenous SRSF3 (SRSF3-endogenous) and the RD/RA phosphomimetics (SRSF3-RA/RD).* ***B)*** *Alignment of the translated sequences from A) for the respective phosphomimetics.*

A

```
SRSF7-endogenous      atgtcacgctacgggcggtatggaggagaaaccaaggtatatgttggtaacctgggaact    60
SRSF7-RA              atgtcacgctacgggcggtatggaggagaaaccaaggtatatgttggtaacctgggaact    60
SRSF7-RD              atgtcacgctacgggcggtatggaggagaaaccaaggtatatgttggtaacctgggaact    60
                      ************************************************************

SRSF7-endogenous      ggtgctggtaaaggagagttagaaagggcattcagttactatgggcccttaagaactgtg    120
SRSF7-RA              ggtgctggtaaaggagagttagaaagggcattcagttactatgggcccttaagaactgtg    120
SRSF7-RD              ggtgctggtaaaggagagttagaaagggcattcagttactatgggcccttaagaactgtg    120
                      ************************************************************

SRSF7-endogenous      tggattgccagaaatcctccaggattcgcctttgtggaatttgaagaccctagagatgca    180
SRSF7-RA              tggattgccagaaatcctccaggattcgcctttgtggaatttgaagaccctagagatgca    180
SRSF7-RD              tggattgccagaaatcctccaggattcgcctttgtggaatttgaagaccctagagatgca    180
                      ************************************************************

SRSF7-endogenous      gaggatgcagttcgaggattggatgggaaagtgatttgtggttctcgagtgagggttgaa    240
SRSF7-RA              gaggatgcagttcgaggattggatgggaaagtgatttgtggttctcgagtgagggttgaa    240
SRSF7-RD              gaggatgcagttcgaggattggatgggaaagtgatttgtggttctcgagtgagggttgaa    240
                      ************************************************************

SRSF7-endogenous      ctatcaacaggcatgcctcggagatctcgttttgataggccacctgcccgtcgtcccttt    300
SRSF7-RA              ctatcaacaggcatgcctcggagatctcgttttgataggccacctgcccgtcgtcccttt    300
SRSF7-RD              ctatcaacaggcatgcctcggagatctcgttttgataggccacctgcccgtcgtcccttt    300
                      ************************************************************

SRSF7-endogenous      gatcctaatgatagatgctatgagtgtggtgaaaagggacattatgcttatgactgtcat    360
SRSF7-RA              gatcctaatgatagatgctatgagtgtggtgaaaagggacattatgcttatgactgtcat    360
SRSF7-RD              gatcctaatgatagatgctatgagtgtggtgaaaagggacattatgcttatgactgtcat    360
                      ************************************************************

SRSF7-endogenous      cgctatagccgacgaagaagaagcaggtcacgatctagatcccattcccgatccagggga    420
SRSF7-RA              cgctatgcccgacgaagaagagccagggcacgagccagagctcatgcccgagccagggga    420
SRSF7-RD              cgctatgatcgacgaagaagagacagggatcgagacagagatcatgaccgagatagggga    420
                      ******   ************  ****   ***   ***   ***  ****   ******

SRSF7-endogenous      aggcgatactctcgctcccgcagcaggagccgaggacggaggtcaagatcagcatctcct    480
SRSF7-RA              aggcgatacgcacgcgcccgcgctagggcccgaggacggagggctagagccgcagcacct    480
SRSF7-RD              aggcgatacgaccgcgatcgcgacagggatcgaggacggagggacagagatgcagaccct    480
                      *********   ***   ***   ***   ************   ***   ***   ***

SRSF7-endogenous      cgccgatcaaggtctgtgtctcttcgtagatcaagatcagcttcactcagaagatctagg    540
SRSF7-RA              cgccgagccagggccgtggctcttcgtagagccagagccgctgccctcagaagagcaagg    540
SRSF7-RD              cgccgagatagggacgtggaccttcgtagagatagagacgctgatctcagaagagacagg    540
                      ******   ***   ***   *********   ***   ***   *********   ***

SRSF7-endogenous      tctggttctataataggatcgaggtatttccaatcccgctcaaggtcgagatcaagatcc    600
SRSF7-RA              gccggtgccataataggagctaggtatttccaagcccgcgccagggccagagcaagagcc    600
SRSF7-RD              gacggtgatataataggagacaggtatttccaagatgacgatagggacagagacagagac    600
                         ***   *********   ************     *   ***   ***   ***  *

SRSF7-endogenous      aggtctatttcacgaccaagaagcagccgatcaaaatccagatctccatctcctaaaaga    660
SRSF7-RA              agggccattgctcgaccaagagccgcacgagccaaagccagagctccagcccctaaaaga    660
SRSF7-RD              agggatattgaccgaccaagagacgatcgagacaaagatagagacccagatcctaaaaga    660
                      ***   ***   *********  *   ***   ***   ***   ***   *********

SRSF7-endogenous      agtcgttccccatcaggaagtccacacagaagtgcaagtccagaaagaatggactga       717
SRSF7-RA              gcccgtgccccagcaggagccccacacagagccgcagctccagaaagaatggactga       717
SRSF7-RD              gaccgtgatccagatggagacccacacagagacgcagacccagaaagaatggac---       714
                         ***   ***   ***   *********   ***   ***************
```

B

```
SRSF7-RA              MSRYGRYGGETKVYVGNLGTGAGKGELERAFSYYGPLRTVWIARNPPGFAFVEFEDPRDA    60
SRSF7-endogenous      MSRYGRYGGETKVYVGNLGTGAGKGELERAFSYYGPLRTVWIARNPPGFAFVEFEDPRDA    60
SRSF7-RD              MSRYGRYGGETKVYVGNLGTGAGKGELERAFSYYGPLRTVWIARNPPGFAFVEFEDPRDA    60
                      ************************************************************

SRSF7-RA              EDAVRGLDGKVICGSRVRVELSTGMPRRSRFDRPPARRPFDPNDRCYECGEKGHYAYDCH    120
SRSF7-endogenous      EDAVRGLDGKVICGSRVRVELSTGMPRRSRFDRPPARRPFDPNDRCYECGEKGHYAYDCH    120
SRSF7-RD              EDAVRGLDGKVICGSRVRVELSTGMPRRSRFDRPPARRPFDPNDRCYECGEKGHYAYDCH    120
                      ************************************************************

SRSF7-RA              RYARRRRARARARAHARARGRRYARARARARGRRARAAAPRRARAVALRRARAAALRRAR    180
SRSF7-endogenous      RYSRRRRSRSRSRSHSRSRGRRYSRSRSRSRGRRSRSASPRRSRSVSLRRSRSASLRRSR    180
SRSF7-RD              RYDRRRRDRDRDRDHDRDRGRRYDRDRDRDRGRRDRDADPRRDRDVDLRRDRDADLRRDR    180
                      ** **** * * * * * ***** * * * **** * * *** * * *** * * *** *

SRSF7-RA              AGAIIGARYFQARARARARARAIARPRAARAKARAPAPKRARAPAGAPHRAAAPERMD*     238
SRSF7-endogenous      SGSIIGSRYFQSRSRSRSRSRSISRPRSSRSKSRSPSPKRSRSPSGSPHRSASPERMD*    238
SRSF7-RD              DGDIIGDRYFQDDDRDRDRDRDIDRPRDDRDKDRDPDPKRDRDPDGDPHRDADPERMD*    238
                       * *** ****   * * * * * ***  * * * * *** * * * *** * ******
```

SuppFigure 9: Nucleotide and amino acid sequence alignment of phosphomimetics derived from full length SRSF7. ***A)*** *Nucleotide sequence alignment of the endogenous SRSF7 (SRSF7-endogenous) and the RD/RA phosphomimetics (SRSF7-RA/RD).* ***B)*** *Alignment of the translated sequences from A) for the respective phosphomimetics.*

9.3 Supplemental tables

SuppTable 2: Genes with significant changes in 3'UTR length upon KD of Srsf3 and Srsf7 quantified with DaPARS. *ΔPDUI > 5%, false discovery rate (FDR) < 10%. The complete data is provided on the attached CD-ROM.*

SuppTable 3: Genes with significant changes in 3'UTR length upon KD of Srsf3 or Srsf7 quantified with MISO. *Bayes factor (BF) > 5. The complete data is provided on the attached CD-ROM.*

SuppTable 4: Quantitative interactome of SRSF3-GFP. *No RNase treatment. 3 biological replicates. Control GFP-NLS. Normalized to GFP. The complete data is provided on the attached CD-ROM.*

SuppTable 5: Nucleotide sequences of RS domains of SRSF3 and SRSF7 and the respective phosphomimetics RA and RD. *The mutated codons are highlighted in red.*

Construct	Nucleotide sequence (5' – 3')
RS3	aggagtcctccacctcggcgcagatccccaagaaggagaagctttcccgaagccggagcaggtcactttctagagatag gagaagagaaaggtctctgtctcgtgagagaaatcacaagccgtctcgatccttctctaggtctcgtagccgatctaggtc aaatgaaaggaaatag
RA3	gggcaagggcccttgctagagataggagaagagaaagggcccctggcccgtgagagaaatcacaagccggcccgagcat tcgccagggcccgtgcccgagctagggccaatgaaaggaaatag
RD3	agggacctccacctcggcgcagagacccaagaaggagagactttgaccgagaccgggacagggaccttgacagagat aggagaagagaaagggacctggaccgtgagagaaatcacaagccggaccgagacttcgacagggaccgtgaccgaga cagggacaatgaaaggaaatag
RS7	agccgacgaagaagaagcaggtcacgatctagatcccattcccgatccaggggaaggcgatactctcgctcccgcagca ggagccgaggacggaggtcaagatcagcatctcctcgccgatcaaggtctgtgtctcttcgtagatcaagatcagcttcac tcagaagatctaggtctggttctataataggatcgaggtatttccaatcccgctcaaggtcgagatcaagatccaggtctat ttcacgaccaagaagcagccgatcaaaatccagatctccatctcctaaaagaagtcgttccccatcaggaagtccacaca gaagtgcaagtccagaaagaatggactga
RA7	gcccgacgaagaagagcaagggcccgagctagagcccatgcccgagcaaggggaaggcgatacgcccgcgcacgcgc taggcccgaggacggaggctagagccgcagcacctcgccgagccagggccgtggctcttcgtagagccagagccgct gccctcagaagagcaagggccggtgccataataggagctaggtatttccaagcccgcgccagggccagagcaagagcca gggccattgctcgaccaagagccgcacgagccaaagccagagctccagccctaaaagagcccgtgcccagcaggag ccccacacagagccgcagctccagaaagaatggactga
RD7	gaccgacgaagaagagacagggaccgagacagagaccatgaccgagacaggggaaggcgatacgaccgcgaccgcg acagggaccgaggacggaggacagagacgcagacgctcgccgagacagggacgtggaccttcgtagagacagagac gctgacctcagaagagacagggacggtgacataataggagacaggtatttccaagaccgcgacagggacagagacaga gacagggacattgaccgaccaagagacgaccgagacaaagacagagaccagaccctaaaagagaccgtgaccccaga cggagaccccacacagagacgcagaccagaaagaatggactga

SuppTable 6: Nucleotide sequences of SRSF3, SRSF3-RA, SRSF3-RS, SRSF7, SRSF7-RA, SRSF7-RD. *The mutated codons are highlighted in red.*

Construct	Nucleotide sequence (5' – 3')
SRSF3-RS	atgcatcgtgattcctgtcccttggattgtaaggtttatgtaggtaatcttggaaataatggaaacaagactgaattagaac gggcttttggctattatggaccactcagaagtgtgtgggttgctcgaaaccctcctggctttgctttcgtcgaatttgaggatc cccgagatgctgctgatgctgtccgggaactagatggaagaacactgtgtggctgccgtgtaagagtggaactgtcgaat ggtgaaaagagaagtcggaatcgtgggccgcctccctcttggggtcgtcgtcctcgagatgattaccgcaggaggagtcct ccacctcggcgcagatccccaagaaggagaagcttttcccgaagccggagcaggtcactttctagagataggagaagag aaaggtctctgtctcgtgagagaaatcacaagccgtctcgatccttctctaggtctcgtagccgatctaggtcaaatgaaag gaaatag
SRSF3-RA	atgcatcgtgattcctgtcccttggattgtaaggtttatgtaggtaatcttggaaataatggaaacaagactgaattagaac gggcttttggctattatggaccactcagaagtgtgtgggttgctcgaaaccctcctggctttgctttcgtcgaatttgaggatc cccgagatgctgctgatgctgtccgggaactagatggaagaacactgtgtggctgccgtgtaagagtggaactgtcgaat ggtgaaaagagaagtcggaatcgtgggccgcctccctcttggggtcgtcgtcctcgagatgattaccgcaggagggccct ccacctcggcgcagagcaccaagaaggagagcctttgctcgagcccgggcaagggcccttgctagagataggagaaga gaaagggccctggcccgtgagagaaatcacaagccggcccgagcattcgccagggcccgtgcccgagctagggccaatg aaaggaaatag
SRSF3-RD	atgcatcgtgattcctgtcccttggattgtaaggtttatgtaggtaatcttggaaataatggaaacaagactgaattagaac gggcttttggctattatggaccactcagaagtgtgtgggttgctcgaaaccctcctggctttgctttcgtcgaatttgaggatc cccgagatgctgctgatgctgtccgggaactagatggaagaacactgtgtggctgccgtgtaagagtggaactgtcgaat ggtgaaaagagaagtcggaatcgtgggccgcctccctcttggggtcgtcgtcctcgagatgattaccgcaggagggaccc tccacctcggcgcagagaccaagaaggagagactttgaccgagaccgggacagggaccttgacagagataggagaag agaaagggacctggaccgtgagagaaatcacaagccggaccgagacttcgacagggaccgtgaccgagacagggaca atgaaaggaaatag
SRSF7-RS	atgtcacgctacgggcggtatggaggagaaaccaaggtatatgttggtaacctgggaactggtgctggtaaaggagagtt agaaagggcattcagttactatgggcccttaagaactgtgtggattgccagaaatcctccaggattcgcctttgtggaattt gaagaccctagagatgcagaggatgcagttcgaggattggatgggaaagtgatttgtggttctcgagtgagggttgaact atcaacaggcatgcctcggagatctcgttttgataggccacctgcccgtcgtccctttgatcctaatgatagatgctatgagt gtggtgaaaagggacattatgcttatgactgtcatcgctatagccgacgaagaagaagcaggtcacgatctagatcccatt cccgatccaggggaaggcgatactctcgctcccgcagcaggagccgaggacggaggtcaagatcagcatctcctcgccg atcaaggtctgtgtctcttcgtagatcaagatcagcttcactcagaagatctaggtctggttctataataggatcgaggtatt tccaatcccgctcaaggtcgagatcaagatccaggtctatttcacgaccaagaagcagccgatcaaaatccagatctccat ctcctaaaagaagtcgttccccatcaggaagtccacacagaagtgcaagtccagaaagaatggactga
SRSF7-RA	atgtcacgctacgggcggtatggaggagaaaccaaggtatatgttggtaacctgggaactggtgctggtaaaggagagtt agaaagggcattcagttactatgggcccttaagaactgtgtggattgccagaaatcctccaggattcgcctttgtggaattt

	gaagaccctagagatgcagaggatgcagttcgaggattggatgggaaagtgatttgtggttctcgagtgagggttgaact atcaacaggcatgcctcggagatctcgttttgataggccacctgcccgtcgtccctttgatcctaatgatagatgctatgagt gtggtgaaaagggacattatgcttatgactgtcatcgctatgcccgacgaagaagagccagggcacgagccagagctcat gcccgagccaggggaaggcgatacgcacgcgcccgcgctagggcccgaggacggagggctagagccgcagcacctcgc cgagccagggccgtggctcttcgtagagccagagccgctgccctcagaagagcaagggccggtgccataataggagcta ggtatttccaagcccgcgccagggccagagcaagagccagggccattgctcgaccaagagccgcacgagccaaagcca gagctccagccctaaaagagcccgtgcccccagcaggagcccccacacagagccgcagctccagaaagaatggactga
SRSF7-RD	atgtcacgctacgggcggtatggaggagaaaccaaggtatatgttggtaacctgggaactggtgctggtaaaggagagtt agaaagggcattcagttactatgggcccttaagaactgtgtggattgccagaaatcctccaggattcgcctttgtggaattt gaagaccctagagatgcagaggatgcagttcgaggattggatgggaaagtgatttgtggttctcgagtgagggttgaact atcaacaggcatgcctcggagatctcgttttgataggccacctgcccgtcgtccctttgatcctaatgatagatgctatgagt gtggtgaaaagggacattatgcttatgactgtcatcgctatgatcgacgaagaagagacagggatcgagacagagatcat gaccgagataggggaaggcgatacgaccgcgatcgcgacagggatcgaggacggagggacagagatgcagaccctcg ccgagataggggacgtggacgcttcgtagagatagagacgctgatctcagaagagacagggacggtgatataataggagac aggtatttccaagatgacgatagggacagagacagagacagggatattgaccgaccaagagacgatcgagacaaagat agagaccagatcctaaaagagaccgtgatccagatggagaccccacacagagacgcagacccagaaagaatggac

SuppTable 7: Nucleotide sequences of the SRSF3-Chimera constructs. *The Zn-knuckle domain is highlighted in light green and the 27aa stretch is highlighted in grey.*

Construct	Nucleotide sequence (5' – 3')
SRSF3-Zn	tcctgtcccttggattgtaaggtttatgtaggtaatcttggaaataatggaaacaagactgaattagaacgggcttttggct attatggaccactcagaagtgtgtgggttgctcgaaaccctcctggctttgctttcgtcgaatttgaggatccccgagatgct gctgatgctgtccgggaactagatggaagaacactgtgtggctgccgtgtaagagtggaactgtcgaatggtgaaaagag aagtcggaatcgtgggccgcctccctcttggggtcgtcgtcctcgagatgattaccgcaggaggagtcctccacctcggcg cagacccctttgatcctaatgatagatgctatgagtgtggtgaaaagggacattatgcttatgactgtcatcgctattccccaa gaaggagaagcttttcccgaagccggagcaggtcactttctagagataggagaagagaaaggtctctgtctcgtgagaga aatcacaagccgtctcgatccttctctaggtctcgtagccgatctaggtcaaatgaaaggaaa
SRSF3-27aa	tcctgtcccttggattgtaaggtttatgtaggtaatcttggaaataatggaaacaagactgaattagaacgggcttttggct attatggaccactcagaagtgtgtgggttgctcgaaaccctcctggctttgctttcgtcgaatttgaggatccccgagatgct gctgatgctgtccgggaactagatggaagaacactgtgtggctgccgtgtaagagtggaactgtcgaatggtgaaaagag aagtcggaatcgtgggccgcctccctcttggggtcgtcgtcctcgagatgattaccgcaggaggagtcctccacctcggcg cagatccccaagaaggagaagcttttcccgaagccggagcaggtcactttctagagataggagaagagaaaggtctctgt ctcgtgagagaaataggtctgtgtctcttcgtagatcaagatcagcttcactcagaagatctaggtctggttctataatagg atcgaggtatttccaacacaagccgtctcgatccttctctaggtctcgtagccgatctaggtcaaatgaaaggaaa
SRSF3-Zn+27aa	tcctgtcccttggattgtaaggtttatgtaggtaatcttggaaataatggaaacaagactgaattagaacgggcttttggct attatggaccactcagaagtgtgtgggttgctcgaaaccctcctggctttgctttcgtcgaatttgaggatccccgagatgct gctgatgctgtccgggaactagatggaagaacactgtgtggctgccgtgtaagagtggaactgtcgaatggtgaaaagag aagtcggaatcgtgggccgcctccctcttggggtcgtcgtcctcgagatgattaccgcaggaggagtcctccacctcggcg cagacccctttgatcctaatgatagatgctatgagtgtggtgaaaagggacattatgcttatgactgtcatcgctattccccaa gaaggagaagcttttcccgaagccggagcaggtcactttctagagataggagaagagaaaggtctctgtctcgtgagaga aataggtctgtgtctcttcgtagatcaagatcagcttcactcagaagatctaggtctggttctataataggatcgaggtattt ccaacacaagccgtctcgatccttctctaggtctcgtagccgatctaggtcaaatgaaaggaaa

SuppTable 8: Genes with significant change in gene expression upon KD of Srsf3 or Srsf7 quantified with DESeq2. *Adjusted P-value > 0.1. The complete data is provided on the attached CD-ROM.*

SuppTable 9: Genes with significant changes in 3'UTR length upon KD of Cpsf6 quantified with DaPARS. *ΔPDUI > 5%, false discovery rate (FDR) < 10%. The complete data is provided on the attached CD-ROM.*

SuppTable 10: Genes with significant changes in 3'UTR length upon neural differentiation quantified with DaPARS. *ΔPDUI > 5%, false discovery rate (FDR) < 10%. The complete data is provided on the attached CD-ROM.*

SuppTable 11: Genes with significant change in gene expression upon neuronal differentiation quantified with DESeq2. *Adjusted P-value (padj) > 0.1. The complete data is provided on the attached CD-ROM.*

10 Danksagung

An allererster Stelle möchte ich mich bei Prof. Michaela Müller-McNicoll, PhD bedanken, dass sie mir die Chance geboten hat an diesem Projekt in ihrer frisch gegründeten Arbeitsgruppe zu forschen und meine Dissertation anzufertigen. Ich bedanke mich für das Vertrauen, dass sie mir entgegengebracht hat, um ihr bei dem Aufbau ihres Labors und Teams beiseite zu stehen. Michi, am tiefsten beeindruckt hat mich dein unermesslicher wissenschaftlicher Elan und die unendliche Faszination für RNA-bindende Proteine und alle damit verbundenen Methoden. Von diesem Wissen habe ich während unserer Diskussionen und Meeting jederzeit profitiert und gelernt. Ich möchte dir auch danken, dass du immer ein offenes Ohr für meine Ideen hattest und diese gefördert hast. Ich möchte mich auch dafür bedanken, dass du es mir ermöglicht hast meine Kenntnisse in Fortbildungen und auf Konferenzen zu erweitern und meine Ergebnisse und Hypothesen der RNA-Community zu präsentieren und zu diskutieren.

Bei Dr. Kathi Zarnack möchte ich mich zunächst herzlichst für die zusätzliche Betreuung meines Projekts und die Zweitbegutachtung dieser Thesis bedanken. Darüber hinaus danke ich ihr auch für den konstruktiven Input bei der Gestaltung und Auswertung der Experimente sowie ihrem unermüdlichen Elan mich in die Tiefen der Bioinformatik und Statistik einweisen zu wollen, was leider nur bedingt funktioniert hat. Vielen Dank auch für die unzähligen Erklärungen aller bioinformatischen Plots und Analysen, die uns in den Journal Clubs begegnet sind und uns - ohne dich - bis heute ein Rätsel geblieben wären.

Prof. Dr. Beatrix Suess möchte ich dafür danken, mir im Rahmen des SFB IRTG 902 Graduiertenkollegs als dritte Betreuerin beiseite zu stehen und den Fortschritt meiner Promotionszeit während der jährlichen Reports und im Rahmen der Summer/Winter-Schools zu bewerten.

Ich möchte mich ebenfalls bei Francois McNicoll, PhD dafür bedanken, dass er mir jederzeit mit seiner Erfahrung sowie mit Rat und Tat zur Seite stand, wenn es galt ein aufwendiges Experiment akkurat zu planen und alle notwendigen Kontrollen zu berücksichtigen. Die kritischen Fragen in den Labormeetings und Journal Clubs haben mir geholfen meine eigenen sowie fremde, publizierte Daten gewissenhafter zu betrachten und zu hinterfragen. Vielen Dank für deine Unterstützung bei den verschiedenen iCLIP-Experimenten und den unzähligen Stunden im Hot-Lab. Der Kampf um den höchsten monatlichen Kaffeekonsum mit dir fehlt ein wenig.

Ein ganz besonderer Dank geht an Camila, Marius und Benjamin: Ohne euch wären die Jahre im Labor deutlich langsamer und trister vergangen. Ich danke euch für die Diskussionen von Daten und experimentellen Ideen, sowie eure ehrliche Meinung während der Meetings. Besonders aber danke ich euch für die gegenseitige moralische Unterstützung in den stressigsten Phasen vor Konferenzen

und, dass ihr immer ein offenes Ohr hattet, wenn dieses benötigt wurde. Auch nach dem Ende unserer gemeinsamen Zeit im Labor haben Camila und Marius mich regelmäßig motiviert und angetrieben diese Thesis weiterzuschreiben, Danke hierfür. Und Marius, danke, dass wir hin und wieder in deiner Sauna die Seele baumeln lassen konnten, wenn auch viel zu selten.

Mein Dank gilt auch all unseren Labor-ManagerInnen: Anfisa, Sandra und Michal. Ihr habt mir ein Umfeld geschaffen, in dem ich problemfrei meine Experimente planen und ohne große Verzögerungen durchführen konnte. Der Dank gilt ebenfalls für alle StudentInnen, die mich in dieser Zeit begleitet haben und die ich begleiten sowie wissenschaftlich anleiten durfte. Durch jeden einzelnen von euch habe ich mein Projekt und die einzelnen Aspekte selbst besser verstanden. Dieses gilt insbesondere für Nicole, die ich auf diesem Weg über lange Zeit begleiten durfte.

An dieser Stelle möchte ich bei Susan, Samarth und vor allem Mario aus der Arbeitsgruppe von Dr. Kathi Zarnack bedanken. Mit eurem fundierten bioinformatischen Wissen habt ihr mir bei allen fortgeschrittenen Analysen und Fragestellungen in diesem Projekt sehr geholfen. Mir haben die Meetings und Sessions mit euch immer große Freude bereitet. Besonders bedanken möchte ich mich hierzu bei Mario. Ohne dich hätten wir es nicht geschafft, die verschiedenen Datensätze sinnvoll miteinander zu verknüpfen und die hier präsentierten Lehren zu ziehen.

Neben der Arbeit an meinem Projekt durfte ich mich noch als Student Speaker der IRTG 902 betätigen. Dass ich diese zusätzlichen Aufgaben mit voller Freude und sehr viel Spaß wahrnehmen konnte lag vor allen an der großartigen Zusammenarbeit mit Bianca, Anna, Marco und Michaela. Ich möchte mich bei Bianca bedanken, dass ich mich während der Vorbereitung und Durchführung der Veranstaltungen immer auf dich verlassen konnte. Bei Anna und Marco möchte ich mich für die Diskussionen auf dem kurzen Dienstweg, die Zusammenarbeit auf Augenhöhe und die breite Unterstützung bei den gemeinsam organsierten Events bedanken.

An dieser Stelle möchte ich auch der Deutschen Forschungsgemeinschaft und insbesondere dem Sonderforschungsbereich 902 „Molekulare Mechanismen der RNA-basierten Regulation“ und seinen Mitgliedern für die Finanzierung dieser Forschungsarbeit danken. Zusammen mit dem IRTG 902 wurde mir eine großartige Plattform geboten, mein Wissen in den verschiedensten Aspekten der RNA-Biologie und zugehörigen Technologien zu erweitern und Kontakt zu anderen WissenschaftlerInnen zu schließen.

Der größte Dank geht an meine Familie. Ihr habt mich während des ganzen Studiums und der Promotion bedingungslos unterstützt und an mich geglaubt. Ohne euch hätte ich diesen Weg niemals bis zu diesem Punkt gehen können. Leider können nicht mehr alle das Ende der Promotion miterleben, was ich mich sehr gewünscht hätte. Aus diesem Grund möchte ich diese Thesis Jürgen, Sigrid und Ela

widmen. Hierin einschließen und hervorheben möchte ich Tanja. Du hast mich die ganzen Jahre motiviert weiterzumachen und mich aufgebaut, wenn es nötig war. Ohne dich hätte ich diese Thesis nicht beenden können. Ich danke dir, dass du einfach immer für mich da bist und es auch die ganzen Jahre auf großer Distanz warst. Zusätzlich möchte ich dir danken, dass du diese Thesis lektoriert hast.

Von ganzem Herzen möchte ich mich bei meiner liebsten Laborpartnerin Theresa bedanken, auch wenn wir schon seit mehr als sieben Jahren nicht mehr gemeinsam an der Laborbank standen. Wir haben gemeinsam das Biotechnologie-Studium in Braunschweig begonnen und beendet, und haben auch danach nie den Kontakt verloren. Ich danke dir dafür, dass du immer ein offenes Ohr für meine Probleme hattest und mich motivieren konntest diesen Weg bis zum Ende zu gehen.

Ebenfalls seit Beginn des Biotechnologie-Studiums begleitet mich die biotechnologische Studenteninitiative, die mir besonders in der Zeit bis zum Master-Abschluss die Möglichkeit geboten hat, mich persönlich und vor allem meine organisatorischen Soft-Skills weiter zu entwickeln. Darüber hinaus habe ich hierdurch viele neue Freundschaften schließen können. Stellvertretend möchte ich mich hier bei Gundula, Micha und Till bedanken.

Zum Abschluss möchte ich mich noch kurz bei Frank und seinem Team aus dem „Whisky For Life“ sowie den Whisky Enthusiasts Frankfurt für die herzliche Aufnahme in der Gruppe bedanken. Jeder Dram(-Talk) war etwas Besonderes und bot mir die Möglichkeit über etwas außerhalb der Forschung zu fachsimpeln sowie viel tiefer in die Materie des wohl schmackhaftesten Getränks auf Erden einzutauchen. Slàinte!

11 Erklärung

Ich erkläre hiermit, dass ich mich bisher keiner Doktorprüfung im mathematisch-naturwissenschaftlichem Bereich unterzogen habe.

Frankfurt am Main, den _____

Oliver Daniel Schwich

12 Versicherung

Ich versichere hiermit, dass die vorgelegte Doktorarbeit mit dem Titel „Same same but different: A comprehensive functional analysis of SRSF3 and SRSF7 in the regulation of alternative polyadenylation within the 3' untranslated region" selbständig und ohne unzulässige fremde Hilfe verfasst wurde, andere als die in ihr angegebene Literatur nicht benutzt wurde und, dass ich alle ganzen oder annährend übernommenen Textstellen, sowie verwendete Grafiken, Tabellen und Auswertungsprogramme gekennzeichnet habe. Außerdem versichere ich, dass die vorgelegte elektronische mit der schriftlichen Version der Doktorarbeit übereinstimmt.

Frankfurt am Main, den _____

Oliver Daniel Schwich