Investigating the mechanism of alternative splicing regulation of the RNA-binding proteins T-STAR and Sam68

Marina Danilenko

Institute of Genetic Medicine

A thesis submitted for the degree of

Doctor of Philosophy

July 2015

Declaration

I, Marina Danilenko, declare that no portion of the work compiled in this thesis has been submitted in support of another degree or qualification at this or any other University or Institute of Learning. This thesis includes nothing which is the work of others, nor the outcomes of work done in collaboration, except where otherwise stated.

……..

Marina Danilenko

Acknowledgements

First and foremost, I would like to thank my project supervisor, Professor David Elliott for his guidance and valuable support during whole time of my project. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist. I would also like to thank my second supervisor, Doctor Alison Tyson-Capper for her encouragement and helpful suggestions.

I would like to thank all members of David's lab, including Dr Sushma Grellsheid, Mrs Caroline Dalgliesh, Doctor Ingrid Ehrmann, Doctor Jennifer Munkley for helping and assisting me with the techniques. I would like to thank my friend Doctor Andrew Best, for his presence in my life in these years of my PhD.

I am also grateful to all collaborators including Doctor Cyril Dominguez (Department of Biochemistry, University of Leicester), Professor Chris Smith (Department of Biochemistry, University of Cambridge), Doctor Claudio Sette (Department of Biomedicine and Prevention, University of Rome Tor Vergata), Doctor Brian Wilson (Institute of Genetic Medicine, Newcastle University), Doctor Julian Venables (Institute of Genetic Medicine, Newcastle University) and Professor Tomaz Curk (Faculty of Computer and Information Science, University of Ljubljana).

My special thank you is to my family. Thank you to my dear parents. Their infinite love and care helped me to stay happy and never give up. Thank you to my dear brother Dmitrij for understanding and care. A special big thank you is to my dear boyfriend Denis for his love, care and attention. Throughout these years I always felt united with these four people and this thesis is for them.

Table of Contents

Declaration ..... **I**

Acknowledgements ..... **II**

Table of Contents ..... **III**

List of Figures ..... **VIII**

List of Tables ..... **XVII**

List of Abbreviations ..... **XVIII**

Abstract ..... **XX**

Chapter 1: Introduction **1**

1.1 Pre-mRNA splicing 1

1.1.1 Single gene can give rise to multiple mRNA molecules 1

1.1.2 Overview of pre-mRNA processing 2

1.1.3 Splicing and splice sites 3

1.1.4 Splice site recognition by the spliceosome 4

1.1.5 Splicing reaction and spliceosome cycle 5

1.1.6 The minor spliceosome 7

1.1.7 Alternative splicing 8

1.1.8 Sophisticated regulation of alternative splicing 9

1.2 Alternative splicing in the brain 15

1.2.1 The brain has a high frequency and a unique pattern of alternative splicing events 15

1.2.2 Splicing changes in neurological diseases 16

1.2.3 RNA-binding proteins in the brain 17

1.3 Techniques to study protein-RNA interactions 25

1.3.1 An overview 25

1.3.2 Minigene assays 26

1.3.3 In-vitro splicing assays....26
1.3.4 SELEX....27
1.3.5 Yeast three-hybrid....27
1.3.6 RNA-binding protein immunoprecipitation and Affinity tag assay....27
1.3.7 Alternative splicing microarrays....28
1.3.8 UV cross-linking and Immunoprecipitation (CLIP)....28
1.4 Research aims and objectives....29

Chapter 2: Materials and Methods....31

2.1 Electromobility Shift Assay (EMSA)....31
2.1.1 Construct cloning and protein preparation....31
2.1.2 Bradford assay to quantify protein concentrations....32
2.1.3 Verifying protein concentrations on the gel....33
2.1.4 Restriction digest....33
2.1.5 Transcription and labelling....33
2.1.6 RNA Precipitation....34
2.1.7 Checking RNA quality....35
2.1.8 Electromobility Shift Assay....36
2.2 Cell culture....36
2.2.1 HEK-293 cells....36
2.2.2 Cell culture conditions and maintenance....37
2.2.3 Cryopreservation....37
2.3 *Tomosyn-2* minigene cloning....37
2.3.1 Primer design....37
2.3.2 PCR amplification....38
2.3.3 Gel electrophoresis....39
2.3.4 Restriction digests....39
2.3.5 Ligation and Re-cleavage....40

2.3.6 Cloning and Sequencing 40
2.4 Site-directed mutagenesis 42
2.4.1 Primer design 42
2.4.2 PCR amplification 44
2.4.3 Gel Electrophoresis 44
2.5 Minigene splicing assays 45
2.5.1 Transfection 45
2.5.2 RNA extraction 45
2.5.3 RT-PCR 46
2.5.4 DNA size verification and product quantification 47
2.5.5 Western Immunoblotting 47
2.6 *Tomosyn-2* splicing analysis in brain tissues by RT-PCR 47

Chapter 3: Dissecting the mechanism of splicing of *Nrxn2* cassette exon by T-STAR 50

3.1 Introduction 50
3.2 The aims of this Chapter were: 52
3.3 Results 53
3.3.1 Both T-STAR and Sam68 bind *Nrxn2* target sequence directly 53
3.3.2 Splicing repression of the *Nrxn2* cassette exon depends on multiple redundant binding sites, as well as a general repressor element 56
3.3.3 Loss of two key groups of AU-rich sequence elements inhibits splicing control by T-STAR. 61
3.3.4 Splicing repression of *Nrxn2* can occur from upstream and downstream intronic locations 65
3.4 Discussion 69
3.4.1 Do T-STAR and Sam68 bind *Nrxn2* directly? 69

3.4.2 Splicing repression of the *Nrxn2* cassette exon depends on multiple redundant binding sites, as well as a general repressor element....71

3.4.3 Splicing repression of *Nrxn2* can occur from upstream and downstream intronic locations....74

3.5 Chapter Summary....77

Chapter 4: Dissecting the mechanism of splicing control of *Tomosyn-2* cassette exon by T-STAR and Sam68....78

4.1 Introduction....78

4.2 The aims of this Chapter are:....80

4.3 Results....80

4.3.1 T-STAR protein expression establishes regional splicing regulation of *Tomosyn-2* exon in the mouse brain....80

4.3.2 *Tomosyn-2* splicing is repressed by T-STAR and Sam68....83

4.3.3 *Tomosyn-2* sites are redundant for splicing regulation by T-STAR and Sam68....85

4.3.4 *Tomosyn-2* splicing is repressed by T-STAR and Sam68 with binding sites being upstream of the exon....88

4.4 Discussion....90

4.4.1 T-STAR splicing is regionally repressed in the mouse brain....90

4.4.2 *Tomosyn-2* is a direct target for T-STAR and Sam68....91

4.4.3 *Tomosyn-2* splicing is highly repressed by T-STAR from an upstream and downstream intronic locations....92

4.5 Chapter Summary....92

Chapter 5: Why is there T-STAR specific regulation of *Neurexin-2* splicing?....93

5.1 Introduction....93

5.2 The aims of this Chapter are:....94

5.3 Results 95

5.3.1 Sam68 regulates *Nrxn2* splicing when *Nrxn2* sequence is replaced by a different A/U-rich regulatory element 95

5.3.2 Can the *Tomosyn-2* downstream response element confer Sam68 regulation on T-STAR? 101

5.3.3 Other STAR proteins also regulate *Nrxn2* splicing 104

5.4 Discussion 107

5.4.1 Can Sam68 regulate *Nrxn2* splicing if the sequence is altered? 107

5.4.2 Will Sam68 regulate *Nrxn2* splicing if the *Nrxn2* sequence is replaced by the downstream region of *Tomosyn-2*? 108

5.4.3 *Nrxn2* AS4 splicing is repressed by Quaking 109

5.5 Chapter summary 110

Chapter 6: General discussion and future work 111

6.1 General discussion 111

6.2 Future work 116

Bibliography 118

Appendix I 130

Appendix II 135

List of Figures

FIGURE 1.1 – POST-TRANSCRIPTIONAL MRNA PROCESSING RESULT IN MULTIPLE MRNAS FROM A SINGLE GENE. A SCHEMATIC, SHOWING POST-TRANSCRIPTIONAL RNA PROCESSING MECHANISMS A) ALTERNATIVE SPLICING, B) ALTERNATIVE POLYADENYLATION, C) RNA EDITING. IMAGE IS ADAPTED FROM (SIOMI AND DREYFUSS, 1997). 2

FIGURE 1.2 – CONSERVED SEQUENCE ELEMENTS RECOGNIZED BY THE MAJOR SPLICEOSOME. FIGURE IS SHOWING FOUR CONSENSUS ELEMENTS REQUIRED FOR INTRON RECOGNITION: THE 5' SPLICE SITE, BRANCH POINT, POLYPYRIMIDINE TRACT AND THE 3' SPLICE SITE. R CORRESPONDS TO PURINE, Y – TO PYRIMIDINE AND N TO ANY NUCLEOTIDE. IMAGE IS ADAPTED FROM (WAHL *ET AL.*, 2009). 4

FIGURE 1.3 – TWO-STEP TRANSESTERIFICATION REACTIONS, HAPPENING DURING SPLICING. STEP 1: THE 2'OH OF THE BRANCHPOINT ADENOSINE (BLUE RHOMBUS) IS ATTACKING THE PHOSPHODIESTER BOND OF THE 5' SPLICE SITE. AS A RESULT, LARIAT INTRON STRUCTURE IS FORMED. STEP 2: THE 3'OH GROUP OF THE EXCISED EXON ATTACKS THE 3' SPLICE SITE. EXONS ARE JOINED TOGETHER BY SPLICING, WHILE THE INTRON LARIAT PRODUCT IS DEGRADED. FIGURE IS ADAPTED FROM (FABRIZIO AND LÜHRMANN, 2012). 5

FIGURE 1.4 – SPLICEOSOME ASSEMBLY CYCLE. THIS SCHEMATIC SHOWS HOW THE MAJOR SPLICEOSOME ASSEMBLES AND DISASSEMBLES IN A SERIES OF STEPS. IN ADDITION TO SMALL RIBONUCLEOPROTEINS, SPECIFIC ENZYMES ARE SHOWN TO BE INVOLVED IN THE PROCESS. FIGURE IS TAKEN FROM WAHL ET AL, 2009. 6

FIGURE 1.5 – DIFFERENT TYPES OF ALTERNATIVE SPLICING EVENTS. CONSTITUTIVE EXONS ARE ALWAYS INCLUDED INTO THE FINAL MRNA. CASSETTE EXONS (ALSO TERMED ALTERNATIVE EXONS) CAN BE INDEPENDENTLY INCLUDED OR EXCLUDED FROM THE FINAL TRANSCRIPT. ALTERNATIVE 3' AND 5' SPLICE SITES ARE COMPETITIVE WITH ONE BEING SELECTED BY THE SPLICEOSOME. ALTERNATIVE SPLICE SITE SELECTION CAN ALTER EXON LENGTH. RETAINED INTRONS ARE NOT EXCISED FROM THE PRE-MRNA. MUTUALLY EXCLUSIVE EXONS ARE EXONS THAT ARE INCLUDED OR EXCLUDED IN A COORDINATED MANNER. THE USE OF ALTERNATIVE PROMOTERS AND ALTERNATIVE POLYADENYLATION CAN ALSO ALTER THE EXONS INCLUDED AND THE LENGTH OF THE FINAL TRANSCRIPT. ALTERNATIVELY SPLICED EXONS ARE MARKED IN ORANGE, WHILE CONSTITUTIVE EXONS ARE COLOURED IN BLUE (FIGURE IS ADOPTED FROM (MATLIN *ET AL.*, 2005; MILLS AND JANITZ, 2012). 9

FIGURE 1.6 – SPLICING REGULATORY ELEMENTS CAN ENHANCE AND REPRESS THE SPLICE SITE SELECTION BY THE SPLICEOSOME. ESE/ISE (EXONIC SPLICING ENHANCERS AND INTRONIC SPLICING ENHANCERS) RECRUIT THE SPLICING ACTIVATORS, WHICH IN TURN PROMOTE SPLICEOSOME ASSEMBLY ON SPLICE SITES. ESS/ISE (EXONIC SPLICING SILENCES AND INTRONIC SPLICING SILENCERS) ATTRACT SPLICING REPRESSORS AND INHIBIT SPLICEOSOME ASSEMBLY. FIGURE IS ADOPTED FROM (SREBROW AND KORNBLIHTT, 2006). 11

FIGURE 1.7 – DOMAIN STRUCTURE OF STAR PROTEINS. ALL STAR PROTEINS HAVE A STAR DOMAIN THAT CONSISTS OF KH DOMAIN, FLANKED BY QUA1 AND QUA2 DOMAINS AND REQUIRED FOR RNA BINDING. IN ADDITION, OTHER MOTIFS ARE PRESENT. SF1 DOES NOT HAVE QUA1 DOMAIN. INSTEAD, U2AF DOMAIN (SHADED LIGHT BLUE) IS REQUIRED FOR RNA BINDING. PROLINE RICH SEQUENCES (LABELLED AS PRO) FOR PROTEIN-PROTEIN INTERACTIONS. THE RG REGIONS (IN GREY) ARE REQUIRED FOR METHYLATION. THE TYROSINE-RICH DOMAIN IS SUBJECT TO PHOSPHORYLATION. THERE IS ALSO A NUCLEAR LOCALIZATION DOMAIN IS PRESENT IN THE C TERMINUS OF ALL PROTEINS (NOT SHOWN ON THE FIGURE). THIS FIGURE IS ADAPTED FROM (STOSS *ET AL.*, 2004). 18

FIGURE 1.8 – *NEUREXIN-2* SPLICING PATTERN. *NEUREXIN-2* IS ALTERNATIVELY SPLICED INTO 6 SPLICE VARIANTS GENERATED BY 13 SPLICING EVENTS. CASSETTE EXON (AS4) REGULATED BY T-STAR IS LABELLED BY THE RED RECTANGLE. THE FIGURE IS ADAPTED FROM UCSC GENOME BROWSER. ... 23

FIGURE 1.9 – *TOMOSYN-2* SPLICING PATTERN. *TOMOSYN-2* IS ALTERNATIVELY SPLICED INTO 12 SPLICE VARIANTS CREATED BY 11 SPLICING EVENTS. CASSETTE EXON REGULATED BY T-STAR IS LABELLED BY THE RED RECTANGLE. THE FIGURE IS ADAPTED FROM UCSC GENOME BROWSER. 25

FIGURE 2.1 – THE DISTRIBUTION OF *NEUREXIN* SEQUENCES USED FOR GEL SHIFT CONSTRUCTS TO MAKE. A – *NEUREXIN-1* CASSETTE EXON IS HIGHLIGHTED IN BLACK, REGION 1 USED IN GEL SHIFT EXPERIMENT IS HIGHLIGHTED IN GREY. B - *NEUREXIN-2* CASSETTE EXON IS HIGHLIGHTED IN BLACK, REGION 2 AND A CONTROL REGION USED IN GEL SHIFT EXPERIMENT ARE HIGHLIGHTED IN GREY. REGION 2 PARTIALLY COVERS THE EXON. 31

FIGURE 2.2 - *TOMOSYN-2* SEQUENCE USED FOR MINIGENE CLONING. FIGURE SHOWS THE *TOMOSYN-2* REGION THAT WAS PREDICTED TO RESPOND TO T-STAR AND WAS CLONED INTO PXJ41 VECTOR. CASSETTE EXON IS HIGHLIGHTED IN BLACK. SEQUENCES FOR CLONING PRIMERS ARE UNDERLINED. 38

FIGURE 2.3 - SEQUENCE REGIONS ALTERED THROUGH SITE-DIRECTED MUTAGENESIS. A – *NEUREXIN-2* SEQUENCE REGION THAT WAS CLONED INTO THE PXJ41 VECTOR AND SUBSEQUENTLY SUBJECTED FOR MUTAGENESIS. EXON IS HIGHLIGHTED IN BLACK. T-STAR TARGET REGION IS HIGHLIGHTED IN BLUE. NUMBERS HIGHLIGHTED IN BLUE SHOW THE DOWNSTREAM (163 AND 241 NUCLEOTIDES DOWNSTREAM THE TARGET EXON) AND UPSTREAM (42 AND 100 NUCLEOTIDES UPSTREAM THE TARGET EXON) POSITIONS FOR THE TARGET REGION INSERTION. B – *TOMOSYN-2* SEQUENCE REGION THAT WAS CLONED INTO THE PXJ41 VECTOR AND SUBSEQUENTLY SUBJECTED FOR MUTAGENESIS. EXON IS HIGHLIGHTED IN BLACK. POTENTIAL T-STAR TARGET REGIONS ARE HIGHLIGHTED IN BLUE AND PURPLE. NUMBER HIGHLIGHTED IN BLUE SHOW AN UPSTREAM POSITION FOR THE FIRST (BLUE) TARGET REGION INSERTION (54 BASES UPSTREAM THE EXON). . 42

FIGURE 2.4 - SITE-DIRECTED MUTAGENESIS. A SCHEMATIC SHOWING SITE-DIRECTED MUTAGENESIS BY OVERLAP-EXTENSION USING PCR. MUTAGENIC PRIMERS WITH A NUMBER OF SINGLE BASE CHANGES WERE DESIGNED TO CREATE SPECIFIC MUTATIONS. THESE PRIMERS WERE USED IN PCR 1 AND PCR 2 REACTIONS, WHICH PRODUCED TWO OVERLAPPING DNA FRAGMENTS WITH COMPLEMENTARY MUTATIONS IN OPPOSITE DNA STRANDS. BOTH PRODUCTS WERE THEN

COMBINED AND THEN RE-AMPLIFIED BY PCR3. THE FINAL PRODUCT WAS THEN CLONED INTO PXJ41 VECTOR. 43

FIGURE 3.1 – *NEUREXIN-2* SPLICING IS SPECIFICALLY REPRESSED BY T-STAR BUT NOT BY SAM68. DIFFERENT VERSIONS OF *NEUREXIN-2* MINIGENE EXON WERE CO-TRANSFECTED WITH GFP-ONLY, T-STAR, SAM68, T-STAR AND SAM68 MUTANTS, AS WELL AS T-STAR AND SAM68 TOGETHER. TRANSFECTION EFFICIENCY WAS ASSESSED BY THE PRESENCE OF FLUORESCENCE IN HEK293 CELL NUCLEI, USING THE FLUORESCENT MICROSCOPE. TRANSFECTED CELLS WERE HARVESTED AND SUBJECTED TO RNA EXTRACTION. RNA WAS REVERSE-TRANSCRIBED INTO CDNA. TRANSFECTED *NEUREXIN-2* CDNA WERE AMPLIFIED APPLYING PXJ41 VECTOR PRIMERS. THE PCRS WERE LOADED ON QIAXCEL, AS A RESULT, THE GEL PICTURES AND CONCENTRATIONS FOR INCLUDED/EXCLUDED BANDS WERE OBSERVED. FIGURE SHOWS THE GEL PICTURE AND THE GRAPHS FOR THE PERCENTAGE OF SPLICING EXCLUSION (PSE%) OF AS4 IN *NEUREXIN-2* TRANSFECTED WITH DIFFERENT CONSTRUCTS. STATISTICAL SIGNIFICANCE WAS CALCULATED APPLYING THE INDEPENDENT TWO-SAMPLE T-TEST. ERROR BARS SHOW STANDARD ERRORS OF THE MEAN. FIGURE IS ADOPTED FROM (EHRMANN *ET AL.*, 2013). 51

FIGURE 3.2 – EMSA ASSAY IS PERFORMED IN A SERIES OF STEPS. A – IN THE EMSA ASSAY A RADIOLABELLED RNA PROBE IS MIXED WITH PURIFIED PROTEIN. THEN, THE NAKED RNA (AS A CONTROL) OR RNA-PROTEIN MIXTURE IS SUBJECTED TO GEL ELECTROPHORESIS. UNBOUND RNA IS EXPECTED TO MIGRATE FURTHER DOWN IN THE GEL, WHILE RNA BOUND BY THE PROTEIN WILL GENERATE A HEAVIER COMPLEX THAT WILL STOP EARLIER, PRODUCING A HIGHER SIZE BAND ON A GEL. B – THE SEQUENCES OF THE RNA PROBES USED FOR EMSA. SEQUENCES CODING FOR *NEUREXIN-1* AND *NEUREXIN-2* RESPONSE ELEMENTS EACH CONTAINING POTENTIAL SITES FOR PROTEIN BINDING (HIGHLIGHTED IN GREEN). A NEGATIVE CONTROL PROBE IS TAKEN FROM A FURTHER DOWNSTREAM SEQUENCE OF *NEUREXIN-2* THAT DOES NOT CONTAIN ANY POTENTIAL STAR PROTEIN BINDING SITES. 53

FIGURE 3.3. T-STAR AND SAM68 BIND THE *NEUREXIN-1* AND *NEUREXIN-2* RNAS DIRECTLY. A- A COOMASSIE-STAINED SDS-PAGE GEL, CONFIRMING THE EQUAL AMOUNTS OF T-STAR AND SAM68 WERE USED IN FOR THE GEL SHIFT ASSAY. B – A RADIOGRAPH FOR *NEUREXIN-2* SHIFT GEL SHIFT EXPERIMENT. LANE 0 CORRESPONDS TO THE NAKED RNA, LANES 1 TO 3 TO DIFFERENT CONCENTRATIONS OF SAM68 AND LANES 4 TO 6 TO DIFFERENT T-STAR CONCENTRATIONS. NAKED RNA DOES NOT PRODUCE ANY SHIFT, FORMING A BAND AT THE GEL BOTTOM, CORRESPONDING TO MIGRATION OF UNBOUND PROBE. SAM68 SHIFTS *NEUREXIN-2* EVEN AT 50 NG OF THE PROTEIN ADDED, WITH A COMPLETE SHIFT AT 200 NG OF THE PROTEIN, WHEN THE COMPLEX IS TOO HEAVY TO MOVE AND GETS STUCK IN THE WELL. T-STAR DETECTABLY SHIFTS THE PROBE WITH 50 NG OF THE PROTEIN, WHILE A COMPLETE SHIFT OCCURS WITH 100 NG. C - A RADIOGRAPH FOR *NEUREXIN-1* SHIFT GEL SHIFT EXPERIMENT; THE RESULTS ARE SIMILAR AS FOR *NEUREXIN-2*. D – EMSA EXPERIMENTS WERE CARRIED OUT WITH THE NEGATIVE CONTROL PROBE. 55

FIGURE 3.4 – A DIAGRAM OF *NRXN2* CASSETTE EXON AND FLANKING INTRON SEQUENCE, TAKEN FROM THE UCSC GENOME BROWSER. THE REGION HIGHLIGHTED IN BLACK AND LIMITED WITH RED LINES CORRESPONDS TO THE DOWNSTREAM SEQUENCE ELEMENT THAT CONTAINS MULTIPLE SITES FOR T-STAR BINDING. MUTATED NUCLEOTIDES ARE SHOWN IN RED. THE MUTANT CALLED "BIG MUTATION" WAS CREATED BY (EHRMANN *ET AL.*, 2013) AND SHOWED TO COMPLETELY ABOLISH T-STAR-MEDIATED EXON EXCLUSION. SEVEN DIFFERENT MUTANT MINIGENES 1-7 HAVE BEEN PRODUCED FROM THIS REGION. 57

FIGURE 3.5 – INDIVIDUAL MUTATIONS DOWNSTREAM OF THE *NRXN2* AS4 INDICATE THE REDUNDANCY BETWEEN SINGLE SITES FOR *NRXN2* SPLICING REPRESSION BY T-STAR. GRAPHS SHOW THE PERCENTAGE OF SPLICING EXCLUSION (PSE%) IN DIFFERENT VERSIONS OF *NRXN2* MINIGENE EXON CO-TRANSFECTED WITH GFP-ONLY, T-STAR AND SAM68. DATA REPRESENTS THE MEAN OF MINIMUM THREE BIOLOGICAL REPLICATES. IN EACH CASE, STATISTICAL SIGNIFICANCE WAS COMPARED BETWEEN HEK293 CELLS EXPRESSING GFP AND HEK293 CELLS EXPRESSING THE GFP FUSION PROTEINS BY MEANS OF INDEPENDENT TWO-SAMPLE T-TEST, WHERE *P<0.05, **P<0.01, ***P<0.001. 59

FIGURE 3.6 – WITHOUT CO-TRANSFECTION OF STAR PROTEINS LEVELS OF SPLICING REPRESSION OF MOST MUTANTS ARE NOT SIGNIFICANTLY DIFFERENT FROM THE WILD TYPE CONTROL. GRAPHS SHOW THE PERCENTAGE OF SPLICING EXCLUSION FOR DIFFERENT VERSIONS OF THE NRXN2 MINIGENE EXON CO-TRANSFECTED WITH GFP-ONLY. THE DATA REPRESENTS THE MEAN OF MINIMUM THREE BIOLOGICAL REPLICATES. STATISTICAL SIGNIFICANCE WAS COMPARED BETWEEN HEK293 CELLS EXPRESSING GFP IN WILD TYPE NRXN2 AND HEK293 CELLS EXPRESSING GFP IN MUTANT NRXN2 BY MEANS OF INDEPENDENT TWO-SAMPLE T-TEST, WHERE *P<0.05, **P<0.01, ***P<0.001. 60

FIGURE 3.7 – THE WESTERN BLOT CONFIRMS THAT GFP, T-STAR AND SAM68 SHOWED SIMILAR LEVELS OF EXPRESSION AFTER TRANSFECTION. FIGURE SHOWS WESTERN BLOT ANALYSIS OF HEK293 CELLS CO-TRANSFECTED WITH GFP-COUPLED FUSION PROTEINS AND WILD TYPE *NRXN2* MINIGENE. PROBING WAS PERFORMED FOR GFP AND ACTIN. 60

FIGURE 3.8 – A COMBINED MUTANT 1-4 IS STILL EFFICIENTLY REPRESSED BY T-STAR. GRAPHS SHOW THE PERCENTAGE OF SPLICING EXCLUSION IN A COMBINED MUTANT 1-4 OF *NRXN2* EXON. THE DATA REPRESENTS THE MEAN OF MINIMUM THREE BIOLOGICAL REPLICATES. STATISTICAL SIGNIFICANCE WAS COMPARED BETWEEN HEK293 CELLS EXPRESSING GFP AND HEK293 CELLS EXPRESSING THE GFP FUSION PROTEINS BY MEANS OF INDEPENDENT TWO-SAMPLE T-TEST,WHERE *P<0.05, **P<0.01, ***P<0.001. 62

FIGURE 3.9 – THE LOSS OF SITES 1-4 AND 5 INHIBITS THE *NRXN2* SPLICING REPRESSION BY T-STAR. GRAPHS SHOW THE PERCENTAGE OF SPLICING EXCLUSION IN WILD TYPE, AS WELL AS COMBINED MUTANTS M57 AND M145 OF *NRXN2* MINIGENE EXON TRANSFECTED WITH T-STAR, SAM68 AND GFP-ONLY CONTROL. THE DATA REPRESENTS THE MEAN OF MINIMUM THREE BIOLOGICAL REPLICATES. IN EACH CASE, STATISTICAL SIGNIFICANCE WAS COMPARED BETWEEN HEK293 CELLS

EXPRESSING GFP AND HEK293 CELLS EXPRESSING THE GFP FUSION PROTEINS BY MEANS OF INDEPENDENT TWO-SAMPLE T-TEST, WHERE *P<0.05, **P<0.01, ***P<0.001. 63

FIGURE 3.10 – FUSION-PROTEIN CONSTRUCTS CO-TRANSFECTED WITH *NRXN2* MUTANT 145 INTO HEK293 CELLS WERE EQUALLY EXPRESSED. WESTERN BLOT ANALYSIS OF HEK293 CELLS CO-TRANSFECTED WITH GFP-COUPLED FUSION PROTEINS AND *NRXN2* M145 MINIGENE. THE WESTERN BLOT WAS PROBED FOR GFP AND ACTIN, AND SHOWS EQUAL EXPRESSION LEVELS OF EACH FUSION PROTEIN. 64

FIGURE 3.11 – THE DELETION OF CUAA REPEAT REGION RESULTS IN A STRONGER *NRXN2* SPLICING REPRESSION BY T-STAR. GRAPHS SHOW THE PERCENTAGE OF SPLICING EXCLUSION IN WILD TYPE, AS WELL AS MUTANT WITH REPETITIVE REPRESSOR SITE DELETED, TRANSFECTED WITH T-STAR, SAM68 AND GFP-ONLY CONTROL. THE DATA REPRESENTS THE MEAN OF MINIMUM THREE BIOLOGICAL REPLICATES. IN EACH CASE, STATISTICAL SIGNIFICANCE WAS COMPARED BETWEEN HEK293 CELLS EXPRESSING GFP AND HEK293 CELLS EXPRESSING THE GFP FUSION PROTEINS BY MEANS OF INDEPENDENT TWO-SAMPLE T-TEST, WHERE *P<0.05, **P<0.01, ***P<0.001. 65

FIGURE 3.12 – T-STAR RESPONSE ELEMENT CAN STILL POTENTLY REPRESS SPLICING EVEN FROM MORE DISTANT DOWNSTREAM LOCATIONS. GRAPHS SHOW THE PERCENTAGE OF SPLICING EXCLUSION IN TRANSCRIPTS MADE FROM THREE MUTANTS WHERE THE *NRXN2* BINDING SITE IS MOVED: A) 100 BASES, B) 150 BASES, C) 250 BASE DOWNSTREAM THE TARGET *NRXN2* EXON, COMPARED TO THE WILD TYPE MINIGENE CONTROL. THE DATA REPRESENTS THE MEAN OF MINIMUM THREE BIOLOGICAL REPLICATES. IN EACH CASE, STATISTICAL SIGNIFICANCE WAS COMPARED BETWEEN HEK293 CELLS EXPRESSING GFP AND HEK293 CELLS EXPRESSING THE GFP FUSION PROTEINS BY MEANS OF INDEPENDENT TWO-SAMPLE T-TEST, WHERE *P<0.05, **P<0.01, ***P<0.001. 67

FIGURE 3.13 – OUTPUT OF A WILD TYPE *NRXN*2 SEQUENCE IN THE SROOGLE PROGRAM. A - FOR THE EXON START AND AN UPSTREAM INTRON; B - FOR THE EXON END AND THE DOWNSTREAM INTRON). SPLICE SITES, BRANCH SITE AND POLYPYRIMIDINE TRACT ARE MARKED WITH COLOURED DASH LINES, THAT MATCH BY COLOUR WITH THE NAMES IN A TABLE BELOW. THE T-STAR BINDING REGION IS MARKED BY A CONTINUOUS BLUE LINE ON A PICTURE B. 68

FIGURE 3.14 – T-STAR RESPONSE ELEMENT CAN STILL POTENTLY REPRESS SPLICING EVEN WHEN PLACED UPSTREAM OF THE REGULATED *NRXN2* EXON. GRAPHS SHOW THE PERCENTAGE OF SPLICING EXCLUSION FOR MOVING MUTANT: WHEN THE BINDING SITE IS MOVED A) 42 BASES, B) 100 BASES UPSTREAM OF THE TARGET *NRXN2* EXON, COMPARED WITH THE WILD TYPE MINIGENE CONTROL. THE DATA REPRESENTS THE MEAN OF MINIMUM THREE BIOLOGICAL REPLICATES. IN EACH CASE, STATISTICAL SIGNIFICANCE WAS COMPARED BETWEEN HEK293 CELLS EXPRESSING GFP AND HEK293 CELLS EXPRESSING THE GFP FUSION PROTEINS BY MEANS OF INDEPENDENT TWO-SAMPLE T-TEST, WHERE *P<0.05, **P<0.01, ***P<0.001. 69

FIGURE 3.15 – *NEUREXIN-2* HAS A CLUSTER OF T-STAR BINDING SITES, WHILE IN *NEUREXIN-1* AND *NEUREXIN-3* TARGET SITES ARE MORE DISPERSED. FIGURE SHOWS THE SEQUENCES OF THE TARGET EXONS REGULATED BY T-STAR (AND BY SAM68 FOR *NEUREXIN-1* AND *NEUREXIN-3*) AND

THE FLANKING INTRONS. EXONS ARE HIGHLIGHTED IN BLACK, WHILE THE POTENTIAL T-STAR AND SAM68 TARGET SITES IN THE DOWNSTREAM INTRONS ARE HIGHLIGHTED IN YELLOW. 71

FIGURE 3.16 – *NRXN2* TARGET EXON (IN RED FONT) WITH 150 BASES OF UPSTREAM AND DOWNSTREAM FLANKING EXON. DOWNSTREAM TARGET REGION IS HIGHLIGHTED IN GREY; POTENTIAL T-STAR BINDING SITES UPSTREAM ARE HIGHLIGHTED IN YELLOW. 75

FIGURE 4.1 – T-STAR REGULATES REGION-SPECIFIC SPLICING OF *NRXN2* AS4 IN THE BRAIN. FIGURES ADOPTED FROM PAPER BY EHRMANN ET AL 2013. A – THE WESTERN BLOT PERFORMED EARLIER IN THE LAB, SHOWING THE EXPRESSION PATTERN OF T-STAR AND SAM68 PROTEINS IN DIFFERENT WILD TYPE AND T-STAR KNOCKOUT MOUSE BRAIN REGIONS. B – SCHEMATIC REPRESENTING THE AREAS OF MOUSE BRAIN. C – *NRXN2* SPLICING PATTERN IN DIFFERENT BRAIN REGIONS. CLOSED COLUMNS LABELLED BY "+" SHOW THE *NEUREXIN-2* SPLICING PATTERN IN THE WILD TYPE MOUSE BRAIN. OPENED COLUMNS LABELLED BY "-" SHOW THE *NEUREXIN-2* SPLICING PATTERN IN THE T-STAR KNOCKOUT MOUSE BRAIN. DATA REPRESENTS THE MEAN OF MINIMUM THREE BIOLOGICAL REPLICATES. D - *TOMOSYN-2* SPLICING PATTERN IN WILD TYPE (WT), HETEROZYGOUS (HET) AND T-STAR KNOCKOUT (KO) MOUSE BRAIN. GRAPHS SHOW THE PERCENTAGE OF SPLICING EXCLUSION. DATA REPRESENTS THE MEAN OF MINIMUM THREE BIOLOGICAL REPLICATES. IN EACH CASE, STATISTICAL SIGNIFICANCE WAS COMPARED BETWEEN WILD TYPE BRAIN CELLS AND HETEROZYGOUS OR T-STAR KNOCKOUT BRAIN CELLS BY MEANS OF INDEPENDENT TWO-SAMPLE T-TEST, WHERE *P<0.05, **P<0.01, ***P<0.001..... 79

FIGURE 4.2 – T-STAR REGULATES REGION-SPECIFIC SPLICING OF *TOMOSYN-2* EXON 23 IN THE BRAIN. . 82

FIGURE 4.3 – T-STAR AND SAM68 REPRESS THE SPLICING OF *TOMOSYN-2* EXON 23 WHEN EXPRESSED FROM A MINIGENE. A – GRAPHS SHOW THE PERCENTAGE OF SPLICING EXCLUSION IN THE WILD TYPE *TOMOSYN-2* MINIGENE CO-TRANSFECTED WITH GFP-ONLY CONTROL AS WELL AS GFP-COUPLED T-STAR AND SAM68 FUSION PROTEINS. THE DATA REPRESENTS THE MEAN OF MINIMUM THREE BIOLOGICAL REPLICATES. IN EACH CASE, STATISTICAL SIGNIFICANCE WAS COMPARED BETWEEN HEK293 CELLS EXPRESSING GFP AND HEK293 CELLS EXPRESSING THE GFP FUSION PROTEINS BY MEANS OF INDEPENDENT TWO-SAMPLE T-TEST, WHERE *P<0.05, **P<0.01, ***P<0.001. B - WESTERN BLOT ANALYSIS OF HEK293 CELLS CO-TRANSFECTED WITH GFP-COUPLED FUSION PROTEINS AND WILD TYPE *NRXN2* MINIGENE. THE WESTERN BLOT WAS PROBED FOR GFP AND ACTIN AND SHOWS EQUAL EXPRESSION LEVELS OF EACH FUSION PROTEIN. 84

FIGURE 4.4 – DIAGRAM OF *TOMOSYN-2* CASSETTE EXON AND FLANKING INTRONS FROM THE UCSC GENOME BROWSER. 85

FIGURE 4.5 – *TOMOSYN-2* MUTANTS 1 AND 2 SHOW THE REDUCED PATTERNS OF SPLICING EXCLUSION BY T-STAR AND SAM68. GRAPHS SHOW THE PERCENTAGE OF EXON EXCLUSION FOR TWO *TOMOSYN-2* SEQUENCE MUTANTS AND THE WILD TYPE SEQUENCE. THE DATA REPRESENTS THE MEAN OF MINIMUM THREE BIOLOGICAL REPLICATES. IN EACH CASE, STATISTICAL SIGNIFICANCE WAS COMPARED BETWEEN HEK293 CELLS EXPRESSING GFP AND HEK293 CELLS EXPRESSING THE

GFP FUSION PROTEINS BY MEANS OF INDEPENDENT TWO-SAMPLE T-TEST, WHERE *P<0.05, **P<0.01, ***P<0.001. 86

FIGURE 4.6 – THE COMBINED *TOMOSYN-2* MUTANT HAD MUCH LESS EXON REPRESSION THAN EITHER SINGLE MUTANT, SUGGESTING THAT THE TWO MOTIFS DO COOPERATE. GRAPHS SHOW THE PERCENTAGE OF EXON EXCLUSION FOR *TOMOSYN-2* MRNAS ENCODED BY A CONSTRUCT IN WHICH SITES 1 AND 2 WERE JOINTLY MUTATED, COMPARED TO A WILD TYPE. EXPERIMENT FOR THE JOINT MUTANT WAS PERFORMED ONCE AND NEEDS REPLICATION. FOR THE WILD TYPE, STATISTICAL SIGNIFICANCE WAS COMPARED BETWEEN HEK293 CELLS EXPRESSING GFP AND HEK293 CELLS EXPRESSING THE GFP FUSION PROTEINS BY MEANS OF INDEPENDENT TWO-SAMPLE T-TEST, WHERE *P<0.05, **P<0.01, ***P<0.001..... 87

FIGURE 4.7 – A - OUTPUT OF A WILD TYPE *TOMOSYN-2* SEQUENCE IN A SROOGLE PROGRAM. SPLICE SITES, BRANCH SITE AND POLYPIRIMIDINE TRACT ARE MARKED WITH COLOURED DASH LINES THAT MATCH BY COLOUR WITH THE NAMES IN A TABLE BELOW. 88

FIGURE 4.8 – T-STAR AND SAM68 RESPONSE ELEMENT CAN STILL POTENTLY REPRESS SPLICING EVEN WHEN PLACED UPSTREAM OF THE REGULATED *TOMOSYN-2* EXON. THE PERCENTAGE OF SPLICING EXCLUSION IN MUTANT WITH TARGET SITES MOVED 54 BASES UPSTREAM OF THE TARGET *TOMOSYN-2* EXON, COMPARING WITH THE WILD TYPE MINIGENE CONTROL. THE DATA REPRESENTS THE MEAN OF MINIMUM THREE BIOLOGICAL REPLICATES. IN EACH CASE, STATISTICAL SIGNIFICANCE WAS COMPARED BETWEEN HEK293 CELLS EXPRESSING GFP AND HEK293 CELLS EXPRESSING THE GFP FUSION PROTEINS BY MEANS OF INDEPENDENT TWO-SAMPLE T-TEST, WHERE *P<0.05, **P<0.01, ***P<0.001..... 89

FIGURE 5.1 – SEQUENCE COMBINATIONS THAT MIGHT RESPOND TO SAM68. A SCHEMATIC REPRESENTING FIVE SEQUENCES (1-5) WHICH WERE INSERTED INTO *NRXN2* MUTANT WITH A BIG MUTATION TO TEST IF THESE WOULD ENABLE A RESPONSE TO SAM68. THE INSERTED SEQUENCE IS HIGHLIGHTED IN YELLOW AND THE OTHER MUTATIONS ARE SHOWN IN RED..... 95

FIGURE 5.2 – SAM68 STARTS REGULATING *NRXN2* SPLICING WHEN *NRXN2* SEQUENCE IS REPLACED BY THE STRING OF EIGHT UAAA MOTIFS. DOWNSTREAM BINDING SITES WERE INSERTED COMPRISING A) UAAA X 4 AND B) UAAA X 8 THAT WERE DESIGNED TO BE POTENTIALLY REGULATED BY SAM68. GRAPHS SHOW THE PERCENTAGE OF SPLICING EXCLUSION (PSE%) IN DIFFERENT VERSIONS OF THE *NRXN2* MINIGENE EXON CO-TRANSFECTED WITH GFP-ONLY, T-STAR AND SAM68. DATA REPRESENTS THE MEAN OF MINIMUM THREE BIOLOGICAL REPLICATES. IN EACH CASE, STATISTICAL SIGNIFICANCE WAS COMPARED BETWEEN HEK293 CELLS EXPRESSING GFP AND HEK293 CELLS EXPRESSING THE GFP FUSION PROTEINS BY MEANS OF INDEPENDENT TWO-SAMPLE T-TEST, WHERE *P<0.05, **P<0.01, ***P<0.001. 97

FIGURE 5.3 – SAM68 STARTS REGULATING *NRXN2* SPLICING WHEN *NRXN2* SEQUENCE IS REPLACED BY THE STRING OF FOUR UAAAA MOTIFS. DOWNSTREAM BINDING SITES WERE INSERTED COMPRISING A) UAAAA, AND B) UAAAA X 4 COMBINED WITH THE WILD TYPE *NRXN2* SEQUENCE. THESE WERE DESIGNED TO BE POTENTIALLY REGULATED BY SAM68. GRAPHS SHOW THE PERCENTAGE OF SPLICING EXCLUSION (PSE%) IN DIFFERENT VERSIONS OF *NRXN2* MINIGENE EXON

CO-TRANSFECTED WITH GFP-ONLY, T-STAR AND SAM68. DATA REPRESENTS THE MEAN OF MINIMUM THREE BIOLOGICAL REPLICATES. IN EACH CASE, STATISTICAL SIGNIFICANCE WAS COMPARED BETWEEN HEK293 CELLS EXPRESSING GFP AND HEK293 CELLS EXPRESSING THE GFP FUSION PROTEINS BY MEANS OF INDEPENDENT TWO-SAMPLE T-TEST, WHERE *P<0.05, **P<0.01, ***P<0.001. 99

FIGURE 5.4 – SAM68 DOES NOT REPRESS *NRXN2* THROUGH A STRING OF 22 ADENINE NUCLEOTIDES. A DOWNSTREAM BINDING SITE OF 22 ADENINES WAS INSERTED AS A POTENTIAL SAM68 REGULATORY SEQUENCE. GRAPHS SHOW THE PERCENTAGE OF SPLICING EXCLUSION (PSE%) IN DIFFERENT VERSIONS OF *NRXN2* A X 22 MUTANT MINIGENE CO-TRANSFECTED WITH GFP-ONLY, T-STAR AND SAM68. DATA REPRESENTS THE MEAN OF MINIMUM THREE BIOLOGICAL REPLICATES. STATISTICAL SIGNIFICANCE WAS COMPARED BETWEEN HEK293 CELLS EXPRESSING GFP AND HEK293 CELLS EXPRESSING THE GFP FUSION PROTEINS BY MEANS OF INDEPENDENT TWO-SAMPLE T-TEST, WHERE *P<0.05, **P<0.01, ***P<0.001..... 101

FIGURE 5.5 – A SCHEMATIC TO SHOW HOW *NRXN2* SEQUENCE WAS REPLACED BY *TOMOSYN-2* SEQUENCES. T-STAR AND SAM68 TARGET SITES IN *TOMOSYN-2* (LABELLED 1 AND 2) WERE INSERTED INTO *NRXN2* TARGET REGION MARKED BY RED BRACKETS. *NRXN2* MUTANT WITH A BIG MUTATION WAS USED A TEMPLATE IN CLONING. 102

FIGURE 5.6 – *NRXN2* DOES NOT SIGNIFICANTLY RESPOND TO SAM68 WHEN REPLACING *NRXN2* SEQUENCE BY T-STAR AND SAM68 REGULATORY ELEMENTS FROM *TOMOSYN-2*. GRAPHS SHOW THE PERCENTAGE OF SPLICING EXCLUSION IN *NRXN*2 WILD TYPE AND TWO MUTANTS WITH TARGET SEQUENCES REPLACED BY THE *TOMOSYN-2* TARGET A – REGION 1 AND B - REGION 2. DATA REPRESENTS THE MEAN OF MINIMUM THREE BIOLOGICAL REPLICATES. IN EACH CASE, STATISTICAL SIGNIFICANCE WAS COMPARED BETWEEN HEK293 CELLS EXPRESSING GFP AND HEK293 CELLS EXPRESSING THE GFP FUSION PROTEINS BY MEANS OF INDEPENDENT TWO-SAMPLE T-TEST, WHERE *P<0.05, **P<0.01, ***P<0.001. C - WESTERN BLOT ANALYSIS OF HEK293 CELLS CO-TRANSFECTED WITH GFP-COUPLED FUSION PROTEINS AND *NRXN2* MUTANT WITH TARGET REGION REPLACED BY SITE 1 FROM *TOMOSYN-2.* THE WESTERN BLOT WAS PROBED FOR GFP AND ACTIN AND SHOWS EQUAL EXPRESSION LEVELS OF EACH FUSION PROTEIN. 103

FIGURE 5.7 – QKI REPRESSES THE SPLICING OF *NRXN2* AS4 FROM THE DOWNSTREAM INTRONIC LOCATION. GRAPHS SHOW THE PERCENTAGE OF SPLICING EXCLUSION IN WILD TYPE *NRXN2* CO-TRANSFECTIONS WITH SLM1, SF1 AND QKI PROTEINS, AS WELL AS IN A CO-TRANSFECTION OF *NRXN*2 MUTANT WITH A BIG MUTATION WITH QKI. DATA REPRESENTS A SINGLE EXPERIMENTS, THUS RESULTS ARE PRELIMINARY. MORE REPLICATES IS NEEDED TO CALCULATE STATISTICAL SIGNIFICANCE. 105

FIGURE 5.8 – QKI REGULATION OF INDIVIDUAL MOTIFS IN *NRXN*2 AS4 IS SIMILAR TO T-STAR REGULATION OF THESE MOTIFS. A - GRAPHS SHOW THE PERCENTAGE OF SPLICING EXCLUSION (PSE%) IN DIFFERENT VERSIONS OF *NRXN2* MINIGENE CO-TRANSFECTED WITH GFP-ONLY AND QKI. DATA REPRESENTS THE MEAN OF MINIMUM THREE BIOLOGICAL REPLICATES. IN EACH CASE, STATISTICAL SIGNIFICANCE WAS COMPARED BETWEEN HEK293 CELLS EXPRESSING GFP AND

HEK293 CELLS EXPRESSING THE GFP FUSION PROTEINS BY MEANS OF INDEPENDENT TWO-SAMPLE T-TEST, WHERE *P<0.05, **P<0.01, ***P<0.001. B - WESTERN BLOT ANALYSIS OF HEK293 CELLS CO-TRANSFECTED WITH GFP-COUPLED QKI AND WILD TYPE *NRXN2* MINIGENE. THE WESTERN BLOT WAS PROBED FOR GFP AND ACTIN AND SHOWS EQUAL EXPRESSION LEVELS OF EACH FUSION PROTEIN. 106

List of Tables

TABLE 2.1: COMPONENTS OF BUFFER D 32
TABLE 2.2: CONDITIONS FOR RESTRICTION DIGEST OF *NEUREXIN* CONSTRUCTS FOR EMSA 33
TABLE 2.3: TRANSCRIPTION MIX FOR EMSA 33
TABLE 2.4: DNASE MIXTURE CONTENTS FOR EMSA 34
TABLE 2.5: COMPONENTS OF TRNA MIXTURE 34
TABLE 2.6: COMPONENTS OF 2X PK BUFFER 35
TABLE 2.7: POLYACRYLAMIDE GEL COMPONENTS FOR EMSA 35
TABLE 2.8: CONSTITUENTS OF 10X TBE BUFFER 35
TABLE 2.9: CONSTITUENTS OF BINDING MIX FOR EMSA 36
TABLE 2.10: PRIMERS USED FOR *TOMOSYN-2* CLONING 38
TABLE 2.11: PCR PROTOCOL 38
TABLE 2.12: THERMOCYCLER PCR PROGRAM 39
TABLE 2.13: COMPONENTS OF RESTRICTION DIGESTION REACTION 39
TABLE 2.14: COMPONENTS OF VECTOR DIGEST BY ANTARCTIC PHOSPHATASE 40
TABLE 2.15: STANDARD LIGATION PROTOCOL 40
TABLE 2.16: COMPONENTS OF RE-CLEAVAGE REACTION 40
TABLE 2.17: PXJ41 VECTOR PRIMERS USED FOR COLONY SCREENING AND FOR SEQUENCING 41
TABLE 2.18: THERMOCYCLER PCR PROGRAM FOR COLONY SCREENING 41
TABLE 2.19: PCR 1 AND PCR 2 PROTOCOL 44
TABLE 2.20: PCR 3 PROTOCOL 44
TABLE 2.21: TRANSFECTION PROTOCOL 45
TABLE 2.22: RT-PCR REACTION CONSTITUENTS 46
TABLE 2.23: RT-PCR REACTION STEPS 46
TABLE 2.24: RT-PCR REACTION STEPS AND CONSTITUENTS 48
TABLE 2.25: *TOMOSYN-2* PRIMERS USED FOR RT-PCRS IN DIFFERENT MOUSE BRAIN REGIONS 48
TABLE 2.26: PCR REACTION CONSTITUENTS 48
TABLE 2.27: THERMO CYCLER PCR PROGRAM 49

List of Abbreviations

BPS	Branch Point Sequence
cDNA	complementary DNA
CLIP	UV Cross-Linking and ImmunoPrecipitation
C-terminus	Carboxy-terminus
DEPC	diethylpyrocarbonate
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTPS	deoxynucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
EST	Expressed Sequence Tag
FBS	Fetal Bovine Serum
GFP	Green Fluorescent Protein
HEK-293	human embryonic kidney 293 cell line
HITS-CLIP	high-throughput sequencing of RNA isolated by CLIP
hnRNP	heterogeneous nuclear ribonucleoprotein
iCLIP	individual nucleotide resolution CLIP
IP	Immunoprecipitation
ISE	Intronic Splicing Enhancer
ISS	Intronic Splicing Silencer
MBNL1	Muscle-blind like protein
mRNA	messenger RNA
NMD	Nonsense-Mediated Decay
NMR	Nuclear Magnetic Resonance
ORF	Open Reading Frame
PAR-CLIP	PhotoActivatable-Ribonucleoside-enhanced CLIP
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PK	Proteinase K
PPT	Polypyrimidine Tract

PSE	Percentage Splicing Exclusion
PTB	Polypyrimidine Tract Binding protein
QKI	Quaking
RBP	RNA-binding protein
RNA	ribonucleic acid
RNA-seq	RNA sequencing
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
Sam68	Src-associated substrate during mitosis of 68 kDA
SDS	sodium dodecyl sulphate
SELEX	Systematic Evolution of Ligands by Exponential enrichment
snRNP	small nuclear ribonucleoprotein
SROOGLE	Splicing RegulatiOn Online GraphicaL Engine
STAR	Signal Transduction and Activation of RNA
TAE	Tris/Acetate/EDTA buffer
TBE	Tris/Borate/EDTA
TBST	Tris-Buffered Saline and Tween-20
UCSC	University of California, Santa Cruz
UTR	Untranslated region

Abstract

Alternative splicing is an important mechanism of pre-mRNA processing, regulated by many splicing factors. Some splicing factors as T-STAR, are poorly characterised due to absence of RNA targets identified. The first main aim of this study was to identify the principles for splicing regulation by T-STAR, by characterising its first identified physiological targets *Neurexin-2* and *Tomosyn-2*. The second important aim was to compare the T-STAR and Sam68 splicing regulation on example of *Neurexin-2* being a T-STAR specific target, and *Tomosyn-2* being regulated by both T-STAR and Sam68. By EMSA analyses, I identified that *Neurexin-2* is a direct target of both T-STAR and Sam68. By minigene assays and mutagenesis, I found that T-STAR response element in *Neurexin-2* is composite. Individually each AU-rich region is redundant for splicing regulation, but loss of two key groups of AU-rich sequence elements inhibits splicing control by T-STAR. Several splicing regulator proteins have been shown to follow a pattern where binding upstream of a regulated exon leads to splicing repression and binding downstream leads to splicing activation. The T-STAR response element starts just 13 nucleotides downstream of the regulated *Neurexin-2* exon, suggesting T-STAR protein binding might physically occlude the 5´splice site. Instead, I found this T-STAR response element can still potently repress splicing even from more distant downstream locations, and surprisingly even when placed upstream of the regulated exon. To find out how general this form of position-independent splicing regulation is I identified AU-rich downstream sequence elements that control splicing patterns of *Tomosyn-2* in response to both T-STAR and Sam68. These *Tomosyn-2* splicing response sequences similarly repressed splicing when moved upstream of the regulated exon. The T-STAR response element in *Neurexin-2* predominantly contained UUAA sequences, while the T-STAR/Sam68 response element in *Tomosyn-2* contained UAAA repeats. Conversion of the *Neurexin* response element to UAAA and UAAAA placed this exon under control of both Sam68 and T-STAR. Current data suggest that T-STAR and Sam68 proteins are unusual in that they repress splicing from either side of the exon. These proteins have subtly different target sequences that can enable them to control distinct patterns of target exons in the cell.

Chapter 1: Introduction

1.1 Pre-mRNA splicing

1.1.1 Single gene can give rise to multiple mRNA molecules

For many years it has been an enigma why the biological complexity of different organisms does not overlap with the number of genes they code. The human genome is known to contain approximately 21 000 protein-coding genes, while the nematode *Caenorhabditis elegans* with much simpler physiology has a genome of 19 000 protein-coding genes. Even the genome of a single-celled yeast *Saccharomyces cerevisiae* contains at least 6 000 protein-coding genes (Pennisi, 2005; Gerstein *et al.*, 2007).

This so-called gene number paradox was explained with discovery of how abundant post-transcriptional gene modifications are. It was known, that once DNA is started being transcribed to messenger RNA (mRNA), the generated primary transcript (pre-mRNA) immediately undergoes a series of processing events, including 5' capping, splicing, transcription termination and 3' polyadenylation (Matlin *et al.*, 2005). High frequencies of alternative splicing, alternative polyadenylation and RNA editing allow generating multiple RNAs from just one gene across the whole genome and thus highly expand the transcriptome (Figure 1.1). Generated mRNA variants encode functionally and structurally different proteins, causing an expansion of the proteome (Nilsen and Graveley, 2010). Alternative translation initiation and posttranslational modifications (such as phosphorylation, methylation, ubiquitylation, glycosylation, etc.) also help to increase the protein coding capacity of the genome (Wilhelm *et al.*, 2014).

Alternative splicing is proposed to be the major multiplication factor to expand the proteome. Around 95% of human genes are subjected to alternative splicing (Pan *et al.*, 2008). Majority of human genes generate two to three different mRNA isoforms (Djebali *et al.*, 2012). However, some produce much more, e.g. the *CD44*, encoding human CD44 glycoprotein, has 10 alternatively spliced exons, and might produce more than a 1000 alternative transcripts (Screaton *et al.*, 1992; Bell *et al.*, 1998).

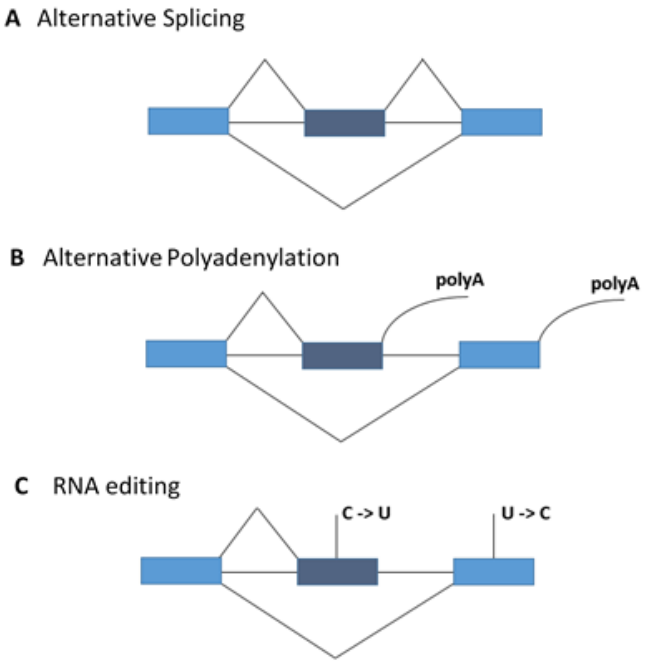


Figure 1.1 – Post-transcriptional mRNA processing result in multiple mRNAs from a single gene. A schematic, showing post-transcriptional RNA processing mechanisms A) alternative splicing, B) alternative polyadenylation, C) RNA editing. Image is adapted from (Siomi and Dreyfuss, 1997).

1.1.2 Overview of pre-mRNA processing

For DNA encoded information to be expressed, several important steps have to be completed. Initially, DNA is transcribed by RNA polymerase II to make a pre-mRNA (Hurwitz, 2005). This is followed by pre-mRNA maturation, by processing events including 5' capping, splicing, transcription termination and 3' polyadenylation. During 5' capping, which takes place after the first 25 nucleotides of the transcript have been generated, a guanine base and a ribose sugar are attached to the 5' end of the transcript. The guanine base is subsequently methylated. The function of 5' cap is to protect transcript from degradation by exonuclease enzymes (Mandal *et al.*, 2004). Following capping, splicing is initiated. Many proteins are involved in the splicing process, including small ribonucleoproteins (snRNPs), heterogenous nuclear ribonucleoproteins (hnRNPs) and other additional proteins (Matlin *et al.*, 2005). To finalize the processing the 3' end of the pre-mRNA is defined by cleavage and a poly(A) tail is added. Poly(A) tail is protecting RNA from being degraded and to provide RNA transfer to cytoplasm for subsequent translation (Lutz, 2008).

Initially, mRNA processing was thought to happen post-transcriptionally. However, in reality transcription and processing are not consecutive, but simultaneous and interdependent (Neugebauer, 2002). It takes just several minutes to transcribe the gene of an average length. The RNA sequencing of human B lymphocytes has shown, that splicing was by 65% completed after 5 minutes since transcription start (Windhager *et al.*, 2012). This, and many other evidence suggest, that mRNA processing is happening co-transcriptionally (Bentley, 2014). After RNA processing events, mature mRNA transcripts are transported from the nucleus to the cytoplasm, where they are translated into proteins. mRNA export is highly dependent on correct processing, i.e. transfer through nuclear pores requires recognition of a cap-binding complex (CBC), added during capping (Mandal *et al.*, 2004).

1.1.3 Splicing and splice sites

During transcription, both coding exons and introns are transcribed together into pre-mRNA. However, only exons are joined into mature mRNAs. Approximately 80% of human genome is transcribed. However, less than 1.5% are exons and as high as 26% correspond to introns (Gregory, 2005).

Intron transcription is an energy consuming process for the cell. For instance, *Dystrophin*, which is the largest human gene is transcribed for 16 hours, however, around 99% of *Dystrophin* primary transcript is comprised of introns (Tennyson *et al.*, 1995). Nevertheless, introns are crucial for the genome. Early experiments show that introns-containing viral constructs increase the gene expression up to 400 times higher, than intron-lacking ones (Buchman and Berg, 1988). Similar results were obtained in yeast and mammals (Juneau *et al.*, 2006; Shabalina *et al.*, 2010). Introns are also important in transcription, genome organization, NMD, nuclear export, translation yield, etc. (Carmel and Chorev, 2012).

The removal of non-coding introns and joining of exons into a final transcript is called splicing. Pre-mRNAs are cut at exon-intron junctions, termed 5' and 3' splice sites. In humans, the 5' splice site has a consensus sequence CAGGURAGU, where R is a purine (guanine or adenine) (Figure 1.2). Even though the consensus exists, splice site sequences are degenerate. In rare cases splice site sequence is represented by CAGGCRAGU, or by other nucleotide combinations. The 3' splice site is comprised of three elements, placed

within around 40 nucleotides upstream the intron/exon junction (Reed, 1996). The 3' splice site itself has a consensus YAGG , where Y is a pyrimidine (uracil or cytosine). Upstream of the 3' splice site there is a polypyrimidine tract (PPT). Commonly, the PPT is a string of uracil bases; rarely it is a run of cytosines. Further upstream there is a branchpoint sequence (BPS) YNYURAY, where adenine (A) is surrounded by a variable sequence (N can be any nucleotide) (Figure 1.2) (Valadkhan, 2007).

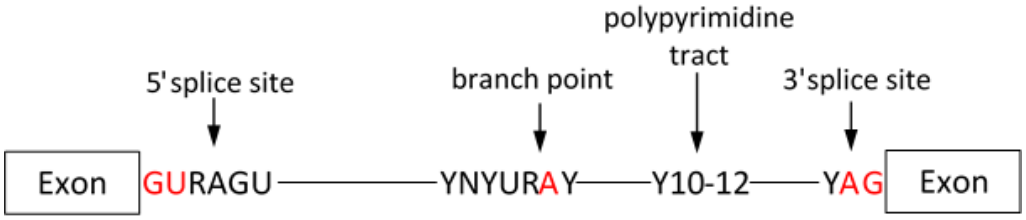


Figure 1.2 – Conserved sequence elements recognized by the major spliceosome. Figure is showing four consensus elements required for intron recognition: the 5' splice site, branch point, polypyrimidine tract and the 3' splice site. R corresponds to purine, Y – to pyrimidine and N to any nucleotide. Image is adapted from (Wahl *et al.*, 2009).

1.1.4 Splice site recognition by the spliceosome

Splicing is performed by a macromolecular complex, called the spliceosome. The spliceosome is a dynamic machinery that consists of small nuclear ribonucleoproteins (snRNPs), and a large number of additional splicing proteins and factors. SnRNPs are composed of Sm proteins (B/B', D1, D2, D3, E, F, G) and one or two small nuclear RNAs (snRNAs). U1 and U2 SnRNPs recognize the splice sites, and initiate splicing. By being complementary to pre-mRNA and interacting with each other, snRNPs also provide a framework on the pre-mRNA, holding the other spliceosome components together. There are five snRNPs: U1, U2, U4, U5 and U6 (Will and Lührmann, 2011). U1 is partially complementary to the 5' splice site, while U2 recognizes the branchpoint sequence (Wahl *et al.*, 2009).

1.1.5 Splicing reaction and spliceosome cycle

Briefly, the splicing reaction starts with exon/intron recognition, followed by spliceosome assembly and two transesterification reactions, by which the intron is spliced out (Wahl *et al.*, 2009) (Figure 1.3).

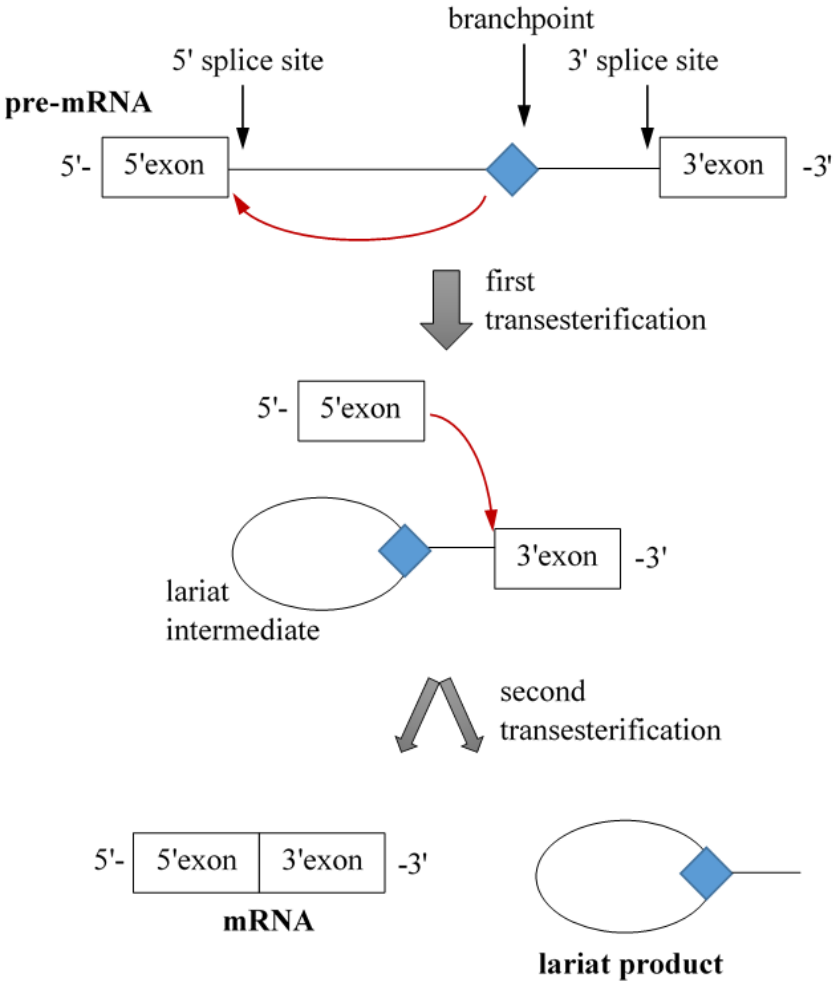


Figure 1.3 – Two-step transesterification reactions, happening during splicing. Step 1: the 2'OH of the branchpoint adenosine (blue rhombus) is attacking the phosphodiester bond of the 5' splice site. As a result, lariat intron structure is formed. Step 2: the 3'OH group of the excised exon attacks the 3' splice site. Exons are joined together by splicing, while the intron lariat product is degraded. Figure is adapted from (Fabrizio and Lührmann, 2012).

Within an assembled spliceosome, splicing occurs by two transesterification reactions. In the first reaction, the 2'OH group of an adenosine nucleotide in the branch point sequence (situated close to 3' splice site) attacks the phosphodiester bond of the 5' splice site, and forms a new phosphodiester bond between the 2'OH and 5' splice site. As a result, the 5' exon is released and the 3'exon forms a lariat intermediate with the intron. In the second

reaction, the free 3'OH group of the released 5'exon attacks the phosphodiester bond at the 3'splice site. Consequently, both free exons are ligated with a new phosphodiester bond and intron is released as a lariat. The spliceosome components dissociate and are recycled (Moore and Sharp, 1993; Sharp, 1994; Wahl *et al.*, 2009).

Splicing is a stepwise reaction. In different steps of the splicing reaction, different protein complexes form. These are the: H, E, A, B and C complexes, listed in order of their assembly (Figure 1.4). Straight after transcription, the H complex forms. It consists of pre-mRNA, bound by a group of RNA-binding proteins, called heterogeneous nuclear ribonucleoproteins (hnRNPs). Splicing starts when the E complex forms, when the 5' consensus splice site is recognized by the U1 snRNP. This is an ATP-dependent reaction. The 3' splice site is recognized by a set of proteins: SF1 (splicing factor 1) interacts with the branchpoint; U2AF65 (U2 auxiliary factor 65kDa subunit) recognizes polypyrimidine tract and U2AF35 (U2 auxiliary factor 35kDa subunit) recognizes the 3' splice site consensus sequence.

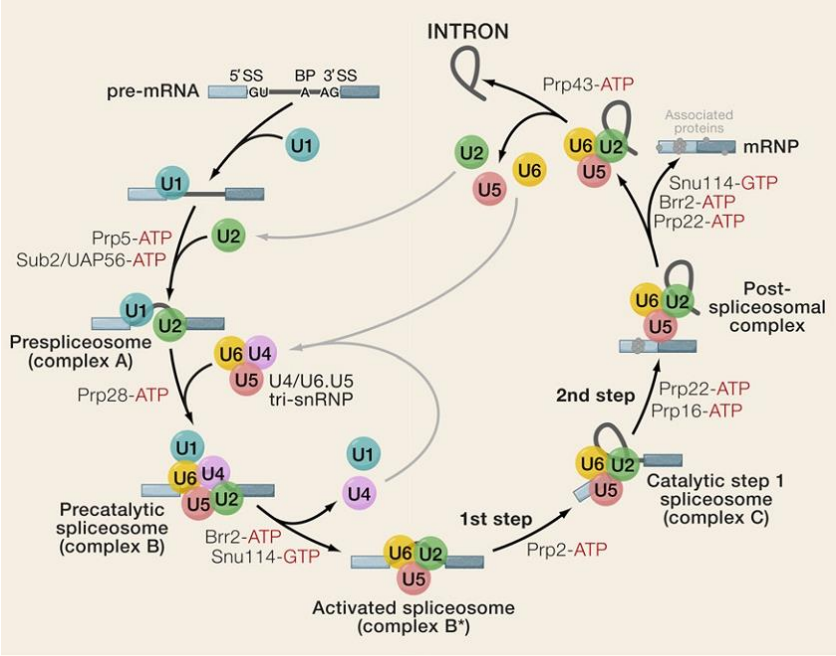


Figure 1.4 – Spliceosome assembly cycle. This schematic shows how the major spliceosome assembles and disassembles in a series of steps. In addition to small ribonucleoproteins, specific enzymes are shown to be involved in the process. Figure is taken from Wahl et al, 2009.

Further on, the A complex is generated. To make this A complex, SF1 is dissociated, and the branchpoint sequence in the 3' splice site is bound by the U2 snRNP. This is an ATP-dependent reaction. U2 snRNP binding is stabilized by U2AF65 interactions with

polypyrimidine tract. An important consequence of complex A formation, is to make the branchpoint adenosine to jut out, ready to attack the 5' splice site. Following this, complex B forms. It is characterized by tri-snRNP U4/U6 and U5 binding the spliceosome. The role of U5 snRNP is to bring into the spliceosome the RNA helicases which activate the catalytic site of the spliceosome. The catalytic core of the spliceosome consists of many proteins, including U2 and U6 snRNPs. Immediately prior to this stage of cycle, U1 and U4 snRNPs dissociate from the spliceosome, triggering the first transesterification reaction. Prp2 RNA helicase remodels the spliceosome to catalytically activates the first transesterification reaction. During the first reaction, complex C is formed, creating the intron lariat intermediate. During the second transesterification reaction, when two exons are joined together, the lariat structure is released.

U2, U5 and U6 snRNPs, as well as around 100 other splicing factors are important on the last stages of the spliceosome cycle (Black, 2003; Valadkhan, 2007; Wahl *et al.*, 2009). Interactions within the spliceosome are weak. For this reason, about 170 supportive proteins perform supplementary stabilization : i.e. ATPases, helicases and DExD/H-box proteins function as molecular switches, catalysts, perform proofreading functions, etc. (Burgess and Guthrie, 1993). Once splicing is over, a mature transcript is released, while the intron lariat structure is degraded and snRNPs are recycled and used in the subsequent splicing reactions (Black, 2003; Valadkhan, 2007; Wahl *et al.*, 2009).

1.1.6 The minor spliceosome

Most eukaryotic introns are removed by the spliceosome in a U2 snRNP-dependent way described in a previous section. This spliceosome is termed the major spliceosome. However, a small number of introns are excised by another machinery, called the minor spliceosome. The minor spliceosome has a set of specific snRNPs: U11, U12, $U4_{atac}$ and $U6_{atac}$. These are functionally analogous to U1, U2, U4 and U6 snRNPs of the major spliceosome. Both spliceosome have U5 snRNP (Turunen *et al.*, 2013). Introns, recognized by the major spliceosome usually start with GT nucleotide, and finish with AG bases. They are termed U2-dependent introns, while introns defined by the minor spliceosome typically start with AT sequence, finish with AC nucleotides, and are called U12-dependent. An interesting feature of U12-dependent introns is that they are spliced slower than U2-dependent introns (Burge *et al.*, 1998; Patel and Steitz, 2003). Even

though the minor spliceosome is less abundant in cells, it is essential: *Drosophila* flies, missing U12 snRNA have severe defects during embryonic development (Patel and Steitz, 2003).

1.1.7 Alternative splicing

Splicing can be constitutive – when all introns are spliced out and all exons are included into a final mRNA. However, splicing can also be regulated alternatively: introns might stay retained (intron retention), while exons might get completely or partially spliced out (alternative exon) (Figure 1.5) (Srebrow and Kornblihtt, 2006).

Alternative splicing can add and remove coding information into mRNAs and might give rise to different amino acids. This might bring changes to protein structure and function i.e. changes in enzyme activation domains, alterations in receptor ligand binding, differences in subcellular localization for signaling proteins, etc. (Black, 2000). An interesting example of the alteration of a protein's role by alternative splicing is in the *WT1* gene (*WILMS TUMOUR1*). WT1 protein can function as a splicing factor or transcription regulator depending on alternative 5' splice site selection in its encoding mRNA (Bor *et al.*, 2006).

Approximately 35% of alternative splicing events cause the alterations in the open reading frame (ORF). For that reason, generated transcripts might have premature stop codons (PTCs) and to prevent the production of shortened proteins are targets to degradation by nonsense - mediated decay (NMD) (Lewis *et al.*, 2003; Soergel *et al.*, 2006). RNA is known to be an unstable molecule. Alternative splicing of 3'-unranslated regions (UTRs) was found to regulate transcript stability (Matlin *et al.*, 2005).

Alternative splicing is important in many cellular pathways, including cell death and differentiation, meiosis, cell signalling, etc. (Black, 2000). Alterations of mRNA splicing levels by mutations (i.e. exon skipping, production of aberrant transcripts, activation of cryptic splice sites, etc.) or mis-expression of splicing regulators can lead to medical conditions including cancer, neurodegeneration, infertility, premature ageing, muscular dystrophies, autoimmune diseases, etc. (Pajares *et al.*, 2007; Wang and Cooper, 2007; Baralle *et al.*, 2009).

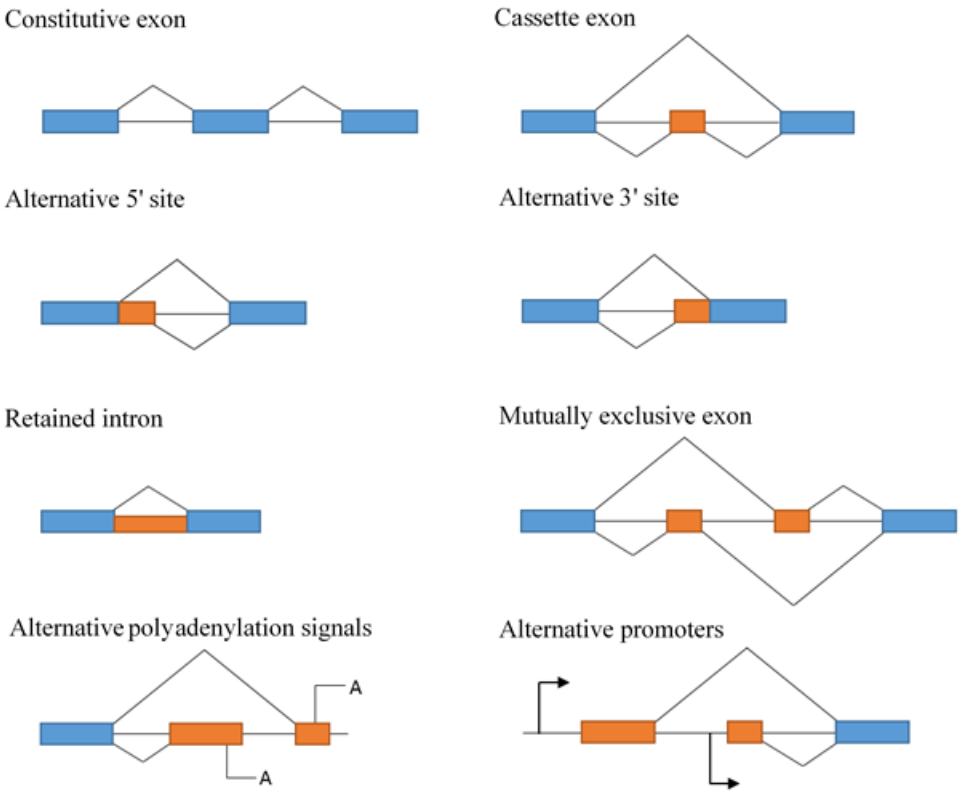


Figure 1.5 – Different types of alternative splicing events. Constitutive exons are always included into the final mRNA. Cassette exons (also termed alternative exons) can be independently included or excluded from the final transcript. Alternative 3' and 5' splice sites are competitive with one being selected by the spliceosome. Alternative splice site selection can alter exon length. Retained introns are not excised from the pre-mRNA. Mutually exclusive exons are exons that are included or excluded in a coordinated manner. The use of alternative promoters and alternative polyadenylation can also alter the exons included and the length of the final transcript. Alternatively spliced exons are marked in orange, while constitutive exons are coloured in blue (Figure is adopted from (Matlin *et al.*, 2005; Mills and Janitz, 2012).

1.1.8 Sophisticated regulation of alternative splicing

As seen from the sections above, splicing is a complex process, involving hundreds of regulatory factors and proteins. However, spliceosomes function with high fidelity and produce a great variety of transcripts depending on the cellular context. Splicing is tightly controlled on different stages of development and in different tissues, and its disruption can lead to severe disorders. In addition to spliceosome function in selecting splice sites, splicing is regulated through a "splicing code", which encompasses all binding sites for RNA-binding splicing repressors and activators within a pre-mRNA. The sequences for RNA-binding proteins in the pre-mRNA are called *cis*-acting elements, while the proteins

that bind these sequences are termed *trans*-acting factors. One current objective in the splicing field is to decipher this "code" and to be able to foresee the splicing outcome of any pre-mRNA form its sequence (Wang and Burge, 2008).

1.1.8.1 Cis-acting elements

1.1.8.1.1 Splice site strength and position

Splice site strength depends on complementarity between splice site sequences and U1 and U2 snRNPs binding to them. Typically, when snRNPs are highly complementary to splice sites, splice sites as considered strong. When the extent of sequence complementarity is less, splice sites are termed weak. As a tendency, constitutive exons have stronger splice sites that are efficiently recognised by the spliceosome. Alternative exons have weak splice sites, that are thus recognized less efficiently (Hertel, 2008). Interestingly, recent studies show, that there are many variations of splice site sequences present, i.e. the human 5' splice site was found to have more than 9000 sequence variants (Roca *et al.*, 2012). Furthermore, apart from U1 and U2 snRNPs, other snRNPs and splicing factors affect splice site choice (Roca *et al.*, 2013). Relative positions of splice sites also affect the splicing outcome. Splice sites, which are closer to each other, tend to be favoured and paired quicker (Nogués *et al.*, 2003). However, when being too close, splice sites prevent neighbouring exons from being spliced together, often leading to mutually exclusive exon splicing (Smith and Nadal-Ginard, 1989).

1.1.8.1.2 Splicing regulatory elements

Not all splice sites completely match the consensus. Additionally, competing or pseudo splice sites might confound the splicing machinery. For precise splicing regulation, additional short RNA sequences termed splicing regulatory elements (SREs) are present. SREs consist of intronic splicing enhancer/silencer elements (ISEs/ISSs) and exonic splicing enhancer/silencer elements (ESEs/ESSs) (Hertel, 2008). These elements are important for choosing between the real exon and pseudoexons, for splicing of constitutive exons and for choosing between competing splice sites. Splicing regulatory elements recruit many splicing factors for binding, and can positively or negatively influence exon inclusion (Figure 1.6) (Srebrow and Kornblihtt, 2006). The effect of SREs on splicing can be tissue-specific, and position-dependent (Dredge *et al.*, 2005; Goers *et*

al., 2010; Llorian *et al.*, 2010). Splicing regulatory elements are usually present in close vicinity of exon/intron junctions. Studies of splicing regulation by SR (serine/arginine-rich) proteins show that with increasing distance between regulatory sequence elements and splice sites, the probability of exon inclusion decreases (Graveley *et al.*, 1998). However, there are some examples of elements efficiently regulating the splicing on a distance from the exon, i.e. sequence motifs bound by NOVA protein (Ule *et al.*, 2006).

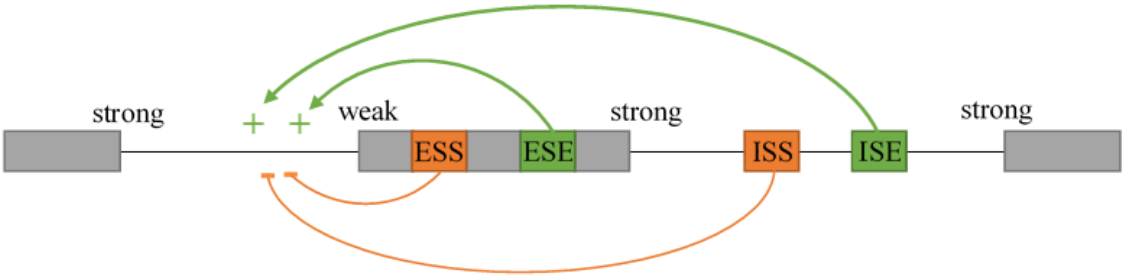


Figure 1.6 – Splicing regulatory elements can enhance and repress the splice site selection by the spliceosome. ESE/ISE (exonic splicing enhancers and intronic splicing enhancers) recruit the splicing activators, which in turn promote spliceosome assembly on splice sites. ESS/ISE (exonic splicing silences and intronic splicing silencers) attract splicing repressors and inhibit spliceosome assembly. Figure is adopted from (Srebrow and Kornblihtt, 2006).

Finally, recognition of each exon depends on a specific interplay between splicing enhancers/silencers. Studies of *SMN* (survival of motor neuron) pre-mRNA splicing uncovered the multiple enhancer and silencer elements within *SMN* exon 7 and in the flanking introns. However, even a single base alteration in a regulatory element changed the balance between enhancers and silencers, leading to different levels of exon inclusion (Lorson and Androphy, 2000; Singh *et al.*, 2004).

1.1.8.2 Trans-acting elements

Splicing regulatory elements are bound by *trans*-acting proteins which can either activate splicing or repress it. Proteins belonging to SR protein family are known to enhance splicing. SR proteins have one or two RRM (RNA recognition motif) protein domains and an RS (arginine-serine rich) domain. The RRM domain mediates SR protein binding with ESEs on the RNA, while RS domains promote spliceosome assembly and activate the splicing. In addition to the SR protein family, there are other SR-related proteins, that also activate splicing, e.g. these include Tra2α and Tra2β proteins (Tacke *et al.*, 1998).

Witten and Ule et al 2011 described several possible mechanisms how splicing activation might happen. One mechanism is when splicing regulator proteins interact with and stabilize binding of U1 snRNP or other spliceosome components on pre-mRNA, that initiate spliceosome assembly. As another possibility, splicing activators can interact with other splicing factors (like TIA proteins), and indirectly stimulate their binding to U1 snRNP and splicing. Splicing can also be activated when regulatory proteins interfere with intronic or exonic silencer elements (Witten and Ule, 2011).

Proteins belonging to the hnRNP protein family often repress splicing. hnRNP proteins are characterized as having up to four RRM domains, or RGG (arginine-glycine-glycine) domain, or KH (hnRNP K homology) domain. In addition to hnRNP proteins, there are other known examples of splicing repressors, e.g. NOVA and MBNL proteins. Each of these repressor proteins recognizes a specific RNA binding sequence (Shen and Green, 2006). A common mechanism of splicing repression happens through blocking the branch point sequence and inhibiting spliceosome assembly. This mechanism has been described for RBPs NOVA, Fox and PTB proteins (Ule *et al.*, 2006; Zhang *et al.*, 2008a; Xue *et al.*, 2009). Another mechanism explaining how splicing can be repressed is through influencing the kinetics of the splicing reaction. For instance, binding of some proteins (e.g. TIA proteins) to a preceding exon are thought to increase the speed of transition between the different stages of splicing reaction. This reduces the time for other splicing factors to bind and results in splicing repression (Witten and Ule, 2011). Splicing repression can also happen if a regulatory protein interferes with splicing enhancer sequences, as in the study by Wu et al on splicing repression of the *MAG* gene by Quaking protein (Wu *et al.*, 2002).

Even though SR proteins are classic splicing activators, and hnRNPs are classic repressors, exceptions exist. Some SR proteins can inhibit splicing when they bind to an intronic sequence in adenovirus IIIa pre-mRNA (Kanopka *et al.*, 1996). Also, some hnRNP proteins activate splicing, including hnRNPL (Hui *et al.*, 2005). The splicing effect of proteins on individual exons varies. The splicing outcome depends on the position of splicing regulatory elements relative to the target exon. For instance, MBLN and Nova proteins can activate splicing through binding downstream of a target exon and repress it when bound upstream (Goren *et al.*, 2006). Some splicing factors are expressed in a tissue-specific manner, e.g. expression of the RNA-binding protein NOVA is brain-specific. The study by Grosso et al 2008 has identified more than 100 cases of tissue-

specific splicing regulation by splicing proteins (Grosso *et al.*, 2008). Splicing outcomes can also be determined by splicing factors interacting with each other.

1.1.8.3 Other factors that affect splicing

1.1.8.3.1 Exon/Intron Architecture

In addition to the splice site sequence, splice site recognition is dependent on exon size. For small exons, splice sites are recognized across the intron, in what is called intron definition. Intron definition is especially typical for eukaryotes with smaller introns, such as yeast (Berget, 1995). However, in higher eukaryotes, introns are large. In human the average exon length is 170 base pairs, while an average intron length reaches 5419 base pairs. 10% of introns stretch over more than 11 000 base pairs and around 5% of introns are longer than 200 000 base pairs (Sakharkar *et al.*, 2004). For these long introns, splice sites are recognized across the exons, in what is called exon definition. During exon definition, U1 snRNP associates with the 5' splice site and stimulates the recognition of upstream 3' splice site across the exon (Robberson *et al.*, 1990). This exon defined complex switches into an intron defined complex at the early stages of spliceosome assembly (Sharma *et al.*, 2008). Splice site definition across the intron is much more efficient than splice site definition across the exon

Splicing happens best when splice sites are separated by an optimal nucleotide length. Exons that are shorter than 50 nucleotides in length are not very efficiently spliced, as splicing factors assembling on their splice sites might suffer from “steric clashing”, and so do not function efficiently. However, exons exceeding 300 nucleotides are not easy to define and splice out. These longer exons might require additional splicing factors, thus are often skipped (Fox-Walsh *et al.*, 2005). Intron size is also important for splicing efficiency. Splicing is more efficient when introns are short, as splice sites are more easily recognized. Computational analyses have shown that the length of the upstream intron is more important for splicing efficiency, than the length of the downstream intron (Hertel, 2008).

1.1.8.3.2 RNA secondary structure

RNA is a single-stranded molecule, but it can fold back against itself to produce secondary and tertiary structures. Several genes show splicing changes due to RNA secondary structure formation. In *TPM2* (tropomyosin 2 (beta)) secondary structure

formation leads to higher splicing efficiency (Sirand-Pugnet *et al.*, 1995), while in *GH1* (growth hormone 1) (Estes *et al.*, 1992) it results in higher expression of a minor transcript. RNA secondary structure can influence splice site selection and spliceosome assembly. When RNA structures mask splice sites and splicing enhancer elements, they can interfere with spliceosome assembly and repress splicing. In other cases, local RNA structures can hide the splicing silencer elements, and promote spliceosome assembly (Hertel, 2008). For long introns, secondary structures might serve to bring distal splice sites closer to each other (Shepard and Hertel, 2009). It was estimated that increased GC content of RNA sequence (usually associated with more stable structures) around the splice sites stimulates the formation of secondary structures (Zhang *et al.*, 2011).

1.1.8.3.3 Co-transcriptional splicing regulation

Splicing is known to often happen co-transcriptionally (Neugebauer, 2002). Thus, transcription is another factor influencing splicing. To begin with, the transcription speed is important. If transcription elongation is rapid, alternative exons with weak splice sites might be not recognised by spliceosome machinery because of the synthesis of stronger downstream exons, and hence, will be skipped. When transcription elongation is slow, it favours weak exon inclusion into mature transcripts (Kornblihtt *et al.*, 2004). Secondly, the structure of the promoters that recruit Polymerase II for transcription is important. Differences in promoter structure can lead to different splicing factors recruited (de la Mata *et al.*, 2003). Thirdly, the aberrations in transcription termination can affect the splicing of the last intron in the transcript (Niwa and Berget, 1991).

1.1.8.3.4 Effect of chromatin structure and modifications

Splicing is also known to be facilitated by the chromatin state. Two models exist to explain the effect of chromatin on splicing. The “Kinetic model” suggests that chromatin affects the rate of transcription elongation, and thus, affects the splicing. As evidence for this model, exons of more highly expressed genes are associated with increased number of nucleosomes. Nucleosomes are thought to produce a barrier for RNA polymerase II and to slow down the transcription, promoting exon inclusion. The “Recruitment model” states that histone modifications can regulate the splicing through attracting the splicing factors. Some histone H3 methyltransferases (as Kmt2a, Kmt3a and Kmt4) are linked with splicing factors, including hnRNPM and hnRNPL. Inhibition of splicing leads to a decrease in histone methyltransferase recruitment (Brown *et al.*, 2012).

1.2 Alternative splicing in the brain

My thesis research is specifically focused on splicing regulation by several RNA-binding proteins in the brain. In this section, I will review the recent findings about the splicing in the brain.

1.2.1 The brain has a high frequency and a unique pattern of alternative splicing events

Brain is the tissue with the highest amount of alternative splicing in human body. According to EST (expressed sequence tag) analyses, more than 40% of genes in brain are alternatively spliced. Some of them have more than one splicing event (Yeo *et al.*, 2004). There are several explanations for the high pattern of alternative splicing in the brain. First, some splicing factors in the brain need to be synthesized, to form memories. Second, some mRNAs in the brain, after being transcribed in the nucleus need to be spatially localized along the axons to synapses. Thirdly, to increase the complexity of the brain, some pre-mRNAs have a different processing pattern in the brain, compared to other tissues (i.e. *PMCA* and *AChE*). Fourthly, some RNA-binding proteins in the brain are neuron-specific and absent in other tissues (i.e. NOVA, PTB2, etc.) (Darnell, 2013).

Alternative splicing is associated with many important processes in the brain, including brain development. For instance, neuron migration across the neocortex requires exon 7 b and 7c skipping in the *Dab1* transcript controlled by the NOVA2 splicing factor. NOVA2 knockout mice have shown defects in neuronal migration (Grabowski, 2011). Nova proteins also regulate RNA targets involved in synaptogenesis (Ule *et al.*, 2005). Alternative splicing also modifies synaptic transmission. For example, the *SAP-97* (synapse associated protein-97) splicing adjusts SAP-97 capacity to bind AMPA receptors and, hence, controls activity of glutamatergic synapses. Similarly, splicing of the NMDA receptor C2-C2’ domain allows constant neuronal activity and supports synaptic plasticity (Lipscombe, 2005). Splicing in the brain is also linked to ageing and age-related diseases. Microarray studies identified exon 11 skipping in lamin A gene that leads to premature ageing syndrome (Hutchinson-Gilford progeria); *PSEN1* mutations in the 5' splice site of intron 4 were detected in Alzheimer’s disease; mutations in splice sites of *Park2* introns cause Parkinson’s disease, etc. (Tollervey *et al.*, 2011; Mills and Janitz, 2012).

The most frequent alternative splicing event in brain is exon skipping. Skipping of exons 15 and 16 in *FXR1* gene results in fragile X mental retardation (Yeo *et al.*, 2004). Interestingly, total RNA sequencing of human brain has shown up to 40% of reads to localize within introns, especially in fetal brain. These intronic RNAs mainly represent unprocessed pre-mRNA transcripts. Several genes with the highest intronic RNA scores (e.g. *Nrxn1* and *Auts2*) are involved in autism and schizophrenia. Thus, it can be suggested, that intronic RNAs correlate with alternative splicing changes in developing fetal brain (Ameur *et al.*, 2011).

1.2.2 Splicing changes in neurological diseases

Neurological diseases occur when a certain number of neurons lose their functional ability and connections, do not respond to external and internal signals and degenerate (Courtney *et al.*, 2010). A review by Anthony and Gallo 2010 provides a list of neurological conditions linked with defects in alternative splicing. These conditions include spinal muscular atrophy, amyotrophic lateral sclerosis, myotonic dystrophy, Rett syndrome, ataxias, Schizophrenia and sporadic Alzheimer's disease (Anthony and Gallo, 2010). Alternative splicing can become aberrant due to malfunction of either *cis*-splicing elements or *trans*-acting factors. *Cis*-mutations in splicing enhancer or silencer sequences can affect splicing by weakening or strengthening the regulatory elements, as well as creating or destroying splice sites (Licatalosi and Darnell, 2006). A point mutation in *SMN2* gene results in *SMN2* exon 7 skipping. (Zhang *et al.*, 2008b). Mutations in the exon 10 and intron 10 of the *MAPT* gene that codes for tau protein impair the *MAPT* splicing. Consequently, tau aggregates form and impair the axonal function of tau (Hong *et al.*, 1998).

Splicing can be abnormally regulated due to the defects in the RNA-binding proteins. For instance, mutations in the gene coding for the RNA-binding protein TDP-43 lead to production of TPD-43 with an abnormal C-terminal domain that fails to interact with hnRNP proteins. Aberrant TDP-43 is associated with amyotrophic lateral sclerosis (ALS) (Lagier-Tourenne and Cleveland, 2009; Anthony and Gallo, 2010). Other aberrations of splicing proteins in the nervous system include abnormal posttranslational modifications, mislocalization, aggregation, degradation, sequestration, etc. (Ule, 2008).

1.2.3 RNA-binding proteins in the brain

Many splicing factors are expressed in the brain. These include Sam68, T-STAR and QUAKING (QKI) proteins that are the subject of my research. These belong to the STAR (signal transduction and activation of RNA) protein family. Thus, in the following sections I will review in detail the features of these RNA binding proteins.

1.2.3.1 The STAR protein family

The STAR protein family is characterised by the presence of the STAR (signal transduction and activation of RNA) protein domain and gained their name due to being linked with signal transduction pathways and RNA metabolism. STAR proteins use the KH (heterogeneous nuclear ribonucleoproteins K homology) domain for RNA binding, however compared to other KH proteins they have several specific features. STAR proteins have only one KH domain, which is flanked QUA1 and QUA2 domains on either side (Artzt and Wu, 2010). The KH domain is involved in RNA binding. The QUA1 domain is involved in protein dimerization (Chen and Richard, 1998), while the QUA2 domain participates in RNA binding (Liu *et al.*, 2001). The KH, QUA1 and QUA2 domains together comprise the STAR domain of the proteins.

STAR proteins are widely expressed and are involved in a variety of developmental events including neurogenesis, spermatogenesis, cardiovascular development and bone metabolism. STAR proteins have been shown to regulate pre-mRNA splicing, RNA localization, mRNA stability and efficiency of translation. They are also involved in cell signaling and post-translational modifications. STAR proteins show high conservation among eukaryotes. There are eight STAR proteins known: Sam68 (Src-associated in mitosis, 68Kda), T-STAR (testis – signal transduction and activation of RNA) (also called SLM2 – Sam68-like mammalian protein 2), SLM1 (Sam68-like mammalian protein 1), QKI (quaking) and SF (splicing factor 1) are mammalian proteins (QKI is mouse-specific); How (held out of wings) protein is found in *Drosophila*; Asd-2 (alternative splicing defective 2) and GLD-1 (defective in germline development 1) are both found in *C. elegans* (Artzt and Wu, 2010). My PhD was mainly focused on Sam68, T-STAR and QKI proteins.

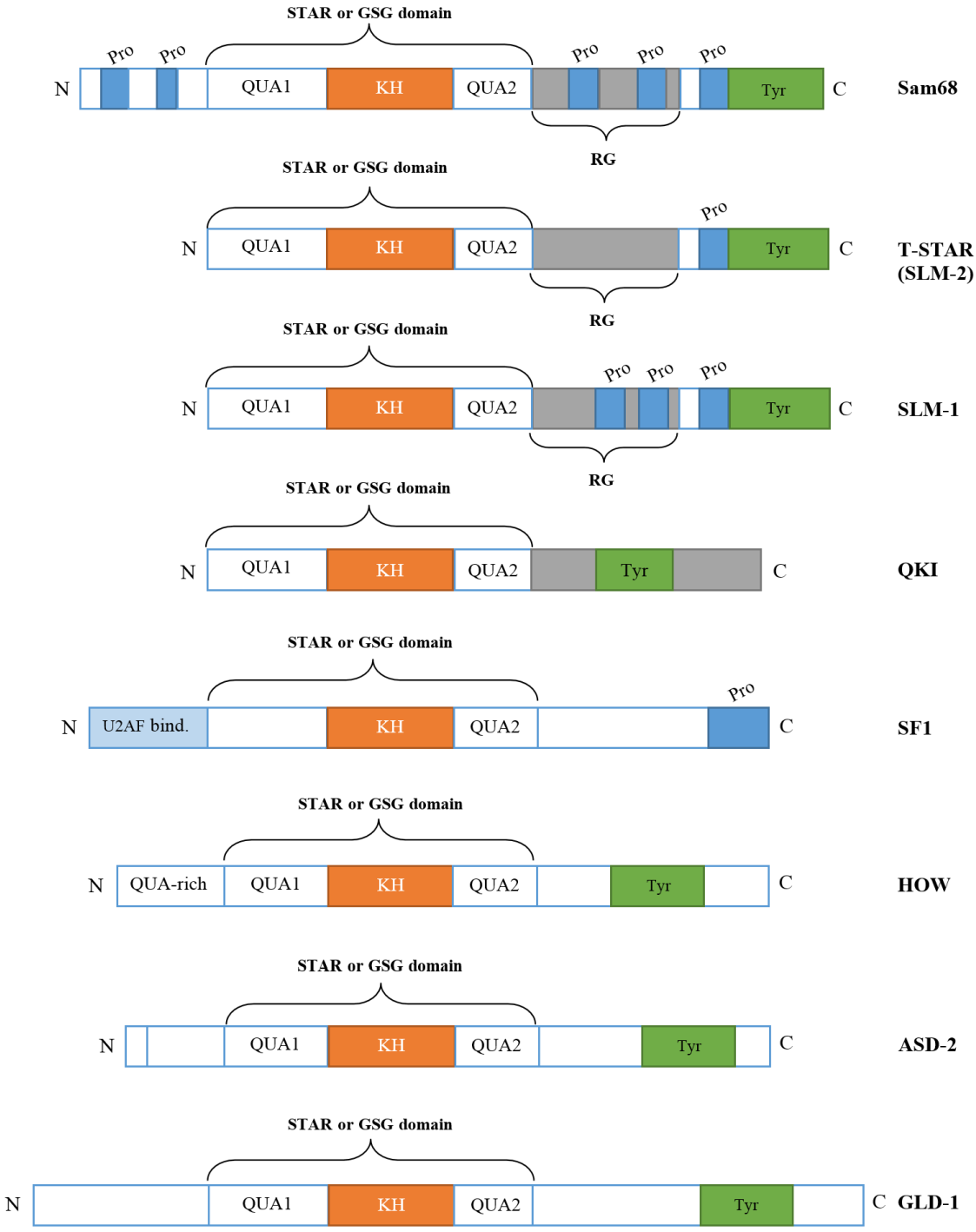


Figure 1.7 – Domain structure of STAR proteins. All STAR proteins have a STAR domain that consists of KH domain, flanked by QUA1 and QUA2 domains and required for RNA binding. In addition, other motifs are present. SF1 does not have QUA1 domain. Instead, U2AF domain (shaded light blue) is required for RNA binding. Proline rich sequences (labelled as Pro) for protein-protein interactions. The RG regions (in grey) are required for methylation. The tyrosine-rich domain is subject to phosphorylation. There is also a nuclear localization domain is present in the C terminus of all proteins (not shown on the figure). This figure is adapted from (Stoss *et al.*, 2004).

1.2.3.2 Sam68

Sam68 (Src-associated in mitosis, 68Kda) is a ubiquitously expressed protein, with high expression levels in the testis and brain (Fumagalli *et al.*, 1994). The STAR (also called GSG) domain of Sam68 is represented by a KH (K homology domain), flanked by QUA1 and QUA2 domains (Figure 1.7). The 100 amino-acid KH domain is required for RNA binding, while the flanking domains provide RNA-binding affinity and enable proteins to homodimerize. Six proline-rich sequences (P0-P5), surrounding the STAR domain, allow Sam68 to interact with SH3 (SRC Homology 3) and WW (tryptophan rich) domains of other proteins. Two RG (arginine-glycine) regions that flank the proline-rich domains are targets for methylation by PRMT1 (protein methytransferase 1) (Fumagalli *et al.*, 1994; Taylor and Shalloway, 1994).

Sam68 methylation alters its RNA binding capacity and localization (Bedford *et al.*, 2000; Côté *et al.*, 2003). The tyrosine rich domain of Sam68 is subject to phosphorylation by tyrosine kinases (e.g. Src and Brk kinases) and binds proteins with SH2 (Src homology 2) domains (Wang *et al.*, 1995; Derry *et al.*, 2000). Phosphorylation usually affects Sam68 RNA-binding abilities (Sánchez-Jiménez and Sánchez-Margalet, 2013). Sam68 is also a target for SUMOylation, acetylation and serine/threonine phosphorylation (Sette *et al.*, 2002; Babic *et al.*, 2004; Babic *et al.*, 2006). Sam68 was shown to be involved in transcription, alternative splicing, translation, signal transduction and RNA export, cell cycle progression, etc. (Rajan *et al.*, 2008; Sette *et al.*, 2010). By SELEX (systematic evolution of ligands by exponential enrichment) and microarray experiments, Sam68 was shown to bind A/U rich RNA sequences. Out of them, 5'-UAAA-3' and 5'-UUUA-3' motifs bound Sam68 with highest affinity (Lin *et al.*, 1997; Chawla *et al.*, 2009; Pedrotti *et al.*, 2010; Feracci *et al.*, 2014).

Many studies were performed to uncover the RNA targets and to study the tissue-specific mechanism of splicing by Sam68. In muscle Sam68 was found to bind intronic *Tropomyosin-β* sequence and affect alternative splicing of *SMN2* exon 7. These events link Sam68 to muscle contraction and spinal muscular atrophy, respectively; (Pedrotti *et al.*, 2010). By means of splicing-sensitive microarrays Sam68 was found to regulate the neurogenic splicing targets *Numa1, Sorbs1* and *Kifap3/Kap3* RNAs. Sam68 knockout inhibited cellular differentiation in the mouse neocortex. Sam68 knockout mice also developed bone and motor coordination abnormalities (Chawla *et al.*, 2009). Recent

studies by Iijima et al 2011 and Ehrmann et al 2013 revealed Sam68 to repress *Neurexin-1* and *Neurexin-3* splicing in neurons (Iijima *et al.*, 2011; Ehrmann *et al.*, 2013).

Sam68 levels are elevated in breast, prostate and kidney cancers. Sam68 was also found to be associated with the lung and colorectal cancers, as well as with neuroblastoma. Because of its increased expression, Sam68 could be used as prognostic marker in cancer diagnosis (Liao *et al.*, 2013; Zhang *et al.*, 2014; Zhao *et al.*, 2014). Sam68 affects alternative splicing of exon v5 in *CD44*, intron retention in the proto-oncogenic cyclin D1b and alternative splicing of the apoptotic regulator *Bcl-x* (making antiapoptotic *Bcl-x* isoform) (Cheng and Sharp, 2006; Paronetto *et al.*, 2007; Paronetto *et al.*, 2010). Hence, Sam68 is linked to migration, adhesion and proliferation of cancer cells (Bielli *et al.*, 2011).

Sam68 is also linked to adipogenesis. A study by Huot et al showed Sam68 to regulate alternative splicing of *mTOR* through binding its' intron 5. Sam68 depletion caused reduced mTOR levels and caused defects in adipogenesis (Huot *et al.*, 2012).

Sam68 is also abundant in the testis. Recently, Sam68 was found to regulate alternative splicing of the *Sgce* gene exon 8 in the testis (Chawla *et al.*, 2009). Sam68 was also shown to interact with other RNA-binding proteins. In germ cells, Sam68 mainly follows a nuclear mode of expression, with some cytoplasmic expression during meiosis II. As a result of this cytoplasmic expression, Sam68 regulates translation during germ cell development and spermatogenesis. Sam68 knockout mice are sterile. Hence, Sam68 is linked to male fertility (Paronetto *et al.*, 2011).

1.2.3.3 T-STAR

T-STAR (testis – signal transduction and activation of RNA) also called SLM2 (Sam68-like mammalian protein 2) is another STAR family protein, closely related to Sam68 by structure and function. T-STAR is less ubiquitously expressed than Sam68, with highest expression in the testis, and weaker expression in the brain (Venables *et al.*, 1999). T-STAR and Sam68 have similar protein domains. T-STAR is shorter than Sam68, as it is lacking the first hundred amino acids of the N-terminal region of Sam68. Furthermore, T-STAR contains just a single proline-rich domain and has a larger RG domain (Figure 1.7) (Venables *et al.*, 1999). T-STAR is also known to be phosphorylated and methylated, to suppress its binding efficiency (Artzt and Wu, 2010). T-STAR is subjected for degradation by the proteasome *via* the E3 ubiquitin ligase (Venables *et al.*, 1999). T-STAR interacts with testis-specific RBMY and hnRNPG-T proteins, suggesting a role in germ cell development. T-STAR also interacts with Sam68 (Venables *et al.*, 1999). SELEX studies has identified a 5'-U(U/A)AA-3' motif as RNA target for T-STAR (Galarneau and Richard, 2009a).

The role of T-STAR in splicing regulation is not yet extensively studied. However, recent studies uncover some T-STAR functions in splicing. A recent study by Ehrmann et al 2013 found that T-STAR represses the splicing of exons in the *Neurexin 1, 2 and 3* and *Tomosyn-2* gene in the mouse brain (Ehrmann *et al.*, 2013). A study by Zhang et al, 2009 has identified *Fabp9, Crisp2* and *Tssk2* genes as potential targets for T-STAR binding in mouse testis. These targets were shown to be alternatively spliced and involved in spermatogenesis. However, they have not yet been physiologically validated (Zhang *et al.*, 2009). T-STAR was also shown to affect alternative splicing of exon v5 in *CD44* minigene, and hence, it might have a role in cancer (Matter *et al.*, 2002).

1.2.3.4 QKI

The *Quaking* gene was first discovered in mice. Mice expressing aberrant QKI had an inability to form a compact myelin sheath in their Schwann cells and oligodendrocytes, and developed tremors when moving. QKI is highly expressed in glial cells (including astrocytes and oligodendrocytes) as well as in many other cell-types and tissues (Hardy, 1998).

There are three QKI isoforms: QKI5 which is nuclear, as well as QKI6 and QKI7 which are cytoplasmic (Wu *et al.*, 1999). These QKI protein isoforms only differ in their C-

terminal region and 3'-UTRs. All have the same STAR domain, consisting of KH, QUA1 and QUA2 domains, a tyrosine domain for phosphorylation and, possibly, several putative SRC-homology (SH3)-binding domains for mediating protein-protein interactions. QKI was found to be phosphorylated in their C-terminal tyrosines in the early mouse brain development: this was suggested to repress QKI affinity to RNAs and to be associated with its role in myelinogenesis. QKI5 was also shown to be weakly methylated in arginine residues.

The RNA binding motif of QKI was determined by mutagenesis experiments as 5'-NA(A/C)UAA-3' (where N is any nucleotide) (Ryder and Williamson, 2004). Similar motifs were identified by SELEX: 5'-NACUAAY-3' and 5'-YAAY-3' (where Y is any pyrimidine) (Galarneau and Richard, 2005). QKI is known to be involved in alternative splicing, translation and mRNA stability. In addition to myelin formation, it has a role in embryogenesis, where it participates in the blood vessel formation and cardiovascular development (Noveroske *et al.*, 2002). Furthermore, QKI was recently identified as a tumour suppressor in prostate, gastric and other cancers (Bian *et al.*, 2012; Zhao *et al.*, 2014).

There are many splicing targets of QKI identified. QKI5 was shown to regulate the splicing of exon 12 of the *MAG* gene (myelin-associated glycoprotein) (Wu *et al.*, 2002). QKI5 was recently shown to regulate a *NUMB* gene in lung cancer (Zong *et al.*, 2014). QKI has also been found to regulate *SOX2* gene expression through binding to its 3'-UTR region in oral cancers (Lu *et al.*, 2014). QKI was also shown to interact with the 3'-UTR region of β-catenin in colon cancers (Yang *et al.*, 2010). Furthermore, QKI promotes exon 9 inclusion of the *Capzb* gene in muscle cells (Hall *et al.*, 2013). Finally, QKI was shown to regulate the splicing of exon a2 of *Myocardin* pre-mRNA in vascular smooth muscle (van der Veer *et al.*, 2013).

1.2.3.5 RNA splicing targets: Neurexin-2 and Tomosyn-2

My PhD has specifically focused on neurological splicing targets for T-STAR and Sam68 proteins. In this section I am introducing my studied targets in more detail. Four genes have been recently identified as physiological splicing targets of T-STAR in mouse brain. These are: *Tomosyn-2*, *Neurexin-1*, *Neurexin-2* and *Neurexin-3* (one alternative cassette exon in each) (Ehrmann *et al.*, 2013).

Neurexins are cell adhesion proteins highly distributed in synapses. Neurexins are located presynaptically. Forming complexes with Neuroligin proteins, Neurexins are involved in synapse formation, maturation and maintenance. Mutations in Neurexins are associated with several neurological conditions, including Alzheimer's disease and autism (Sindi *et al.*, 2014). There are two isoforms of Neurexin proteins, the longer α-Neurexins and the shorter β-Neurexins. Neurexin function and distribution to excitatory and inhibitory synapses is regulated by a complicated pattern of alternative splicing in *Neurexin* genes. α-*Neurexin* genes contain alternative splice sites (SS1-5), two of which are also present in the *β-Neurexin* genes (SS4 and SS5). Differences in splicing result in differences in the Neurexin ectodomain, altering Neurexin-Neuroligin interactions. Current study has focused on *Neurexin-2* splicing regulation by T-STAR. As visible from the Figure 1.8, *Neurexin-2* has different alternative splicing events, including four cassette exons. T-STAR specifically regulates AS4 (alternatively spliced segment 4, generated through alternative splicing of SS4) (Figure 1.8) altering the edge of the Neuroligin binding domain in the Neurexin protein. As a consequence, an α-helix is generated that might result in a steric clash or interfere with the salt bridge, crucial for Neurexin-Neuroligin interactions (Araç *et al.*, 2007);(Krueger *et al.*, 2012).

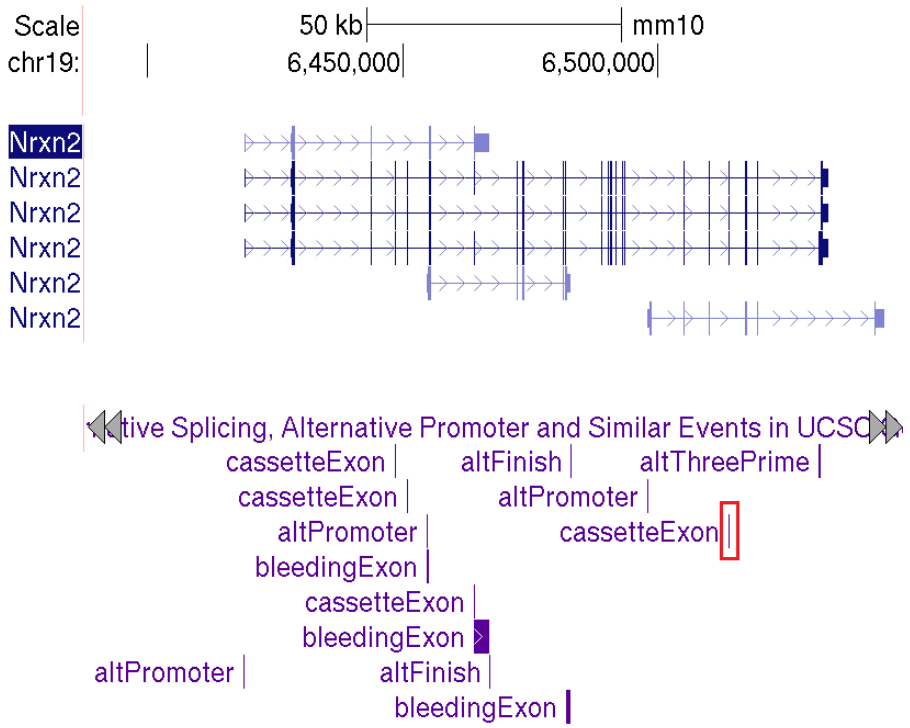


Figure 1.8 – *Neurexin-2* splicing pattern. *Neurexin-2* is alternatively spliced into 6 splice variants generated by 13 splicing events. Cassette exon (AS4) regulated by T-STAR is labelled by the red rectangle. The figure is adapted from UCSC genome browser.

Tomosyn-2 encodes a SNARE protein (soluble n-ethylmaleimide-sensitive-factor attachment receptor) protein expressed all over the brain, but mainly in the hippocampus and cerebellum. Tomosyn-2 is involved in neuronal exocytosis and is thought to direct and inhibit vesicle fusion in the synapse. Tomosyn-2 is important in postnatal stages of mouse brain development (Groffen *et al.*, 2005). Recently, a *Tomosyn-2* knockout mouse was generated. This mouse had a phenotype of impaired motor performance, with synaptic defects in the neuromuscular junction. Based on this study, Tomosyn-2 was shown to inhibit the release of neurotransmitter acetylcholine from axon terminals. This normally helps to avoid synaptic fatigue and maintain a normal motor function (Geerts *et al.*, 2014).

The splicing of *Tomosyn-2* gene results in production of twelve splice variants (Figure 1.9). However, four of them (called s, m, b and xb) are better described in the literature and are linked to differences in protein structure. Each isoform encodes the Tomosyn-2 protein with its hypervariable domain of a different length. The hypervariable domain separates the WD40 repeat domain from the coiled coil domain of Tomosyn-2. The hypervariable domain is important for Tomosyn-2 pairing with other SNARE complexes. That defines the fusogenic capacity of the synapse. Repressing *Tomosyn-2* exon 23 splicing (Figure 1.9), T-STAR promotes the production of *m-Tomosyn-2* isoform (McNew *et al.*, 1999; Hatsuzawa *et al.*, 2003). The study by Williams et al 2011 showed that expression of *m-Tomosyn-2* isoform results in a higher concentration of Tomosyn-2 protein (Williams *et al.*, 2011). Thus, it is highly possible that the role of T-STAR splicing repression of *Tomosyn-2* is to change the size of hypervariable domain, to promote the higher expression of Tomosyn-2 and to regulate the vesicle fusion in the synapse.

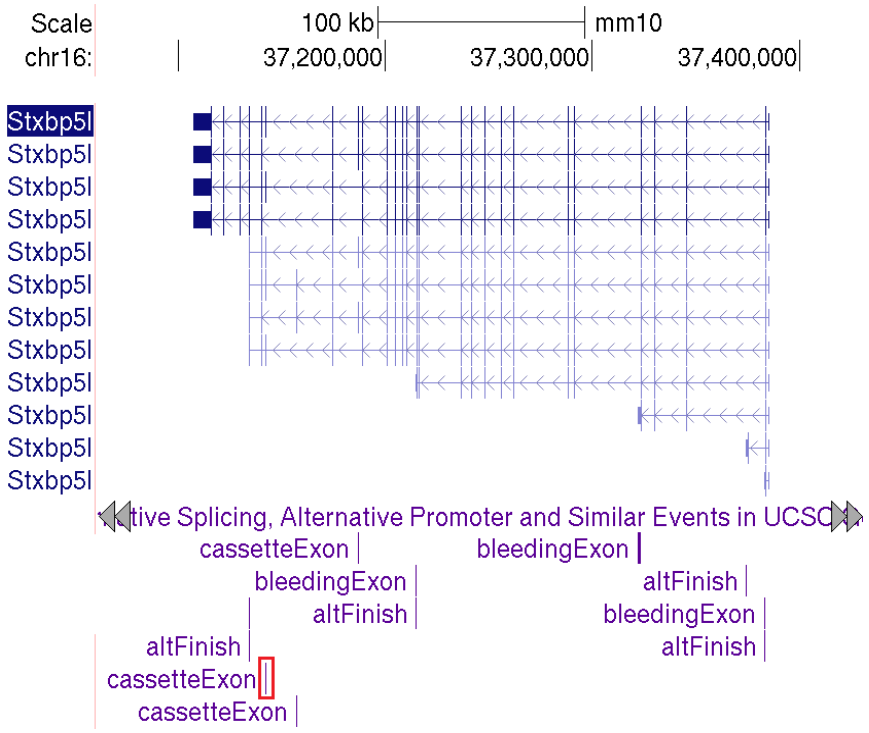


Figure 1.9 – *Tomosyn-2* splicing pattern. *Tomosyn-2* is alternatively spliced into 12 splice variants created by 11 splicing events. Cassette exon regulated by T-STAR is labelled by the red rectangle. The figure is adapted from UCSC genome browser.

1.3 Techniques to study protein-RNA interactions

1.3.1 An overview

Due to the biological and biomedical importance of alternative splicing, a set of experimental techniques has been developed to study the alternative splicing on a genome wide scale. They can be subdivided into the *in vivo*, *in vitro* and the *in silico* splicing assays. *In vivo* methods are used to model the mechanism of splicing regulation by different *cis*- and *trans*-acting factors in living cells and tissues. An example of *in vivo* methods are minigene assays. Recently, powerful high-throughput techniques as CLIP (UV cross-linking and Immunoprecipitation) and RNA-sequencing have been developed. In *vitro* methods aim to recapitulate the RNA-protein interactions happening *in vivo*. One of them is SELEX, which helps to identify the RNA-binding sequences for a specific *trans*-acting factor. *In silico* methods allow us to answer many questions on splicing through computational techniques. *In silico* methods are helpful in predicting splice site selection, the frequency of alternative events in a specific tissue, the consequences of mutations on splicing, or in modeling the structures of RNA-protein complexes, etc. To

conclude, the most efficient way to study alternative splicing is by combining the different *in vivo*, *in vitro* and *in silico* approaches (Garcia-Blanco, 2005; Witten and Ule, 2011).

1.3.2 Minigene assays

It is very common to study the splicing *in vivo* by using minigene splicing assays. To generate a minigene, a genomic region of interest (that usually encompasses the alternatively spliced exon with the flanking intronic sequence) is cloned into a plasmid vector. In my PhD, selected DNA fragments were cloned into the pXJ41 exon trap vector, in between the constitutively spliced β-globin exons of the vector. Subsequently, plasmids, containing the cloned fragments are transfected into the cells. Transcription in the vector was driven by the eukaryotic human cytomegalovirus promoter. Splicing patterns form the minigene are assessed by RT-PCR. The minigene assay is the most common technique nowadays to test the effect of a mutation on the splicing outcome. Splicing factor binding motifs, splice sites, or auxiliary regions are often mutagenized to replace, move, delete or insert the regulatory sequence elements and to study their effect on the splicing.

1.3.3 In-vitro splicing assays

In-vitro splicing assays are often used to characterize cis-acting elements and trans-acting splicing factors. During the assay, minigenes containing the constructs of interests are cloned. Consequently, they are digested. DNA is reverse-transcribed, radiolabelled and incubated with the whole nuclear extract from the cell or a specific protein of interest. The splicing products, resulting in the reaction are analyzed on the denaturing gel. EMSA (electrophoretic mobility shift assay) used in the current study is based on electrophoretic separation of protein-RNA complexes from the unbound RNA. EMSA do not require transfection and long maintenance of the cell lines. One EMSA limitation is that isotope decay time is limiting the experiment time. Furthermore, EMSA restricts the size of the tested fragment up maximum 250 nucleotides (Hicks *et al.*, 2005).

1.3.4 SELEX

SELEX (systematic evolution of ligands by exponential enrichment) is an *in vitro* screening assay that allows potential RNA sites bound by a specific protein to be selected. SELEX starts with a DNA library - up to 10^{16} oligonucleotides with fixed ends (Wang and Burge, 2008). This is transcribed into an RNA library, and incubated with a protein of interest. Unbound targets are removed by affinity chromatography, while bound ones are reverse-transcribed to DNA and amplified. Through a series of experimental cycles, only strongly bound targets are selected. Genomic SELEX technique was developed (Glisovic *et al.*, 2008). Through addition of adaptor sequences to DNA fragments in the library, Genomic SELEX uses genome as a starting point, rather than being randomized. One of the disadvantages of SELEX are non-specific bindings which result in identification of some non-functional RNA targets as reviewed by (Ule *et al.*, 2005; Elliott and Rajan, 2010). Furthermore, the data generated by SELEX always requires verification by *in vivo* techniques.

1.3.5 Yeast three-hybrid

The yeast three-hybrid system is based on two non-interacting proteins that are expected to bind a target RNA molecule, which will connect them as a linker. Once RNA is bound to one protein, the second protein activates transcription of a reporter gene and a visible output is produced. Yeast-three hybrid is inappropriate for proteins that are involved in transcription activation themselves (Hook *et al.*, 2005).

1.3.6 RNA-binding protein immunoprecipitation and Affinity tag assay

RBP immunoprecipitation (RIP) creates protein-target RNA cross-links in live cells by means of formaldehyde. Cells are then lysed and proteins of interest in complex with target RNAs are immunoprecepitated. However, the RIP technique is highly dependent on antibody specificity. Affinity tag assay is similar to RIP, but applies a designed tag, specific to a certain antibody, to attach and link a precipitated protein to antibody. Thus, no antibody specificity is required. However, the tag on itself might interfere with protein-RNA binding (Glisovic *et al.*, 2008).

1.3.7 Alternative splicing microarrays

The development of splice-sensitive microarrays allowed identification of a variety of splicing events in one experiment. Microarrays use multiple oligonucleotide probes (designed to recognize the exons, introns or exon-intron junctions) to hybridize with the labelled target RNA (Matlin *et al.*, 2005). The development of isoform-specific microarrays allowed different splice isoforms generated from one gene to be distinguished (Johnson *et al.*, 2003). Microarrays have some limitations, including cross-hybridization of probes and a high percentage of unspecific binding. Furthermore, as microarrays require pre-designed probes, they can only be applied for the organisms with the sequenced genomes (Mills and Janitz, 2012).

1.3.8 UV cross-linking and Immunoprecipitation (CLIP)

High throughput sequencing of RNA isolated by UV cross-linking and Immunoprecipitation (CLIP) can detect real physiological RNA targets. CLIP applies UV radiation to introduce irreversible RNA-protein cross-links in live cells/tissue homogenate. RNA-protein complexes are immunoprecipitated, cleaved by Proteinase K and separated by SDS-PAGE gel electrophoresis. RNAs then are reverse-transcribed and a cDNA library is generated. As a result, short CLIP tags are produced, which correspond to regions of original protein binding. CLIP tags are subsequently amplified and subjected to high-throughput sequencing. Thus, RNA targets from the whole transcriptome are identified. However, they still have to be functionally validated by subject protein depletion (e.g. by mouse knockout or by small interfering RNA (siRNA).

HITS-CLIP was successfully used to identify the splicing targets of FOX2 (Yeo *et al.*, 2004), NOVA (Ule *et al.*, 2006) and Tra2β (Grellscheid *et al.*, 2011). Recently, new types of CLIP, called iCLIP (individual nucleotide resolution CLIP) and PAR-CLIP (Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation) were developed. In contrast to older CLIP version, by means of RNA circularization iCLIP allows the detection of exact nucleotides cross-linked to protein. iCLIP also prevents the loss of prematurely truncated RNAs, thus providing more data (Konig *et al.*, 2011). As for PAR-CLIP, by applying photoactivatable nucleoside analogs that require a longer-wavelenght UV for the crosslinking, it has higher crosslinking efficiency compared to the other CLIP versions. One of the limitations of CLIP techniques are that

they do not distinguish between stable as and transient RNA-protein interactions. Also, the RNA binding sequences identified by CLIP are often degenerate, and differ from the natural motifs bound by the protein (Foot *et al.*, 2014). Furthermore, CLIP data requires extensive bioinformatics analyses, including ranking the targets by the number of CLIP tags, by the type of a regulated alternative even, or by their physiological functions (Ascano *et al.*, 2012).

1.4 Research aims and objectives

The study by Ehrmann et al 2013 has identified the first physiological splicing targets for T-STAR proteins, in the *Neurexin-1*, *Neurexin-2* and *Neurexin-3* and *Tomosyn-2* genes (Ehrmann et al., 2013). The splicing of *Neurexin-1 and Neurexin-3* genes was also shown to be regulated by Sam68 protein.

The key aim of my project was to identify the mechanism of splicing regulation of *Neurexin-2* and *Tomosyn-2* by T-STAR and Sam68 and to compare the T-STAR and Sam68 functions in splicing regulation.

The third Chapter of my thesis is focused on *Neurexin-2*. To uncover the mechanism of splicing by T-STAR and Sam68 I firstly aimed to test if T-STAR and Sam68 proteins bind to *Neurexin* target sequences directly. To complete this aim, I applied the EMSA technique. Subsequently, I aimed to identify the RNA binding sites for T-STAR and Sam68 in the *Neurexin* RNA. Also, my aim was to uncover the positional effect of T-STAR and Sam68 RNA binding motifs on the splicing outcome. To complete these aims, I performed multiple cloning experiments, applied the site-directed mutagenesis technique and assessed splicing patterns by the minigene splicing assays.

The fourth Chapter is focused on studying *Tomosyn-2* splicing regulation by T-STAR and Sam68. By site-directed mutagenesis experiments I was aiming to dissect T-STAR and Sam68 binding sites and to uncover the positional effect of T-STAR and Sam68 RNA binding motifs on the splicing outcome. I also aimed to identify if *Tomosyn-2* splicing is regionally regulated by T-STAR in mouse brain using RT-PCR experiments.

In the fifth Chapter, I aimed to compare the binding sequence specificity between T-STAR and Sam6 applying the site-directed mutagenesis technique and assessed splicing patterns by the minigene splicing assays.

In addition to my study on T-STAR and Sam68 proteins, as described in Chapter five, by using a series of mutagenesis and minigene splicing assays I aimed to uncover the mechanism of splicing regulation of *Neurexin-2* more generally, and whether it involved other STAR proteins.

Chapter 2: Materials and Methods

2.1 Electromobility Shift Assay (EMSA)

2.1.1 Construct cloning and protein preparation

GST-tagged T-STAR and Sam68 fusion proteins, required for EMSAs, were produced and purified by Mrs. Caroline Dalgliesh (Newcastle University). One region from the *Nrxn1* gene (92bp subject sequence) (Figure 2.1 A) and two short regions from *Nrxn2* (95bp subject sequence and 113bp negative control sequence) (Figure 2.1 B) were cloned in front of a T7 promoter in pBSK- (Bluescript) vector. Cloning was also performed by Mrs. Caroline Dalgliesh (Newcastle University).

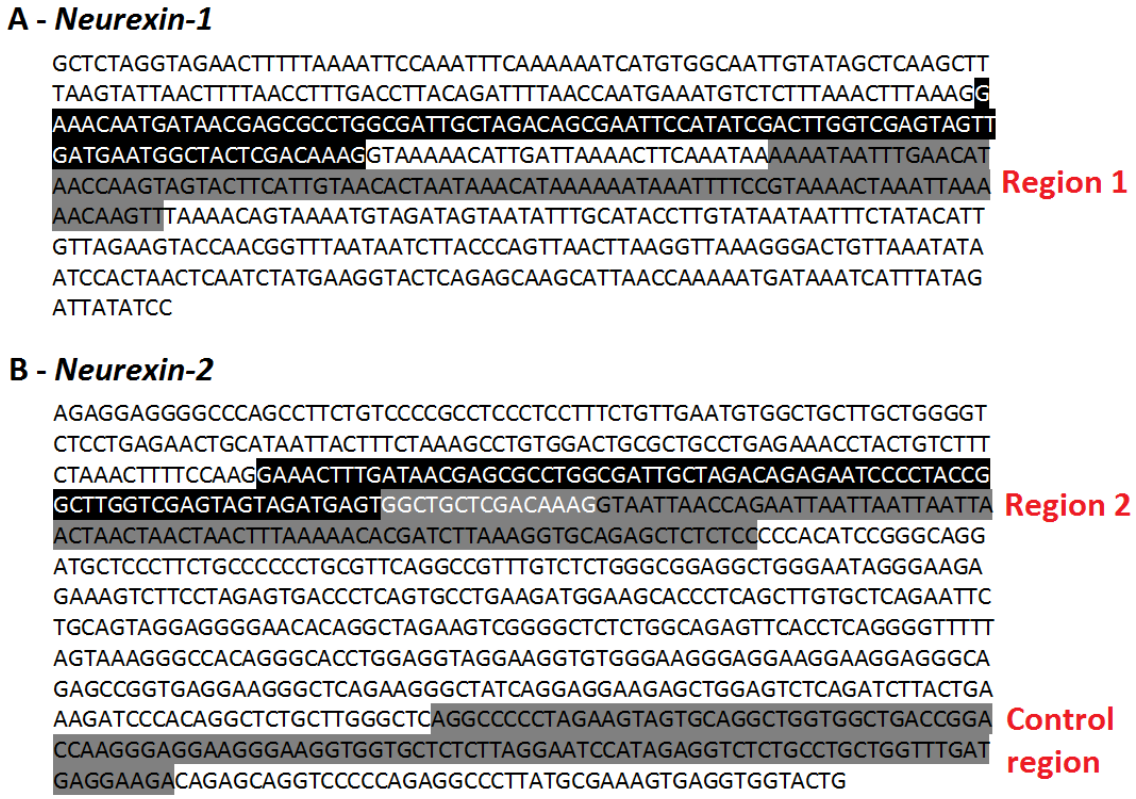


Figure 2.1 – The distribution of *Neurexin* sequences used for gel shift constructs to make. A – *Neurexin-1* cassette exon is highlighted in black, Region 1 used in gel shift experiment is highlighted in grey. B - *Neurexin-2* cassette exon is highlighted in black, Region 2 and a Control region used in gel shift experiment are highlighted in grey. Region 2 partially covers the exon.

2.1.2 Bradford assay to quantify protein concentrations

Bradford assay is a spectroscopy-based method for quantification protein concentrations. BSA was used as a protein standard. A set of eight standard protein concentrations was prepared: 0.625 ug/ml; 1.25 ug/ml; 2.5 ug/ml; 5 ug/ml; 10 ug/ml; 15 ug/ml; 20 ug/ml and 25 ug/ml. Buffer D was used to dilute BSA samples. Coomassie Blue stain was added to standard samples, 1:1. Absorbance of each standard was read on a spectrophotometer to a wavelength of 595 nm. A standard curve was produced from standard concentrations (absorbance versus the concentration). Absorbance for T-STAR-GST and Sam68-GST proteins was also measured. A standard curve was used to calculate the concentrations for T-STAR-GST and Sam68-GST.

Table 2.1: Components of Buffer D

Reagent	Volume
0.5M EDTA	0.2 ml
2.5M KCl	20 ml
Glycerol	25 ml
DEPC-treated H_2O	198.75 ml
1M DTT	0.25 ml
1M HEPES pH 7.9	20 ml

2.1.3 Verifying protein concentrations on the gel

To verify concentrations calculated in Bradford assay, equal concentrations of proteins were loaded on 10% SDS-PAGE gel. The gel was washed in Coomassie Blue satin. By visual analysis of stained bands the accuracy of protein concentrations was assessed.

2.1.4 Restriction digest

Cloned gel shift constructs were subjected to restriction digest in order to linearize them. Constructs were digested by *Eco*RI (following the instructions in Table 2.2). To look for digest quality, all constructs were run on 1% agarose gel. Non-digested samples were run nearby as controls. Digested samples were column purified using QIAGEN PCR Purification kit by manufacturer's protocol.

Table 2.2: Conditions for restriction digest of *Neurexin* constructs for EMSA

Constituent	Volume (µl) / Concentration
pBSK vector or PCR product	5.0 in 10x reaction buffer (2 µg of DNA)
*Eco*RI buffer	1.0 (1x)
*Eco*RI	3.0 (30 units)
BSA (100x)	6.0 (10x)

2.1.5 Transcription and labelling

To transcribe linearized DNA into RNA, 500ng of each sample was added into reaction mix described in Table 2.3.

Table 2.3: Transcription mix for EMSA

Reagent	Volume (µl)
T7 RNA polymerase buffer	3.5
AUC 10 mM	1.5
BSA 2 µg/µl	0.75
DTT 500mM	0.3
Cold GTP 0.5 mM	1.0
32P-GTP 0.5 mM	1.0
RNAsin 40U/µl	1.0
T7 RNA polymerase 20 U/µl	1.0
DEPC-treated H_2O	3.45
+ 500ng target DNA	1.0 to 2.0

Mixtures were vortexed and incubated at 37°C for 1.5 hours. To prevent vapour spreading after incubation, samples were briefly centrifuged. Following that, each sample was combined with DNase mix (see contents in Table 2.4; contents for DNAse buffer (1x): 2ml of 1M Tris; 1ml of 1M $MgCl_2$; DEPC-treated water up to 100ml), vortexed, incubated at 37°C for 25 minutes and briefly centrifuged.

Table 2.4: DNase mixture contents for EMSA

Reagent	Volume (µl)
DNase 10 U/µl	2.0
RNAsin 40 U/µl	2.0
DNase buffer (1x)	250.0

2.1.6 RNA Precipitation

Radiolabelled samples were mixed with 200µl of tRNA buffer (see contents in Table 2.5; components of 2xPK buffer see in table 2.6). Precipitation was performed in Manual Phase Lock gel column (PLGs) (purchased from '5 PRIME' company). Prior to experiment start, columns were centrifuged for 50 seconds at 13500rcf to move the gel to the tube bottom. 200µl Phenol-Chloroform and target samples were added to the columns and mixed by pipetting. Columns were centrifuged at 4°C, 12000rcf for 5 minutes. 250µl Chloroform was added, followed by another centrifugation step at 4°C, 12000rcf for 5 minutes. An upper aqueous phase (corresponding to RNA) was transferred to new eppendorfs. RNA was washed by 12.5µl 5M Ammonium Acetate for 45 minutes. Subsequently, RNA pellet was washed by 1) 750µl 100% DEPC Ethanol and 2) by 500µl 80% DEPC ethanol by centrifugation at 4°C, each time at 12000rcf for 5 minutes. All solution was removed from the tubes, and RNA pellets were left open to dry for 10 minutes. Afterwards, they were re-dissolved in 50µl DEPC-treated water, and RNA was denatured by incubating tubes at 80°C for 3 minutes.

Table 2.5: Components of tRNA mixture

Reagent	Volume (µl)
2xPK buffer	123.0
E.coli tRNA carrier	1.0
DEPC-treated water	72.0

Table 2.6: Components of 2x PK buffer

Reagent	Volume
1M Tris-HCl pH 7.4	20 ml
0.2M EDTA	5 ml
10% SDS	5 ml
5M NaCl	2 ml
DEPC-treated water	68 ml

2.1.7 Checking RNA quality

1µl from each RNA sample was diluted in 2µl of DEPC-treated water and 2µl of blue loading dye. Resulting mixes were electrophoresed on 6% Urea polyacrylamide gel (see gel constituents in Table 2.7; contents of 10x TBE in Table 2.8) at 650V for 15 minutes. Following electrophoresis, gel was dried on a vacuum gel drier for 1 hour and exposed to a storage phosphor screen overnight. The screen was scanned with a Typhoon Imager. After image analysis, the RNA dilution factor was decided. RNAs were diluted in DEPC-treated water aiming to have equal RNA concentrations in each tube.

Table 2.7: Polyacrylamide gel components for EMSA

Constituent	For 6% UREA polyacrylamide gel (6%, 10X TBE, Urea, 20ml volume)	For 6% polyacrylamide gel (6%, 10X TBE, 30ml volume)
Reagent	Amount	Reagent
Urea	8.4 g	-
Acrylamide 40%	3 ml	4.5 ml
10x TBE	2 ml	1.5 ml
dH_2O	Up to 20 ml	Up to 30ml
APS	200 µl	160 µl
TEMED	7.5 µl	30 µl

Table 2.8: Constituents of 10x TBE buffer

Reagent	Volume / Amount
Tris-NaOH	108g
Boric Acid	55g
0.5M EDTA pH 8.0	40 l

2.1.8 Electromobility Shift Assay

50, 100, 200 and 400ng of proteins (Sam68-GST and T-STAR-GST) were mixed with each RNA sample to check the RNA binding. Mentioned protein amounts were mixed with buffer D, adjusting buffer volume to a final volume of 12.5µl. A binding mix was also prepared (see Table 2.9 for constituents).

Table 2.9: Constituents of Binding mix for EMSA

Reagent	Volume (µl)
KCl 2.5M	1.5
NaCl 1M	0.5
NP-40 1%	2.5
BSA 2µg/µl	0.375
Glycerol 87%	0.375
DTT 100mM	0.25
RNAsin 40U/µl	1.0
DEPC-treated H20	4.0
tRNA E.coli 200ng/µl	1.0

11.5µl of Binding mix was combined with 12.5 µl of protein mix and 1 µl of diluted and denatured RNA. Resulting mixtures were vortexed, briefly centrifuged and incubated for 20 minutes at room temperature. Following that, mixes were run on a 6% Polyacrylamide gel (See Table 2.7 for constituents) for 2 hours at 3W. Gels were dried for 2 hours and exposed to the storage phosphor screen overnight. The screen was scanned with a Typhoon Imager. Images were analysed and suggestions about protein effect were made.

2.2 Cell culture

2.2.1 HEK-293 cells

The HEK-293 cell line was used for all transfections in this thesis. HEK293 is a human embryonic kidney cell line (Graham *et al.*, 1977). It has a high transfection efficiency and thus is often used in splicing assays. HEK293 cells were initially purchased from the American Type Culture collection and LGC Standards, Europe. Catalogue number is: ATCC-CRL-1573.

2.2.2 Cell culture conditions and maintenance

Culture was performed in a Class II laminar flow microbiological safety cabinet. Cells were grown in 25cm^2/75cm^2 flasks in 37°C incubator with 5% CO_2. DMEM (Dulbecco's Modified Eagle's Media) with phenol red (purchased from PAA) with 10% FBS added (purchased from Sigma-Aldrich) was used as a culture medium. Cells were passaged once 70-80% confluency was reached in the flask. To passage, DMEM was removed and cells were washed with 1xPBS (phosphate buffered Saline). This was followed by incubation with 2ml of 2mM trypsin-EDTA (Sigma-Aldrich) for 5 minutes at 37°C. To stop the effect of trypsin, the culture media was added. Cells were centrifuged at 180xg for 5 minutes. The supernatant was discarded, while cell pellets were dissolved in DMEM with 10% FBS and placed into new flasks for culture with a ratio 1:5 (diluted cells to new culture media).

2.2.3 Cryopreservation

Cells were regularly frozen to maintain a cell stock. After reaching a high confluency in the flask, cells were passaged as normally and aliquoted in 1 ml of cryoprotective media (95% FBS, 5% DMSO (dimethyl sulphoxide; purchased from Sigma-Aldrich)). Frozen stocks were kept at -80°C in cryovials (Sigma-Aldrich). To thaw the cells, cryovials were incubated in 37°C water bath for 5 minutes. To replace the cryoprotective media by DMEM cells were centrifuged at 180xg for 5 minutes. Freezing media was discarded. Cells were resuspended in DMEM with 10% FBS and placed for culturing into new flasks.

2.3 *Tomosyn-2* minigene cloning

2.3.1 Primer design

The *Tomosyn-2* alternative exon and approximately 300 nucleotides of flanking upstream and downstream intronic sequences were selected for cloning into the exon trap pXJ41 vector. Primers were designed in the Primer3 program available at: http://bioinfo.ut.ee/primer3-0.4.0/primer3/. An *Eco*RI restriction site and a string of six adenine bases was added to the 5' end of each primer. Primers are listed in Table 2.10. The sequence used for *Tomosyn-2* cloning is shown on Figure 2.2.

Table 2.10: Primers used for *Tomosyn-2* cloning

Primer	Sequence
Tomosyn-2 cloning primer F	AAAAAACAATTGGCATATTTCATATTG CCATCCA
Tomosyn-2 cloning primer R	AAAAAACAATTGCAGTAGAGGAAATTAAGGTTGCAG

GCATATTTCATATTGCCATCCAacaatacattaaataaatagtagggaagtatatgacatttcaagaaatgttaattaga
taaattatgaaaatatgtcctatacataataagtatcttgatttttttataatctggaattgtaactaaatcttaaaataacaaaaaa
aaatagtcatattcaaatgaattggaagaagaggaattcctgttcctccatttcttaacctatgcctttggatttcctcatctgaagac
ttgattatgatatttattctgccttcttcctcttggtacagACCATGTAAATGGACACTGCACAAGTCCAACCTCTCA
GAGTTGCAGTTCTGGAAAACGTCTTTCCAGTGCTGACGTTTCAAAAGTAAATCGCTGGGGTCCTG
GAAGACCACCATTTAGAAAGGCACAGTCAGCTGCTTGCATGGAGATTTCTTTACCAGTTACAACT
GAAGgtaggataaacatatcaatttctgtgagtacagtttaaaatttgataaaatttcgtttatcatggtaacaggtgctgggaga
caaagaaatggaagttacatttaaaagatgatttaaaaacaaaaacaaaaacaaaactacttgacattgaaaaattaagatcct
gtttgactgtggccataacttcaattctattacttactcacttccttccttatctctaaaaataaaatgatcccactgtaggtagccatg
caaggagataatctgtaaCTGCAACCTTAATTTCCTCCTG

Figure 2.2 - *Tomosyn-2* sequence used for minigene cloning. Figure shows the *Tomosyn-2* region that was predicted to respond to T-STAR and was cloned into pXJ41 vector. Cassette exon is highlighted in black. Sequences for cloning primers are underlined.

2.3.2 PCR amplification

50µl reactions were performed by the protocol shown in Table 2.11. PCR Phusion Kit from *Thermo Scientific* was used. Amplification was carried out in SensoQuest thermo cyclers by the program shown in

Table 2.12. All PCR products were column purified using the QIAquick PCR Purification Kit following a standard manufacturer's protocol.

Table 2.11: PCR protocol

Reagent	Volume (µl) / Concentration
5xHF buffer	10.0 (1x)
dNPs (10mM)	1.0 (200uM)
10µM Forward Primer	2.5 (0.5 uM)
10µM Reverse Primer	2.5 (0.5 uM)
DMSO	1.0 (3%)
Template (mouse genomic DNA)	1.0 (< 250 ng)
Phusion DNA Polymerase	0.5 (1.0 units/50 µl PCR)
dH_2O	31.5

Table 2.12: Thermocycler PCR program

Step	Temperature °C	Time	Cycle Number
Heat activation	98°C	30 seconds	1
Denaturation	98°C	10 seconds	32
Annealing	60°C	20 seconds	32
Extension	72°C	45 seconds	32
End Stage Extension	72°C	4 minutes	1
Cooling	15°C	∞	1

2.3.3 Gel electrophoresis

PCR products were run on 1.2% agarose gels for product size estimation. Lanes corresponding to amplified products were cut out. Constructs were purified from the gel (using QIAquick Gel Extraction Kit by manufacturer's instruction).

2.3.4 Restriction digests

To create a single cut in the vector, the pXJ41 vector stock was digested by *Mfe*I restriction enzyme (New England Biolabs). Samples were digested by *Eco*RI enzyme (New England Biolabs) (Table 2.13). Digestion mixes were incubated for 3 hours at 37°C followed by a 15 minute heat inactivation at 65°C. Afterwards, digestion mixes were column purified using QIAquick PCR Purification Kits (by standard manufacturer's protocol). Following digestion, to prevent the vector re-circularization, the vector was also treated by Antarctic Phosphatase (New England Biolabs) (Table 2.14). Vector digest was incubated at 37°C for 45 minutes.

Table 2.13: Components of restriction digestion reaction

Constituent	Volume (µl) / Concentration
pXJ41 vector or PCR	50.0 in 1x reaction buffer (2-3 ug of DNA)
*Mfe*I/*Eco*RI buffer (10x)	1.0 (1x)
*Mfe*I/*Eco*RI	3.0 (30 units)
BSA (100x)	6.0 (10x)

Table 2.14: Components of vector digest by Antarctic Phosphatase

Constituent	Volume (µl) / Concentration
pXJ41 vector	60.0 (2-3 ug of DNA)
Antarctic Phosphatase buffer	7.0 (1x)
Antarctic Phosphatase	3.0 (30 units)

2.3.5 Ligation and Re-cleavage

Ligation reactions were performed in steps described in Table 2.15, at 16°C overnight. To remove the non-digested/re-circularized vector background, ligation mixes were re-cleaved with *Mfe*I restriction enzyme (Table 2.16). Re-cleavage mixes were incubated at 37°C for 2 hours, followed by 15-minute heat inactivation at 65°C.

Table 2.15: Standard ligation protocol

Component	Volume (µl) / Concentration
Insert	8.0 (0.5 ug of DNA)
Vector	3.0 (0.2 ug of DNA)
T4 Buffer (10x)	2.0 (1x)
T4 DNA Ligase (Promega)	2.0 (20 units)
dH_2O	5.0

Table 2.16: Components of re-cleavage reaction

Component	Volume (µl)
Ligation mix	20.0 (0.7 ug of DNA)
*Mfe*I buffer (10x)	5.0 (1x)
*Mfe*I	2.0 (20 units)
dH_2O	23.0

2.3.6 Cloning and Sequencing

4µl of inserts were transformed into 25µl of commercial competent cells (α-Select Chemically Competent Cells from BIOLINE) using the BIOLINE heat shock protocol. Transformants were allowed to settle on LB + Ampicillin (1µl ampicillin per 1ml LB) plates overnight. Subsequently, 8 to 15 colonies were picked from each plate and grown overnight (in separate tubes) in LB + Ampicillin, in the incubated shaker at 37°C. Bacteria from each colony were subjected to colony screening by PCR, when one primer of vector (see vector primers in Table 2.17) and one cloning primer specific to the insert was used to make sure that bacteria contained a correct plasmid (Table 2.18 for PCR conditions). Bacteria, amplified overnight in a shaker, were centrifuged for 1.5minute at 11 000rcf. Plasmids were extracted from a resulting pellet by means of QIAprep Miniprep Kit,

following QIAGEN instructions. Plasmid DNA was sent for sequencing with vector primers (pXJB1 and pXJRTF, Table 2.17) to Source Bioscience, Oxford. Sequences were verified by aligning the expected and received sequence versions using the CLUSTALW2 program.

Table 2.17: pXJ41 vector primers used for colony screening and for sequencing

Primer	Sequence
pXJB1	GCTCCGGATCGATCCTGAGAACT
pXJRTF	GCTGCAATAAACAAGTTCTGCT

Table 2.18: Thermocycler PCR program for colony screening

Step	Temperature	Time	Cycle Number
Heat activation	95°C	2 minutes	1
Denaturation	95°C	1 minute	35
Annealing	58°C	1 minute	35
Extension	72°C	1 minute	35
End Stage Extension	72°C	5 minutes	1
Cooling	4°C	∞	1

2.4 Site-directed mutagenesis

2.4.1 Primer design

Site-directed mutagenesis by overlap extension PCR was performed to introduce mutations into potential T-STAR and Sam68 binding regions of the *Nrxn2* and *Tomosyn-2* sequence. All primers and templates used to make mutants are listed in Appendix I. Figure 2.3 shows the regions of *Neurexin-2* (Figure 2.3 A) and *Tomosyn-2* (Figure 2.3 B) mutagenized in deferent experiments. The method of mutagenesis is described in Figure 2.4.

Figure 2.3 - Sequence regions altered through site-directed mutagenesis. A – *Neurexin-2* sequence region that was cloned into the pXJ41 vector and subsequently subjected for mutagenesis. Exon is highlighted in black. T-STAR target region is highlighted in blue. Numbers highlighted in blue show the downstream (163 and 241 nucleotides downstream the target exon) and upstream (42 and 100 nucleotides upstream the target exon) positions for the target region insertion. B – *Tomosyn-2* sequence region that was cloned into the pXJ41 vector and subsequently subjected for mutagenesis. Exon is highlighted in black. Potential T-STAR target regions are highlighted in blue and purple. Number highlighted in blue show an upstream position for the first (blue) target region insertion (54 bases upstream the exon).

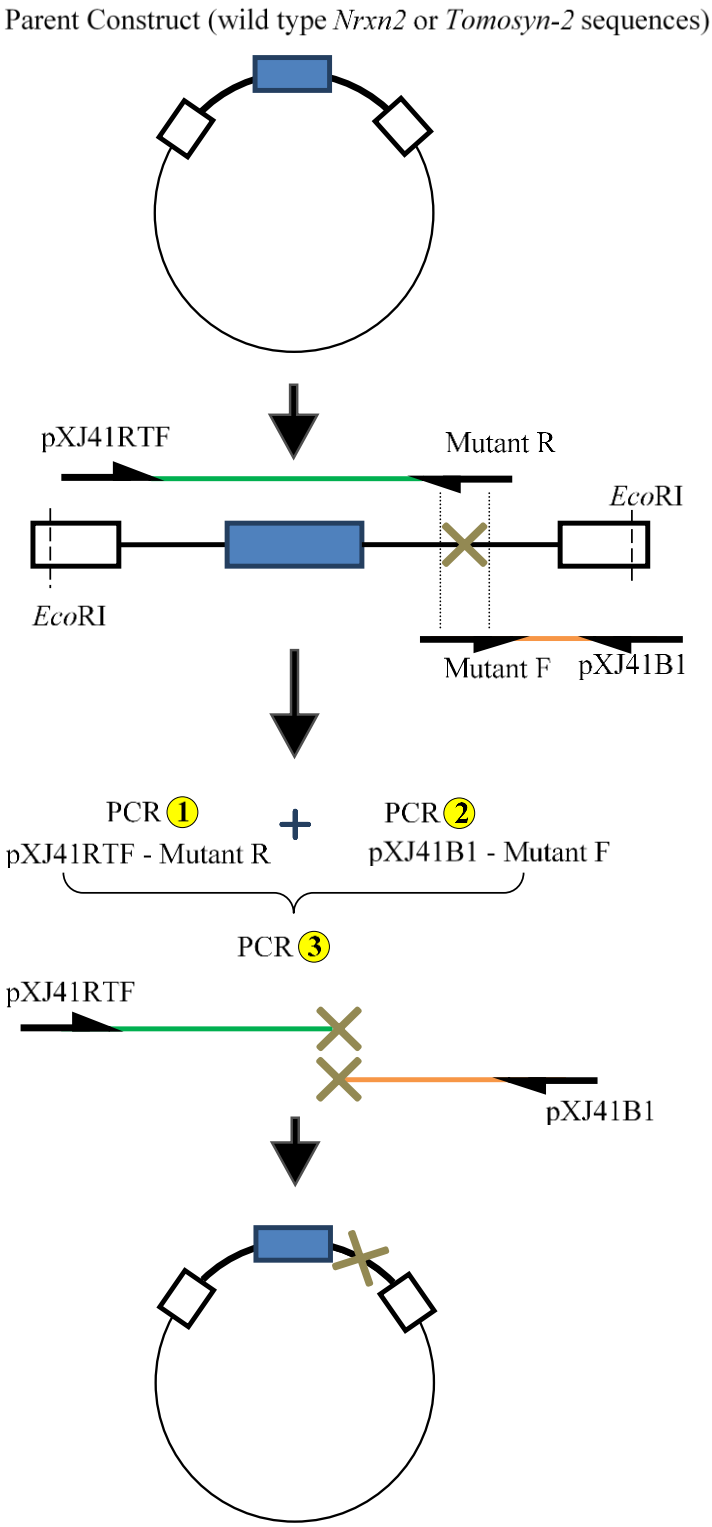


Figure 2.4 - Site-directed mutagenesis. A schematic showing site-directed mutagenesis by overlap-extension using PCR. Mutagenic primers with a number of single base changes were designed to create specific mutations. These primers were used in PCR 1 and PCR 2 reactions, which produced two overlapping DNA fragments with complementary mutations in opposite DNA strands. Both products were then combined and then re-amplified by PCR3. The final product was then cloned into pXJ41 vector.

2.4.2 PCR amplification

50µl reactions were performed using the protocol stated in Table 2.19 (PCR 1 and 2) and Table 2.20 (PCR 3). PCR Phusion Kit from *Thermo Scientific* was used. Amplification was carried out in SensoQuest thermo cyclers by the program stated in Table 2.12. All PCR products were column purified applying QIAquick PCR Purification Kit (following the standard manufacturer's protocol).

Table 2.19: PCR 1 and PCR 2 protocol

Reagent	Volume (µl) / Concentration
5xHF buffer	10.0 (1x)
dNPs	1.0 (200 uM)
Mutant Reverse (PCR 1) / Mutant Forward primer (PCR 2)	2.5 (0.5 uM)
pXJRTF (PCR 1) /pXJB1 (PCR 2)	2.5 (0.5 uM)
DMSO	1.0 (2%)
Template (see table 2.16)	1.0 (< 250 ng)
Phusion DNA Polymerase	0.5 (1.0 unit)
dH_2O	31.5

Table 2.20: PCR 3 protocol

Reagent	Volume (µl) / Concentration
5xHF buffer	10.0 (1x)
dNPs	1.0 (200 uM)
pXJB1	2.5 (0.5 uM)
pXJRTF	2.5 (0.5 uM)
DMSO	1.0 (2%)
Template 1 – (PCR 1)	1.0 (< 250 ng)
Template 2 – (PCR 2)	1.0 (< 250 ng)
Phusion DNA Polymerase	0.5 (1.0 unit)
dH_2O	30.5

2.4.3 Gel Electrophoresis

PCR 1 and 2 products were run on a 1.2% agarose gels for product size estimation only. PCR 3 products were run on 2% agarose gel. Lanes corresponding to amplified products were cut out. Constructs were purified from the gel (using QIAquick Gel Extraction Kits according to manufacturer's instructions).

All the rest of the steps for site-directed mutagenesis were performed exactly the same as described in parts 2.3.4, 2.3.5 and 2.3.6.

2.5 Minigene splicing assays

2.5.1 Transfection

Verified plasmids and plasmids encoding GFP fusion proteins (e.g. T-STAR-GFP and Sam68-GFP; cloned and provided to me by Mrs. Caroline Dalgliesh, Newcastle University) were co-transfected into HEK-293 cells. Transfections were performed using GeneJammer by a protocol in Table 2.21. In every transfection experiment in addition to the samples of interest, *Neurexin-2* (or *Tomosyn-2*) minigene was co-transfected together with: 1) GFP plasmid, 2) T-STAR-GFP, 3) Sam68-GFP, as control samples.

Table 2.21: Transfection protocol

Reaction Mix	Constituents	Volumes	Incubation time
Mix 1	GeneJammer + DMEM	3µl of GeneJammer, 97µl DMEM	13 minutes
Mix 2	Mix 1 + Plasmids + expression vectors	200ng plasmids + 500ng expression vectors + 100µl Mix1	30 minutes
Short Centrifugation			
Mix 3	Mix 2 + HEK-293 (pre-plated for 24 hours in 6-well plates)		37°C, overnight

Following overnight incubation, the transfection efficiency was checked using a fluorescent microscope – cells that had accepted GFP-linked constructs were fluorescent. T-STAR and Sam68 proteins are normally localized in the nucleus. Efficiency was assessed by eye. When 60-70% of cells (or more) had fluorescent nuclei, experiment was continued. Cells were equally divided into 2 tubes and centrifuged at for 5min at 7000rcf. Resulting pellet aliquots were subjected to RNA extraction and Western blotting.

2.5.2 RNA extraction

Each pellet was completely dissolved in 100µl Trizol (by vortexing). 20µl of chloroform reagent was added. Samples were then vortexed and centrifuged at 13000rcf at 4°C for 15 minutes. The upper aqueous phase corresponding to RNA was moved to separate tubes, and 50µl Isopropanol was added. Samples were vortexed, followed by a 10 minute incubation on the bench. Samples were subsequently centrifuged at 13000rcf at 4°C for

10 minutes. All supernatant was removed. Pellets were washed using 75% ethanol in DEPC-treated water, and centrifuged at 6700rcf at 4°C for 5 minutes. As previously, all solution was removed and pellets were left to air-dry in open tubes for 10 minutes. Samples were then diluted in 50µl DEPC-treated water, warmed in a hot block at 65°C and vortexed. 50ng/µl RNA concentrations were prepared for reverse transcription PCR (RT-RCR).

2.5.3 RT-PCR

RT-PCR was performed using QIAGEN 1-step RT-PCR kit using the recommended QIAGEN protocol. Reaction components and amplification steps are shown in Table 2.22 and Table 2.23, respectively.

Table 2.22: RT-PCR reaction constituents

Reaction component	Volume (µl) / Concentration
Q solution	1.0 (1x)
5X buffer	1.0 (1x)
dNTPs	0.2 (200 uM)
pXJRTF (10µM)	0.3 (0.3 uM)
pXJB1 (10µM)	0.3 (0.3 uM)
RNA (50ng/µl)	2.0 (0.1 ug)
Enzyme mix	0.2

Table 2.23: RT-PCR reaction steps

Reaction Step	Temperature	Time	Cycle №
Reverse transcription	50°C	30 minutes	1
Heating	95°C	15 minutes	1
Denaturation	94°C	30 seconds	32
Annealing	56°C	30 seconds	32
Primer extension	72°C	1 minute	32
Final extension	72°C	10 minutes	1
Incubation	4°C	∞	1

2.5.4 DNA size verification and product quantification

DNA was electrophoresed on a QIAxcel multi-capillary gel electrophoresis (Qiagen) device. The QIAxcel Bio calculator software (Qiagen) was used to measure DNA size (bp) and concentration (ng/µl). The percentage of splicing exclusion (PSE) was calculated by the formula:

$$PSE\ (\%) = \frac{\text{Concentration of PCR band for exon exclusion (ng/ul)}}{\text{A sum of oncentrations for PCR bands for exon inclusion and exon exclusion (ng/ul)}} \times 100$$

2.5.5 Western Immunoblotting

Half of each transfected sample was tested for protein expression levels using Western blotting, applying a primary mouse polyclonal α-GFP antibody (Abcam, ab1218) and a secondary sheep α-mouse HRP-linked secondary antibody (Amersham, NA931VS). Loading and transfection efficiency was verified by actin immunoblotting using a primary rabbit α-actin rabbit polyclonal antibody (Sigma) and a secondary α-rabbit antibody (Jacson Lab). All antibodies used were diluted 1:2000. Briefly, cell samples were lysed in 2x SDS loading buffer and boiled at 100°C for 5 minutes. Proteins were separated on 10% SDS-PAGE gels, followed by transfer to PVDF membrane (Hybond-P, GE). The membranes were blocked in blocking solution (TBST (Tris-Buffered Saline with Tween 20) with 5% non-fat dry milk) for one hour and incubated with a primary antibody for one hour. Membranes were subsequently washed with TBST (three times, 5 minutes each time). Membranes were further probed with a secondary antibody for 1 hour, followed by washing with TBST (three times, 5 minutes each time). Prior to exposing membranes to a film, ECL (enhanced chemiluminescent) Prime Western Blotting Detection Kit (Amersham) was applied.

2.6 *Tomosyn-2* splicing analysis in brain tissues by RT-PCR

The splicing pattern of the T-STAR regulated *Tomosyn-2* exon was analysed in 11 brain regions of three adult wild type and T-STAR knockout mice. The regions tested were: olfactory bulb, cortex, hippocampus, basal ganglia, thalamus, hypothalamus, midbrain, cerebellum, pons, medulla and spinal cord. Brain dissection and tissue preparation, as well as RNA extraction was performed by Dr Ingrid Ehrmann (Newcastle University). For some samples, RNA was converted by me to cDNA and amplified by PCR using a 2-

step RT-PCR reaction (Superscript III Reverse Transcriptase kit, Invitrogen) (Table 2.24) shows the first cDNA strand synthesis reaction protocol. Primers used for reaction were designed in Primer 3 program (http://bioinfo.ut.ee/primer3-0.4.0/primer3/) and are listed in a Table 2.25. However, majority of brain cDNA samples were kindly provided to me ready made by Dr Ingrid Ehrmann (Newcastle University).

Table 2.24: RT-PCR reaction steps and constituents

Reaction component	Volume (µl) / Concentration
Primers (random 6-mers)	1.0 (0.5 uM)
RNA template	1.0 (0.1-0.3 ug)
10x dNTPs	1.0 (200 uM)
DEPC-treated H20	10.0
Incubate at 65°C for 5 minutes. Incubate on ice for 1 minute. Centrifuge.	
5x First Strand Buffer	4.0
0.1 M DTT	1.0
RNAse OUT	1.0 (40 units)
Superscript III RT	1.0 (200 units)
Incubate at 25°C for 5 minutes. Incubate at 50°C for 1 hour. Inactivate at 70°C for 15 minutes. Store at -20°C.	

Table 2.25: *Tomosyn-2* primers used for RT-PCRs in different mouse brain regions

Primer	Sequence
Tomosyn-2 in brain F	TATCACCGAAGGGACTGTCG
Tomosyn-2 in brain R	TGGCAAGCTTAATTTCCTTGA

The quality of the produced cDNA was first amplified by PCR and assessed on agarose gel for a house keeping gene *Hprt*, which is known to be equally expressed in all brain regions. Following that, cDNA from different brain regions was tested with primers for the *Tomosyn-2* target exon. Reaction components and PCR conditions for the *Hprt* and *Tomosyn-2* cDNA amplification are shown in Table 2.26 and Table 2.27, respectively.

Table 2.26: PCR reaction constituents

Reaction component	Volume (µl)
5x GoTaq buffer (Promega)	5.0 (1x)
10mM dNTPs	0.5 (0.5 uM)
Forward primer	0.5 (0.2 uM)
Reverse primer	0.5 (0.2 uM)
cDNA	1.0 (<0.5µg)
GoTaq (Promega)	0.15 (0.75 units)
dH_2O	17.35

Table 2.27: Thermo cycler PCR program

Step	Temperature °C	Time	Cycle Number
Heat activation	94°C	2 minutes	1
Denaturation	94°C	1 minute	30
Annealing	60°C	1 minute	30
Extension	72°C	1 minute	30
End Stage Extension	72°C	10 minutes	1
Cooling	15°C	∞	1

Chapter 3: Dissecting the mechanism of splicing of *Nrxn2* cassette exon by T-STAR

3.1 Introduction

T-STAR (testis – signal transduction and activation of RNA) and Sam68 (Src-associated in mitosis, 68Kda) are two splicing factors with an important role in alternative splicing. They both belong to the same STAR protein family and have a similar domain structure. Sam68 is ubiquitously expressed, while T-STAR expression is specific to brain, testis and kidney. Both proteins have been predicted to bind AU-rich motifs on pre-mRNA (Lin *et al.*, 1997; Chawla *et al.*, 2009; Pedrotti *et al.*, 2010; Feracci *et al.*, 2014; Galarneau and Richard, 2009a) . Sam68 was shown to regulate alternative exons in a number of targets: *CD44* (in cancer), *SMN2* (in muscle), *Numa1* and *Sorbs1* (in neurogenesis) (Bielli *et al.*, 2011; Iijima *et al.*, 2011). However, this response to Sam68 was only detected in minigene assays, and has not been physiologically validated. For a long time, there were no T-STAR targets known.

As a mentioned in Introduction sections, four genes have been recently identified as physiological splicing targets of T-STAR in mouse brain. These are: *Tomosyn-2*, *Neurexin-1*, *Neurexin-2* and *Neurexin-3* (one alternative cassette exon in each) (Ehrmann *et al.*, 2013). These targets opened a possibility to study the mechanism of splicing regulation by T-STAR and to compare it to that of Sam68.

Before I started this project, some work had already been performed to uncover the mechanism of T-STAR and Sam68 splicing regulation of *Neurexin* genes. *Neurexins* showed region-specific splicing repression in mouse brain, with highest repression in forebrain-derived regions. This correlated with the T-STAR expression pattern in the brain. Also, minigene assays confirmed that T-STAR and Sam68 both repress the splicing of a target cassette exon 20 of *Neurexin-1* and *Neurexin-3* (equivalent to AS4). *Neurexin-2* AS4 was only repressed by T-STAR, and did not respond to Sam68 (Figure 3.1). Thus, Sam68 and T-STAR, even being similar proteins, have some shared and some different splicing targets (Ehrmann *et al.*, 2013). In this Chapter by means of EMSA assay, I investigated whether T-STAR and Sam68 bind *Neurexin-1* and *Neurexin-2 in vivo*.

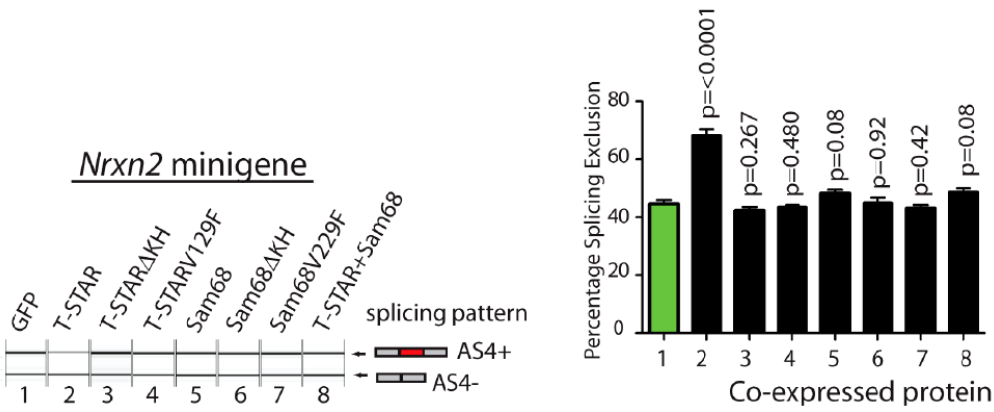


Figure 3.1 – *Neurexin-2* splicing is specifically repressed by T-STAR but not by Sam68. Different versions of *Neurexin-2* minigene exon were co-transfected with GFP-only, T-STAR, Sam68, T-STAR and Sam68 mutants, as well as T-STAR and Sam68 together. Transfection efficiency was assessed by the presence of fluorescence in HEK293 cell nuclei, using the fluorescent microscope. Transfected cells were harvested and subjected to RNA extraction. RNA was reverse-transcribed into cDNA. Transfected *Neurexin-2* cDNA were amplified applying pXJ41 vector primers. The PCRs were loaded on QIAxcel, as a result, the gel pictures and concentrations for included/excluded bands were observed. Figure shows the gel picture and the graphs for the percentage of splicing exclusion (PSE%) of AS4 in *Neurexin-2* transfected with different constructs. Statistical significance was calculated applying the independent two-sample t-test. Error bars show standard errors of the mean. Figure is adopted from (Ehrmann *et al.*, 2013).

Previously, by means of SELEX, and recently by NMR studies, potential binding sites for T-STAR have been identified. Based on SELEX, T-STAR tends to bind UAAA/UUAA motifs (Lin *et al.*, 1997; Galarneau and Richard, 2009b) Based on NMR, T-STAR was shown to bind UAA motif, specifically followed by A and preceded by U (as UUAAA) (Foot *et al.*, 2014). However, these are the predicted sequences from *in vitro* experiments only, not confirmed by any real targets. Here I examine what are the sequences needed for T-STAR regulation of a real physiological target *Nrxn2* (*Neurexin-2*). In the study by Ehrmann et al, to uncover the T-STAR regulatory sequence in *Nrxn2*, five U(U/A)AA motifs downstream from the AS4 were altered by mutagenesis (in this thesis I refer to the mentioned mutant as to the "*Neurexin-2* mutant with a big mutation"). Mutation of U(U/A)AA repeats completely blocked exon exclusion by T-STAR (Ehrmann *et al.*, 2013). In this Chapter I dissected this ISS to find the important and less important sites for T-STAR splicing regulation of *Nrxn2*.

There are many examples in the literature when binding of RNA binding proteins (RBPs) share common positional principles. For instance, for RBPs including Nova, PTB, Fox

and muscleblind, binding near the branch point upstream of the exon, near splice sites or within the exon leads to exon exclusion (splicing repression) (Dredge *et al.*, 2005; Zhang *et al.*, 2008b; Goers *et al.*, 2010; Llorian *et al.*, 2010). Repression is thought to happen because of protein competition with core spliceosome components. As one example, SR proteins were shown to bind an intronic sequence close to the branch point in adenovirus. This blocks 3' splice site usage and represses the splicing (Kanopka *et al.*, 1996; Kanopka *et al.*, 1998). RBPs binding downstream of exons, like TIA and hnRNP L, often activate splicing. There are several mechanisms how splicing activation might happen. One of the common ones is when an RBP interacts with and stabilizes U1 snRNP, which is a splicing initiator (Witten and Ule, 2011). This is a general rule, but the precise mechanism for each splicing event is more sophisticated. Different from the general tendency, *Nrxn2* AS4 splicing is repressed by T-STAR with binding sites downstream of the regulated cassette exon. Here I investigate if *Nrxn2* splicing repression by T-STAR is position dependant and test the hypothesis that T-STAR represses the splicing of *Nrxn2* by blocking the 5' splice site.

3.2 The aims of this Chapter were:

1. To investigate whether T-STAR and Sam68 directly interact with *Nrxn2* response element

2. To dissect the target response element into smaller elements and investigate their importance for *Nrxn2* splicing exclusion and splicing repression by T-STAR

3. To investigate the importance of the upstream and downstream position of response element for splicing regulation

4. To study whether T-STAR is blocking splicing exclusion by means of blocking the 5' splice site

3.3 Results

3.3.1 Both T-STAR and Sam68 bind Nrxn2 target sequence directly

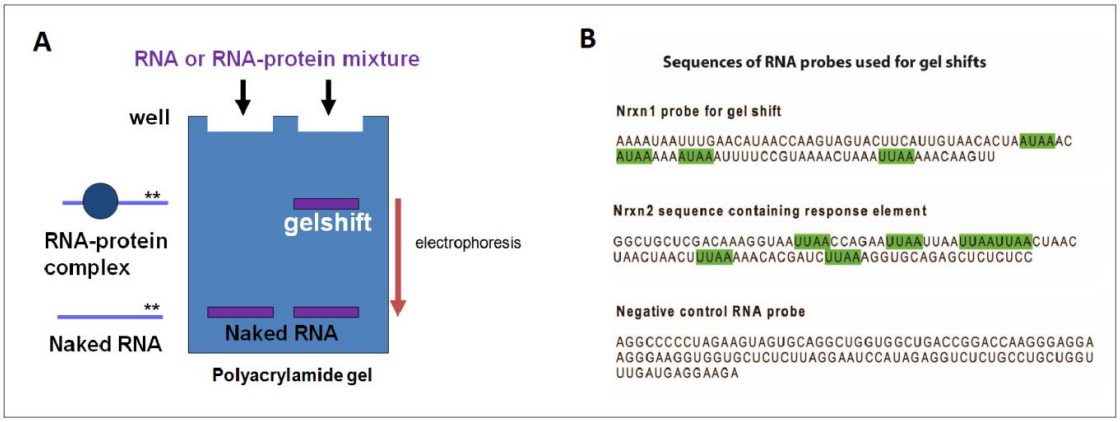


Figure 3.2 – EMSA assay is performed in a series of steps. A – In the EMSA assay a radiolabelled RNA probe is mixed with purified protein. Then, the naked RNA (as a control) or RNA-protein mixture is subjected to gel electrophoresis. Unbound RNA is expected to migrate further down in the gel, while RNA bound by the protein will generate a heavier complex that will stop earlier, producing a higher size band on a gel. B – The sequences of the RNA probes used for EMSA. Sequences coding for *Neurexin-1* and *Neurexin-2* response elements each containing potential sites for protein binding (highlighted in green). A negative control probe is taken from a further downstream sequence of *Neurexin-2* that does not contain any potential STAR protein binding sites.

Minigene studies have shown *Nrxn1* and *Nrxn3* to be regulated by both Sam68 and T-STAR. *Nrxn2* was shown to be regulated by T-STAR only (Ehrmann *et al.*, 2013). As long as T-STAR and Sam68 were both predicted to bind similar sequences, I aimed to see whether both proteins could bind *Nrxn2 and Nrxn1* directly. Alternatively, T-STAR might bind to *Nrxn2* better than Sam68. For this, EMSA assays were performed.

GST-tagged T-STAR and Sam68 fusion proteins, required for EMSAs, were produced by Mrs. Caroline Dalgliesh (Newcastle University). RNA probes comprised of regions of *Nrxn1 and Nrxn2* genes were cloned into pBluescript vector. Two probes contained predicted sequences for T-STAR and Sam68 binding. A third probe was cloned from *Nrxn2* sequence localized further downstream from the target region, known to be important for splicing regulation, to serve as a negative control for the binding assays (Figure 3.2 B). Prior to EMSAs, I adjusted the concentrations of Sam68-GST and T-

STAR-GST fusion proteins to be equal, by running them on a Coomassie stained gel (Figure 3.3 A).

The full experimental details are given in section 2.1 of the Materials and Methods Chapter. Briefly, before the EMSA assays, pBluescript plasmids were transcribed into P^{32} radiolabelled RNA probes. RNAs were diluted, and incubated with purified full length Sam68-GST and T-STAR-GST fusion proteins and analysed by gel electrophoresis. In case of RNA-protein binding, a shift was expected on a gel (Figure 3.2 A for a general EMSA method).

EMSAs were used to test the binding of T-STAR and Sam68 to the labelled RNA probes. Lane 0 of each gel shows how the probe migrates in the absence of T-STAR or Sam68 protein (Figure 3.3 B). Addition of either T-STAR or Sam68 protein prevented the *Nrxn2* RNA probe from moving outside the well into the gel, suggesting the formation of large molecular complexes on this probe. T-STAR protein bound to and shifted the *Nrxn2* probe, with a complete shift observed with 100 ng of added T-STAR protein. Sam68 protein also shifted the *Nrxn2* RNA probe, but with a maximal shift observed at 200 ng of added protein (a higher concentration than for T-STAR, Figure 3.3 B). Similar results were observed for the *Nrxn1* gene (Figure 3.3 C). No binding was observed to a control RNA probe (Figure 3.3 D).

Even though Sam68 did not regulate the splicing of *Nrxn2* in a minigene, these EMSA results suggest that *Nrxn1* and *Nrxn2* are both bound by both T-STAR and Sam68. However, a higher concentration of Sam68 compared to that of T-STAR is required for binding.

The described EMSA results were included in the publication on T-STAR and *Neurexin* by Ehrmann et al 2013, making me a co-author of this paper (Ehrmann *et al.*, 2013).

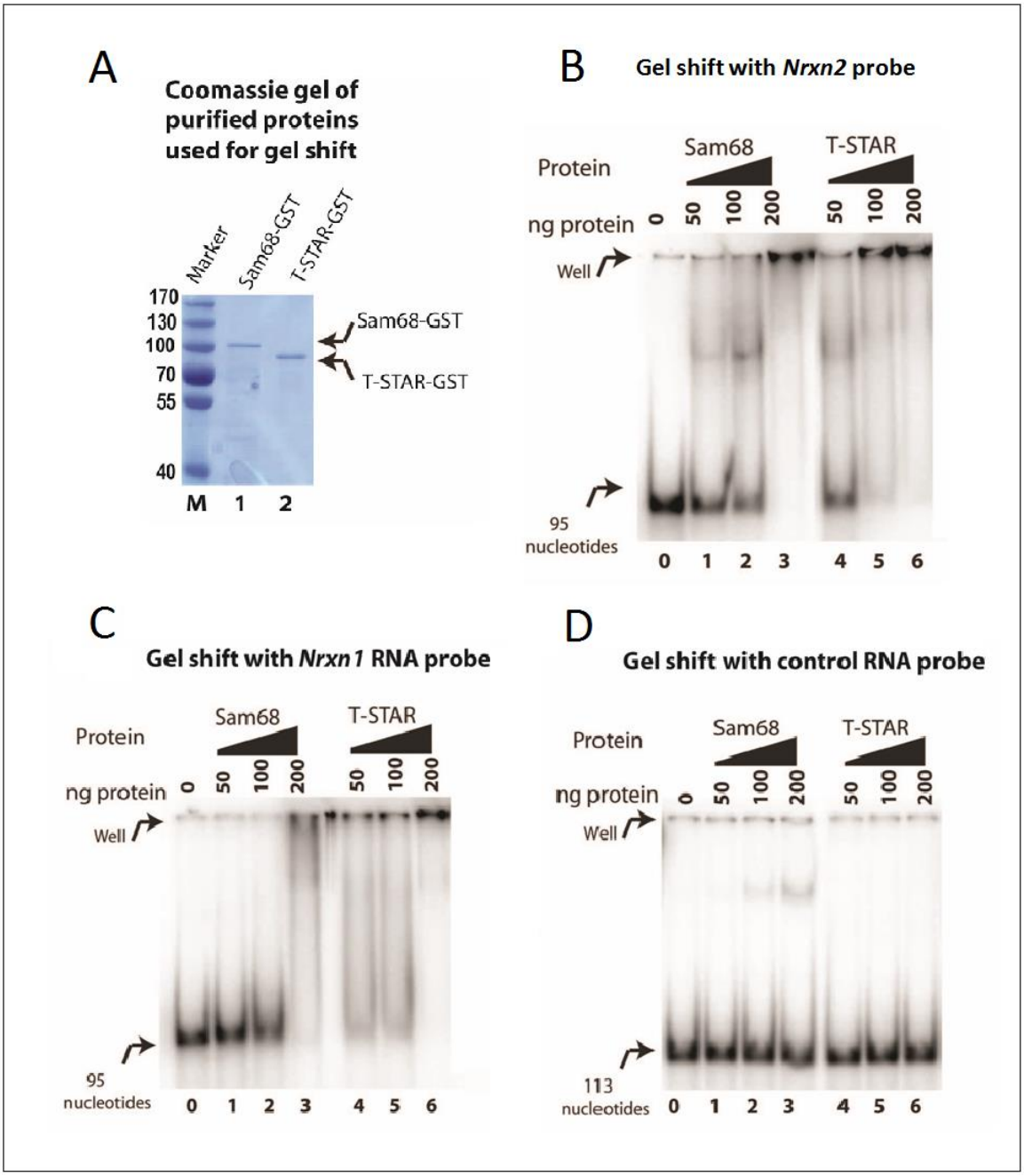


Figure 3.3. T-STAR and Sam68 bind the *Neurexin-1* and *Neurexin-2* RNAs directly. A- A Coomassie-stained SDS-PAGE gel, confirming the equal amounts of T-STAR and Sam68 were used in for the gel shift assay. B – A radiograph for *Neurexin-2* shift gel shift experiment. Lane 0 corresponds to the naked RNA, lanes 1 to 3 to different concentrations of Sam68 and lanes 4 to 6 to different T-STAR concentrations. Naked RNA does not produce any shift, forming a band at the gel bottom, corresponding to migration of unbound probe. Sam68 shifts *Neurexin-2* even at 50 ng of the protein added, with a complete shift at 200 ng of the protein, when the complex is too heavy to move and gets stuck in the well. T-STAR detectably shifts the probe with 50 ng of the protein, while a complete shift occurs with 100 ng. C - A radiograph for *Neurexin-1* shift gel shift experiment; the results are similar as for *Neurexin-2*. D – EMSA experiments were carried out with the negative control probe.

3.3.2 Splicing repression of the Nrxn2 cassette exon depends on multiple redundant binding sites, as well as a general repressor element

The previously described EMSA experiment and mutagenesis of an extended region, with T-STAR binding sites identified a 51 base pair region 12 bases downstream of the *Nrxn2* target exon, within which T-STAR and Sam68 binding occurs (Figure 3.4). There are 7 potential T-STAR binding sites in this region: 6 single UUAA sequences (Sites 1-4 and 5-6) (these were predicted as potential T-STAR binding sequences from SELEX data) and a sequence of 3 CUAA repeats (Site 7) (Figure 3.4).

To test the importance of these binding sites, using site directed mutagenesis, seven mutant minigenes were cloned, where UA was replaced by CC (a base combination that T-STAR is expected not to bind).

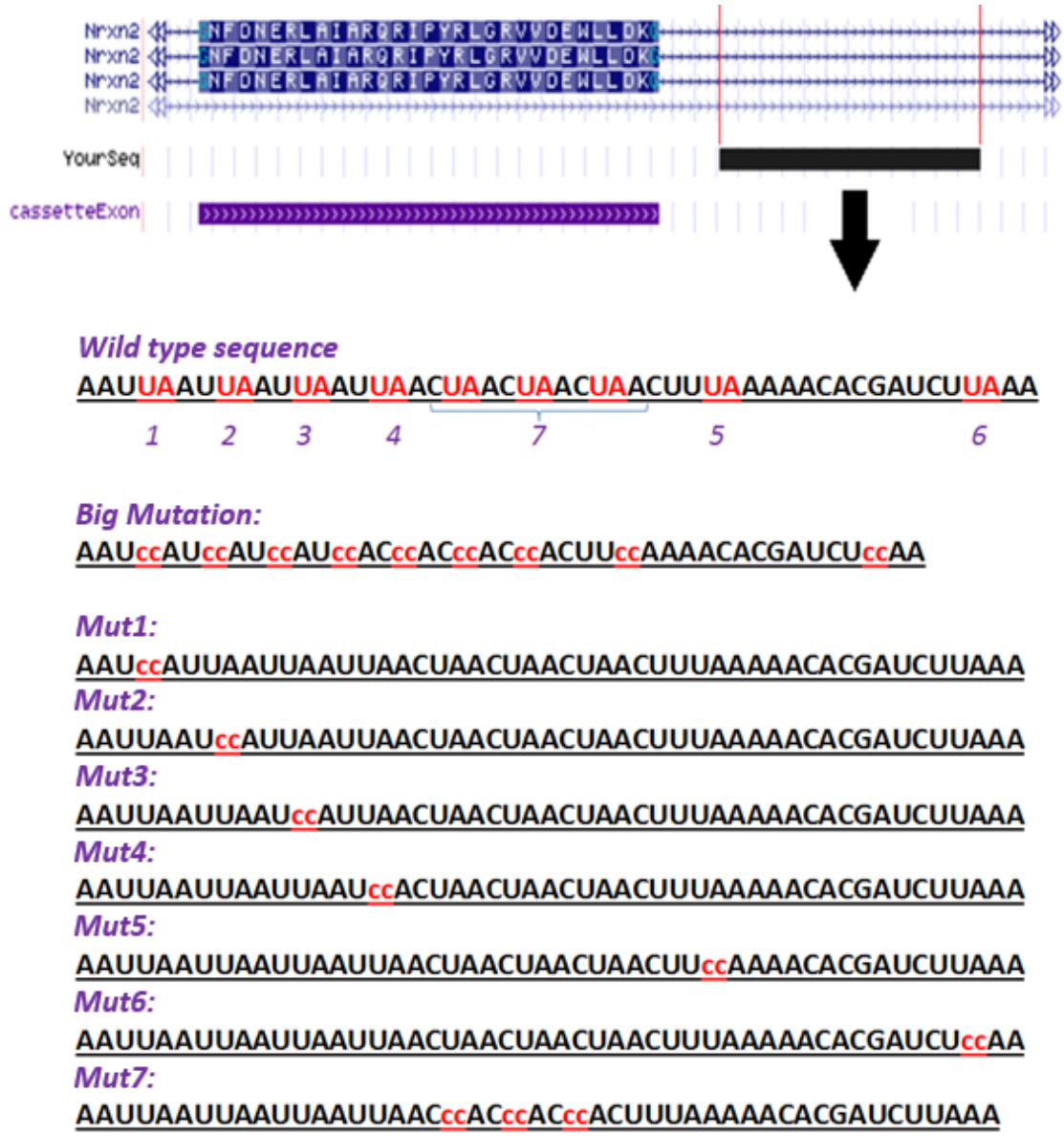


Figure 3.4 – A diagram of *Nrxn2* cassette exon and flanking intron sequence, taken from the UCSC genome browser. The region highlighted in black and limited with red lines corresponds to the downstream sequence element that contains multiple sites for T-STAR binding. Mutated nucleotides are shown in red. The Mutant called "Big Mutation" was created by (Ehrmann *et al.*, 2013) and showed to completely abolish T-STAR-mediated exon exclusion. Seven different mutant minigenes 1-7 have been produced from this region.

Each minigene was co-expressed in HEK293 cells with expression constructs encoding T-STAR and Sam68 proteins. The cells were harvested, RNA was extracted and reverse transcribed. Following capillary gel electrophoresis, splicing patterns were analysed. Two bands were always observed on a gel, one for exon inclusion and one for exon exclusion. Concentration for each band was calculated by a multi-capillary QIAxcel gel electrophoresis device. Based on the concentration ratio between exon exclusion and exon

inclusion, percentages were calculated and plotted on the graphs (Figure 3.5). The full details of how the calculation was performed are given in section 2.5.4 of the Materials and Methods Chapter. The level of splicing exclusion (splicing repression) caused by overexpressed T-STAR and Sam68 was compared with the level of splicing exclusion without co-transfection of STAR proteins (GFP co-transfection) for the same constructs. The level of splicing exclusion in each mutant *Nrxn2* construct was also compared with the level of splicing exclusion in the wild type *Nrxn2.*

The co-expression experiments show that mutations made in T-STAR binding sites of mutants 1, 2, 3, 4 and 6 did not significantly reduce exon exclusion, compared to the wild type minigene (Figure 3.5).

The constructs were analysed after co-expression of STAR proteins. Mutants 1 and 2 did not show any change in splicing repression by T-STAR, compared to wild type. There was a small reduction in splicing repression by T-STAR in mutants 3, 4 and 6. This was significantly different from the GFP-only co-transfection for each mutant. Thus, even when individual sites were mutated, splicing repression still happened. No effect on *Nrxn2* splicing of Sam68 overexpression was detected in any of the mentioned mutants.

In contrast with the other mutations, Mutant 5 showed a substantial reduction in exon exclusion when compared to the wildtype minigene. Furthermore, Mutant 5 transfected with T-STAR and Sam68 was no longer significantly different from the GFP-only control. This suggests that UUUAAAAA is the most important site for T-STAR splicing repression in the whole ISS.

Mutant 7 was still highly repressed by T-STAR, compared to wild type. However, splicing repression has decreased even with endogenous T-STAR levels (GFP-only). This might mean that Mutant 7, which is comprised of a triple CUAA sequence repeat, might represent a general splicing silencer element in *Nrxn2* sequence that is needed to gel full levels of repression of *Nrxn2* AS4.

None of the mutants above had such a strong effect as mutating the whole binding region (*Nrxn2* mutant with big mutation), where: a) exon inclusion is no longer repressed by co-transfected T-STAR; b) exon inclusion is much less repressed even without co-transfection of STAR proteins (i.e. at endogenous protein levels) (Figure 3.5).

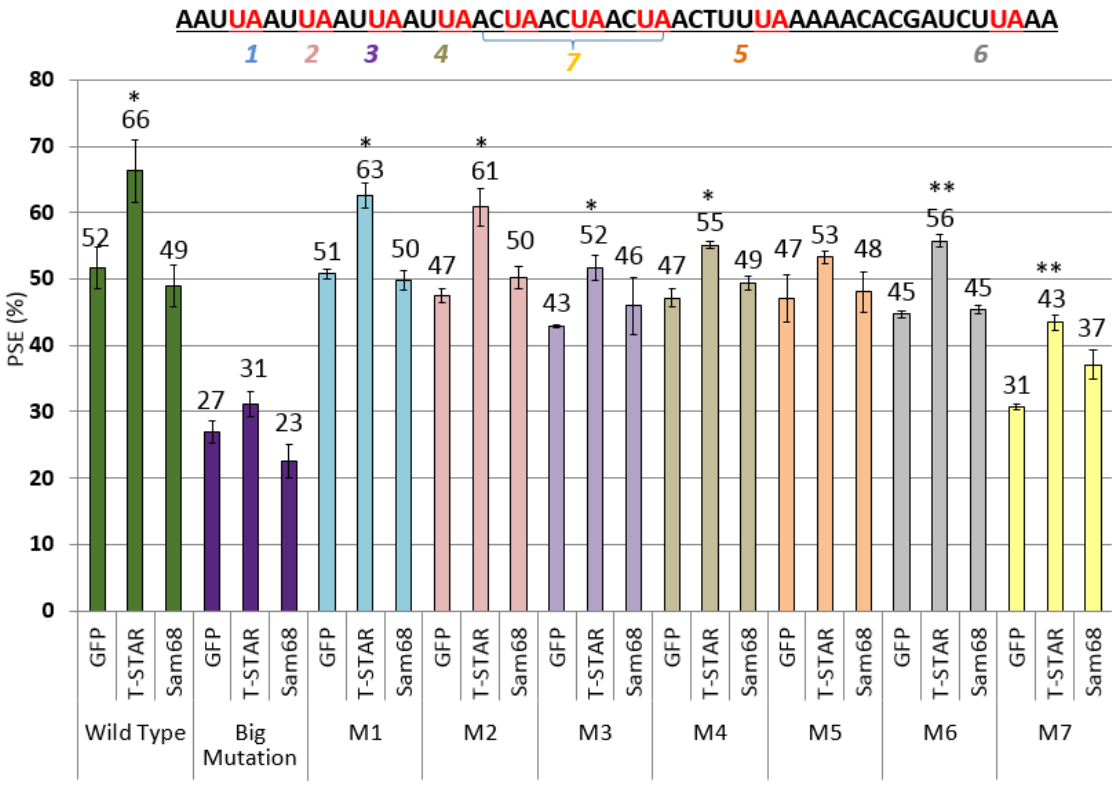


Figure 3.5 – Individual mutations downstream of the *Nrxn2* AS4 indicate the redundancy between single sites for *Nrxn2* splicing repression by T-STAR. Graphs show the percentage of splicing exclusion (PSE%) in different versions of *Nrxn2* minigene exon co-transfected with GFP-only, T-STAR and Sam68. Data represents the mean of minimum three biological replicates. In each case, statistical significance was compared between HEK293 cells expressing GFP and HEK293 cells expressing the GFP fusion proteins by means of independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.001.

Together this data indicated that UUAA sites in Mutants 3, 4, 6 and mainly the UUUAAAAA site in Mutant 5 are important for T-STAR splicing repression. The CUAA triple sequence repeat (Mutant 7) might represent a *Nrxn2* splicing repressor element, important for repression of the exon, but not responsible for the response to T-STAR.

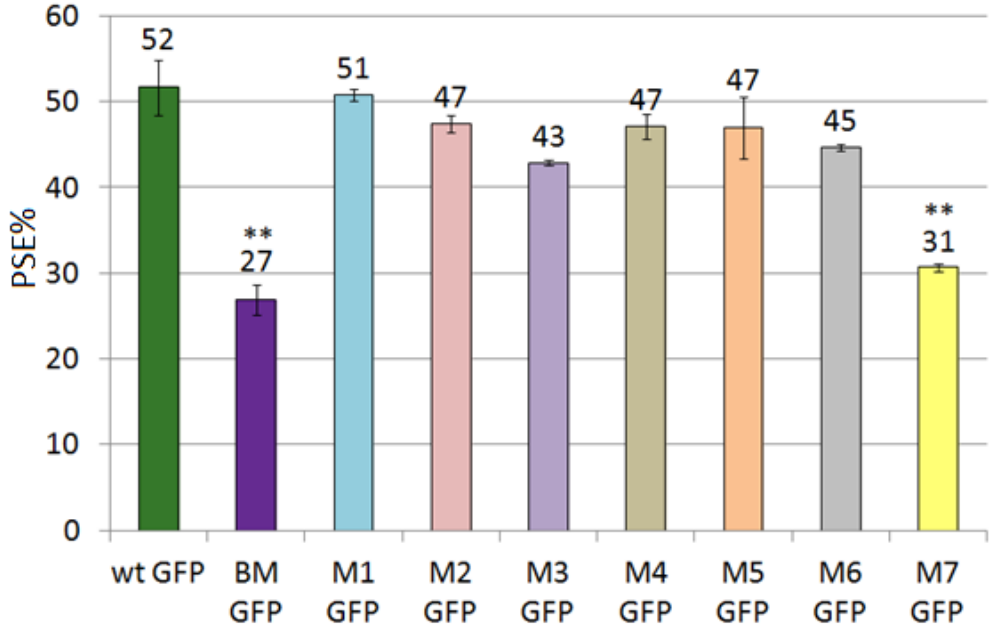


Figure 3.6 – Without co-transfection of STAR proteins levels of splicing repression of most mutants are not significantly different from the wild type control. Graphs show the percentage of splicing exclusion for different versions of the Nrxn2 minigene exon co-transfected with GFP-only. The data represents the mean of minimum three biological replicates. Statistical significance was compared between HEK293 cells expressing GFP in wild type Nrxn2 and HEK293 cells expressing GFP in mutant Nrxn2 by means of independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.001.

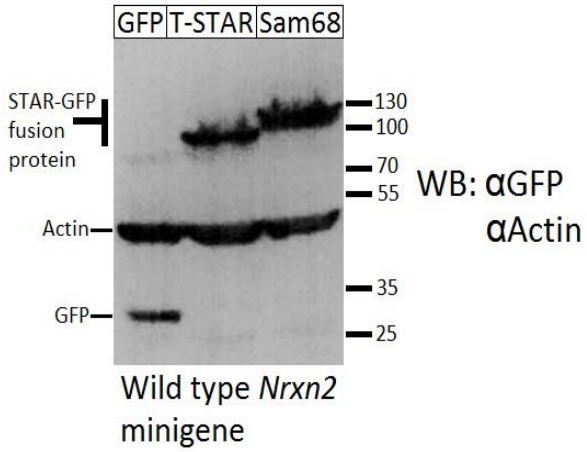


Figure 3.7 – The Western blot confirms that GFP, T-STAR and Sam68 showed similar levels of expression after transfection. Figure shows Western blot analysis of HEK293 cells co-transfected with GFP-coupled fusion proteins and wild type *Nrxn2* minigene. Probing was performed for GFP and actin.

Figure 3.7 is a Western blot for HEK293 cells co-transfected with GFP-coupled fusion proteins and wild type *Nrxn2* minigene. Western blot shows that the expression levels of T-STAR and Sam68 were similar.

From Figure 3.5 it is visible that for some of the mutants the percentage of exon exclusion is changing even for the GFP-only co-transfection. On Figure 3.6

I have plotted the graph for the GFP-only controls in those mutants tested. Neither of the mutants 1 to 6 appeared to be significantly different from the wild type GFP co-transfection. For the *Nrxn2* mutant with the big mutation and *Nrxn2* Mutant 7, percentage of splicing exclusions are both significantly different from the wild type GFP.

3.3.3 Loss of two key groups of AU-rich sequence elements inhibits splicing control by T-STAR.

Even though mutation of individual UUAA sites 1 - 4 in the *Nrxn2* response element were not strong determinants for T-STAR splicing regulation, these sites are predicted as binding sites for T-STAR by SELEX. To test if they are redundantly required for splicing control by a minigene assay, I combined them into a single large mutation 1-4, to check if these sites have an additive or redundant effect. Figure 3.8 shows the effect of these mutations on splicing, after transfection into HEK293. Even the combined mutant 1-4 did not have any effect of splicing repression by T-STAR. T-STAR was still able to repress the exon exclusion of Mutant 1-4 up to a significant level, compared to the GFP-only control.

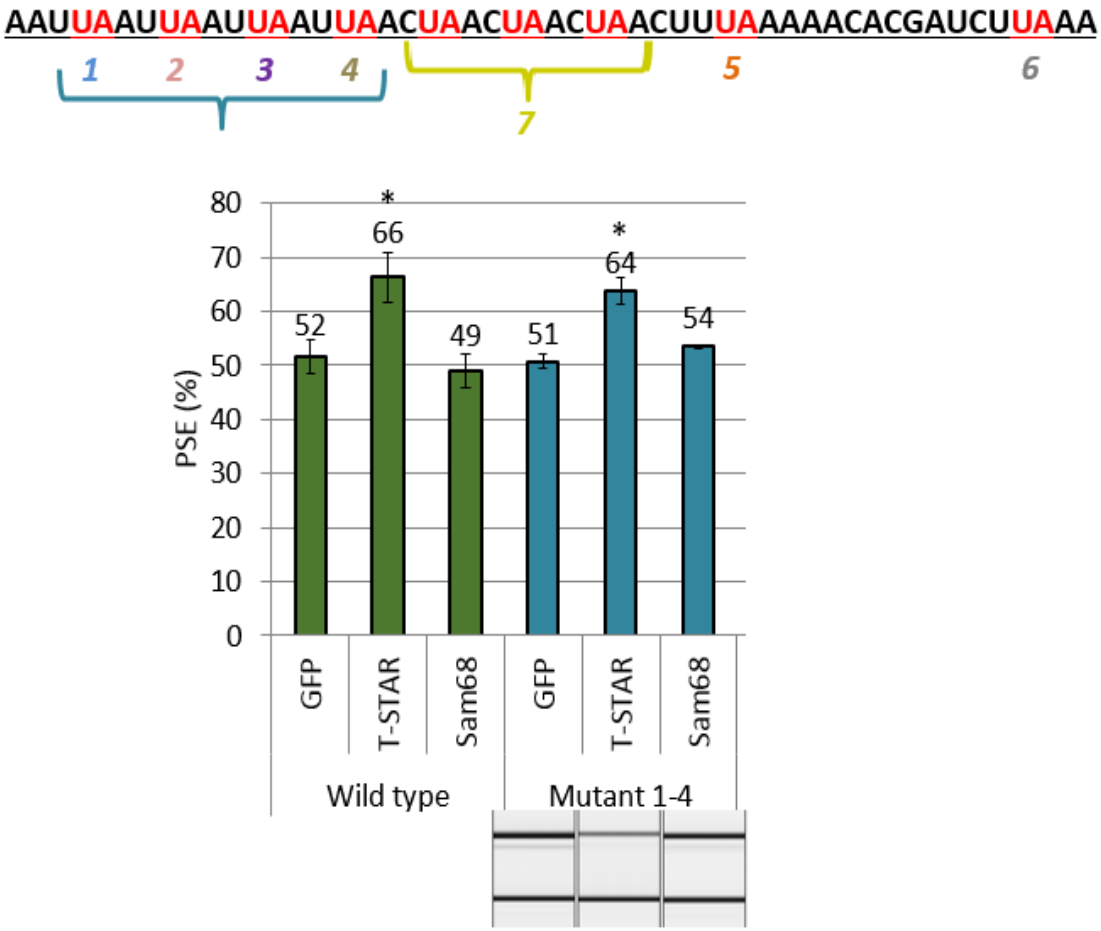


Figure 3.8 – A combined Mutant 1-4 is still efficiently repressed by T-STAR. Graphs show the percentage of splicing exclusion in a combined Mutant 1-4 of *Nrxn2* exon. The data represents the mean of minimum three biological replicates. Statistical significance was compared between HEK293 cells expressing GFP and HEK293 cells expressing the GFP fusion proteins by means of independent two-sample t-test,where *p<0.05, **p<0.01, ***p<0.001.

To further look for redundancy between the sites, mutations 1-4 were combined with Mutation 5 to create a Mutant 145. Interestingly, even at endogenous protein levels in HEK293 cells splicing repression was very high (Figure 3.9). This could have happened if with the sequence deletion a splicing enhancer was removed, thus leading to the high splicing repression with endogenous T-STAR and Sam68 levels. Comparing to GFP control, T-STAR appeared not to repress the splicing of this *Nrxn2* mutant anymore (Figure 3.9), with the percentage of splicing exclusion nearly equal among GFP-only control, T-STAR and Sam68 co-transfections. The additive effect of Mutation 5 together with Mutations 1-4 is stronger, than Mutation 5 on its own. These results suggest that even though sites 1-4 are not necessary for T-STAR repressive function on their own, they still have an important role in the complete T-STAR splicing element. The equal

expression levels of fusion proteins for Mutant 145 co-transfection was confirmed by Western Blot (Figure 3.10).

To further investigate the role of the sequence affected by Mutant 7 for T-STAR splicing repression, a combined Mutant 57 was created. In contrast to Mutant 5 alone, Mutant 57 was highly repressed by T-STAR (Figure 3.9). As the site covered by Mutant 5 is gone, the splicing of the Mutant 57 is repressed by T-STAR through sites 1-4.

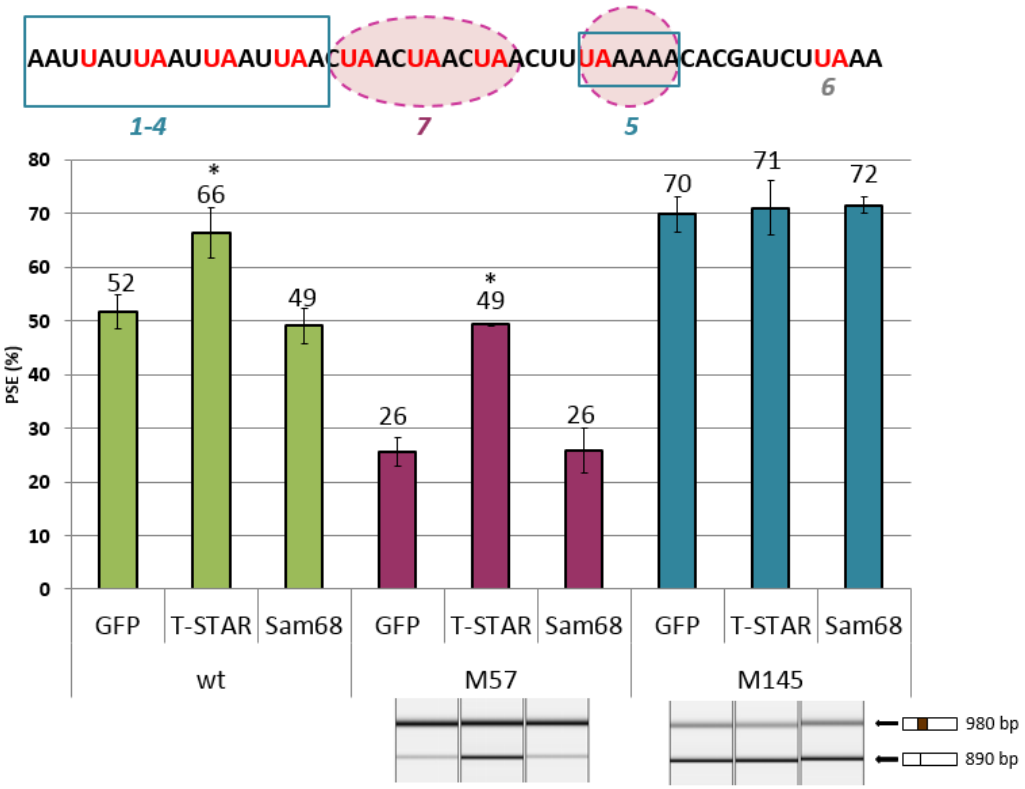


Figure 3.9 – The loss of sites 1-4 and 5 inhibits the *Nrxn2* splicing repression by T-STAR. Graphs show the percentage of splicing exclusion in wild type, as well as combined mutants M57 and M145 of *Nrxn2* minigene exon transfected with T-STAR, Sam68 and GFP-only control. The data represents the mean of minimum three biological replicates. In each case, statistical significance was compared between HEK293 cells expressing GFP and HEK293 cells expressing the GFP fusion proteins by means of independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.001.

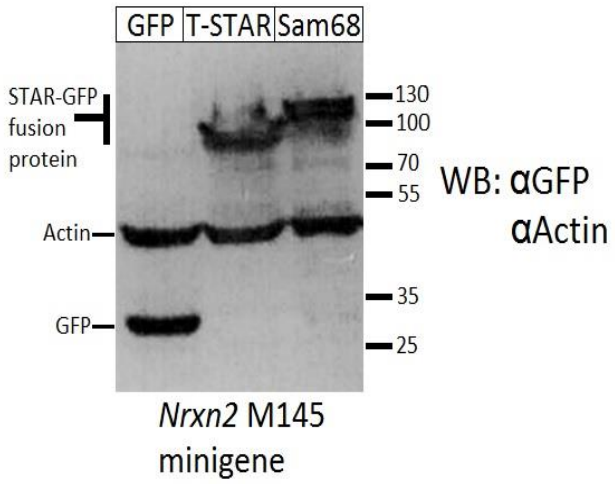


Figure 3.10 – Fusion-protein constructs co-transfected with *Nrxn2* Mutant 145 into HEK293 cells were equally expressed. Western blot analysis of HEK293 cells co-transfected with GFP-coupled fusion proteins and *Nrxn2* M145 minigene. The Western blot was probed for GFP and actin, and shows equal expression levels of each fusion protein.

Thus, CUAACUAACUAA appeared not to be a T-STAR regulatory site, but to some extent prevent T-STAR from regulation. I further aimed to study how CUAACUAACUAA affects T-STAR. In wild type *Nrxn2* CUAACUAACUAA is separating two groups of binding sites: UUAA repeats and UAAAAA sequence. Even when CUAACUAACUAA was mutated, it still creates a 'gap' between the mentioned target motifs. Thus, it was decided to delete CUAACUAACUAA from the sequence completely.

A mutant with deleted CUAACUAACUAA sequence was created (Figure 3.11). Interestingly, after co-transfection with T-STAR, strong repression by T-STAR was observed (25% significant difference comparing to GFP-only control). This repression pattern is significantly higher (by 12%) than repression in wild type *Nrxn2* co-transfection with T-STAR. Sam68 did not show to regulate splicing repression. Thus, a complete deletion of repetitive repressor sequence has increased T-STAR effect on *Nrxn2* splicing repression, compared to the wild type minigene.

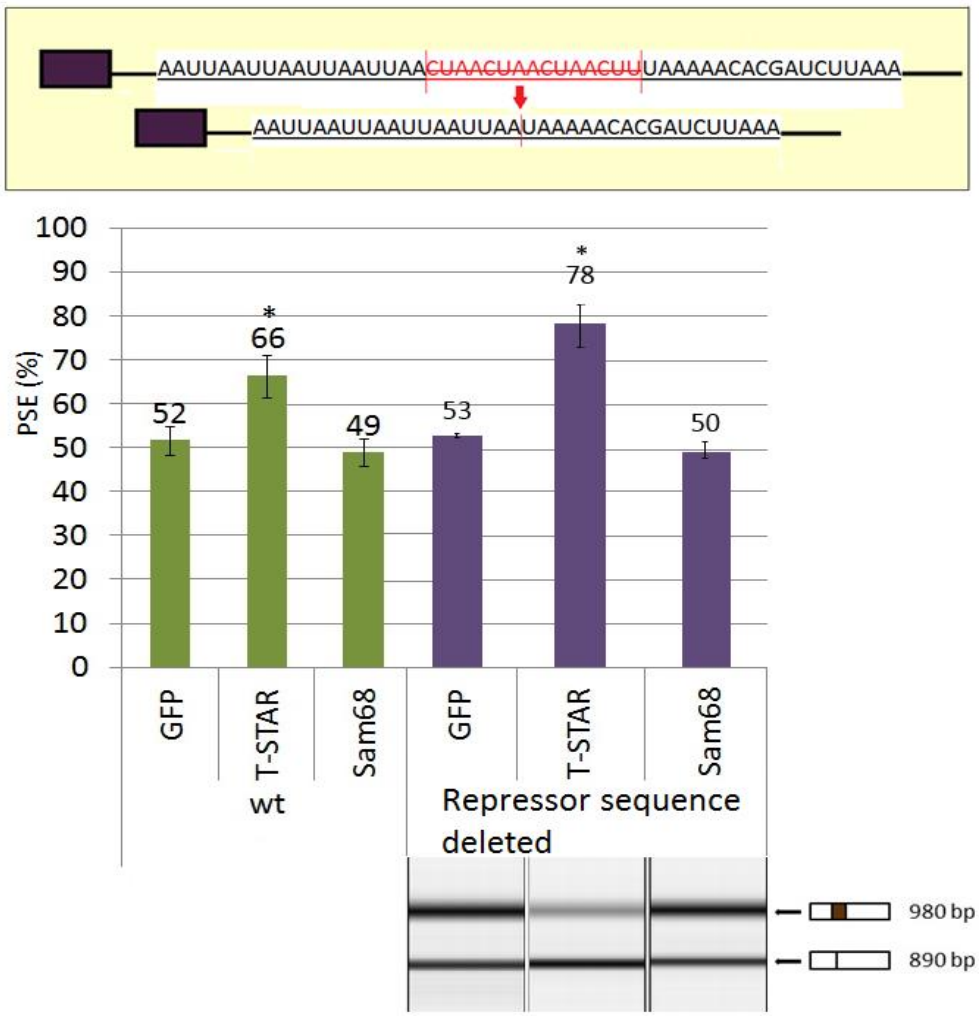


Figure 3.11 – The deletion of CUAA repeat region results in a stronger *Nrxn2* splicing repression by T-STAR. Graphs show the percentage of splicing exclusion in wild type, as well as mutant with repetitive repressor site deleted, transfected with T-STAR, Sam68 and GFP-only control. The data represents the mean of minimum three biological replicates. In each case, statistical significance was compared between HEK293 cells expressing GFP and HEK293 cells expressing the GFP fusion proteins by means of independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.001.

3.3.4 Splicing repression of Nrxn2 can occur from upstream and downstream intronic locations

Splicing factors are known to function differently depending on their binding position regarding to the target exon. The currently accepted rules are that when a splicing factor binds upstream, it represses splicing through blocking the branch point site, while when it binds downstream, splicing is activated through stabilising the U1 snRNP binding (Witten and Ule, 2011). T-STAR response sequences in *Nrxn2* are situated 12 bases

downstream of the target exon. Binding downstream, T-STAR represses the splicing, which is more an exception to the general rules.

Knowing the sequence elements through which T-STAR represses *Nrxn2* splicing, we further addressed the mechanism through which T-STAR causes splicing repression. I hypothesised that T-STAR might be blocking the U1 spliceosome access to the 5' splice site, and hence prevent the spliceosome assembly. I aimed to test if splicing is position dependent and if T-STAR exerts the same effect both upstream and downstream of the target exon. If T-STAR is restricting U1 snRNP access to the 5' splice site, I predict moving the binding site upstream of the exon, or far downstream, might prevent splicing regulation.

Thus, *Nrxn2* mutants were cloned, where the whole T-STAR response sequence was moved 100, 150 and 250 bases further downstream in the intron sequence. The 250 base interval was chosen as the most distant one, as the downstream intron length in the minigene did not allow to move the sites even further downstream. As a template for this cloning the *Nrxn2* mutant with a big mutation was used, in which all T-STAR response elements have been mutated, making it unresponsive to T-STAR overexpression.

After co-transfection with T-STAR, all of the mutants still showed significant repression of exon splicing, compared to a GFP-only control. Even though splicing repression has reduced with endogenous T-STAR levels within HEK293 cells, T-STAR still repressed *Nrxn2* splicing of each of the mutants to a similar level as in the wild type control (Figure 3.12). Hence, even with binding sites being a distance of 100, 150 or 250 bases downstream the *Nrxn2* exon, T-STAR still efficiently repressed *Nrxn2* splicing. Sam68 co-transfection did not have any effect on splicing of any of the mutants - in all three mutants the percentage of splicing exclusion for Sam68 co-transfection is nearly equal to the GFP-only control (Figure 3.12).

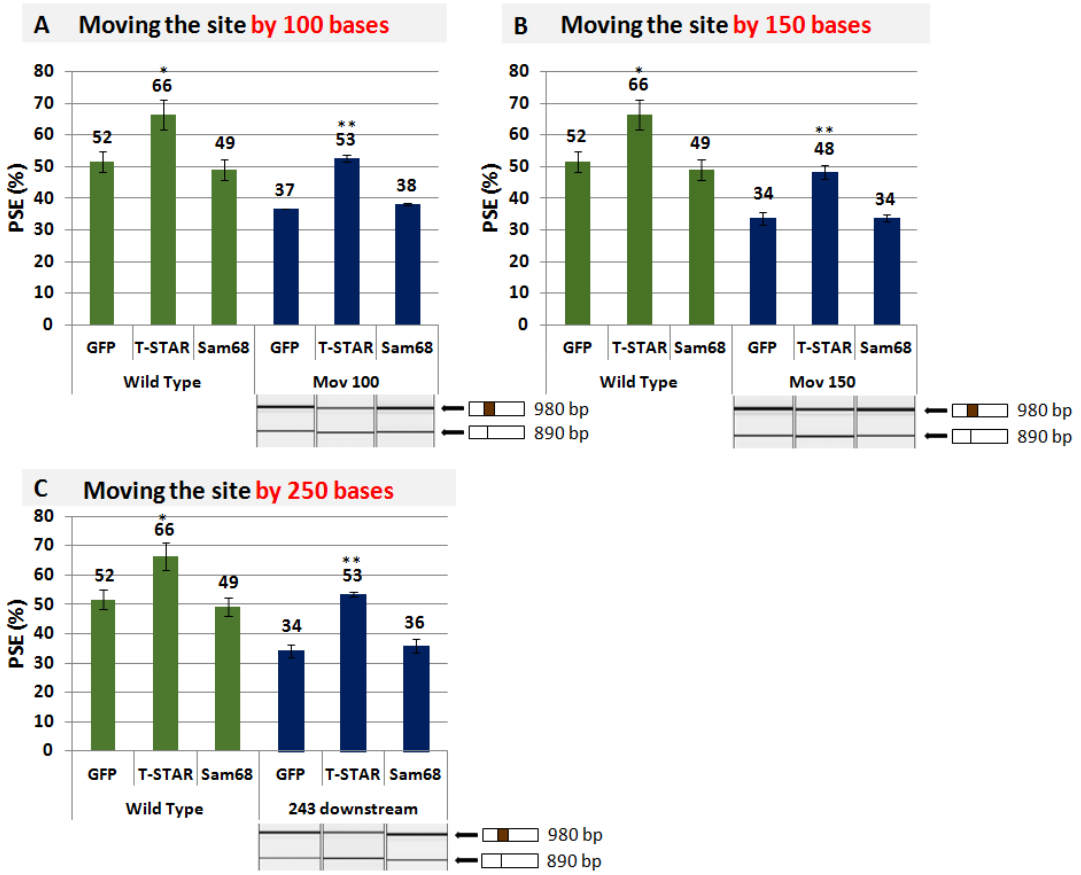
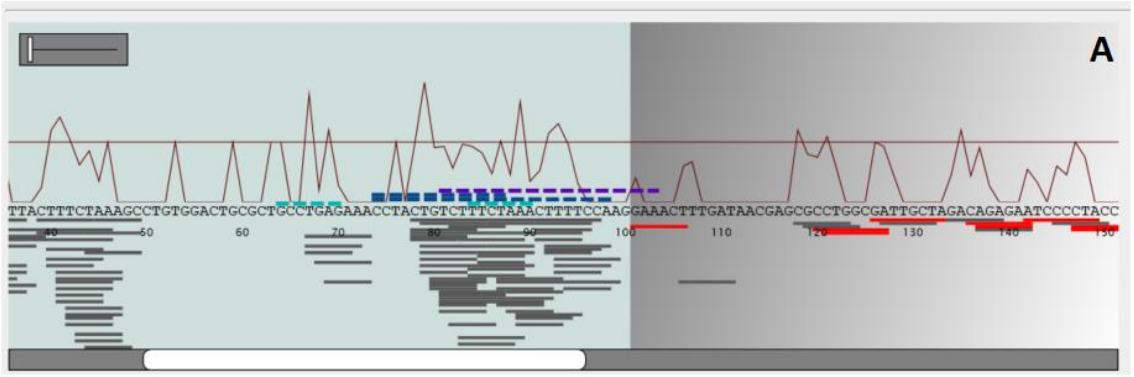
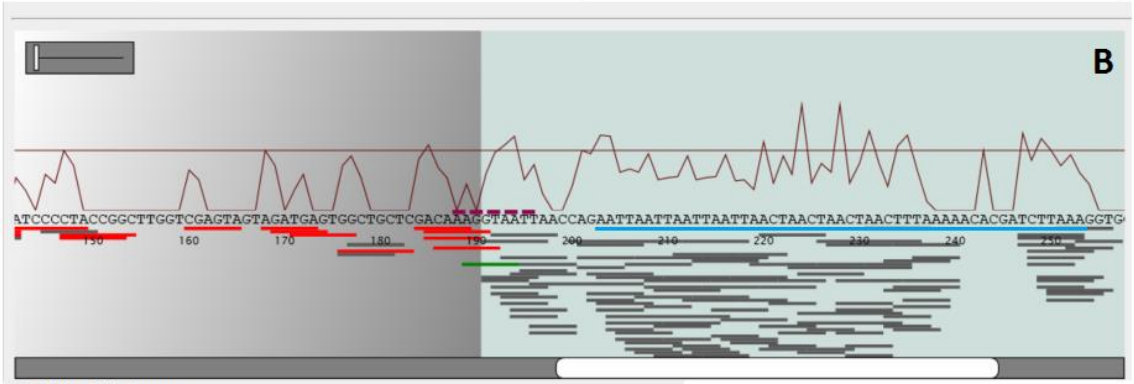


Figure 3.12 – T-STAR response element can still potently repress splicing even from more distant downstream locations. Graphs show the percentage of splicing exclusion in transcripts made from three mutants where the *Nrxn2* binding site is moved: A) 100 bases, B) 150 bases, C) 250 base downstream the target *Nrxn2* exon, compared to the wild type minigene control. The data represents the mean of minimum three biological replicates. In each case, statistical significance was compared between HEK293 cells expressing GFP and HEK293 cells expressing the GFP fusion proteins by means of independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.001.

The experiments above show that *Nrxn2* splicing repression by T-STAR can still be repressed even with the binding sites being further in a downstream intron. To investigate whether splicing is similarly repressed when the binding sites are moved into the upstream intron, two mutants were cloned with the whole T-STAR response sequence moved 42 and 100 bases upstream of the exon. The program called Sroogle (Splicing Regulation Online Graphical Engine) (http://sroogle.tau.ac.il/) was used to identify the Branch Site sequence and Polypyrimidine tract in the sequence upstream. A 42-base distance was chosen as an earliest point upstream to prevent any interference with these important sequence elements (Figure 3.13 A).





Splice Site Scores

Represented by dashed annotations **above** the sequence.

Element	Start	End		Score	Score Percentile (Const Exons)	Score Percentile (Alt Exons)
Branch site	64	70	3.30	(Kol et al)	0.61	0.64
	84	90	2.00	(Schwartz et al)	0.32	0.29
Polypyrimidine tract	74	98	0.79	(Kol et al)	0.49	0.49
	74	87	7.14	(Schwartz et al)	0.22	0.25
3' splice site	81	103	5.51	(Max entropy)	0.1	0.17
			82.98	(PSSM)	0.23	0.3
			-5.20	(Delta-G)	0.39	0.43
5' splice site	188	196	8.83	(Max entropy)	0.52	0.6
			87.85	(Senepathy)	0.72	0.8

Figure 3.13 – Output of a wild type *Nrxn*2 sequence in the Sroogle program. A - for the exon start and an upstream intron; B - for the exon end and the downstream intron). Splice sites, Branch site and Polypyrimidine tract are marked with coloured dash lines, that match by colour with the names in a table below. The T-STAR binding region is marked by a continuous blue line on a picture B.

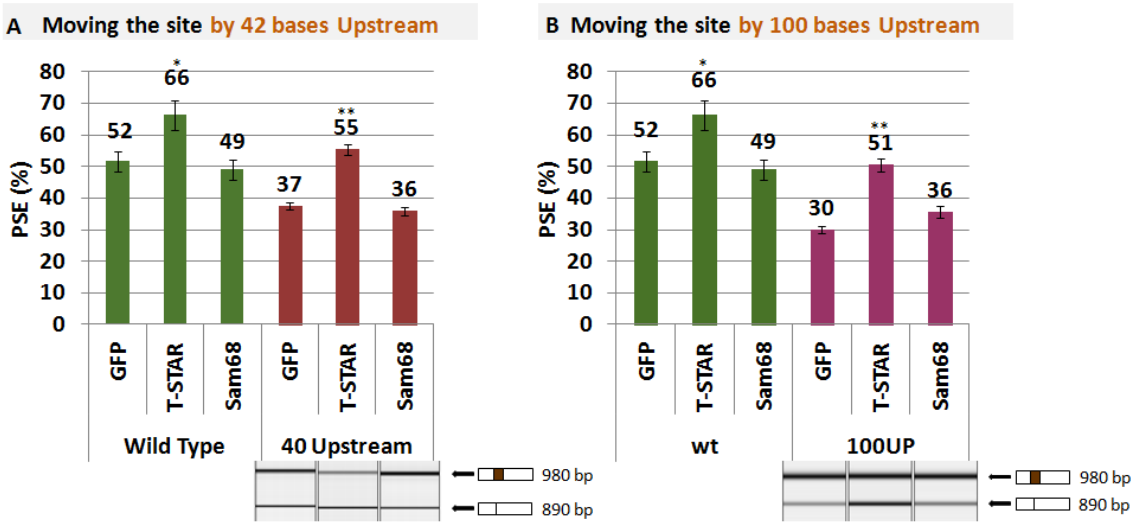


Figure 3.14 – T-STAR response element can still potently repress splicing even when placed upstream of the regulated *Nrxn2* exon. Graphs show the percentage of splicing exclusion for moving mutant: when the binding site is moved A) 42 bases, B) 100 bases upstream of the target *Nrxn2* exon, compared with the wild type minigene control. The data represents the mean of minimum three biological replicates. In each case, statistical significance was compared between HEK293 cells expressing GFP and HEK293 cells expressing the GFP fusion proteins by means of independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.001.

When co-transfected with T-STAR and Sam68, both mutants where the T-STAR binding site was moved upstream were still highly repressed by T-STAR co-expression (Figure 3.14 A and B). There was also some reduction in splicing repression on endogenous protein levels. As previously, Sam68 did not induce splicing repression of these.

3.4 Discussion

3.4.1 Do T-STAR and Sam68 bind Nrxn2 directly?

The first aim of this Chapter was to investigate whether T-STAR and Sam68 directly interact with the *Nrxn2* response element. T-STAR and Sam68 are both very similar proteins, both belonging to the same protein family, having a similar domain composition and UA-rich sequences predicted as target regions for RNA-binding. It thus would be expected that they regulate the same set of targets. However, the study by Ehrmann et al 2013 showed that T-STAR is a specific splicing regulator of *Nrxn2* gene. In this study it was found, that three *Neurexin* genes (*Nrxn1*, *Nrxn2* and *Nrxn3*) are all regulated by T-STAR, but only two of them (*Nrxn1* and *Nrxn3*) were regulated by Sam68. Hence, *Nrxn2*

appeared to be T-STAR specific (Ehrmann *et al.*, 2013). The preliminary target region identified for T-STAR was highly rich in UA elements, but also highly resembled a Sam68 binding site. Thus, it was of interest, whether T-STAR and Sam68 can both bind the target region, even though only T-STAR induces a splicing response. To find this out an EMSA was performed. Both T-STAR and Sam68 proteins appeared to bind the target region of *Nrxn2*; however, T-STAR binding was more efficient, as 100ng was enough for complete shift (comparing to 200ng of Sam68). That might mean that the *Nrxn2* binding site is more T-STAR specific than Sam68 specific. Interestingly, the study by Ehrmann et al also showed that when co-transfecting T-STAR and Sam68 together with *Nrxn2* minigene, the effect of T-STAR is completely abolished by Sam68. This might suggest, that even though Sam68 binding is less efficient than T-STAR, it might be competing with T-STAR for the same binding sites, and blocking the latter from binding. It was previously shown by Venables et al that T-STAR interacts with Sam68 in the yeast two-hybrid system (Venables *et al.*, 1999). Thus, when overexpressed, Sam68 could be interacting with T-STAR, and blocking it from splicing regulation.
As it was shown on Figure 1.7 of the Introduction Chapter, T-STAR is having a shorter N-terminus, comparing to Sam68. Thus, it might be that N-terminus of Sam68 prevents it from regulating *Neurexin-2* splicing. To test that, a mutant Sam68 should be made, with its N-terminus being equal by length and composition to the one of T-STAR. It then should be transfected with *Neurexin-2* minigene. Furthermore, *Neurexin-2* splicing might be specific to T-STAR due to *Neurexin-2* sequence composition. As visible from the Figure, T-STAR sites in *Neurexin-2* are clustered, while the potential Sam68 and T-STAR sites in *Neurexin-1* and *Neurexin-3* are more dispersed across the downstream intron.

Neurexin-1

GCTCTAGGTAGAACTTTTTAAAATTCCAAATTTCAAAAAATCATGTGGCAATTGTATAGCTCAAGCTT
TAAGTATTAACTTTTAACCTTTGACCTTACAGATTTTAACCAATGAAATGTCTCTTTAAACTTTAAAGG
AAACAATGATAACGAGCGCCTGGCGATTGCTAGACAGCGAATTCCATATCGACTTGGTCGAGTAGTT
GATGAATGGCTACTCGACAAAGGTAAAAACATTGATTAAAACTTCAAATAAAAAATAATTTGAACAT
AACCAAGTAGTACTTCATTGTAACACTAATAAACATAAAAAATAAATTTTCCGTAAAACTAAATTAAA
AACAAGTTTAAAACAGTAAAATGTAGATAGTAATATTTGCATACCTTGTATAATAATTTCTATACATT
GTTAGAAGTACCAACGGTTTAATAATCTTACCCAGTTAACTTAAGGTTAAAGGGACTGTTAAATATA
ATCCACTAACTCAATCTATGAAGGTACTCAGAGCAAGCATTAACCAAAAATGATAAATCATTTATAG

Neurexin-2

AGAGGAGGGGCCCAGCCTTCTGTCCCCGCCTCCCTCCTTTCTGTTGAATGTGGCTGCTTGCTGGGGT
CTCCTGAGAACTGCATAATTACTTTCTAAAGCCTGTGGACTGCGCTGCCTGAGAAACCTACTGTCTTT
CTAAACTTTTCCAAGGAAACTTTGATAACGAGCGCCTGGCGATTGCTAGACAGAGAATCCCCTACCG
GCTTGGTCGAGTAGTAGATGAGTGGCTGCTCGACAAAGGTAATTAACCAGAATTAATTAATTAATTA
ACTAACTAACTAACTTTAAAAACACGATCTTAAAGGTGCAGAGCTCTCTCCCCCACATCCGGGCAGG
ATGCTCCCTTCTGCCCCCCTGCGTTCAGGCCGTTTGTCTCTGGGCGGAGGCTGGGAATAGGGAAGA
GAAAGTCTTCCTAGAGTGACCCTCAGTGCCTGAAGATGGAAGCACCCTCAGCTTGTGCTCAGAATTC
TGCAGTAGGAGGGGAACACAGGCTAGAAGTCGGGGCTCTCTGGCAGAGTTCACCTCAGGGGTTTTT

Neurexin-3

ACCTCTACTAACAATAAATTGATATCTATCTCTAAAACTCTTCATAAAACAAACTGAACACCTAGGAT
GTTTTTCTTTAATGACCTATTGACTTTTCAGATTGTTCTTTAAGCTGTTTTTTAAACTTTAAAGGCAAC
ACTGATAATGAACGCCTCCAAATGGTAAAACAGAAAATCCCCTTCAAATATAACCGGCCCGTAGAGG
AGTGGCTGCAGGAAAAAGGTAACGGTCGTTAAACTTTCATTCTAAAATCATTTGATTCTAAACCCTCA
TTTTTAAAGCCTGCAAATAAATGTTCCTTAAATGATAATCGCCTGCTGCATAAATGACTGTTTTAATTT
TCATTATAATTTGTGTCAGCTTCCACCTGCCATATGTTGATAAATCATGCAATAATACATCGCCATTAA
AACACAGTTAAAGGGACTAACATCTAGCCTGCTGTCAACATGGCAGCCACAACCAAGGAATTTCCAG
GAAACTATTTTAAAAGAAAAAGAAAGAAAGAAATGGATTATGGCA

Figure 3.15 – *Neurexin-2* has a cluster of T-STAR binding sites, while in *Neurexin-1* and *Neurexin-3* target sites are more dispersed. Figure shows the sequences of the target exons regulated by T-STAR (and by Sam68 for *Neurexin-1* and *Neurexin-3*) and the flanking introns. Exons are highlighted in black, while the potential T-STAR and Sam68 target sites in the downstream introns are highlighted in yellow.

3.4.2 Splicing repression of the Nrxn2 cassette exon depends on multiple redundant binding sites, as well as a general repressor element

The second aim of this Chapter was to dissect the target response element into smaller elements and investigate their importance for *Nrxn2* splicing repression by T-STAR

In the study by Ehrmann et al 2013, an overall T-STAR target region in *Nrxn2* was identified. This sequence contained several UA-rich sequence repeats that might independently bind T-STAR protein. This region was dissected by me to try to identify

smaller T-STAR binding sequence motifs. Minigene assays were performed to monitor functional response. Even though minigene creation is laborious and does not allow a global analysis, when working on a few targets minigenes provide an ideal mechanism to manipulate the sequence, mutating and rearranging it in required ways. There are many others studies which used minigene assays for splicing e.g. a study by (Steffensen *et al.*, 2014) for the *BRCA1* gene; the study by (Ulzi *et al.*, 2014) for the *CLCN1* gene; the study by (Xiao *et al.,* 2012) on the *Bcl-x* gene; the study by (Grellscheid *et al.*, 2011) on the Tra2β target *HIPK3*; etc. Minigene assays have shown that T-STAR target region in *Nrxn2* is having an interesting structure.

Being a UA-rich sequence, the CUAACUAACUAA region (covered by Mutant 7), consisting of three CUAA repeats was predicted to be a T-STAR binding site. However, when the CUAACUAACUAA region was mutagenized, T-STAR still regulated *Nrxn2* splicing up to similar level as in wild type; but overall levels of splicing exclusion were significantly reduced. This suggested that CUAACUAACUAA is functioning as a general splicing repressor sequence. Previous studies show that there are certain sequences that function as exonic or intronic splicing repressor elements. For instance, CUCUCU functions as an intronic splicing silencer element, while UGCAUG – as a splicing enhancer sequence (Chan and Black, 1997; Modafferi and Black, 1997).

The UUUAAAAA site (covered by Mutant 5) appeared to be the strongest site for T-STAR regulation. When mutated, no significant difference in splicing repression was any more noticeable between GFP-only control and T-STAR co-transfection. However, by mutating site 5 splicing repression by T-STAR was not blocked completely (splicing was still repressed by T-STAR 6%, compared to 14% repression in a wild type control).

The UUAA sites (1-4) were predicted by SELEX as potential T-STAR binding sites (Lin *et al.*, 1997; Galarneau and Richard, 2009a). However, mutagenesis of separate UUAA sites did not show any large effect on T-STAR splicing regulation. Even when combined together into one mutation 1-4, loss of the UUAA sites did not impact T-STAR function in splicing. However, sites 1-4 were important for regulation by T-STAR. When sites 5 and 1-4 were blocked together into a mutant 145, no splicing regulation by T-STAR was visible any more. This fits with the recent unpublished findings by Dr Cyril Dominguez lab (Leicester University) that T-STAR binds the RNA sequence as a dimer with at least 15 nucleotides spacing between the binding sites for each monomer. Sites 1-4 and 5 have

15 nucleotides of CUAA sequence between them, thus might be ideal sites of T-STAR monomers to bind. Hence, with both 1-4 and 5 being blocked, T-STAR fails to function.

A combined together Mutant 5 with Mutant 7 (Mutant 57), still responded to T-STAR, as Mutant 7 does alone. That means, when we block mutant 5 and 7 together, T-STAR is able to function and regulate *Nrxn2* splicing, presumably through sites 1-4.

Thus, to summarize, CUAACUAACUAA functions to repress splicing, but does not mediate a response to transfected T-STAR.

Experiments show, that when CUAACUAACUAA is wild type, site 5 (UAAAAA) serves as a main site to respond to transfected T-STAR, and sites 1-4 (UUAA) recruit T-STAR as secondary sites, having just a small effect on T-STAR regulation.

Removal of CUAACUAACUAA sequence reduces the baseline level of splicing repression, but response to transfected T-STAR keeps still being strong. Splicing repression could be performed through site 5 (UAAAAA). However, when site 5 is mutated with site 7 together into a Mutant 57, T-STAR still highly represses *Nrxn2* splicing. That means, when site 7 is mutated, sites 1-4 still function as T-STAR target motifs.

The above mechanism is a good example of a complex and tight splicing control by the cell, where many factors participate and many compensatory mechanisms function, in case one of the pathways fails to work. Such tight control may be important for regulating the amounts of alternative transcript produced, and eventually, the amounts of the protein.

Interestingly, when repetitive sequence covered by Mutant 7 was deleted completely, the T-STAR repression pattern has significantly raised. CUAACUAACUAA is separating two groups of binding sites: UUAA repeats and UAAAAA sequence. Once it was deleted, the gap has disappeared and separate sequences created a single motif. Thus, a longer continuous AU-rich motif has appeared, and a new sequence combination as AAUAAAAA. Based on the study by Foot et al 2014, AAAUAA sequence is the one of single stranded RNA sequences bound by T-STAR KH domain (Foot *et al.*, 2014). The current new motif AAUAAAAA highly resembles the motif from the study above. Thus, deletion of repressor sequence might be creating a better T-STAR binding site.

3.4.3 Splicing repression of Nrxn2 can occur from upstream and downstream intronic locations

The third aim of this Chapter was to investigate the importance of the upstream and downstream position of response element for splicing regulation. The position of protein-RNA interactions relative to the target exon can play an important role in splicing regulation. The splicing activity of RBPs is linked to their binding site. For instance, TIA protein binding downstream of the target exons results in splicing activation (Wang *et al.*, 2010). Other RBPs are more flexible and show position-dependent splicing activity, i.e Nova, Fox, Mbnl and PTB RBPs (Dredge *et al.*, 2005; Zhang *et al.*, 2008a; Goers *et al.*, 2010; Llorian *et al.*, 2010). For instance, Nova binding downstream of exons activates splicing, while binding within the exons causes splicing repression (Dredge *et al.*, 2005). As the *Neurexin* genes are the first physiological splicing targets identified for T-STAR, it was so far unknown what the positional principles for T-STAR splicing regulation were (Ehrmann *et al.*, 2013). To find out we moved T-STAR binding sites (that in a wild type are twelve nucleotides downstream the exon) further downstream from the exon as well as upstream from the exon. However, all five mutants produced, even with T-STAR binding sequences a distance of 250 bases downstream of the exon and 100 bases upstream of the exon (maximum distances tried), had *Nrxn2* AS4 exon splicing repressed by T-STAR similarly to a wild type control. To summarize, here we show that, surprisingly, *Nrxn2* splicing repression by T-STAR is position independent, and as efficient as on different distances from the exon both downstream and upstream.

Commonly, RBPs binding downstream of the exons activate U1 snRNP and splicing (Förch *et al.*, 2002; Xue *et al.*, 2009; Wang *et al.*, 2010; Witten and Ule, 2011). In the case of T-STAR, *Nrxn2* splicing is repressed with regulation happening through a composite site twelve bases downstream of the exon. This site is very close to the 5' splice site where the spliceosome starts to assemble (Valadkhan, 2007). Thus, I hypothesised, that opposite to activation, splicing might be repressed by T-STAR directly blocking U1 snRNP. In this case, splicing repression would fail to happen with T-STAR binding sites being at a greater distance from the exon. Hence, it is possible that my suggested model of splicing repression through blocking U1 snRNP was wrong and T-STAR may use a different mechanism to repress *Nrxn2* splicing.

T-STAR could also follow the mechanism suggested for hnRNP A1 and PTB proteins, which bind the target sites from both sides of the exon, and loop out the region containing

the target exon (Blanchette and Chabot, 1999; Black, 2003; Spellman and Smith, 2006). In published examples, U1 snRNP fails to interact with the 5' splice site on the looped part. This represses splicing. Here I have dissected the T-STAR target region downstream of *Nrxn2* exon. To follow the hypothesis about the looping model, I have also checked the sequence upstream of *Nrxn2* exon UCSC genome browser, to see if any T-STAR-like binding sites are present. Interestingly, within 150 bases upstream the exon I found two UAAA sites (10 and 63 nucleotides upstream) and one AAUU site (53 nucleotide upstream) (Figure 3.16). Thus, the proposed model might be functional for T-STAR. It would explain why repression is happening with T-STAR binding sites moved on a distance. The presence of RNA loops in *Nrxn2* sequence could be tested by FRET and NMR spectroscopy, as in study by (Lamichhane *et al.*, 2010). However, the simpler way to test the model would be to mutate the upstream and downstream target sequences together.

AGAGGAGGGGCCCAGCCTTCTGTCCCCGCCTCCCTCCTTTCTGTTGAATG
TGGCTGCTTGCTGGGGTCTCCTGAGAACTGCATAATTACTTTCTAAAGCC
TGTGGACTGCGCTGCCTGAGAAACCTACTGTCTTTCTAAACTTTTCCAAG
GAAACTTTGATAACGAGCGCCTGGCGATTGCTAGACAGAGAATCCCCTAC
CGGCTTGGTCGAGTAGTAGATGAGTGGCTGCTCGACAAAGGTAATTAACC
AGAATTAATTAATTAATTAACTAACTAACTAACTTTAAAAACACGATCTT
AAAGGTGCAGAGCTCTCTCCCCCACATCCGGGCAGGATGCTCCCTTCTGC
CCCCCTGCGTTCAGGCCGTTTGTCTCTGGGCGGAGGCTGG

Figure 3.16 – *Nrxn2* target exon (in red font) with 150 bases of upstream and downstream flanking exon. Downstream target region is highlighted in grey; potential T-STAR binding sites upstream are highlighted in yellow.

In the study by Hui et al 2005, a similar technique of moving the hnRNPL target motif on different distances was applied, to find out the positional effect of hnRNPL on splicing of target genes (Hui *et al.*, 2005). In wild type genes, hnRNPL bound downstream of the target exon and activated splicing. When binding sites were moved further downstream of 5' splice site, for some of the tested genes splicing became repressed. However, for some genes, such as *eNOS*, even the binding sites at a distance still functioned as splicing enhancers. It was found that moving the binding sites further downstream in *eNOS* resulted in an activation of a cryptic 5’ splice site further downstream in the wild type sequence, and still enhanced splicing (Hui *et al.*, 2005). However, cryptic downstream splice sites lead to changes in a final product size, as a part of an intronic region is spliced

together with exon. Assessing the T-STAR regulation of moving mutants, I did not see any difference in sizes of final product, thus a cryptic splice site version as a potential mechanism for T-STAR should be excluded, unless a cryptic splice site is activated very close to an original splice site, so that difference in product sizes cannot be detected on a gel.

Witten and Ule et al 2011 described several possible mechanisms of how proteins regulate splicing by binding downstream of alternative exons (Witten and Ule, 2011). If not repressing U1 snRNP directly, T-STAR might interact with other proteins (like TIA proteins) and repress them from binding U1 snRNP, by these means destabilizing U1 and repressing the splicing.

Another interesting explanation about how splicing can be regulated on a distance is by influencing the kinetics of the splicing reaction. For instance, some proteins (as TIA proteins) through binding to a preceding exon are thought to increase the speed of transition between different stages of splicing reaction, and hence to reduce the time for other RBPs to bind, what results in splicing repression (Witten and Ule, 2011).

Splicing repression could also happen if T-STAR is interfering with enhancer elements of sequence, as it was suggested in study by Wu et al on splicing repression of *MAG* gene by Quaking protein (Wu *et al.*, 2002). Splicing repression could also happen through some other means, including epigenetic modification of exons (Brown *et al.*, 2012).

All the mechanisms mentioned above are possible for *Nrxn2* sites placed downstream the exon. For the sites placed upstream splicing repression could happen through blocking branch point sequence and inhibiting spliceosome assembly, as it happens for other RBPs as Nova, Fox and PTB, that bind upstream and repress the splicing (Ule *et al.*, 2006; Zhang *et al.*, 2008a; Xue *et al.*, 2009).

3.5 Chapter Summary

In this Chapter direct binding of T-STAR and Sam68 to *Nrxn2* RNA was confirmed by EMSA analysis. The T-STAR target motif in *Nrxn2* was dissected into separate binding sites, showing UAAAAA site as the one mostly used for T-STAR splicing regulation. Furthermore, the CUAACUAACUAA repeat region was identified as a general repressor site that regulates the availability of T-STAR sites and T-STAR involvement in *Nrxn2* repression. By the series of moving mutants it was found that *Nrxn2* repression by T-STAR is position independent. The exact mechanism by which T-STAR represses splicing in wild type *Nrxn2* was not identified, but several mechanisms by which repression could happen, as looping model, or indirect repression model, or kinetics model, etc. were proposed. To conclude, *Nrxn2* splicing regulation by T-STAR is having a multilevel control and thus might have a fundamental role for the cell.

Chapter 4: Dissecting the mechanism of splicing control of *Tomosyn-2* cassette exon by T-STAR and Sam68

4.1 Introduction

In the study by Ehrmann et al 2013, in addition to *Neurexin* targets, another physiological target of T-STAR was identified. This target exon is within a gene called *Syntaxin-binding protein 5-like* (*Stxbp5l*, also called *Tomosyn-2*).

Ehrmann et al 2013 showed the splicing of *Neurexin* targets to be region-specific in the mouse brain, with high levels of exon AS4 exclusion in the forebrain-derived regions. These regional splicing patterns were in agreement with T-STAR levels in these brain regions, and were totally abolished in the brains of T-STAR null mice (Ehrmann *et al.*, 2013). Figure 4.1 is reproduced from the paper by Ehrmann et al 2013. It shows the Western blot with the expression pattern of T-STAR and Sam68 in different brain regions (4.1 A). The Schematic in Figure 4.1 B shows where each area is situated in a mouse brain. In this Chapter I examine if *Tomosyn-2* splicing repression by T-STAR is similarly differentially regulated in different mouse brain regions.

Tomosyn-2 was shown to be a physiological target of T-STAR, however, it is not known if, in contrast to *Nrxn2*, *Tomosyn-2* might be also regulated by Sam68. In this current Chapter I test this and find that *Tomosyn-2* splicing repression is regulated by T-STAR and Sam68 in a minigene and the regulation sites are dissected.

Based on results from Chapter 3, T-STAR splicing regulation of *Nrxn2* appeared to be position independent, as with T-STAR binding sites being moved further downstream and upstream of the target exon, splicing was still highly repressed. In this current Chapter I similarly test if *Tomosyn-2* splicing regulation depends on the position of protein-RNA binding sites relative to the regulated exon.

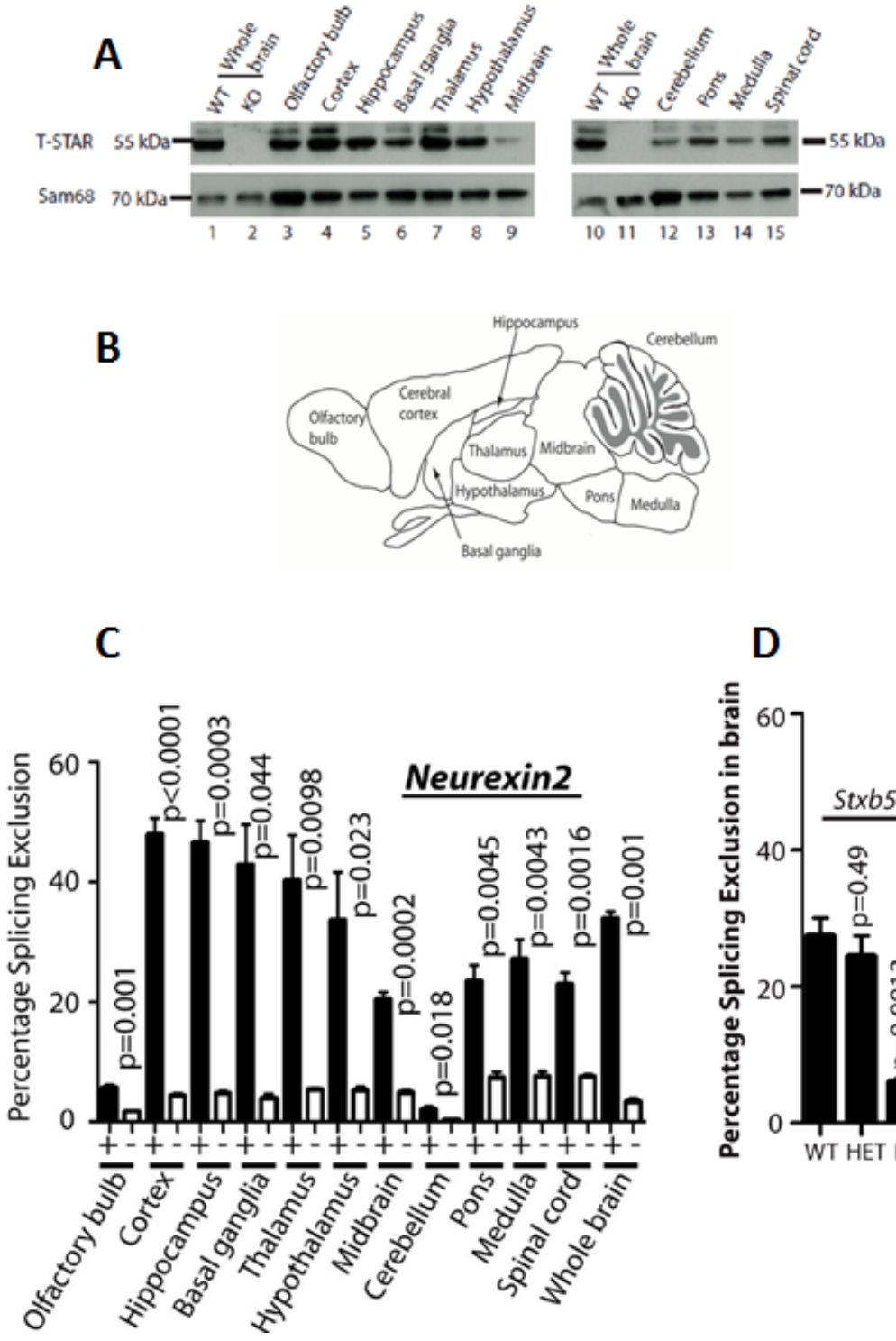


Figure 4.1 – T-STAR regulates region-specific splicing of *Nrxn2* AS4 in the brain. Figures adopted from paper by Ehrmann et al 2013. A – The Western Blot performed earlier in the lab, showing the expression pattern of T-STAR and Sam68 proteins in different wild type and T-STAR knockout mouse brain regions. B – Schematic representing the areas of mouse brain. C – *Nrxn2* splicing pattern in different brain regions. Closed columns labelled by "+" show the *Neurexin-2* splicing pattern in the wild type mouse brain. Opened columns labelled by "-" show

the *Neurexin-2* splicing pattern in the T-STAR knockout mouse brain. Data represents the mean of minimum three biological replicates. D - *Tomosyn-2* splicing pattern in wild type (WT), heterozygous (HET) and T-STAR knockout (KO) mouse brain. Graphs show the percentage of splicing exclusion. Data represents the mean of minimum three biological replicates. In each case, statistical significance was compared between wild type brain cells and heterozygous or T-STAR knockout brain cells by means of independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.001.

4.2 The aims of this Chapter are:

1. To investigate if *Tomosyn-2* splicing repression by T-STAR in the mouse brain is region-specific
2. To investigate whether *Tomosyn-2* is a direct splicing target for T-STAR using a minigene approach
3. To dissect the target response elements and investigate their importance for *Nrxn2* splicing repression by T-STAR and Sam68
4. To investigate the position dependence of response elements for *Tomosyn-2* splicing regulation

4.3 Results

4.3.1 T-STAR protein expression establishes regional splicing regulation of Tomosyn-2 exon in the mouse brain

Tomosyn-2 splicing patterns were assessed in 11 brain regions of three wild type and three T-STAR knockout mice. For this purpose, whole RNAs, previously extracted from different mouse brain areas by Dr Ingrid Ehrmann, were converted by me to cDNAs and used as templates in RT-PCR reactions. The resulting samples were analysed using the capillary electrophoresis system. If T-STAR is a concentration-dependent splicing regulator as predicted, the more there is T-STAR is in each brain area the more exon exclusion (splicing repression) will happen. To test this, the percentages of exon exclusion were analysed for each brain region.

In each brain region, the percentage of *Tomosyn-2* exon exclusion was different (Figure 4.2). Similar to *Nrxn2* (Figure 4.1 C), *Tomosyn-2* splicing was highly repressed in the

forebrain regions i.e. cortex and hippocampus. Splicing repression was also high in the thalamus and pons. Similar to *Nrxn2*, less exon repression was detected in the olfactory bulb and cerebellum. Intermediate splicing repression levels were detected in midbrain and medulla. As for *Nrxn2*, *Tomosyn-2* splicing levels are correlated with T-STAR expression levels in the brain: i.e. T-STAR is highly expressed in cortex, hippocampus and thalamus, but low expression is seen in the cerebellum. Interestingly, in the pons, splicing repression appeared to be high (Figure 4.2), while T-STAR expression is moderate (Figure 4.1 A).

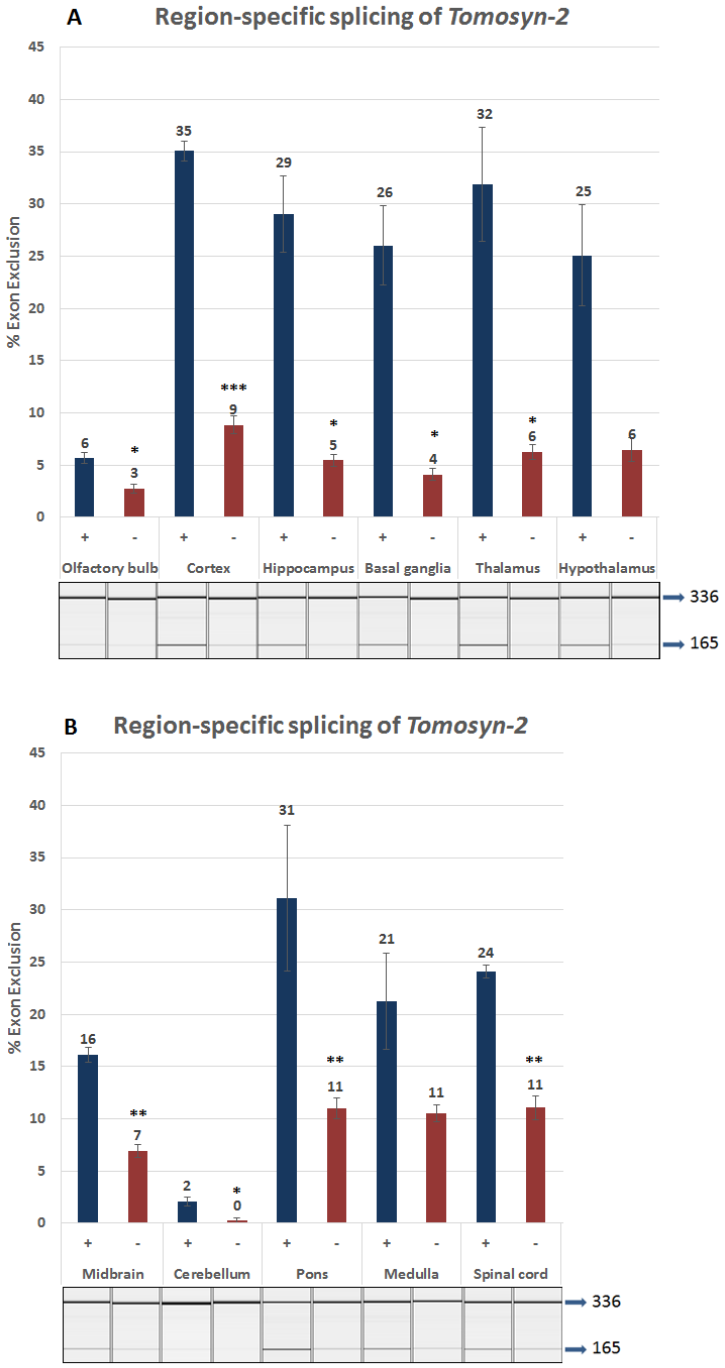


Figure 4.2 – T-STAR regulates region-specific splicing of *Tomosyn-2* exon 23 in the brain. A and B - The graphs show percentages of exon exclusion of *Tomosyn-2* in different areas of the wild type and T-STAR knockout mouse brain. Blue columns labelled by "+" show the *Tomosyn-2* splicing pattern in the wild type mouse brain. Red columns labelled by "-" show the *Tomosyn-2* splicing pattern in the T-STAR knockout mouse brain. Data represents the mean of minimum three biological replicates. In each case, statistical significance was compared between wild type brain cells and T-STAR knockout brain cells by means of independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.001.

In T-STAR knockout mice (Figure 4.2), regional splicing patterns of *Tomosyn-2* were severely abolished, and splicing repression reduced. If for wild type brain parts splicing repression was in the order of 35%, for T-STAR knockout mice, it was not more than 11%. Thus, it can be concluded, that T-STAR is regulating *Tomosyn-2* splicing in different brain regions, with more repression happening in forebrain-derived regions.

4.3.2 Tomosyn-2 splicing is repressed by T-STAR and Sam68

As in Chapter 3, a minigene construct was made to test splicing regulation by T-STAR. The *Tomosyn-2* exon controlled by T-STAR (171bp), together with 300bp of the flanking intron sequences was cloned into a minigene. This minigene was co-expressed in HEK293 cells with expression constructs encoding T-STAR and Sam68 proteins. The cells were harvested, RNA was extracted and reverse transcribed. Following capillary gel electrophoresis, splicing patterns were analysed. Two bands were observed on a gel, one for exon inclusion and one for exon exclusion. The concentration of each band was calculated by a multi-capillary QIAxcel gel electrophoresis device. Based on the concentration ratios between exon exclusion and exon inclusion, percentages were calculated and plotted on the graphs.

Both T-STAR and Sam68 highly repressed exon inclusion of *Tomosyn-2*, compared to endogenous conditions (GFP lane) (Figure 4.3 A). The percentage of exon exclusion has significantly increased by 32% for T-STAR co-transfection and by 41% for Sam68 co-transfection (compared to 22% for transfection with GFP-only). Interestingly, even though *Tomosyn-2* was identified as a T-STAR target initially, Sam68 regulated its splicing in a minigene even more than T-STAR. Western blot confirmed similar levels of transfected fusion proteins (Figure 4.3 B).

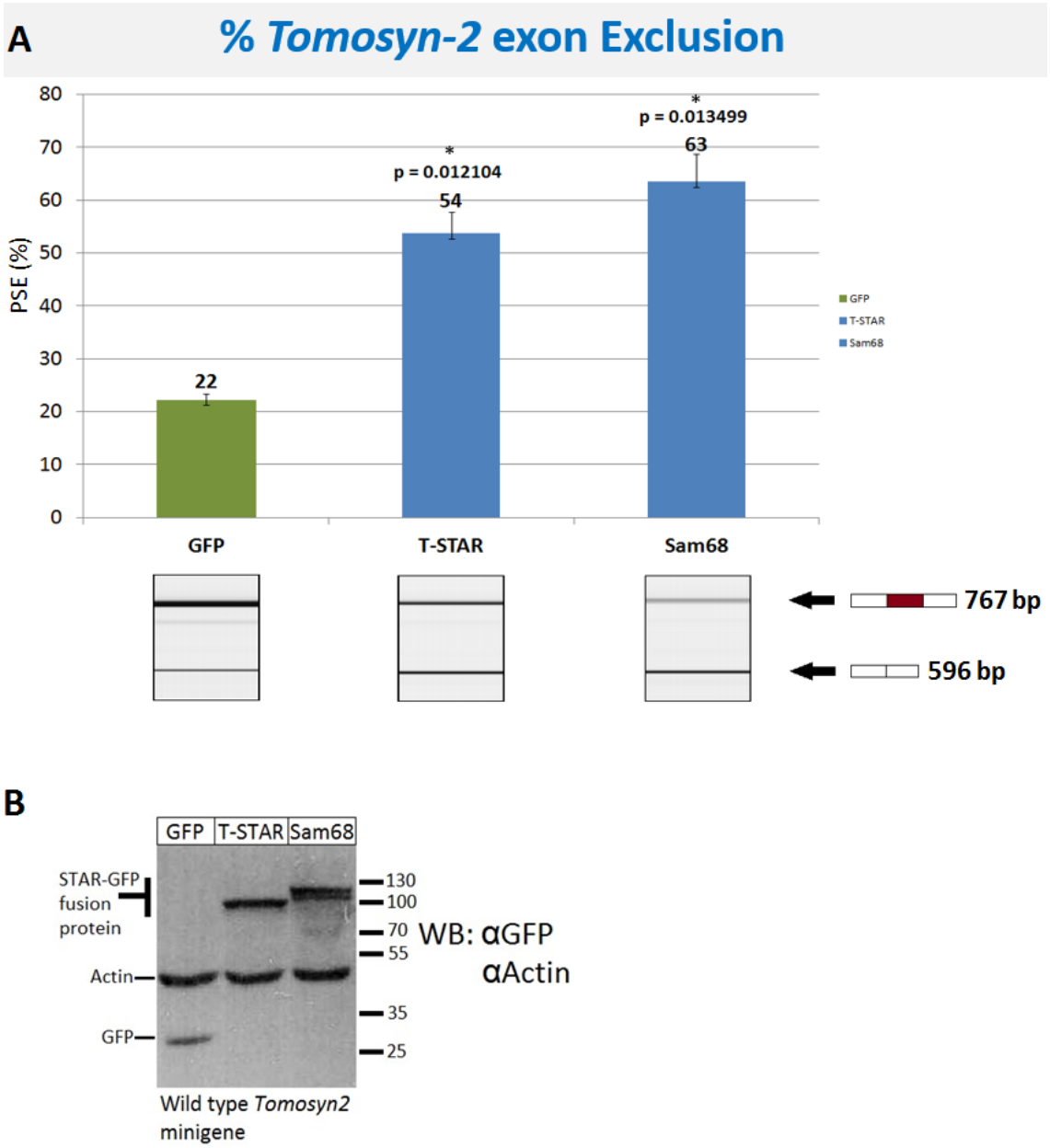


Figure 4.3 – T-STAR and Sam68 repress the splicing of *Tomosyn-2* exon 23 when expressed from a minigene. A – Graphs show the percentage of splicing exclusion in the wild type *Tomosyn-2* minigene co-transfected with GFP-only control as well as GFP-coupled T-STAR and Sam68 fusion proteins. The data represents the mean of minimum three biological replicates. In each case, statistical significance was compared between HEK293 cells expressing GFP and HEK293 cells expressing the GFP fusion proteins by means of independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.001. B - Western blot analysis of HEK293 cells co-transfected with GFP-coupled fusion proteins and wild type *Nrxn2* minigene. The Western blot was probed for GFP and actin and shows equal expression levels of each fusion protein.

4.3.3 Tomosyn-2 sites are redundant for splicing regulation by T-STAR and Sam68

The *Tomosyn-2* sequence was analysed to identify potential T-STAR and Sam68 binding sites. Six potential regions were found in the upstream intron, two inside the exon, and seven in the downstream intron (Figure 4.4). Since *Nrxn2* was repressed by T-STAR, with T-STAR binding sites being downstream the exon, I focused more on the downstream sites. One group of sites was located 38 bases downstream of the regulated exon and another group 106 bases downstream. One potential site was also detected very close to the exon (7 nucleotides downstream) and two potential sites further downstream (156 and 228 nucleotides).

GCATATTTCATATTGCCATCCAacaatacattaaataaatagtagggaag
tatatgacatttcaagaaatgttaattagataaattatgaaaatatgtcc
tatacataataagtatcttgatttttttataatctggaattgtaactaaa
tcttaaaataacaaaaaaaaatagtcatattcaaatgaattggaagaaga
ggaattcctgttcctccatttcttaacctatgcctttggatttcctcatc
tgaagacttgattatgatatttattctgccttcttcctcttggtacagac
catgtaaatggacactgcacaagtccaacctctcagagttgcagttctgg
aaaacgtctttccagtgctgacgtttcaaaagtaaatcgctggggtcctg
gaagaccaccatttagaaaggcacagtcagctgcttgcatggagatttct
ttaccagttacaactgaaggtaggataaacatatcaatttctgtgagtac
agtttaaaatttgataaaatttcgtttatcatggtaacaggtgctgggag
acaaagaaatggaagttacatttaaaagatgatttaaaaacaaaaacaaa
aacaaaactacttgacattgaaaaattaagatcctgtttgactgtggcca
taacttcaattctattacttactcacttccttccttatctctaaaaataa
aatgatcccactgtaggtagccatgcaaggagataatctgtaaCTGCAAC
CTTAATTTCCTCTACTG

Figure 4.4 – Diagram of *Tomosyn-2* cassette exon and flanking introns from the UCSC genome browser.

The exon is highlighted in grey. The rest of the sequence are flanking introns. The target motifs are highlighted in yellow. The underlined sequences were subjected to mutagenesis.

To test the involvement of the sites in splicing control, two groups of sites 38 and 106 nucleotides downstream of the exon were chosen for site directed mutagenesis (underlined on Figure 4.4). TA motifs were replaced by CC sequences. Mutant 1 had 5 potential single motifs mutated, while Mutant 2 had 2. (Figure 4.5).

When co-transfected with T-STAR and Sam68 expression constructs, each mutant had significantly reduced splicing repression by both T-STAR and Sam68 proteins, compared to the wild type sequence (Figure 4.5). In Mutant 1, repression was reduced by 14% for

T-STAR, and by 18% for Sam68, in comparison to wild type. In Mutant 2 there was a reduction by 15% for both T-STAR and Sam68, in comparison to wild type. Thus, these data show the tested groups of sites are target sites for T-STAR and Sam68 splicing regulation.

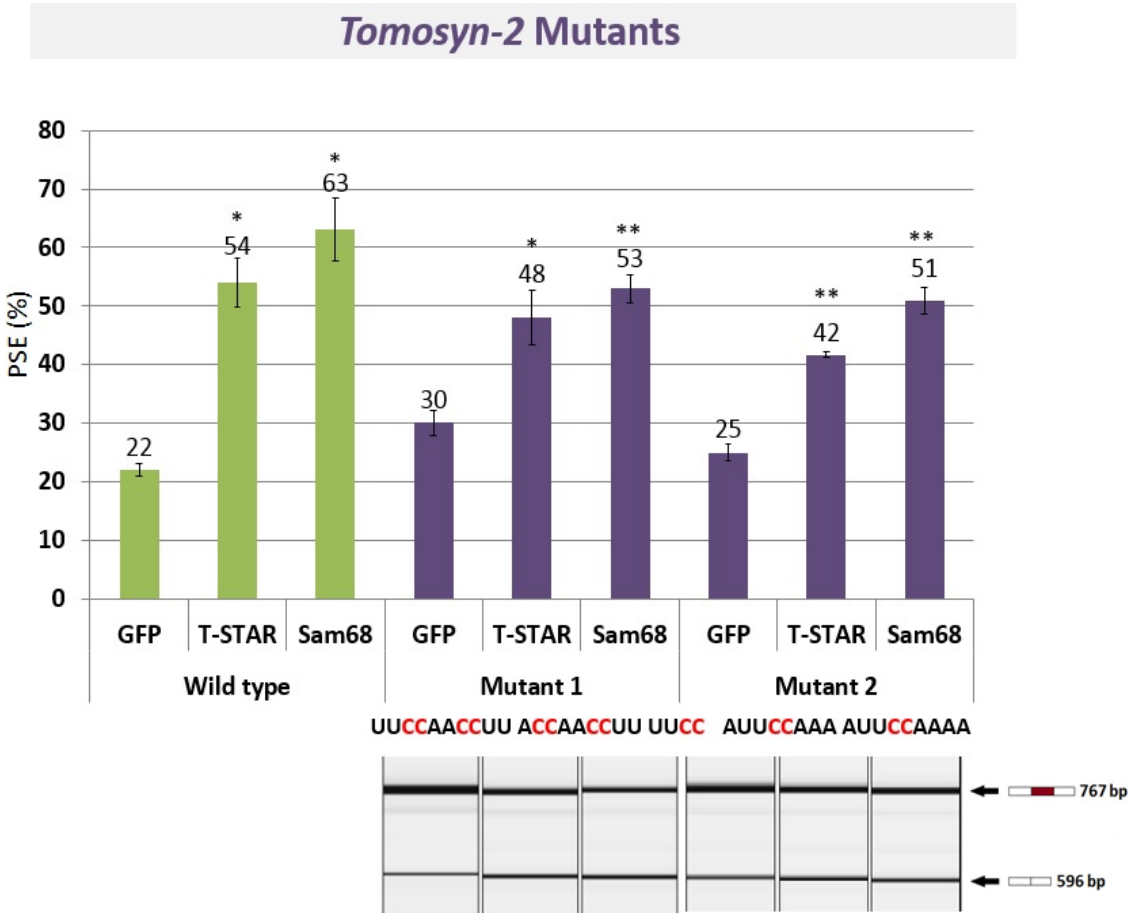


Figure 4.5 – *Tomosyn-2* Mutants 1 and 2 show the reduced patterns of splicing exclusion by T-STAR and Sam68. Graphs show the percentage of exon exclusion for two *Tomosyn-2* sequence mutants and the wild type sequence. The data represents the mean of minimum three biological replicates. In each case, statistical significance was compared between HEK293 cells expressing GFP and HEK293 cells expressing the GFP fusion proteins by means of independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.001.

To examine if mutation of all the sites tested above would abolish the splicing repression completely, a combined mutant was created, with both groups of T-STAR binding sites mutated.

After co-transfection with T-STAR and Sam68 fusion proteins, the combined *Tomosyn-2* mutant had much less exon repression, suggesting that the two sequence elements

covered by Mutant 1 and Mutant 2 do cooperate in achieving normal levels of splicing repression (Figure 4.6). However, joint mutation did not completely abolish *Tomosyn-2* splicing repression by T-STAR and Sam68. Repression by Sam68 was higher, than by T-STAR. This co-transfection was performed only single time, thus the result is preliminary, and needs replication.

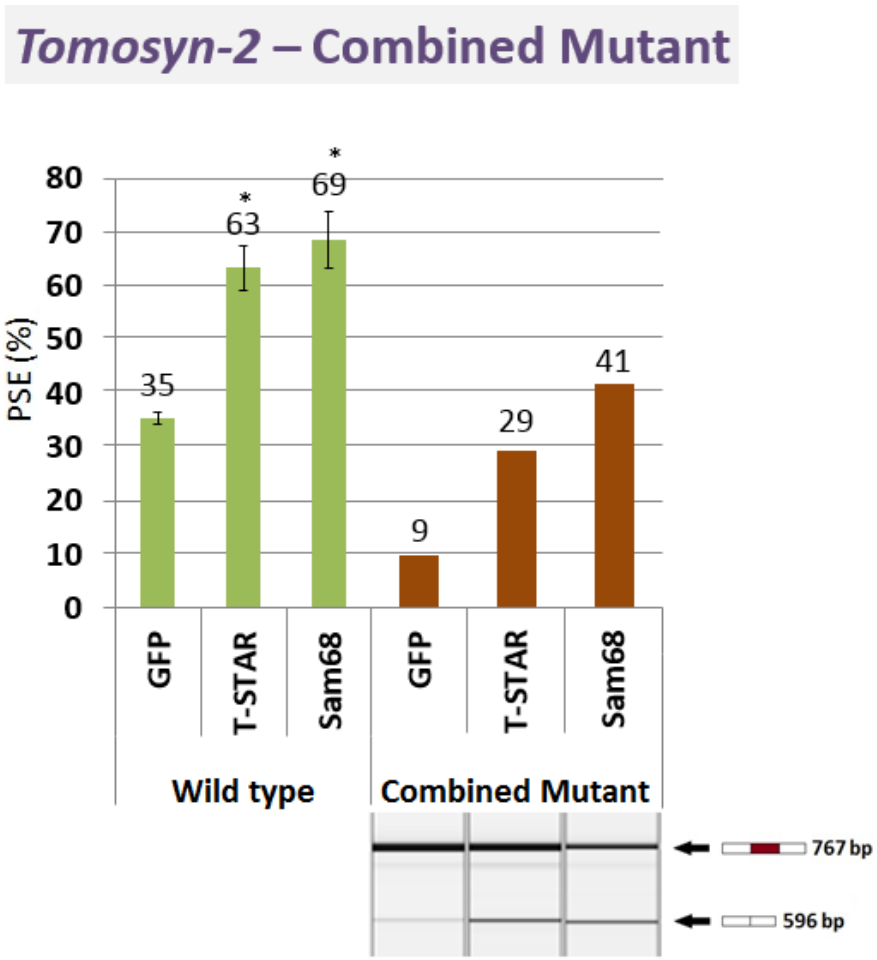
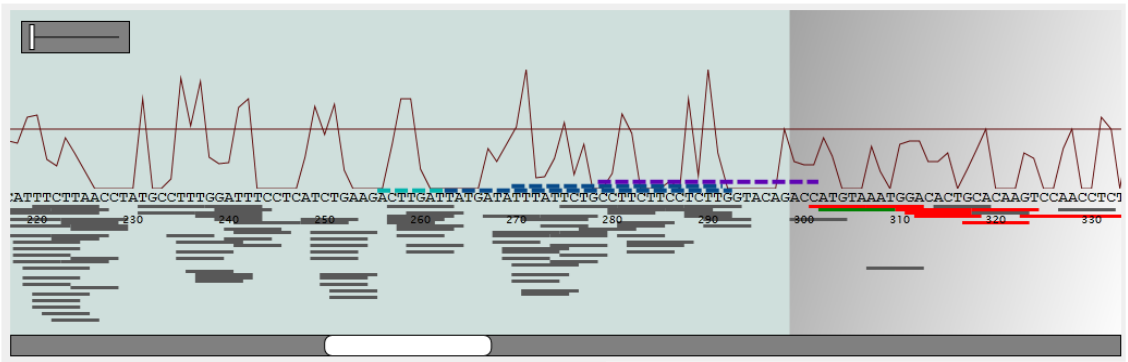


Figure 4.6 – The combined *Tomosyn-2* mutant had much less exon repression than either single mutant, suggesting that the two motifs do cooperate. Graphs show the percentage of exon exclusion for *Tomosyn-2* mRNAs encoded by a construct in which sites 1 and 2 were jointly mutated, compared to a wild type. Experiment for the joint mutant was performed once and needs replication. For the wild type, statistical significance was compared between HEK293 cells expressing GFP and HEK293 cells expressing the GFP fusion proteins by means of independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.001.

4.3.4 Tomosyn-2 splicing is repressed by T-STAR and Sam68 with binding sites being upstream of the exon

In Chapter 3 it was found that splicing repression of *Nrxn2* AS4 by T-STAR is position-independent relative to the regulated exon. Hence, I tested if the same pattern of splicing regulation is observed for *Tomosyn-2*, for both T-STAR and Sam68. A version of the *Tomosyn-2* minigene was cloned with the whole response element covered by site 1 moved 54 nucleotides upstream of the target exon. The Combined *Tomosyn*-2 Mutant with both binding sites downstream the exon mutated (Section 4.3.3) was used as a template for mentioned cloning. Similar as in Chapter 3, a Sroogle (Splicing Regulation Online Graphical Engine) program (http://sroogle.tau.ac.il/) was used to locate the Branch Site sequence and Polypyrimidine tract in the sequence upstream (output is on Figure 4.7) of the regulated exon, so as not to interfere with these regions. A 54-nucleotide distance was chosen as an earliest point upstream the exon to prevent any interference with these important sequence elements.



Splice Site Scores

Represented by dashed annotations **above** the sequence.

Element	Start	End	Score		Score Percentile (Const Exons)	Score Percentile (Alt Exons)
Branch site	256	262	3.00	(Kol et al)	0.29	0.32
	NA	NA	NA	(Schwartz et al)	NA	NA
Polypyrimidine tract	263	292	0.79	(Kol et al)	0.5	0.51
	270	291	14.73	(Schwartz et al)	0.83	0.82
3' splice site	279	301	8.29	(Max entropy)	0.42	0.5
			86.15	(PSSM)	0.53	0.58

Figure 4.7 – A - Output of a wild type *Tomosyn-2* sequence in a Sroogle program. Splice sites, Branch site and Polypirimidine tract are marked with coloured dash lines that match by colour with the names in a table below.

With the sites moved upstream and the downstream sites being blocked, splicing repression was reduced at endogenous protein levels. Comparing the ratios between GFP and T-STAR / GFP and Sam68, splicing repression of *Tomosyn-2* exon by T-STAR and Sam68 in the mutant appeared to be even stronger than using a wild type minigene (Figure 4.8). Thus, it can be suggested that T-STAR and Sam68 repress *Tomosyn-2* splicing both upstream and downstream of the exon.

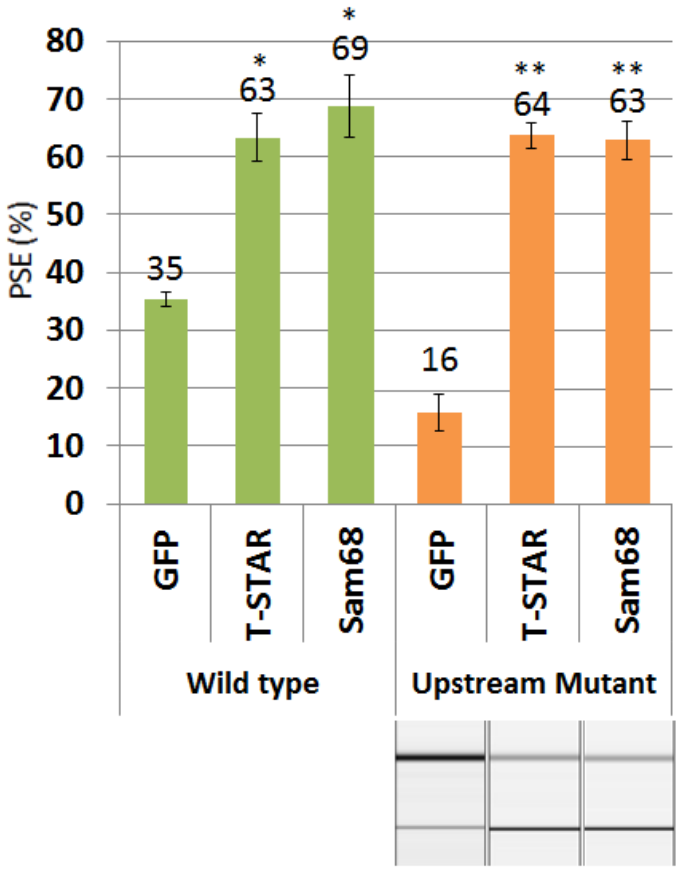


Figure 4.8 – T-STAR and Sam68 response element can still potently repress splicing even when placed upstream of the regulated *Tomosyn-2* exon. The percentage of splicing exclusion in mutant with target sites moved 54 bases upstream of the target *Tomosyn-2* exon, comparing with the wild type minigene control. The data represents the mean of minimum three biological replicates. In each case, statistical significance was compared between HEK293 cells expressing GFP and HEK293 cells expressing the GFP fusion proteins by means of independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.001.

4.4 Discussion

4.4.1 T-STAR splicing is regionally repressed in the mouse brain

The first aim of this Chapter was to investigate if *Tomosyn-2* splicing repression by T-STAR in the mouse brain is region-specific. Splicing of different targets in the brain is coupled to different pathways in the brain. The study by Iijima et al 2011 has found that *Nrxn1* splicing regulation by Sam68 is dependent on neuronal activity. Neuronal activity promotes a shift in *Nrxn1* splice isoform choice by means of calcium/calmodulin-dependent kinase IV signalling (Iijima et al 2011). The study by Ehrmann et al 2013 has suggested a model where T-STAR is providing a concentration-dependent switch to repress *Nrxn2* splicing in different brain regions. According to this model, high concentrations of T-STAR in forebrain repress the splicing of *Neurexin* genes, while low T-STAR concentrations activate *Neurexin* splicing. The prediction of this model is that high T-STAR concentrations in forebrain-derived structures repress the splicing of *Nrxn1*, *Nrxn2* and *Nrxn3* AS4. Lower T-STAR concentrations in the brain regions like cerebellum result in splicing activation of *Nrxn1*, *Nrxn2* and *Nrxn3* AS4 (Ehrmann *et al.*, 2013).

To test this model, in this Chapter I assessed the splicing of *Tomosyn-2* exon 23 in 11 brain regions of wild type and T-STAR knockout mice. The *Tomosyn-2* repression pattern in the mouse brain by T-STAR was found to be similar to that of *Nrxn2*. Splicing was highly repressed by T-STAR in all forebrain regions tested (cerebral cortex, hippocampus, basal ganglia, thalamus and hypothalamus). The highest levels of splicing repression overlap with the highest T-STAR expression levels in these forebrain-derived regions. In the midbrain and hindbrain parts of the brain (cerebellum and medulla), the splicing repression level is less. This overlaps with lower T-STAR expression in these brain areas. In the T-STAR knockout mouse, splicing repression in all brain regions is reduced. These data are consistent with *Tomosyn-2* splicing repression also being regulated by a T-STAR concentration-dependent switch model as suggested by Ehrmann et al 2013 for *Nrxn2*.

4.4.2 Tomosyn-2 is a direct target for T-STAR and Sam68

The second aim of this Chapter was to investigate whether *Tomosyn-2* is a direct splicing target for T-STAR using a minigene approach. *Tomosyn-2* splicing is disrupted in the T-STAR null mice, but *Tomosyn-2* could be a direct or indirect target. To test between these possibilities I constructed a minigene and co-transfected it with T-STAR and Sam68 expression constructs. The endogenous splicing repression of *Tomosyn-2* by T-STAR in mouse brain was confirmed in a minigene. Furthermore, *Tomosyn-2* splicing was shown to be highly repressed by Sam68.

The third objective of this Chapter was to dissect the target response elements and investigate their importance for *Nrxn2* splicing repression by T-STAR and Sam68. T-STAR protein binds to A/U-rich sequences. By visually searching the sequence, I found seven sites downstream of regulated *Tomosyn-2* exon, which appeared to respond to both T-STAR and Sam68. Seven identified motifs were mutated creating two mutants. Mutant 1 covered a group of 5 potential single motifs, while Mutant 2 covered a group of 2 potential single sites. Deletion of these motif groups individually impacted splicing repression by T-STAR and Sam68. When both groups were combined into one mutant, the effect of mutations was additive, showing that motifs cooperate in splicing repression. However, even with all seven downstream motifs mutated, *Tomosyn-2* was still highly repressed by both T-STAR and Sam68. The motifs tested are not the only A/U-rich motifs in this sequence. Thus, I suggest that there is redundancy between T-STAR and Sam68 regulatory motifs. When some sites are removed, others start to function. Alternatively, all of the sites can work all the time in a semi-redundant way.

From Chapter 3 it was seen that *Nrxn2* splicing was not repressed by Sam68. Interestingly, Sam68 repression of *Tomosyn-2* is 9% stronger than repression by T-STAR. Comparing the target motifs in both genes, *Tomosyn-2* mainly has UAAA/UAAAA/UAAAAA repeats, which are the best Sam68 sites based on SELEX (Lin *et al.*, 1997; Galarneau and Richard, 2009b). The *Nrxn2* splicing response region mainly consists of UUAA motifs and is repressed by T-STAR only. Thus, I speculate, that *Nrxn2* target sites have higher affinity for T-STAR, while *Tomosyn-2* motifs are better targets for Sam68.

4.4.3 Tomosyn-2 splicing is highly repressed by T-STAR from an upstream and downstream intronic locations

As the last aim of this Chapter, I investigated the importance of the wild type downstream position of response elements for *Tomosyn-2* splicing regulation. As shown in Chapter 3, *Nrxn2* splicing repression by T-STAR appeared to be position-independent. Here I moved the *Tomosyn-2* response sites from a downstream position to 54 nucleotides upstream of the regulated exon. This experiment showed that with response sites being upstream, *Tomosyn-2* splicing was still significantly repressed. Thus, the splicing repression of *Tomosyn-2* is position-independent.

4.5 Chapter Summary

In this Chapter T-STAR was shown to be a regional regulator of *Tomosyn-2* splicing in mouse brain. *Tomosyn-2* splicing was repressed by T-STAR and Sam68 when expressed from a minigene. Regulatory sites for both T-STAR and Sam68 were uncovered in the downstream intron of *Tomosyn-2*. However, the sites in *Tomosyn-2* appeared to be redundant. With downstream motifs mutated, splicing was still repressed by T-STAR and Sam68, but at lower overall levels. Splicing regulation of *Tomosyn-2* by T-STAR and Sam68 still occurred when the binding sites were both upstream and downstream of the exon.

Chapter 5: Why is there T-STAR specific regulation of *Neurexin-2* splicing?

5.1 Introduction

Based on our EMSA experiments described in a Chapter 3 both T-STAR and Sam68 directly bind to the *Nrxn2* target sequence. That means that the *Nrxn2* sequence composition appears to be enough for Sam68 binding, but not for splicing regulation. This Chapter will investigate why *Nrxn2* splicing is specifically regulated by T-STAR, but not by Sam68. I hypothesize that this regulation specificity is due to the *Nrxn2* sequence composition, elements of which are specifically recognized and required for splicing repression by T-STAR, and not by Sam68.

Based on SELEX experiments, high affinity sites for T-STAR are UAAA/UUAA motifs (Galarneau and Richard, 2009a). By NMR technique, Foot et al has identified an AAAUAA sequence as one single stranded RNA sequence bound by the T-STAR KH domain (Foot *et al.*, 2014). Based on SELEX, Sam68 high affinity sites are comprised of UAAA/UUUA sequences (Galarneau and Richard, 2009a). As shown in Chapter 4, *Tomosyn-2* splicing is regulated by both T-STAR and Sam68 through sequence motifs mainly comprising UAAA/UAAAA/UAAAAA. The UA-rich sequence downstream of the *Nrxn2* exon (AAUUAAUUAAUUAAUUAACUAACUAACUAACUUUAAAA ACACGAUCUUAAA) consists of six UUAA repeats that resemble predicted T-STAR target motifs, and only one UAAAAA sequence that is a proposed target for Sam68 (as well as for T-STAR). Minigene analyses described in Chapter 3 have confirmed T-STAR splicing regulation using these *Nrnx*2 motifs. No effect of Sam68 on *Nrxn2* splicing was observed. Neither was there an effect on *Nrxn2* response to Sam68 after mutating the UAAAAA sequence that is a predicted Sam68 target, nor on any other motifs tested. Thus, based on previous data and on minigene analyses from previous Chapters, the *Nrxn2* target region must be a better T-STAR response sequence, than a Sam68 one. In this current Chapter, I will replace this *Nrxn2* target sequence with several AU-rich motif combinations, which are expected to be a better target for Sam68. I will also investigate whether the length of the potential Sam68-like motif affects the splicing regulation.

In Chapter 4 it was shown that *Tomosyn-2* splicing is repressed by T-STAR and Sam68. Regulatory sites for both T-STAR and Sam68 were uncovered in the downstream intron

of *Tomosyn-2*. In this Chapter I investigate whether it is possible to induce a Sam68 response by moving a *Tomosyn-2* regulatory element downstream of *Nrxn2* AS4.

Both T-STAR and Sam68 are members of STAR protein family that are very similar to each other. However, there are other proteins in these family as well, including Slm1 (Synthetic Lethal with Mss4), SF1 (Splicing Factor 1) and QKI (Quaking). These also belong to the STAR family and share some similarity with T-STAR and Sam68. To further study *Nrxn2* splicing regulation, I investigate if other STAR proteins regulate *Nrxn2* splicing. I also examine through which sequences the splicing regulation is happening.

5.2 The aims of this Chapter are:

1. To test whether *Nrxn2* is being regulated by Sam68 when the *Nrxn2* sequence is changed to be a better Sam68 target
2. To investigate, whether placing Sam68 responsive motifs from *Tomosyn-2* into *Nrxn2* can put *Nrxn2* AS4 under Sam68 control
3. To test if other STAR proteins also regulate *Nrxn2* splicing and if so, through which sequences this regulation is happening

5.3 Results

5.3.1 Sam68 regulates Nrxn2 splicing when Nrxn2 sequence is replaced by a different A/U-rich regulatory element

Based on SELEX, UAAA/UUUA sequences are considered the best Sam68 targets (Galarneau and Richard, 2009a). The *Nrxn2* target region already includes one UAAAAA sequence; however, it is not enough for splicing regulation by Sam68. To find out if Sam68 needs a different A/U-rich motif (i.e. longer, different base combination, etc.) to regulate *Nrxn2* splicing, by means of site-directed mutagenesis, the wild type *Nrxn2* sequence was replaced by four copies of UAAA, with a final sequence being AAUAAAUAAAUAAAUAAA (Figure 5.1-1).

Figure 5.1 – Sequence combinations that might respond to Sam68. A schematic representing five sequences (1-5) which were inserted into *Nrxn2* mutant with a big mutation to test if these would enable a response to Sam68. The inserted sequence is highlighted in yellow and the other mutations are shown in red.

As previously, the *Nrxn2* mutant with a big mutation where all target sites are blocked by mutagenesis was used as a cloning template. The new minigenes were co-expressed in HEK293 cells, with expression constructs encoding T-STAR and Sam68 fusion proteins, as well as a GFP-only control. As for previously described co-transfections, the cells were harvested, RNA was extracted and reverse transcribed. Capillary gel electrophoresis was performed and splicing patterns were analysed.

Two bands were always observed on a gel, one for exon inclusion and one for exon exclusion. Concentrations for each band were calculated using a multi-capillary QIAxcel gel electrophoresis device. Based on the concentration ratio between exon exclusion and exon inclusion, percentages were calculated and plotted on the graphs. The full detail on how the calculation was performed are given in the section 2.5.4 of the Materials and Methods Chapter. T-STAR repressed the splicing of Mutant 4xUAAA up to similar extent as in the wild type minigene (Figure 5.2-A). However, even with four copies of the target sequence inserted, Sam68 still does not regulate *Nrxn2* splicing (only 2% difference was observed after co-transfection of Sam68-GFP compared with the GFP-only control, which is not statistically significant).

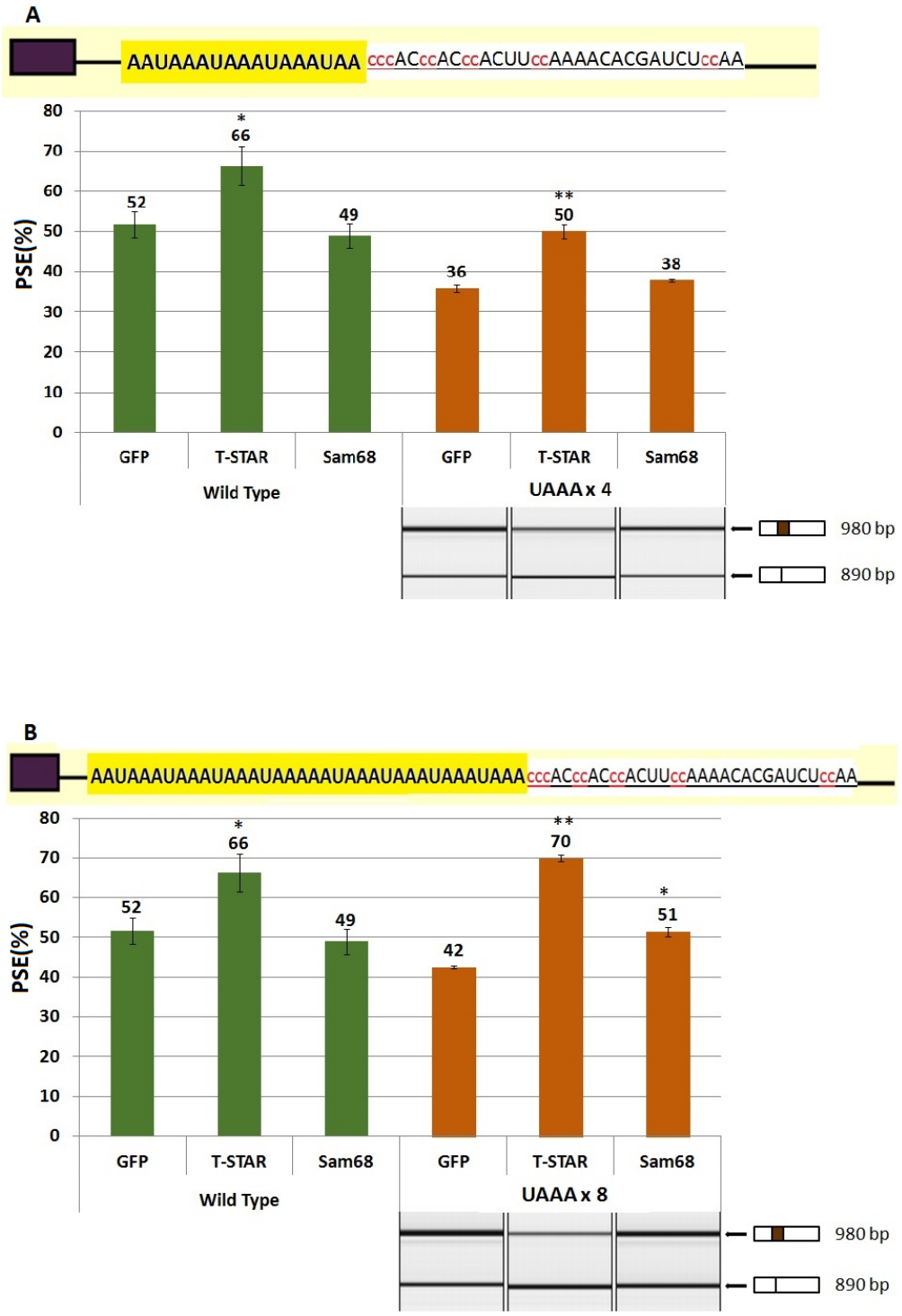


Figure 5.2 – Sam68 starts regulating *Nrxn2* splicing when *Nrxn2* sequence is replaced by the string of eight UAAA motifs. Downstream binding sites were inserted comprising A) UAAA x 4 and B) UAAA x 8 that were designed to be potentially regulated by Sam68. Graphs show the percentage of splicing exclusion (PSE%) in different versions of the *Nrxn2* minigene exon co-transfected with GFP-only, T-STAR and Sam68. Data represents the mean of minimum three biological replicates. In each case, statistical significance was compared between HEK293 cells expressing GFP and HEK293 cells expressing the GFP fusion proteins by means of independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.001.

Following this experiment, it was decided to increase the length of the inserted potential Sam68 target sequence. The hypothesis was that a longer A/U-rich sequence could be a better target for Sam68. Hence, another mutant with 8xUAAAs was produced (with the final target sequence being AAUAAAUAAAUAAAUAAAAAUAAAUAAAUAAA UAAA) (Figure 5.1-2). Interestingly, when co-transfected into HEK293 cells, T-STAR repressed the Mutant 8xUAAA splicing at twice as high levels as the wild type *Nrxn2* minigene (Figure 5.2 - C). Importantly, when Mutant 8xUAAA was co-transfected with Sam68, a 9% higher splicing repression was observed compared to the GFP-only control. This PSE(%) difference was statistically significant (Figure 5.2 B).

As UAAA x 8 did not reach strong repression levels by Sam68, another mutant with a potential Sam68 target sequence was cloned, where the sequence was converted into 4xUAAAA (Figure 5.1-3). The hypothesis was, that the motif with more adenine nucleotides could be a better Sam68 target. When co-transfected with T-STAR and Sam68, a similar effect to that of mutant 8xUAAA was observed (Figure 5.3 – A): splicing was significantly repressed by T-STAR (18% significant difference comparing to GFP control) and to a lesser extent by Sam68 (7% significant difference comparing to GFP control) (Figure 5.3 – A). Thus, both 8xUAAA and 4xUAAAA sequences are apparently equally functional for Sam68 to repress *Nrxn2* splicing.

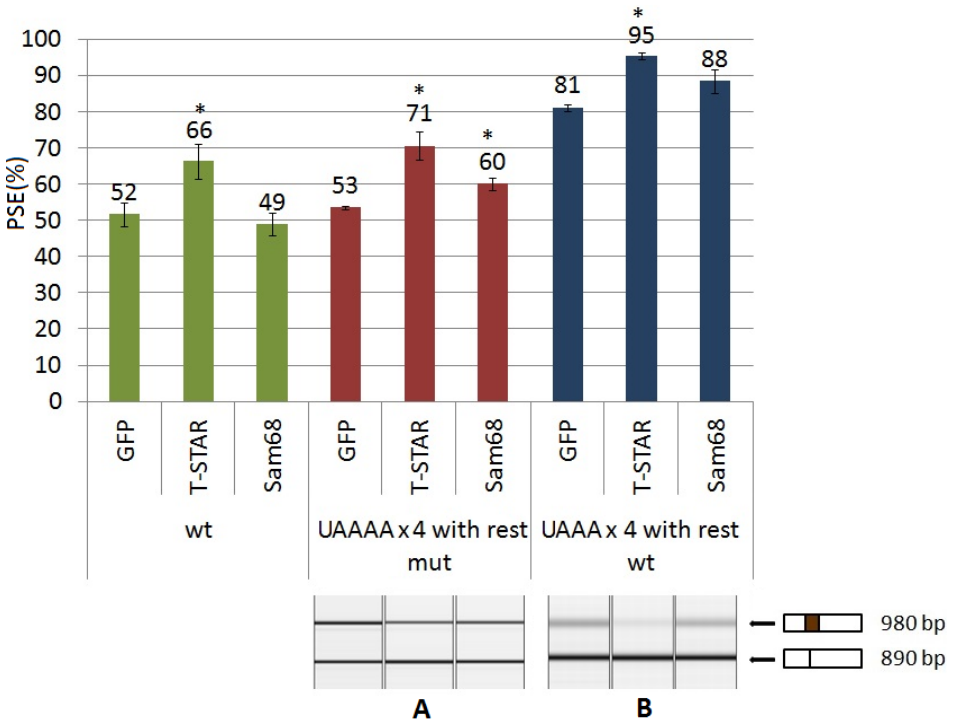


Figure 5.3 – Sam68 starts regulating *Nrxn2* splicing when *Nrxn2* sequence is replaced by the string of four UAAAA motifs. Downstream binding sites were inserted comprising A) UAAAA, and B) UAAAA x 4 combined with the wild type *Nrxn2* sequence. These were designed to be potentially regulated by Sam68. Graphs show the percentage of splicing exclusion (PSE%) in different versions of *Nrxn2* minigene exon co-transfected with GFP-only, T-STAR and Sam68. Data represents the mean of minimum three biological replicates. In each case, statistical significance was compared between HEK293 cells expressing GFP and HEK293 cells expressing the GFP fusion proteins by means of independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.001.

As long as T-STAR splicing repression of 4xUAAAA *Nrxn2* mutant was not strong, I suggested, that the inserted 4xUAAAA sequence combined with UUAA repeats and a UAAAAA motif in the wild type *Nrxn2* sequence could potentially result in stronger splicing repression levels by Sam68. Sam68 is known to bind to RNA as a dimer (Galarneau and Richard, 2009a). In this mutant, the inserted sequence was separated from the wild type *Nrxn*2 sequences by CUAACUAACUAA motif. The hypothesis was, that CUAACUAACUAA sequence will put A/U-rich sequences at a distance from each other, enabling Sam68 requirements for binding as a dimer.

A mutant with 4xUAAAA sequence and the rest of the sequence wild type (containing original regulation sites) was cloned (Figure 5.1 – 4). As before, the new minigene was co-expressed in HEK293 cells, with expression constructs encoding T-STAR and Sam68

fusion proteins, as well as a GFP-only control. As for previously described co-transfections, the cells were harvested, RNA was extracted and reverse transcribed. Capillary gel electrophoresis was performed and splicing patterns were analysed. For this mutant, splicing was highly repressed at endogenous protein levels (reached 81% for GFP-only control) (Figure 5.3 - B). When co-transfected with T-STAR and Sam68, the mutant still appeared to be highly repressed by T-STAR (14% difference between T-STAR and GFP-only control; this difference is statistically significant) but not by Sam68 (even though a 7% difference with GFP control was observed, this appeared to be not statistically significant) (Figure 5.3 - B).

To further investigate sequence requirements, I also manipulated different AU-sequence combinations. According to SELEX, a string of adenine nucleotides (without uracil) is also a target of Sam68 (Galarneau and Richard, 2009a). Thus, a mutant in which the *Nrxn2* sequence was replaced by string of twenty two adenine nucleotides was produced (Figure 5.1 – 5). The aim of this mutant was to test if for *Nrxn2* this could be a better Sam68 target sequence, than those previously tried. However, when Mutant 22xA was analysed, only a small effect of T-STAR was observed (6% between T-STAR and GFP-only control; statistically significant) and no regulation by Sam68 at all (Figure 5.4).

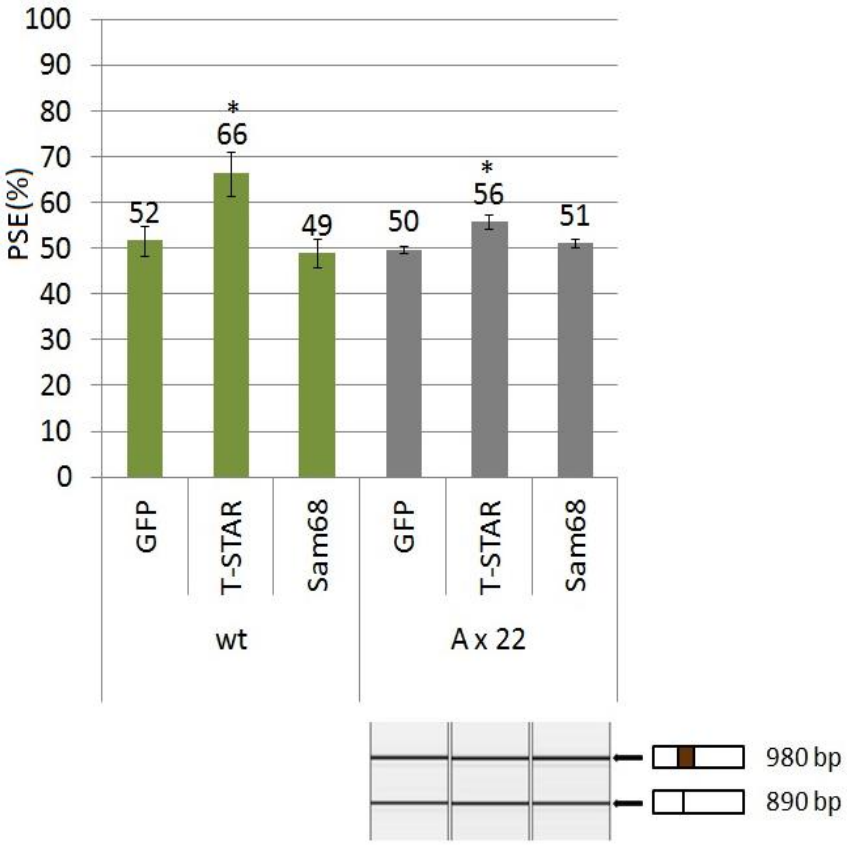


Figure 5.4 – Sam68 does not repress *Nrxn2* through a string of 22 adenine nucleotides. A downstream binding site of 22 adenines was inserted as a potential Sam68 regulatory sequence. Graphs show the percentage of splicing exclusion (PSE%) in different versions of *Nrxn2* A x 22 mutant minigene co-transfected with GFP-only, T-STAR and Sam68. Data represents the mean of minimum three biological replicates. Statistical significance was compared between HEK293 cells expressing GFP and HEK293 cells expressing the GFP fusion proteins by means of independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.001.

5.3.2 Can the Tomosyn-2 downstream response element confer Sam68 regulation on T-STAR?

In Chapter 4 I uncovered the elements in *Tomosyn-2* sequence that respond to both T-STAR and Sam68, from both downstream and upstream intronic locations (labelled 1 and 2 on Figure 5.5). Here I test, if *Nrxn2* responds to Sam68 when replacing the *Nrxn2* sequence by T-STAR and Sam68 regulatory elements from *Tomosyn-2*.

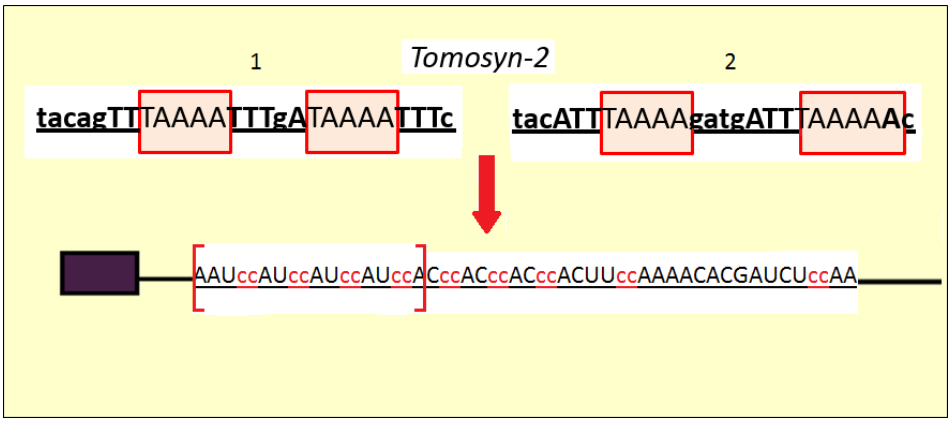


Figure 5.5 – A schematic to show how *Nrxn2* sequence was replaced by *Tomosyn-2* sequences. T-STAR and Sam68 target sites in *Tomosyn-2* (labelled 1 and 2) were inserted into *Nrxn2* target region marked by red brackets. *Nrxn2* mutant with a big mutation was used a template in cloning.

The *Nrxn2* sequence was converted so it contained Sam68 target sequences from *Tomosyn-2*. As known from Chapter 4, *Tomosyn-2* contains two separate regions that were important for Sam68 regulation. Thus, two constructs were produced. Each contained one Sam68 target region from *Tomosyn-2* (Figure 5.5). After co-transfection, T-STAR splicing repression was observed for each construct. Surprisingly, even though the *Tomosyn-2* response regions were transferred into *Nrxn2*, Sam68 still does not regulate splicing of *Nrxn2* (Figure 5.6).

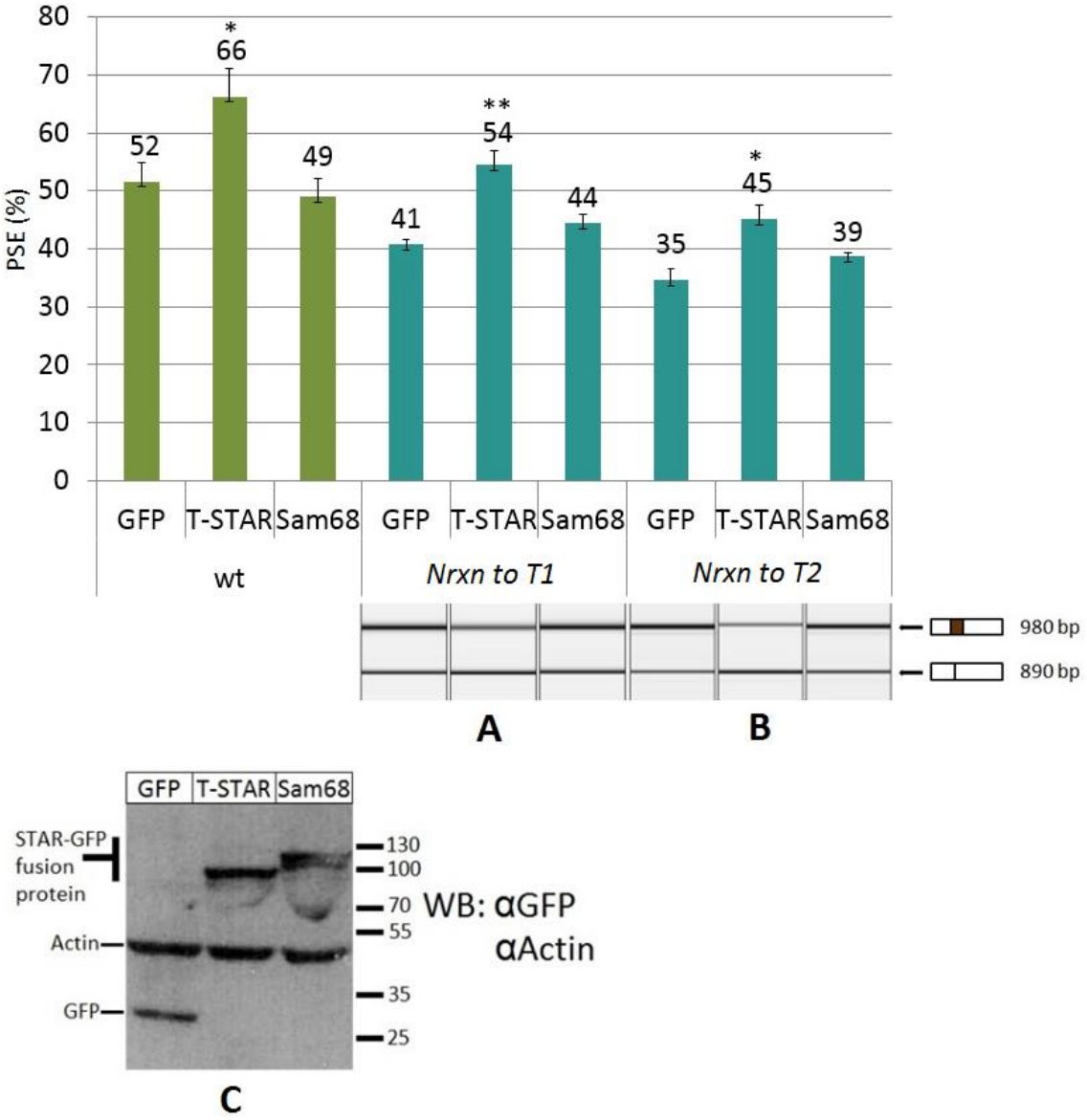


Figure 5.6 – *Nrxn2* does not significantly respond to Sam68 when replacing *Nrxn2* sequence by T-STAR and Sam68 regulatory elements from *Tomosyn-2*. Graphs show the percentage of splicing exclusion in *Nrxn*2 wild type and two mutants with target sequences replaced by the *Tomosyn-2* target A – region 1 and B - region 2. Data represents the mean of minimum three biological replicates. In each case, statistical significance was compared between HEK293 cells expressing GFP and HEK293 cells expressing the GFP fusion proteins by means of independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.001. C - Western blot analysis of HEK293 cells co-transfected with GFP-coupled fusion proteins and *Nrxn2* mutant with target region replaced by site 1 from *Tomosyn-2*. The Western blot was probed for GFP and actin and shows equal expression levels of each fusion protein.

5.3.3 Other STAR proteins also regulate Nrxn2 splicing

To find out if other members of STAR protein family regulate *Nrxn2* splicing, wild type *Nrxn2* was co-transfected with a set of GFP-fusion proteins such as Slm1 (Sam68-like mammalian protein 1), SF1 (splicing factor 1) and QKI (Quaking). GFP-fusion protein plasmids were given to us from our collaborator Prof. Chris Smith (Cambridge) and co-transfected with wild type *Nrxn2*. The GFP-only plasmid was used as a control. Transfections were performed once only, thus the results are preliminary and need replication. SF1 expression did not regulate *Nrxn2* splicing (Figure 5.7). Some splicing repression by Slm1 was detected. This corresponds with the results from a paper published about *Nrxn2* splicing regulation by Slm1 by Iijima et al 2014 (Iijima *et al.*, 2014). The biggest effect was visible from *Nrxn2* co-transfection with QKI protein (Figure 5.7). To test, if QKI represses *Nrxn2* AS4 splicing through the same downstream sequence as T-STAR, QKI was co-transfected with the *Nrxn2* mutant having the big mutation. As before, the big mutation showed half-reduced *Nrxn2* splicing repression at endogenous protein levels. Furthermore, the splicing repression by QKI was completely abolished, suggesting that QKI regulatory motifs are situated within the sequence 12 nucleotides downstream of *Nrxn2* AS4, the same as T-STAR regulatory motifs.

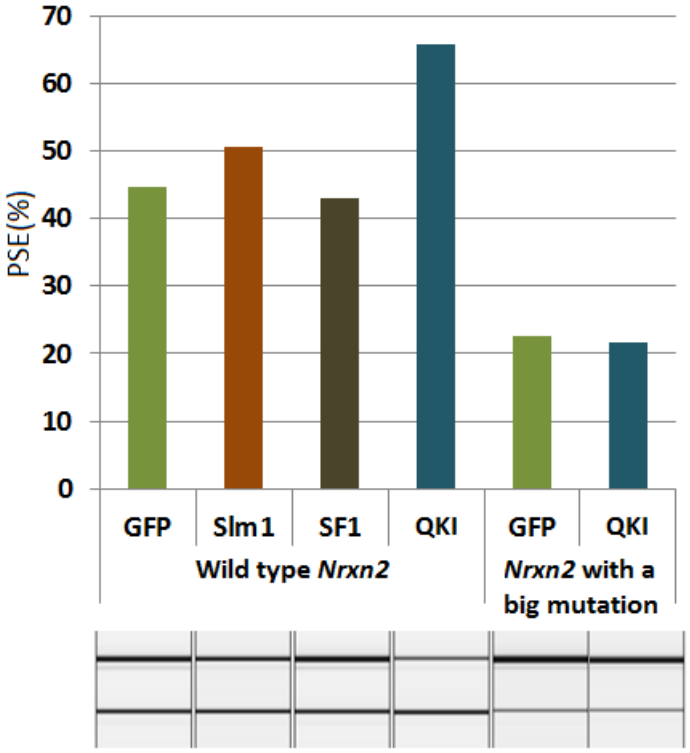


Figure 5.7 – QKI represses the splicing of *Nrxn2* AS4 from the downstream intronic location. Graphs show the percentage of splicing exclusion in wild type *Nrxn2* co-transfections with Slm1, SF1 and QKI proteins, as well as in a co-transfection of *Nrxn*2 mutant with a big mutation with QKI. Data represents a single experiments, thus results are preliminary. More replicates is needed to calculate statistical significance.

Thus, I further focused on studying splicing regulation by QKI protein. To start with, the *Nrxn2* co-transfection with QKI was repeated in triplicate (Figure 5.8 - A) and *Nrxn2* was confirmed to be highly repressed by QKI (39% difference compared to GFP-only control). Expression levels of transfected proteins were assessed by Western blot analysis (Figure 5.8 – B). Western blot confirmed the expression of the GFP-only and QKI fusion constructs.

Subsequently, I decided to look for QKI regulation site in *Nrxn2* minigene. Based on previous studies, the most common predicted binding motifs for QKI are 5'-NACUAAY-$N_{1–20}$-UAAY-3', 5'-NA(A/C)UUA-3' or 5'YAAY-3' (where N is any base, and Y is a pyrimidine) (Ryder and Williamson, 2004; Galarneau and Richard, 2005). These mentioned sequences exactly overlap with the already identified T-STAR responding sequence of *Nrxn2* (Figure 5.8 - A, sequences marked in blue and purple).

Using minigene analyses, I aimed to test if QKI regulates *Nrxn2* splicing though the mentioned sequences. To test whether QKI binds to the sequence mutated previously in

Nrxn2 Mutant 7 (Chapter 3), this mutant was co-transfected with QKI. As before, a reduction of splicing repression of Mutant 7 was observed with endogenous protein levels in HEK293 cells.

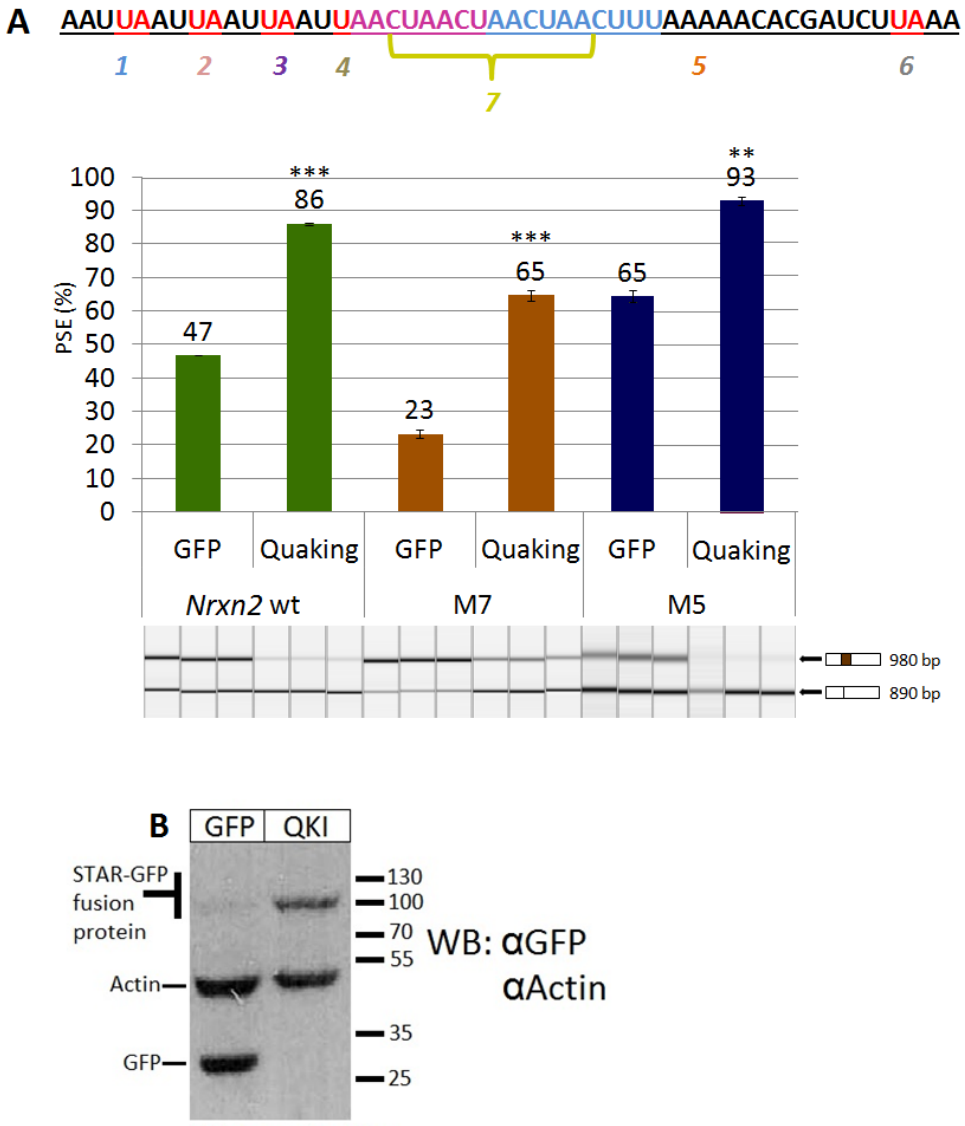


Figure 5.8 – QKI regulation of individual motifs in *Nrxn2* AS4 is similar to T-STAR regulation of these motifs. A - Graphs show the percentage of splicing exclusion (PSE%) in different versions of *Nrxn2* minigene co-transfected with GFP-only and QKI. Data represents the mean of minimum three biological replicates. In each case, statistical significance was compared between HEK293 cells expressing GFP and HEK293 cells expressing the GFP fusion proteins by means of independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.001. B - Western blot analysis of HEK293 cells co-transfected with GFP-coupled QKI and wild type *Nrxn2* minigene. The Western blot was probed for GFP and actin and shows equal expression levels of each fusion protein.

However, as seen from the graphs on Figure 5.8 - A, for Mutant 7 where the CUAACUAACUAA sequence mutated, QKI still represses *Nrxn2* splicing. Thus, this sequence represents a *Nrxn2* splicing repressor element, important for repression of AS4, but not responding to QKI.

*Nrxn*2 Mutant 5 also destroys a possible QKI regulatory site (as it has UAAAAA motif mutated into ccAAAA). Thus, Mutant 5 was also co-transfected with QKI. However, it was still strongly repressed by QKI. Thus, there should be other sites in a downstream intronic sequence of *Nrxn2* for *Nrxn2* splicing repression by QKI.

5.4 Discussion

5.4.1 Can Sam68 regulate Nrxn2 splicing if the sequence is altered?

The first aim of this chapter was to test whether *Nrxn2* is regulated by Sam68 when the *Nrxn2* sequence is changed to be a better Sam68 target.

From the results in Chapter 3, Sam68 did not regulate *Nrxn2* splicing through wild type UUAA motifs or the UAAAAA sequence downstream the exon. Based on the known data regarding the best sequence motifs for Sam68 binding and splicing regulation, I tested seven different sequence combinations, separately replacing wild type sequences by sequence versions that are expected to be Sam68 targets.

Minigene analyses showed that all examined sequences were regulated by T-STAR much stronger, than by Sam68. Thus, all motifs tested in *Nrxn2* minigene appeared to be better targets for T-STAR, rather than for Sam68. *Nrxn2* was twice higher repressed by T-STAR when sequence was replaced by 8xUAAA, than through the wild type T-STAR response element in the *Nrxn2* sequence.

Interestingly, the motif that did not contain any uracil, but just a string of 22 adenines, provoked only a very small splicing repression by T-STAR (6% higher comparing to GFP-only control) in comparison with all AU-rich combinations tested. Thus, adenine motifs without uracil are very moderate targets for T-STAR.

Although wild type *Nrxn2* did not respond at all to Sam68, out of all sequences tested, 8xUAAA (with UAAAAA inside) enabled some repression by Sam68. Some repression by Sam68 was also observed for the 4xUAAAA mutant. No significant effect of Sam68

was seen in the Mutants 4xUAAA, 22xA or 4xAAAA with the rest of sequence wild type, as well as in the Mutants where *Nrxn2* sites were replaced by *Tomosyn-2* sites.

To make *Nrxn2* sequence respond to Sam68 better, further sequence alterations could be performed in future. Currently, I have identified that Sam68 starts to regulate *Nrxn2* splicing through UAAA/UAAAA motifs. I suggest that for Sam68 regulation, motifs with three adenines need to be repeated more than motifs with four adenines. Possibly, the sequence with more adenines could provoke a stronger *Nrxn2* regulation by Sam68. To find out an optimal number of adenines in each motif, by site-directed mutagenesis, a sequence with a single uracil and five adenines (the highest number tested in a current study) can be inserted into *Nrxn2* downstream intron. In each new mutant, one adenine should be added to the mentioned sequence, and the effect of mutation on *Nrxn2* splicing regulation by Sam68 should be assessed. The sequence with highest repression pattern might be considered as a Sam68 consensus motif with an optimal number of adenines.

I found that Sam68 regulates *Nrxn2* splicing through the continuous string of Sam68-like motif repeats. Currently, I tested combinations of four and eight repeats. It would be interesting to find out an optimal number of repeats in a motif, promoting the higher splicing repression pattern by Sam68, than currently observed. By site-directed mutagenesis, different numbers of repeats can be inserted, increasing the length by one motif per each mutant. The sequence with highest repression pattern might be considered as a Sam68 consensus sequence of an optimal length.

5.4.2 Will Sam68 regulate Nrxn2 splicing if the Nrxn2 sequence is replaced by the downstream region of Tomosyn-2?

The next aim of this chapter was to investigate whether placing Sam68 responsive motifs from *Tomosyn-2* into *Nrxn2* can put *Nrxn2* AS4 under Sam68 control. To reach this aim, I replaced the *Nrxn2* sequence by Sam68 regulatory motifs from *Tomosyn-2*. However, when *Tomosyn-2* motifs together with flanking sequences were transferred into *Nrxn2*, Sam68 still did not regulate *Nrxn2* splicing.

One explanation can be, that *Tomosyn-2* splicing repression is mediated by both groups of regulatory sites together. Thus, in future, two or more response regions can be inserted into *Nrxn2* sequence together, what might give a stronger repression pattern by Sam68.

It is also possible, that *Nrxn2* exon has strong splice sites, and thus needs T-STAR to regulate it. *Tomosyn-2* exon might have weaker splice sites than in *Nrxn2*, thus it needs both T-STAR and Sam68 to regulate its splicing. That would explain, why even with *Tomosyn-2* target sites inserted into *Nrxn2*, *Nrxn2* splicing is still not regulated by Sam68. This hypothesis should be tested in a future study.

5.4.3 Nrxn2 AS4 splicing is repressed by Quaking

The last aim of this Chapter was to find out, if there are other proteins regulating *Nrxn2* splicing. Specifically, I was interested to see if any other proteins regulate splicing through the T-STAR and Sam68 binding sequence downstream of *Nrxn2* AS4. That would give more information on which other factors could possibly interact with/control T-STAR and possibly suppress Sam68 from function.

It is known that STAR proteins are highly related to each other by structure and by function (Ryder and Massi, 2010). It was found before that T-STAR and Sam68 proteins interact with each other (Venables *et al.*, 1999). I supposed that other members of the STAR protein family might be also involved in *Nrxn2* splicing regulation. Thus, *Nrxn2* was co-transfected with SLM1, SF1 and QKI GFP-fusion proteins. Some splicing repression was observed with SLM1. However, from my provisional data this effect of SLM1 on splicing was not high. Furthermore, during the time of my study, another group has published their research (Iijima *et al.*, 2014) suggesting splicing regulation of *Nrxn2* by SLM1 in more detail. Based on my preliminary data, SF1 did not appear to regulate *Nrxn2* splicing. However, I observed a strong repression of *Nrxn2* by QKI fusion protein.

Quaking is a STAR protein with its protein domains very similar to T-STAR and Sam68 (Ryder and Massi, 2010). Interestingly, QKI is known to be involved in neural development and myelin formation. Thus, the splicing repression of *Nrxn2* (that is also expressed in neurons) by QKI could have an important effect on the neural function. Mice lacking QKI fail to produce a compact myelin sheet and develop tremors in movement (Ebersole *et al.*, 1996). QKI was also shown to be involved in cancers, including oral cancers (Lu *et al.*, 2014). One of the few previously identified RNA targets for QKI is *MAG* (myelin-associated glycoprotein), where QKI acts as splicing repressor of *MAG* exon 12. Interestingly, similarly to T-STAR and Sam68, Quaking represses *MAG* splicing through regulatory sites in the downstream intron (Wu *et al.*, 2002).

When QKI was co-transfected with the *Nrxn2* mutant with a big mutation, splicing repression by QKI was abolished. Thus, QKI is repressing AS4 splicing through the same downstream intronic sequence region as T-STAR. When QKI was further co-transfected with other *Nrxn2* mutants, a similar regulation pattern as for T-STAR was observed. Similarly to T-STAR, QKI still regulated Mutant 7, with the destroyed repetitive CUAACUAACUAA sequence. Thus, it is likely that QKI and T-STAR repress *Nrxn2* splicing through same motifs. Further study could continue dissecting the intronic sequence region downstream of AS4 to identify QKI regulatory sequences and to investigate the mechanism of splicing regulation by QKI.

5.5 Chapter summary

In this Chapter I show that Sam68 starts repressing *Nrxn2* splicing when the *Nrxn2* sequence is replaced by a different A/U-rich regulatory element, however the repression is not strong, compared to the effect of T-STAR. I also found that all sequences designed to make *Nrxn2* respond to Sam68 respond better to T-STAR than to Sam68. I also found out that single Sam68 target regions from *Tomosyn-2* do not respond to Sam68 when inserted into *Nrxn2*. Presumably, *Tomosyn-2* target sequences should be inserted together, to achieve the splicing repression by Sam68. I also showed in this Chapter, that *Nrxn2* splicing is highly repressed by Quaking protein. Quaking regulatory sites appeared to overlap with the T-STAR response element in *Nrxn2*.

Chapter 6: General discussion and future work

6.1 General discussion

Alternative splicing is an important mechanisms of pre-mRNA processing, regulated by many splicing factors. Aberrant splicing in just a single gene can lead to serious consequences in the cell and whole organism. Cancers, neurological conditions, muscular diseases, and infertility cases are linked to abnormal alternative splicing. To keep a strict control over splicing regulation, each gene has certain rules and mechanisms. To understand the principles behind splicing regulation, it is crucial to identify and study RNA-protein interactions on the resolution of the whole transcriptome. Using the data sets from several transcriptome-wide studies, RNA splicing maps have been built. These show position-dependent effects of protein-RNA interactions (Witten and Ule, 2011). For some RBPs, such as hnRNP proteins, existing splicing maps are already extensive and provide important information about the mechanism of splicing (Huelga *et al.*, 2012). However, for some splicing factors even the general principles of splicing regulations are still unknown, as no splicing targets have yet been identified.

In my PhD I studied the splicing regulation by T-STAR protein of two recently identified physiological targets, *Nrxn2* and *Tomosyn-2*. These were the first physiological splicing targets identified for T-STAR protein. It gave me a unique opportunity to be first person to uncover the principles of T-STAR splicing regulation.

By EMSA experiments I confirmed, that T-STAR is regulating *Nrxn2* splicing through a direct binding. Once I knew that *Nrxn2* is a direct target of T-STAR, I aimed to dissect a binding region for T-STAR on *Nrxn2* RNA. A previous study by Ehrmann et al 2013 uncovered a sequence element downstream of the *Nrxn2* exon as a region involved in T-STAR regulation. I performed sequence analysis of the element, and identified three groups of motifs: one UAAAAA motif, five UUAA repeats and one motif with three CUAA repeats. By site directed mutagenesis, I found that UAAAAA motif is the best T-STAR target in *Nrxn2*. I also found that UUAA repeats together comprise a weaker T-STAR target motif. Individually the UAAAAA sequence and UUAA repeat motif are redundant for splicing regulation. However, the loss of both elements together completely inhibits splicing repression by T-STAR. I also showed, that CUAA sequence repeats do

not respond to T-STAR directly, but comprise a more general element repressing the *Nrxn2* exon.

Studies of Nova, Fox, Mbnl and PTB RBPs splicing factors show, that splicing regulation is dependent on the position where protein binds on RNA (Dredge *et al.*, 2005; Zhang *et al.*, 2008a; Goers *et al.*, 2010; Llorian *et al.*, 2010). In my PhD I aimed to find out, if splicing regulation by T-STAR is dependent on its binding position. In wild type *Nrxn2*, T-STAR binds 12 nucleotides downstream of the exon. Moving the T-STAR binding region by site directed mutagenesis by 100, 150 and 250 nucleotides downstream of the target exon I found, that T-STAR still represses splicing efficiently, even with the binding sites being further downstream in the intron. By moving the T-STAR target region 42 and 100 nucleotides upstream the target exon, I found that T-STAR is also efficiently repressing splicing with binding sites placed upstream the exon. Thus, by my research I found T-STAR splicing regulation of *Nrxn2* is position-independent relative to the regulated exon.

Typically, proteins that bind upstream of exons inhibit splicing through inhibiting the branch point sequence. Proteins that bind downstream activate splicing through activating U1 snRNP spliceosome component. Different from other factors, T-STAR is binding downstream of *Nrxn2* exon and is repressing splicing. Thus, it was of interest to identify what the mechanism of regulation by T-STAR is. Our initial hypothesis was that T-STAR binding might directly inhibit U1 snRNP association with the pre-mRNA, thus de-stabilizing spliceosome assembly. However, as T-STAR is repressing splicing even at a 250 nucleotide distance downstream of the exon, splicing repression may be happening by another mechanism. By analysis of sequence upstream *Nrxn2* gene, in addition to downstream target region, I found AU-rich sequences upstream *Nrxn2* exon. Thus, I hypothesize that splicing could be repressed by T-STAR binding from both sides of the exon, and looping out the target exon, preventing spliceosome assembly. The following model has been already found functional for the PTB protein. It was shown by FRET experiments, that RNA binding domains of PTB protein RRM3 and RRM4 bind to 3' and 5' PTB binding sites (polypyrimidine tracts), respectively, on several model RNAs tested. As a result, 3' and 5' end of the RNA come together, and the exon is looped out. U1 snRNP fails to interact with the 5' splice site on the looped part. This prevents the spliceosome assembly leading to splicing repression (Lamichhane *et al.*, 2010). The study by Lamichhane et al 2010 is a good example of how to test the presence of RNA loops by FRET and NMR spectroscopy. My study could also be continued by testing formation of

RNA loops in *Nrxn2* RNA as a result of T-STAR regulation by the techniques mentioned. However, the easier way would be to mutagenize the target *Neurexin-2* binding sites both upstream and downstream of the target exon, what would inhibit that loop production. Hence, splicing will be no longer repressed. T-STAR could also inhibit U1 snRNP indirectly, or interfere with function of enhancers. All these hypotheses should be tested in a future study.

In addition to *Nrxn2* gene, the study by Ehrmann et al 2013 also identified another physiological target of T-STAR, called *Tomosyn-2*. In my PhD, I also focused on studying this T-STAR target, in order to compare the splicing regulation of T-STAR between *Nrxn2* and *Tomosyn2*.

I confirmed *Tomosyn-2* splicing repression by T-STAR using a minigene. Interestingly, I also found *Tomosyn-2* splicing to be highly repressed by another STAR protein, Sam68 that is highly related to T-STAR. Thus, I aimed to dissect the regulatory sites on *Tomosyn-2* RNA needed for splicing regulation by both proteins. By sequence analysis, I found AU-rich motifs as potential binding sites for T-STAR and Sam68 in *Tomosyn-2*. By site directed mutagenesis, I confirmed two groups of UAAAA/UAAAAA motifs downstream *Tomosyn-2* exon as targets for T-STAR and Sam68 splicing regulation. However, when I combined the identified target sites into one mutant, T-STAR and Sam68 were still highly repressing the splicing of *Tomosyn-2*. Thus, the other motifs in *Tomosyn-2* sequence might also regulate *Tomosyn-2* splicing, what is a subject for further study.

In my study I aimed to find, if *Tomosyn-2* splicing repression by T-STAR and Sam68 is position-dependent. *Tomosyn-2* splicing is repressed by T-STAR and Sam68 that regulate *Tomosyn-2* target sites downstream of the exon (38 and 106 nucleotides downstream). When the target sites were moved 54 nucleotides upstream of the *Tomosyn-2* exon, T-STAR and Sam68 appeared to significantly repress splicing, similarly as with target motifs being downstream. Thus, I found that *Tomosyn-2* splicing repression event is position-independent, however, splicing repression is stronger if binding is upstream of the exon. Therefore, one of the interesting findings in my PhD was that T-STAR and Sam68 proteins are unusual in that they repress splicing from either side of the exon.

By Ehrmann et al 2013 it was found, that *Nrxn2* splicing is regionally repressed in mouse brain. In my study, I have examined *Tomosyn-2* splicing in different 11 mouse brain regions. I found out that *Tomosyn-2* repression correlates with T-STAR expression in the brain and is abolished in the T-STAR knockout mouse. Splicing repression is highest in

the forebrain regions, and lower in midbrain and hindbrain. Thus, T-STAR is a regional splicing regulator of *Tomosyn-2*.

Interestingly, previous studies have shown that the *Tomosyn-2* alternative exon that is a current target for T-STAR and Sam68 regulation, codes for a hypervariable domain of Tomosyn-2 protein. This domain defines the fusogenic capacity in the synapse, through regulating Tomosyn-2 pairing with SNARE (soluble n-ethylmaleimide-sensitive-factor attachment receptor) complexes (McNew *et al.*, 1999; Hatsuzawa *et al.*, 2003). Thus, a functional consequence of *Tomosyn-2* splicing repression by T-STAR and Sam68 might be the amount of neurotransmitter released in synapse, and hence, the overall function of synapse. Sam68 has been already shown before to be important for synaptic functioning (Klein *et al.*, 2013). T-STAR has been associated with neurological conditions, as epilepsy (Sugimoto *et al.*, 2001).

If *Tomosyn-2* splicing was shown to be regulated by both T-STAR and Sam68 proteins, the study by Ehrmann et al 2013 found that *Nrxn2* splicing repression is T-STAR specific. This was a surprising finding, as T-STAR and Sam68 are both very similar proteins, sharing same protein domains and structure, as well as a similar AU-rich predicted target motif for RNA binding by both proteins. I investigated why T-STAR and Sam68 proteins have differences in terms of their function in *Nrxn2* splicing. This would also help to explain why both proteins, even being so similar, were kept by evolution. By EMSA analysis I have confirmed that Sam68 is binding the *Nrxn2* target region directly, but less efficiently than T-STAR.

In my PhD I aimed to uncover why *Nrxn2* splicing is not regulated by Sam68. Our hypothesis was that Sam68 fails to repress *Nrxn2* splicing as the target sequence downstream of the exon is not efficiently bound by Sam68. Thus, in my study I converted the *Nrxn2* downstream target region into four different AU-rich sequence combinations, attempting to make it more Sam68-like. Interestingly, I found that Sam68 did repress *Nrxn2* splicing with continuous strings of four to eight UAAA/UAAAA motif repeats. I suggest that for Sam68 regulation, motifs with three adenines need to be repeated more than motifs with four adenines.

Through converting the *Nrxn2* sequence into the sequence that could be regulated by Sam68, I observed that splicing repression by Sam68 was much less than silencing by T-STAR. It is known that Sam68 binds RNA sequence as a dimer (Galarneau and Richard, 2009a). We hypothesized that to be more Sam68-like, AU-rich sequence might need to

be separated by unspecific sequence, to allow Sam68 to regulate splicing as a dimer. Thus, wild type *Nrxn2* sequence was converted into Sam68-like sequence, leaving a gap between target sites. Also, *Nrxn2* sequence was replaced by the Sam68 target motif (together with flanking sequence) from *Tomosyn-2* sequence. However, no response of Sam68 was detected. I suggest that more *Tomosyn-2* motifs might need to be inserted into *Nrxn2* sequence together, to get the Sam68 response.

It is important to note, that in addition to searching for ideal sequence combination for Sam68 binding, I also tested different AU-rich motifs with T-STAR. Interestingly, the strings of UAAA/UAAAA motifs appeared to be very strong T-STAR targets for *Nrxn2* splicing repression (stronger than wild type *Nrxn2* target motif).

To further compare and contrast splicing regulation of *Nrxn2* by T-STAR and Sam68, it was decided to check if any other STAR family proteins regulate *Nrxn2* splicing. Information about other *Nrxn2* splicing regulators would tell more about the regulation of AS4 splicing in *Nrxn2*. The fusion-plasmids with Slm1, SF1 and QKI proteins were kindly received from our collaborator Professor Chris Smith. After a series of co-transfections, I found the strong splicing repression of *Nrxn2* by QKI (Quaking) protein. I also found some moderate repression by Slm1, and no effect by SF1. Thus, my further research was focused on QKI, as a strongest *Nrxn2* splicing repressor out of three proteins tested. As the proposed target sequence for QKI comprised of similar motifs as those in *Nrxn2* target region for T-STAR, it was decided to test some of the previously generated *Nrxn2* mutants with QKI. When co-transfected with *Nrxn2* mutant having a big mutation, *Nrxn2* splicing repression by QKI got abolished. Hence, QKI represses *Nrxn2* splicing through the mentioned downstream sequence region. Thus, QKI appeared to share some principles for splicing regulation with T-STAR.

To conclude, research during my PhD enabled for the very first time to uncover the principles for splicing regulation by T-STAR protein. In Chapter three of my thesis I have dissected the T-STAR target site in *Neurexin-2*. Importantly, I also found *Neurexin-2* splicing regulation by T-STAR to be position-independent. As for *Neurexin-2* in Chapter three, in Chapter four I have uncovered the regulatory sites for T-STAR and Sam68 in *Tomosyn-2*. Similarly as for *Neurexin-2*, splicing regulation of *Tomosyn-2* appeared to be position-independent. I also observed the functional effect of splicing by T-STAR on a tissue level, in mouse brain. It was similar for both *Neurexin-2* and *Tomosyn-2*. Thus, results in Chapters three and four show the similar pattern of splicing regulation for *Neurexin-2* and *Tomosyn-2*, with the main difference that *Tomosyn-2* is controlled by both

T-STAR and Sam68, while *Neurexin-2* regulation is T-STAR specific. In Chapter five I showed that *Neurexin-2* is a target for T-STAR but not for Sam68 due to sequence specificity. Sam68 starts repressing *Neurexin-2* splicing when the *Neurexin-2* sequence is replaced by a different A/U-rich regulatory element. In Chapter 5 I also showed that *Neurexin-2* splicing is highly repressed by Quaking protein. Quaking regulatory sites appeared to overlap with the T-STAR response element in *Neurexin-2*. Hence it is possible that the splicing regulation mechanism by Quaking is similar to that one by T-STAR

6.2 Future work

Some future questions to be answered in future studies and possible experiments to be performed:

- **Which proteins regulate *Nrxn2* splicing repression through CUAA repeat sequence?**

  In a current study CUAA repeat region was detected as a general *Nrxn2* splicing repressor, and sequence affecting splicing repression by T-STAR and Sam68. It would be interesting to find out which proteins regulate this sequence. Experimentally, this question could be answered by pull down assays.
- **Is *Nrxn2* and *Tomosyn-2* splicing position-independent at even further distances downstream and upstream of the target exons?**

  T-STAR regulation of *Nrxn2* splicing was currently found position-independent. However, it is interesting to follow until which downstream and upstream distance in the introns splicing repression will still be occurring normally. That would help us to further uncover positional principles of T-STAR regulation. This question could be answered by cloning new mutants with target sequence being moved by site directed mutagenesis.
- **Which splicing factors does T-STAR interact with?**

  To find out more about the splicing mechanism of T-STAR, it would be interesting to find out which other splicing factors T-STAR might interact with. Interactions of T-STAR with other proteins can be studied in detail by pull-down assays or by protein immunoprecipitation assays.
- **Are *Tomosyn-2* and *Nrxn2* physiological targets for Sam68 and Quaking, respectively?**

Some of the splicing regulation events in my PhD were only assessed in minigenes. It would be useful to know, if *Tomosyn-2* is a real physiological target of Sam68 and Quaking, as well as if *Nrxn2* is a target for QKI protein in a real tissue. To answer this, Sam68 and QKI should be knocked out in a mouse, or knocked down in cells, and splicing pattern of *Nrxn2* and *Tomosyn-2* should be assessed in a knockout, in comparison to wild type.

- **Does QKI repress *Nrxn2* splicing directly? Do T-STAR and Sam68 repress *Tomosyn-2* splicing directly?**

  Is *Nrxn2* splicing repression by QKI as well as *Tomosyn-2* repression by T-STAR and Sam68 direct or indirect? To test this, *Tomosyn-2* target regions should be cloned into a vector and EMSA essay with T-STAR and Sam68 (or *Nrxn2* region with QKI) protein should be performed.

- **Does QKI regulate *Tomosyn-2* splicing?**

  It would be interesting to test if *Tomosyn-2* splicing is also repressed by QKI. That would give us an idea if T-STAR, Sam68 and QKI are together involved in regulation of other targets apart from *Nrxn2*, and would help to better understand the interplay between these proteins.

- **What are the other target sites for T-STAR and Sam68 splicing regulation in *Tomosyn-2*?**

  By site-directed mutagenesis I identified two groups of target motifs in *Tomosyn-2*. Blocking them reduced the splicing repression by T-STAR and Sam68. However, it did not abolish the splicing repression by T-STAR and Sam68 completely. It would be interesting to see through which other sites to T-STAR and Sam68 regulate *Tomosyn-2* splicing. For this, other potential target sites should be mutagenized, and the effect of mutations on splicing regulation should be assessed.

- **Is *Nrxn2* repression by T-STAR occurring through RNA looping mechanism?**

  One of our hypotheses is that *Nrxn2* splicing repression by T-STAR might occur through the formation of RNA loops, and hence, preventing the spliceosome assembly. To test this, binding sites from both downstream and upstream of the *Neurexin-2* exon should be mutagenized together.

Bibliography

Ameur, A., Zaghlool, A., Halvardson, J., Wetterbom, A., Gyllensten, U., Cavelier, L. and Feuk, L. (2011) 'Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain', *Nat Struct Mol Biol*, 18(12), pp. 1435-1440.

Anthony, K. and Gallo, J.-M. (2010) 'Aberrant RNA processing events in neurological disorders', *Brain Research*, 1338(0), pp. 67-77.

Araç, D., Boucard, A.A., Özkan, E., Strop, P., Newell, E., Südhof, T.C. and Brunger, A.T. (2007) 'Structures of Neuroligin-1 and the Neuroligin-1/Neurexin-1β Complex Reveal Specific Protein-Protein and Protein-Ca2+ Interactions', *Neuron*, 56(6), pp. 992-1003.

Artzt, K. and Wu, J.I. (2010) 'STAR trek: An introduction to STAR family proteins and review of quaking (QKI)', *Adv Exp Med Biol*, 693, pp. 1-24.

Ascano, M., Hafner, M., Cekan, P., Gerstberger, S. and Tuschl, T. (2012) 'Identification of RNA–protein interaction networks using PAR-CLIP', *Wiley interdisciplinary reviews. RNA*, 3(2), pp. 159-177.

Babic, I., Cherry, E. and Fujita, D.J. (2006) 'SUMO modification of Sam68 enhances its ability to repress cyclin D1 expression and inhibits its ability to induce apoptosis', *Oncogene*, 25(36), pp. 4955-4964.

Babic, I., Jakymiw, A. and Fujita, D.J. (2004) 'The RNA binding protein Sam68 is acetylated in tumor cell lines, and its acetylation correlates with enhanced RNA binding activity', *Oncogene*, 23(21), pp. 3781-3789.

Baralle, D., Lucassen, A. and Buratti, E. (2009) *Missed threads*.

Bedford, M.T., Frankel, A., Yaffe, M.B., Clarke, S., Leder, P. and Richard, S. (2000) 'Arginine Methylation Inhibits the Binding of Proline-rich Ligands to Src Homology 3, but Not WW, Domains', *Journal of Biological Chemistry*, 275(21), pp. 16030-16036.

Bell, M.V., Cowper, A.E., Lefranc, M.-P., Bell, J.I. and Screaton, G.R. (1998) 'Influence of Intron Length on Alternative Splicing of CD44', *Molecular and Cellular Biology*, 18(10), pp. 5930-5941.

Bentley, D.L. (2014) 'Coupling mRNA processing with transcription in time and space', *Nat Rev Genet*, 15(3), pp. 163-175.

Berget, S.M. (1995) 'Exon Recognition in Vertebrate Splicing', *Journal of Biological Chemistry*, 270(6), pp. 2411-2414.

Bian, Y., Wang, L., Lu, H., Yang, G., Zhang, Z., Fu, H., Lu, X., Wei, M., Sun, J., Zhao, Q., Dong, G. and Lu, Z. (2012) 'Downregulation of tumor suppressor QKI in gastric cancer and its implication in cancer prognosis', *Biochemical and Biophysical Research Communications*, 422(1), pp. 187-193.

Bielli, P., Busà, R., Paronetto, M.P. and Sette, C. (2011) 'The RNA-binding protein Sam68 is a multifunctional player in human cancer', *Endocrine-Related Cancer*, 18(4), pp. R91-R102.

Black, D.L. (2000) 'Protein Diversity from Alternative Splicing: A Challenge for Bioinformatics and Post-Genome Biology', *Cell*, 103(3), pp. 367-370.

Black, D.L. (2003) 'MECHANISMS OF ALTERNATIVE PRE-MESSENGER RNA SPLICING', *Annual Review of Biochemistry*, 72(1), pp. 291-336.

Blanchette, M. and Chabot, B. (1999) 'Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization', *The EMBO Journal*, 18(7), pp. 1939-1952.

Bor, Y., Swartz, J., Morrison, A., Rekosh, D., Ladomery, M. and Hammarskjöld, M.-L. (2006) 'The Wilms' tumor 1 (WT1) gene (+KTS isoform) functions with a CTE to enhance translation from an unspliced RNA with a retained intron', *Genes & Development*, 20(12), pp. 1597-1608.

Brown, S.J., Stoilov, P. and Xing, Y. (2012) 'Chromatin and epigenetic regulation of pre-mRNA processing', *Human Molecular Genetics*, 21(R1), pp. R90-R96.

Buchman, A.R. and Berg, P. (1988) 'Comparison of intron-dependent and intron-independent gene expression', *Molecular and Cellular Biology*, 8(10), pp. 4395-4405.

Burge, C.B., Padgett, R.A. and Sharp, P.A. (1998) 'Evolutionary Fates and Origins of U12-Type Introns', *Molecular Cell*, 2(6), pp. 773-785.

Burgess, S.M. and Guthrie, C. (1993) 'Beat the clock: paradigms for NTPases in the maintenance of biological fidelity', *Trends in Biochemical Sciences*, 18(10), pp. 381-384.

Carmel, L. and Chorev, M. (2012) 'The function of introns', *Frontiers in Genetics*, 3.

Chan, R.C. and Black, D.L. (1997) 'The polypyrimidine tract binding protein binds upstream of neural cell-specific c-src exon N1 to repress the splicing of the intron downstream', *Molecular and Cellular Biology*, 17(8), pp. 4667-4676.

Chawla, G., Lin, C.-H., Han, A., Shiue, L., Ares, M. and Black, D.L. (2009) 'Sam68 Regulates a Set of Alternatively Spliced Exons during Neurogenesis', *Molecular and Cellular Biology*, 29(1), pp. 201-213.

Chen, T. and Richard, S. (1998) 'Structure-Function Analysis of Qk1: a Lethal Point Mutation in Mouse quaking Prevents Homodimerization', *Molecular and Cellular Biology*, 18(8), pp. 4863-4871.

Cheng, C. and Sharp, P.A. (2006) 'Regulation of CD44 Alternative Splicing by SRm160 and Its Potential Role in Tumor Cell Invasion', *Molecular and Cellular Biology*, 26(1), pp. 362-370.

Côté, J., Boisvert, F.o.-M., Boulanger, M.-C., Bedford, M.T. and Richard, S. (2003) 'Sam68 RNA Binding Protein Is an In Vivo Substrate for Protein Arginine N-Methyltransferase 1', *Molecular Biology of the Cell*, 14(1), pp. 274-287.

Courtney, E., Kornfeld, S., Janitz, K. and Janitz, M. (2010) 'Transcriptome profiling in neurodegenerative disease', *Journal of Neuroscience Methods*, 193(2), pp. 189-202.

Darnell, R.B. (2013) 'RNA Protein Interaction in Neurons', *Annual review of neuroscience*, 36, pp. 243-270.

de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M.a., Pelisch, F., Cramer, P., Bentley, D. and Kornblihtt, A.R. (2003) 'A Slow RNA Polymerase II Affects Alternative Splicing In Vivo', *Molecular Cell*, 12(2), pp. 525-532.

Derry, J.J., Richard, S., Valderrama Carvajal, H., Ye, X., Vasioukhin, V., Cochrane, A.W., Chen, T. and Tyner, A.L. (2000) 'Sik (BRK) Phosphorylates Sam68 in the Nucleus and Negatively Regulates Its RNA Binding Ability', *Molecular and Cellular Biology*, 20(16), pp. 6114-6126.

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., Roder, 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., Guigo, R. and Gingeras, T.R. (2012) 'Landscape of transcription in human cells', *Nature*, 489(7414), pp. 101-108.

Dredge, B.K., Stefani, G., Engelhard, C.C. and Darnell, R.B. (2005) *Nova autoregulation reveals dual functions in neuronal splicing*.

Ebersole, T.A., Chen, Q., Justice, M.J. and Artzt, K. (1996) 'The quaking gene product necessary in embryogenesis and myelination combines features of RNA binding and signal transduction proteins', *Nat Genet*, 12(3), pp. 260-265.

Ehrmann, I., Dalgliesh, C., Liu, Y., Danilenko, M., Crosier, M., Overman, L., Arthur, H.M., Lindsay, S., Clowry, G.J., Venables, J.P., Fort, P. and Elliott, D.J. (2013) 'The Tissue-Specific RNA Binding Protein T-STAR Controls Regional Splicing Patterns of Neurexin Pre-mRNAs in the Brain', *PLoS Genetics*, 9(4), p. e1003474.

Elliott, D.J. and Rajan, P. (2010) 'The role of the RNA-binding protein Sam68 in mammary tumourigenesis', *The Journal of Pathology*, 222(3), pp. 223-226.

Estes, P.A., Cooke, N.E. and Liebhaber, S.A. (1992) 'A native RNA secondary structure controls alternative splice-site selection and generates two human growth hormone isoforms', *Journal of Biological Chemistry*, 267(21), pp. 14902-8.

Fabrizio, P. and Lührmann, R. (2012) 'The Spliceosome in Constitutive Splicing', in *Alternative pre-mRNA Splicing*. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 49-64.

Feracci, M., Foot, J. and Dominguez, C. (2014) 'Structural investigations of the RNA-binding properties of STAR proteins', *Biochem Soc Trans*, 42(4), pp. 1141-6.

Foot, J.N., Feracci, M. and Dominguez, C. (2014) 'Screening protein – Single stranded RNA complexes by NMR spectroscopy for structure determination()', *Methods (San Diego, Calif.)*, 65(3), pp. 288-301.

Förch, P., Puig, O., Martínez, C., Séraphin, B. and Valcárcel, J. (2002) 'The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5′ splice sites', *The EMBO Journal*, 21(24), pp. 6882-6892.

Fox-Walsh, K.L., Dou, Y., Lam, B.J., Hung, S.-p., Baldi, P.F. and Hertel, K.J. (2005) 'The architecture of pre-mRNAs affects mechanisms of splice-site pairing', *Proceedings of the National Academy of Sciences of the United States of America*, 102(45), pp. 16176-16181.

Fumagalli, S., Totty, N.F., Hsuan, J.J. and Courtneidge, S.A. (1994) 'A target for Src in mitosis', *Nature*, 368(6474), pp. 871-874.

Galarneau, A. and Richard, S. (2005) 'Target RNA motif and target mRNAs of the Quaking STAR protein', *Nat Struct Mol Biol*, 12(8), pp. 691-698.

Galarneau, A. and Richard, S. (2009a) 'The STAR RNA binding proteins GLD-1, QKI, SAM68 and SLM-2 bind bipartite RNA motifs', *BMC Molecular Biology*, 10, pp. 47-47.

Galarneau, A. and Richard, S. (2009b) 'The STAR RNA binding proteins GLD-1, QKI, SAM68 and SLM-2 bind bipartite RNA motifs', *BMC Molecular Biology*, 10(1), p. 47.

Garcia-Blanco, M.A. (2005) 'Making antisense of splicing', *Curr Opin Mol Ther*, 7(5), pp. 476-82.

Geerts, C., Plomp, J., Koopmans, B., Loos, M., van der Pijl, E., van der Valk, M., Verhage, M. and Groffen, A.A. (2014) 'Tomosyn-2 is required for normal motor performance in mice and sustains neurotransmission at motor endplates', *Brain Structure and Function*, pp. 1-12.

Gerstein, M.B., Bruce, C., Rozowsky, J.S., Zheng, D., Du, J., Korbel, J.O., Emanuelsson, O., Zhang, Z.D., Weissman, S. and Snyder, M. (2007) 'What is a gene, post-ENCODE? History and updated definition', *Genome Research*, 17(6), pp. 669-681.

Glisovic, T., Bachorik, J.L., Yong, J. and Dreyfuss, G. (2008) 'RNA-binding proteins and post-transcriptional gene regulation', *FEBS Letters*, 582(14), pp. 1977-1986.

Goers, E.S., Purcell, J., Voelker, R.B., Gates, D.P. and Berglund, J.A. (2010) 'MBNL1 binds GC motifs embedded in pyrimidines to regulate alternative splicing', *Nucleic Acids Research*, 38(7), pp. 2467-2484.

Goren, A., Ram, O., Amit, M., Keren, H., Lev-Maor, G., Vig, I., Pupko, T. and Ast, G. (2006) 'Comparative Analysis Identifies Exonic Splicing Regulatory Sequences—The Complex Definition of Enhancers and Silencers', *Molecular Cell*, 22(6), pp. 769-781.

Grabowski, P. (2011) 'Alternative splicing takes shape during neuronal development', *Current Opinion in Genetics & Development*, 21(4), pp. 388-394.

Graham, F.L., Smiley, J., Russell, W.C. and Nairn, R. (1977) 'Characteristics of a Human Cell Line Transformed by DNA from Human Adenovirus Type 5', *Journal of General Virology*, 36(1), pp. 59-72.

Graveley, B.R., Hertel, K.J. and Maniatis, T. (1998) *A systematic analysis of the factors that determine the strength of pre‐mRNA splicing enhancers*.

Gregory, T.R. (2005) 'Synergy between sequence and size in Large-scale genomics', *Nat Rev Genet*, 6(9), pp. 699-708.

Grellscheid, S., Dalgliesh, C., Storbeck, M., Best, A., Liu, Y., Jakubik, M., Mende, Y., Ehrmann, I., Curk, T., Rossbach, K., Bourgeois, C.F., Stévenin, J., Grellscheid, D., Jackson, M.S., Wirth, B. and Elliott, D.J. (2011) 'Identification of Evolutionarily Conserved Exons as Regulated Targets for the Splicing Activator Tra2β in Development', *PLoS Genetics*, 7(12), p. e1002390.

Groffen, A.J.A., Jacobsen, L., Schut, D. and Verhage, M. (2005) 'Two distinct genes drive expression of seven tomosyn isoforms in the mammalian brain, sharing a conserved structure with a unique variable domain', *Journal of Neurochemistry*, 92(3), pp. 554-568.

Grosso, A.R., Gomes, A.Q., Barbosa-Morais, N.L., Caldeira, S., Thorne, N.P., Grech, G., von Lindern, M. and Carmo-Fonseca, M. (2008) 'Tissue-specific splicing factor gene expression signatures', *Nucleic Acids Research*, 36(15), pp. 4823-4832.

Hall, M.P., Nagel, R.J., Fagg, W.S., Shiue, L., Cline, M.S., Perriman, R.J., Donohue, J.P. and Ares, M. (2013) 'Quaking and PTB control overlapping splicing regulatory networks during muscle cell differentiation', *RNA*, 19(5), pp. 627-638.

Hardy, R.J. (1998) 'Molecular defects in the dysmyelinating mutant quaking', *Journal of Neuroscience Research*, 51(4), pp. 417-422.

Hatsuzawa, K., Lang, T., Fasshauer, D., Bruns, D. and Jahn, R. (2003) 'The R-SNARE Motif of Tomosyn Forms SNARE Core Complexes with Syntaxin 1 and SNAP-25 and Down-regulates Exocytosis', *Journal of Biological Chemistry*, 278(33), pp. 31159-31166.

Hertel, K.J. (2008) 'Combinatorial Control of Exon Recognition', *Journal of Biological Chemistry*, 283(3), pp. 1211-1215.

Hicks, M.J., Lam, B.J. and Hertel, K.J. (2005) 'Analyzing mechanisms of alternative pre-mRNA splicing using in vitro splicing assays', *Methods*, 37(4), pp. 306-313.

Hong, M., Zhukareva, V., Vogelsberg-Ragaglia, V., Wszolek, Z., Reed, L., Miller, B.I., Geschwind, D.H., Bird, T.D., McKeel, D., Goate, A., Morris, J.C., Wilhelmsen, K.C., Schellenberg, G.D., Trojanowski, J.Q. and Lee, V.M.Y. (1998) 'Mutation-Specific Functional Impairments in Distinct Tau Isoforms of Hereditary FTDP-17', *Science*, 282(5395), pp. 1914-1917.

Hook, B., Bernstein, D., Zhang, B. and Wickens, M. (2005) 'RNA–protein interactions in the yeast three-hybrid system: Affinity, sensitivity, and enhanced library screening', *RNA*, 11(2), pp. 227-233.

Huelga, S.C., Vu, A.Q., Arnold, J.D., Liang, T.Y., Liu, P.P., Yan, B.Y., Donohue, J.P., Shiue, L., Hoon, S., Brenner, S., Ares, M. and Yeo, G.W. (2012) 'Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins', *Cell reports*, 1(2), pp. 167-178.

Hui, J., Hung, L.H., Heiner, M., Schreiner, S., Neumüller, N., Reither, G., Haas, S.A. and Bindereif, A. (2005) *Intronic CA‐repeat and CA‐rich elements: a new class of regulators of mammalian alternative splicing*.

Huot, M.-É., Vogel, G., Zabarauskas, A., Ngo, Chau T.-A., Coulombe-Huntington, J., Majewski, J. and Richard, S. (2012) 'The Sam68 STAR RNA-Binding Protein Regulates mTOR Alternative Splicing during Adipogenesis', *Molecular Cell*, 46(2), pp. 187-199.

Hurwitz, J. (2005) 'The Discovery of RNA Polymerase', *Journal of Biological Chemistry*, 280(52), pp. 42477-42485.

Iijima, T., Iijima, Y., Witte, H. and Scheiffele, P. (2014) 'Neuronal cell type–specific alternative splicing is regulated by the KH domain protein SLM1', *The Journal of Cell Biology*, 204(3), pp. 331-342.

Iijima, T., Wu, K., Witte, H., Hanno-Iijima, Y., Glatter, T., Richard, S. and Scheiffele, P. (2011) 'SAM68 regulates neuronal activity-dependent alternative splicing of Neurexin-1', *Cell*, 147(7), pp. 1601-1614.

Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R. and Shoemaker, D.D. (2003) 'Genome-Wide Survey of Human

Alternative Pre-mRNA Splicing with Exon Junction Microarrays', *Science*, 302(5653), pp. 2141-2144.

Juneau, K., Miranda, M., Hillenmeyer, M.E., Nislow, C. and Davis, R.W. (2006) 'Introns Regulate RNA and Protein Abundance in Yeast', *Genetics*, 174(1), pp. 511-518.

Kanopka, A., Muhlemann, O. and Akusjarvi, G. (1996) 'Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA', *Nature*, 381(6582), pp. 535-538.

Kanopka, A., Muhlemann, O., Petersen-Mahrt, S., Estmer, C., Ohrmalm, C. and Akusjarvi, G. (1998) 'Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins', *Nature*, 393(6681), pp. 185-187.

Klein, M.E., Younts, T.J., Castillo, P.E. and Jordan, B.A. (2013) 'RNA-binding protein Sam68 controls synapse number and local β-actin mRNA metabolism in dendrites', *Proceedings of the National Academy of Sciences of the United States of America*, 110(8), pp. 3125-3130.

Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M. and Ule, J. (2011) 'iCLIP - Transcriptome-wide Mapping of Protein-RNA Interactions with Individual Nucleotide Resolution', *Journal of Visualized Experiments : JoVE*, (50), p. 2638.

Kornblihtt, A.R., De La Mata, M., Fededa, J.P., MuÑOz, M.J. and NoguÉS, G. (2004) 'Multiple links between transcription and splicing', *RNA*, 10(10), pp. 1489-1498.

Krueger, D.D., Tuffy, L.P., Papadopoulos, T. and Brose, N. (2012) 'The role of neurexins and neuroligins in the formation, maturation, and function of vertebrate synapses', *Current Opinion in Neurobiology*, 22(3), pp. 412-422.

Lagier-Tourenne, C. and Cleveland, D.W. (2009) 'Rethinking ALS: the FUS about TDP-43', *Cell*, 136(6), pp. 1001-1004.

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(9), pp. 4105-4110.

Lewis, B.P., Green, R.E. and Brenner, S.E. (2003) 'Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans', *Proceedings of the National Academy of Sciences*, 100(1), pp. 189-192.

Liao, W.-T., Liu, J.-L., Wang, Z.-G., Cui, Y.-M., Shi, L., Li, T.-T., Zhao, X.-H., Chen, X.-T., Ding, Y.-Q. and Song, L.-B. (2013) 'High expression level and nuclear localization of Sam68 are associated with progression and poor prognosis in colorectal cancer', *BMC Gastroenterology*, 13, pp. 126-126.

Licatalosi, D.D. and Darnell, R.B. (2006) 'Splicing Regulation in Neurologic Disease', *Neuron*, 52(1), pp. 93-101.

Lin, Q., Taylor, S.J. and Shalloway, D. (1997) 'Specificity and Determinants of Sam68 RNA Binding: IMPLICATIONS FOR THE BIOLOGICAL FUNCTION OF K HOMOLOGY DOMAINS', *Journal of Biological Chemistry*, 272(43), pp. 27274-27280.

Lipscombe, D. (2005) 'Neuronal proteins custom designed by alternative splicing', *Current Opinion in Neurobiology*, 15(3), pp. 358-363.

Liu, Z., Luyten, I., Bottomley, M.J., Messias, A.C., Houngninou-Molango, S., Sprangers, R., Zanier, K., Krämer, A. and Sattler, M. (2001) 'Structural Basis for Recognition of the Intron Branch Site RNA by Splicing Factor 1', *Science*, 294(5544), pp. 1098-1102.

Llorian, M., Schwartz, S., Clark, T.A., Hollander, D., Tan, L.-Y., Spellman, R., Gordon, A., Schweitzer, A.C., de la Grange, P., Ast, G. and Smith, C.W.J. (2010) 'Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB', *Nat Struct Mol Biol*, 17(99), pp. 1114-1123.

Lorson, C.L. and Androphy, E.J. (2000) 'An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene SMN', *Human Molecular Genetics*, 9(2), pp. 259-265.

Lu, W., Feng, F., Xu, J., Lu, X., Wang, S., Wang, L., Lu, H., Wei, M., Yang, G., Wang, L., Lu, Z., Liu, Y. and Lei, X. (2014) 'QKI impairs self-renewal and tumorigenicity of oral cancer cells via repression of SOX2', *Cancer Biology & Therapy*, 15(9), pp. 1174-1184.

Lutz, C.S. (2008) 'Alternative Polyadenylation: A Twist on mRNA 3′ End Formation', *ACS Chemical Biology*, 3(10), pp. 609-617.

Mandal, S.S., Chu, C., Wada, T., Handa, H., Shatkin, A.J. and Reinberg, D. (2004) 'Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II', *Proceedings of the National Academy of Sciences of the United States of America*, 101(20), pp. 7572-7577.

Matlin, A.J., Clark, F. and Smith, C.W.J. (2005) 'Understanding alternative splicing: towards a cellular code', *Nat Rev Mol Cell Biol*, 6(5), pp. 386-398.

Matter, N., Herrlich, P. and Konig, H. (2002) 'Signal-dependent regulation of splicing via phosphorylation of Sam68', *Nature*, 420(6916), pp. 691-695.

McNew, J.A., Weber, T., Engelman, D.M., Söllner, T.H. and Rothman, J.E. (1999) 'The Length of the Flexible SNAREpin Juxtamembrane Region Is a Critical Determinant of SNARE-Dependent Fusion', *Molecular Cell*, 4(3), pp. 415-421.

Mills, J.D. and Janitz, M. (2012) 'Alternative splicing of mRNA in the molecular pathology of neurodegenerative diseases', *Neurobiology of Aging*, 33(5), pp. 1012.e11-1012.e24.

Modafferi, E.F. and Black, D.L. (1997) 'A complex intronic splicing enhancer from the c-src pre-mRNA activates inclusion of a heterologous exon', *Molecular and Cellular Biology*, 17(11), pp. 6537-6545.

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(6444), pp. 364-368.

Neugebauer, K.M. (2002) 'On the importance of being co-transcriptional', *Journal of Cell Science*, 115(20), pp. 3865-3871.

Nilsen, T.W. and Graveley, B.R. (2010) 'Expansion of the eukaryotic proteome by alternative splicing', *Nature*, 463(7280), pp. 457-463.

Niwa, M. and Berget, S.M. (1991) 'Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns', *Genes & Development*, 5(11), pp. 2086-2095.

Nogués, G., Muñoz, M.J. and Kornblihtt, A.R. (2003) 'Influence of Polymerase II Processivity on Alternative Splicing Depends on Splice Site Strength', *Journal of Biological Chemistry*, 278(52), pp. 52166-52171.

Noveroske, J.K., Lai, L., Gaussin, V., Northrop, J.L., Nakamura, H., Hirschi, K.K. and Justice, M.J. (2002) 'Quaking is essential for blood vessel development', *genesis*, 32(3), pp. 218-230.

Pajares, M.J., Ezponda, T., Catena, R., Calvo, A., Pio, R. and Montuenga, L.M. (2007) 'Alternative splicing: an emerging topic in molecular and clinical oncology', *The Lancet Oncology*, 8(4), pp. 349-357.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J. and Blencowe, B.J. (2008) 'Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing', *Nat Genet*, 40(12), pp. 1413-1415.

Paronetto, M.P., Achsel, T., Massiello, A., Chalfant, C.E. and Sette, C. (2007) 'The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x', *The Journal of Cell Biology*, 176(7), pp. 929-939.

Paronetto, M.P., Cappellari, M., Busà, R., Pedrotti, S., Vitali, R., Comstock, C., Hyslop, T., Knudsen, K.E. and Sette, C. (2010) 'Alternative Splicing of the Cyclin D1 Proto-Oncogene Is Regulated by the RNA-Binding Protein Sam68', *Cancer research*, 70(1), pp. 229-239.

Paronetto, M.P., Messina, V., Barchi, M., Geremia, R., Richard, S. and Sette, C. (2011) 'Sam68 marks the transcriptionally active stages of spermatogenesis and modulates alternative splicing in male germ cells', *Nucleic Acids Research*, 39(12), pp. 4961-4974.

Patel, A.A. and Steitz, J.A. (2003) 'Splicing double: insights from the second spliceosome', *Nat Rev Mol Cell Biol*, 4(12), pp. 960-970.

Pedrotti, S., Bielli, P., Paronetto, M.P., Ciccosanti, F., Fimia, G.M., Stamm, S., Manley, J.L. and Sette, C. (2010) 'The splicing regulator Sam68 binds to a novel exonic splicing silencer and functions in SMN2 alternative splicing in spinal muscular atrophy', *The EMBO Journal*, 29(7), pp. 1235-1247.

Pennisi, E. (2005) 'Why Do Humans Have So Few Genes?', *Science*, 309(5731), p. 80.

Rajan, P., Gaughan, L., Dalgliesh, C., El-Sherif, A., Robson, C.N., Leung, H.Y. and Elliott, D.J. (2008) 'The RNA-binding and adaptor protein Sam68 modulates signal-dependent splicing and transcriptional activity of the androgen receptor', *The Journal of Pathology*, 215(1), pp. 67-77.

Reed, R. (1996) 'Initial splice-site recognition and pairing during pre-mRNA splicing', *Current Opinion in Genetics & Development*, 6(2), pp. 215-220.

Robberson, B.L., Cote, G.J. and Berget, S.M. (1990) 'Exon definition may facilitate splice site selection in RNAs with multiple exons', *Molecular and Cellular Biology*, 10(1), pp. 84-94.

Roca, X., Akerman, M., Gaus, H., Berdeja, A., Bennett, C.F. and Krainer, A.R. (2012) 'Widespread recognition of 5′ splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides', *Genes & Development*, 26(10), pp. 1098-1109.

Roca, X., Krainer, A.R. and Eperon, I.C. (2013) 'Pick one, but be quick: 5′ splice sites and the problems of too many choices', *Genes & Development*, 27(2), pp. 129-144.

Ryder, S.P. and Massi, F. (2010) 'Insights into the structural basis of RNA recognition by STAR domain proteins', *Adv Exp Med Biol*, 693, pp. 37-53.

Ryder, S.P. and Williamson, J.R. (2004) 'Specificity of the STAR/GSG domain protein Qk1: Implications for the regulation of myelination', *RNA*, 10(9), pp. 1449-1458.

Sakharkar, M.K., Chow, V.T.K. and Kangueane, P. (2004) 'Distributions of Exons and Introns in the Human Genome', *In Silico Biology*, 4(4), pp. 387-393.

Sánchez-Jiménez, F. and Sánchez-Margalet, V. (2013) 'Role of Sam68 in Post-Transcriptional Gene Regulation', *International Journal of Molecular Sciences*, 14(12), pp. 23402-23419.

Screaton, G.R., Bell, M.V., Jackson, D.G., Cornelis, F.B., Gerth, U. and Bell, J.I. (1992) 'Genomic structure of DNA encoding the lymphocyte homing receptor CD44 reveals at least 12 alternatively spliced exons', *Proceedings of the National Academy of Sciences*, 89(24), pp. 12160-12164.

Sette, C., Messina, V. and Paronetto, M.P. (2010) 'Sam68: A New STAR in the Male Fertility Firmament', *Journal of Andrology*, 31(1), pp. 66-74.

Sette, C., Paronetto, M.P., Barchi, M., Bevilacqua, A., Geremia, R. and Rossi, P. (2002) *Tr‐kit‐induced resumption of the cell cycle in mouse eggs requires activation of a Src‐like kinase*.

Shabalina, S.A., Ogurtsov, A.Y., Spiridonov, A.N., Novichkov, P.S., Spiridonov, N.A. and Koonin, E.V. (2010) 'Distinct Patterns of Expression and Evolution of Intronless and Intron-Containing Mammalian Genes', *Molecular Biology and Evolution*, 27(8), pp. 1745-1749.

Sharma, S., Kohlstaedt, L.A., Damianov, A., Rio, D.C. and Black, D.L. (2008) 'Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome', *Nat Struct Mol Biol*, 15(2), pp. 183-191.

Sharp, P.A. (1994) 'Split genes and RNA splicing', *Cell*, 77(6), pp. 805-815.

Shen, H. and Green, M.R. (2006) 'RS domains contact splicing signalsand promote splicing by a commonmechanism in yeast through humans', *Genes & Development*, 20(13), pp. 1755-1765.

Shepard, P.J. and Hertel, K.J. (2009) 'The SR protein family', *Genome Biology*, 10(10), pp. 242-242.

Sindi, I.A., Tannenberg, R.K. and Dodd, P.R. (2014) 'A role for the neurexin–neuroligin complex in Alzheimer's disease', *Neurobiology of Aging*, 35(4), pp. 746-756.

Singh, N.N., Androphy, E.J. and Singh, R.N. (2004) 'The Regulation and Regulatory Activities of Alternative Splicing of the SMN Gene', 14(4), p. 16.

Siomi, H. and Dreyfuss, G. (1997) 'RNA-binding proteins as regulators of gene expression', *Current Opinion in Genetics & Development*, 7(3), pp. 345-353.

Sirand-Pugnet, P., Durosay, P., Brody, E. and Marie, J. (1995) 'An intronic (A/U)GGG repeat enhances the splicing of an alternative intron of the chicken β-tropomyosin pre-mRNA', *Nucleic Acids Research*, 23(17), pp. 3501-3507.

Smith, C.W.J. and Nadal-Ginard, B. (1989) 'Mutually exclusive splicing of α-tropomyosin exons enforced by an unusual lariat branch point location: Implications for constitutive splicing', *Cell*, 56(5), pp. 749-758.

Soergel, D.A.W., Lareau, L.F. and Brenner, S.E. (2006) 'Regulation of gene expression by coupling of alternative splicing and NMD', *Nonsense-mediated mRNA decay, L. Maquat, ed (Georgetown, Texas: Landes bioscience)*, pp. 175-190.

Spellman, R. and Smith, C.W.J. (2006) 'Novel modes of splicing repression by PTB', *Trends in Biochemical Sciences*, 31(2), pp. 73-76.

Srebrow, A. and Kornblihtt, A.R. (2006) 'The connection between splicing and cancer', *Journal of Cell Science*, 119(13), pp. 2635-2641.

Steffensen, A.Y., Dandanell, M., Jønson, L., Ejlertsen, B., Gerdes, A.-M., Nielsen, F.C. and Hansen, T.v. (2014) 'Functional characterization of BRCA1 gene variants by mini-gene splicing assay', *European Journal of Human Genetics*, 22(12), pp. 1362-1368.

Stoss, O., Novoyatleva, T., Gencheva, M., Olbrich, M., Benderska, N. and Stamm, S. (2004) 'p59fyn-mediated phosphorylation regulates the activity of the tissue-specific splicing factor rSLM-1', *Molecular and Cellular Neuroscience*, 27(1), pp. 8-21.

Sugimoto, Y., Morita, R., Amano, K., Shah, P.U., Pascual-Castroviejo, I., Khan, S., Delgado-Escueta, A.V. and Yamakawa, K. (2001) 'T-STAR gene: fine mapping in the candidate region for childhood absence epilepsy on 8q24 and mutational analysis in patients', *Epilepsy Research*, 46(2), pp. 139-144.

Tacke, R., Tohyama, M., Ogawa, S. and Manley, J.L. (1998) 'Human Tra2 Proteins Are Sequence-Specific Activators of Pre-mRNA Splicing', *Cell*, 93(1), pp. 139-148.

Taylor, S.J. and Shalloway, D. (1994) 'An RNA-binding protein associated with Src through its SH2 and SH3 domains in mitosis', *Nature*, 368(6474), pp. 867-871.

Tennyson, C.N., Klamut, H.J. and Worton, R.G. (1995) 'The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced', *Nat Genet*, 9(2), pp. 184-190.

Tollervey, J.R., Wang, Z., Hortobágyi, T., Witten, J.T., Zarnack, K., Kayikci, M., Clark, T.A., Schweitzer, A.C., Rot, G., Curk, T., Zupan, B., Rogelj, B., Shaw, C.E. and Ule, J. (2011) 'Analysis of alternative splicing associated with aging and neurodegeneration in the human brain', *Genome Research*, 21(10), pp. 1572-1582.

Turunen, J.J., Niemelä, E.H., Verma, B. and Frilander, M.J. (2013) 'The significant other: splicing by the minor spliceosome', *Wiley Interdisciplinary Reviews: RNA*, 4(1), pp. 61-76.

Ule, J. (2008) 'Ribonucleoprotein complexes in neurologic diseases', *Current Opinion in Neurobiology*, 18(5), pp. 516-523.

Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B.J. and Darnell, R.B. (2006) 'An RNA map predicting Nova-dependent splicing regulation', *Nature*, 444(7119), pp. 580-586.

Ule, J., Ule, A., Spencer, J., Williams, A., Hu, J.-S., Cline, M., Wang, H., Clark, T., Fraser, C., Ruggiu, M., Zeeberg, B.R., Kane, D., Weinstein, J.N., Blume, J. and Darnell, R.B. (2005) 'Nova regulates brain-specific splicing to shape the synapse', *Nat Genet*, 37(8), pp. 844-852.

Ulzi, G., Sansone, V., Magri, F., Corti, S., Bresolin, N., Comi, G. and Lucchiari, S. (2014) 'In vitro analysis of splice site mutations in the CLCN1 gene using the minigene assay', *Molecular Biology Reports*, 41(5), pp. 2865-2874.

Valadkhan, S. (2007) 'The spliceosome: caught in a web of shifting interactions', *Current Opinion in Structural Biology*, 17(3), pp. 310-315.

van der Veer, E.P., de Bruin, R.G., Kraaijeveld, A.O., de Vries, M.R., Bot, I., Pera, T., Segers, F.M., Trompet, S., van Gils, J.M., Roeten, M.K., Beckers, C.M., van Santbrink, P.J., Janssen, A., van Solingen, C., Swildens, J., de Boer, H.C., Peters, E.A., Bijkerk, R., Rousch, M., Doop, M., Kuiper, J., Schalij, M.J., van der Wal, A.C., Richard, S., van Berkel, T.J.C., Pickering, J.G., Hiemstra, P.S., Goumans, M.J., Rabelink, T.J., de Vries, A.A.F., Quax, P.H.A., Jukema, J.W., Biessen, E.A.L. and van Zonneveld, A.J. (2013) 'Quaking, an RNA-Binding Protein, Is a Critical Regulator of Vascular Smooth Muscle Cell Phenotype', *Circulation Research*, 113(9), pp. 1065-1075.

Venables, J.P., Vernet, C., Chew, S.L., Elliott, D.J., Cowmeadow, R.B., Wu, J., Cooke, H.J., Artzt, K. and Eperon, I.C. (1999) 'T-STAR/ÉTOILE: A Novel Relative of SAM68 That Interacts with an RNA-Binding Protein Implicated in Spermatogenesis', *Human Molecular Genetics*, 8(6), pp. 959-969.

Wahl, M.C., Will, C.L. and Lührmann, R. (2009) 'The Spliceosome: Design Principles of a Dynamic RNP Machine', *Cell*, 136(4), pp. 701-718.

Wang, G.-S. and Cooper, T.A. (2007) 'Splicing in disease: disruption of the splicing code and the decoding machinery', *Nat Rev Genet*, 8(10), pp. 749-761.

Wang, L.L., Richard, S. and Shaw, A.S. (1995) 'p62 Association with RNA Is Regulated by Tyrosine Phosphorylation', *Journal of Biological Chemistry*, 270(5), pp. 2010-2013.

Wang, Z. and Burge, C.B. (2008) 'Splicing regulation: From a parts list of regulatory elements to an integrated splicing code', *RNA*, 14(5), pp. 802-813.

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(10), p. e1000530.

Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A.M., Lieberenz, M., Savitski, M.M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J.-H., Bantscheff, M., Gerstmair, A., Faerber, F. and Kuster, B. (2014) 'Mass-spectrometry-based draft of the human proteome', *Nature*, 509(7502), pp. 582-587.

Will, C.L. and Lührmann, R. (2011) 'Spliceosome Structure and Function', *Cold Spring Harbor Perspectives in Biology*, 3(7).

Williams, A.L., Bielopolski, N., Meroz, D., Lam, A.D., Passmore, D.R., Ben-Tal, N., Ernst, S.A., Ashery, U. and Stuenkel, E.L. (2011) 'Structural and Functional Analysis of Tomosyn Identifies Domains Important in Exocytotic Regulation', *The Journal of Biological Chemistry*, 286(16), pp. 14542-14553.

Windhager, L., Bonfert, T., Burger, K., Ruzsics, Z., Krebs, S., Kaufmann, S., Malterer, G., L'Hernault, A., Schilhabel, M., Schreiber, S., Rosenstiel, P., Zimmer, R., Eick, D., Friedel, C.C. and Dölken, L. (2012) 'Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution', *Genome Research*, 22(10), pp. 2031-2042.

Witten, J.T. and Ule, J. (2011) 'Understanding splicing regulation through RNA splicing maps', *Trends in Genetics*, 27(3-2), pp. 89-97.

Wu, J., Zhou, L., Tonissen, K., Tee, R. and Artzt, K. (1999) 'The Quaking I-5 Protein (QKI-5) Has a Novel Nuclear Localization Signal and Shuttles between the Nucleus and the Cytoplasm', *Journal of Biological Chemistry*, 274(41), pp. 29202-29210.

Wu, J.I., Reed, R.B., Grabowski, P.J. and Artzt, K. (2002) 'Function of quaking in myelination: Regulation of alternative splicing', *Proceedings of the National Academy of Sciences*, 99(7), pp. 4233-4238.

Xiao, Q., Ford, A.L., Xu, J., Yan, P., Lee, K.-Y., Gonzales, E., West, T., Holtzman, D.M. and Lee, J.-M. (2012) 'Bcl-x pre-mRNA Splicing Regulates Brain injury after Neonatal Hypoxia-Ischemia', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 32(39), pp. 13587-13596.

Xue, Y., Zhou, Y., Wu, T., Zhu, T., Ji, X., Kwon, Y.-S., Zhang, C., Yeo, G., Black, D.L., Sun, H., Xiang-Dong, F. and Zhang, Y. (2009) 'Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping', *Molecular Cell*, 36(6), pp. 996-1006.

Yang, G., Fu, H., Zhang, J., Lu, X., Yu, F., Jin, L., Bai, L., Huang, B., Shen, L., Feng, Y., Yao, L. and Lu, Z. (2010) 'RNA binding protein Quaking, a critical regulator of colon epithelial differentiation and a suppressor of colon cancer', *Gastroenterology*, 138(1), pp. 231-40.e1-5.

Yeo, G., Holste, D., Kreiman, G. and Burge, C.B. (2004) 'Variation in alternative splicing across human tissues', *Genome Biology*, 5(10), pp. R74-R74.

Zhang, C., Zhang, Z., Castle, J., Sun, S., Johnson, J., Krainer, A.R. and Zhang, M.Q. (2008a) 'Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2', *Genes & Development*, 22(18), pp. 2550-2563.

Zhang, J., Kuo, C.C.J. and Chen, L. (2011) 'GC content around splice sites affects splicing through pre-mRNA secondary structures', *BMC Genomics*, 12, pp. 90-90.

Zhang, L.Y., Zeng, M., Chen, P., Sun, H.Q., Tao, D.C., Liu, Y.Q., Lin, L., Yang, Y., Zhang, S.Z. and Ma, Y.X. (2009) 'Identification of messenger RNA substrates for mouse T-STAR', *Biochemistry*, 74(11), pp. 1270-7.

Zhang, Z., Lotti, F., Dittmar, K., Younis, I., Wan, L., Kasim, M. and Dreyfuss, G. (2008b) 'SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing', *Cell*, 133(4), pp. 585-600.

Zhang, Z., Xu, Y., Sun, N., Zhang, M., Xie, J. and Jiang, Z. (2014) 'High Sam68 expression predicts poor prognosis in non-small cell lung cancer', *Clinical and Translational Oncology*, 16(10), pp. 886-891.

Zhao, Y., Zhang, G., Wei, M., Lu, X., Fu, H., Feng, F., Wang, S., Lu, W., Wu, N., Lu, Z. and Yuan, J. (2014) 'The tumor suppressing effects of QKI-5 in prostate cancer: A novel diagnostic and prognostic protein', *Cancer Biology & Therapy*, 15(1), pp. 108-118.

Zong, F.-Y., Fu, X., Wei, W.-J., Luo, Y.-G., Heiner, M., Cao, L.-J., Fang, Z., Fang, R., Lu, D., Ji, H. and Hui, J. (2014) 'The RNA-Binding Protein QKI Suppresses Cancer-Associated Aberrant Splicing', *PLoS Genetics*, 10(4), p. e1004289.

Appendix I

Primer	Sequence	Template used for PCR in cloning
Nrxn2 Mutant 1 F	AATCCATTAATTAATTAACTAACTAACT AACTTTAAAAACACGATCTTAAAAAGG TGCAGAGCTCTCTC	Wild type *Nrxn2* minigene
Nrxn2 Mutant 1 R	TTTAAGATCGTGTTTTTAAAGTTAGTTA GTTAGTTAATTAATTAATGGATTCTGGT TAATTACCTTTGTC	
Nrxn2 Mutant 2 F	AATTAATCCATTAATTAACTAACTAACT AACTTTAAAAACACGATCTTAAAAAGG TGCAGAGCTCTCTC	Wild type *Nrxn2* minigene
Nrxn2 Mutant 2 R	TTTAAGATCGTGTTTTTAAAGTTAGTTA GTTAGTTAATTAATGGATTAATTCTGGT TAATTACCTTTGTC	
Nrxn2 Mutant 3 F	AATTAATTAATCCATTAACTAACTAACT AACTTTAAAAACACGATCTTAAAAAGG TGCAGAGCTCTCTC	Wild type *Nrxn2* minigene
Nrxn2 Mutant 3 R	TTTAAGATCGTGTTTTTAAAGTTAGTTA GTTAGTTAATGGATTAATTAATTCTGGT TAATTACCTTTGTC	
Nrxn2 Mutant 4 F	AATTAATTAATTAATCCACTAACTAACT AACTTTAAAAACACGATCTTAAAAAGG TGCAGAGCTCTCTC	Wild type *Nrxn2* minigene
Nrxn2 Mutant 4 R	TTTAAGATCGTGTTTTTAAAGTTAGTTA GTTAGTGGATTAATTAATTAATTCTGGT TAATTACCTTTGTC	
Nrxn2 Mutant 5 F	AATTAATTAATTAATTAACTAACTAACT AACTTCCAAAACACGATCTTAAAAAGG TGCAGAGCTCTCTC	Wild type *Nrxn2* minigene
Nrxn2 Mutant 5 R	TTTAAGATCGTGTTTTGGAAGTTAGTTA GTTAGTTAATTAATTAATTAATTCTGGT TAATTACCTTTGTC	
Nrxn2 Mutant 6 F	AATTAATTAATTAATTAACTAACTAACT AACTTTAAAAACACGATCTCCAAGGTGC AGAGCTCTCTC	Wild type *Nrxn2* minigene

Nrxn2 Mutant 6 R	CTTGGAGATCGTGTTTTTAAAGTTAGTT AGTTAGTTAATTAATTAATTAATTCTGG TTAATTACCTTTGTC	
Nrxn2 Mutant 7 F	AATTAATTAATTAATTAACCCACCCACC CACTTTAAAAACACGATCTTAAAGGTGC AGAGCTCTCTC	Wild type *Nrxn2* minigene
Nrxn2 Mutant 7 R	TTTAAGATCGTGTTTTTAAAGTGGGTGG GTGGGTTAATTAATTAATTAATTCTGGT TAATTACCTTTGTC	
Nrxn2 Mutant 1-4 F	AATCCATCCATCCATCCACTAACTAACT AACTT	Wild type *Nrxn2* minigene
Nrxn2 Mutant 1-4 R	TGGATGGATGGATGGATTCTGGTTAATT ACCTTTG	
Nrxn2 Mutant 57 F	TAATTAATTAACCCACCCACCCACTTCC AAAACACGATCTTAAAGGTGCAG	Wild type *Nrxn2* minigene
Nrxn2 Mutant 57 R	GTTTTGGAAGTGGGTGGGTGGGTTAATT AATTAATTAATTCTGGTTA	
Nrxn2 Mutant 145 F	ACTAACTAACTAACTTCCAAAACACGAT CTTAAAGGTG	*Nrxn2* Mutant 1-4 minigene
Nrxn2 Mutant 145 R	CTTTAAGATCGTGTTTTGGAAGTTAGTT AGTTAGT	
Nrxn2 Mutant with repressor sequence deleted F	CCAGAATTAATTAATTAATTAATAAAAA CACGATCTTAAAGGTG	*Nrxn2* minigene with big mutation
Nrxn2 Mutant with repressor sequence deleted R	CACCTTTAAGATCGTGTTTTTATTAATTA ATTAATTAATTCTGG	
Nrxn2 Mutant with sequence moved 100bp downstream F	AATTAATTAATTAATTAACTAACTAACT AACTTTAAAAACACGATCTTAAACCCTG CGTTCAGGCCGTTTG	*Nrxn2* minigene with big mutation
Nrxn2 Mutant with sequence moved 100bp downstream R	TTTAAGATCGTGTTTTTAAAGTTAGTTA GTTAGTTAATTAATTAATTAATTGGGCA GAAGGGAGCATCCTG	
Nrxn2 Mutant with sequence moved 150bp downstream	AATTAATTAATTAATTAACTAACTAACT AACTTTAAAAACACGATCTTAAAAAAG TCTTCCTAGAGTG	*Nrxn2* minigene with big mutation

Nrxn2 Mutant with sequence moved 150bp downstream F	TTTAAGATCGTGTTTTTAAAGTTAGTTA GTTAGTTAATTAATTAATTAATTCTCTTC CCTATTCCCAGC	
Nrxn2 Mutant with sequence moved 250bp downstream	AATTAATTAATTAATTAACTAACTAACT AACTTTAAAAACACGATCTTAAAGGAA CACAGGCTAGAA	*Nrxn2* minigene with big mutation
Nrxn2 Mutant with sequence moved 250bp downstream R	TTTAAGATCGTGTTTTTAAAGTTAGTTA GTTAGTTAATTAATTAATTAATTCCTCCT ACTGCAGAATT	
Nrxn2 Mutant with sequence moved 42bp upstream F	AATTAATTAATTAATTAACTAACTAACT AACTTTAAAAACACGATCTTAAAGCGCT GCCTGAGAAAC	*Nrxn2* minigene with big mutation
Nrxn2 Mutant with sequence moved 42bp upstream R	TTTAAGATCGTGTTTTTAAAGTTAGTTA GTTAGTTAATTAATTAATTAATTAGTCC ACAGGCTTTAGAAAG	
Nrxn2 Mutant with sequence moved 100bp upstream F	CTTTCTGTTGAATGAATTAATTAATTAA TTAACTAACTAACTAACTTTAAAAACAC GATCTTAAATGGCTGCTTGCTGGG	*Nrxn2* minigene with big mutation
Nrxn2 Mutant with sequence moved 100bp upstream R	CCAGCAAGCAGCCATTTAAGATCGTGTT TTTAAAGTTAGTTAGTTAGTTAATTAAT TAATTAATTCATTCAACAGAAAGGAG	
Tomosyn-2 Mutant 1 F	CAGTTCCAACCTTGACCAACCTTCGTTC CTCATGGTAACAGGTGCTGGGAG	Wild type *Tomosyn-2* minigene
Tomosyn-2 Mutant 1 R	TGAGGAACGAAGGTTGGTCAAGGTTGG AACTGTACTCACAGAAATTGATAT	
Tomosyn-2 Mutant 2 F	CATTCCAAAGATGATTCCAAAACAAAA ACAAAAACAAAACTAC	Wild type *Tomosyn-2* minigene
Tomosyn-2 Mutant 2 R	GTTTTGGAATCATCTTTGGAATGTAACT TCCATTTCTTTGTCT	
Tomosyn-2 Combined Mutant F	CAGTTCCAACCTTGACCAACCTTCGTTC CTCATGGTAACAGGTGCTGGGAG	*Tomosyn-2* Mutant 2 minigene
Tomosyn-2 Combined Mutant R	TGAGGAACGAAGGTTGGTCAAGGTTGG AACTGTACTCACAGAAATTGATAT	
Tomosyn-2 Mutant with target sites moved 54bp upstream F	GAATTGGAAGAACATATCAATTTCTGTG AGTACAGTTTAAAATTTGATAAAATTTC GTTTATCATGGTAACAGGTGCTGGGAGG AATTCCTGTTCCTC	*Tomosyn-2* Combined Mutant minigene

Tomosyn-2 Mutant with target sites moved 54bp upstream R	CAGGAATTCCTCCCAGCACCTGTTACCATGATAAACGAAATTTTATCAAATTTTAAACTGTACTCACAGAAATTGATATGTTCTTCCAATTCATTTG	
Nrxn2 Mutant 4xUAAA F	AATAAATAAATAAATAAACCCACCCACCCACTTCCAAAAC	*Nrxn2* minigene with big mutation
Nrxn2 Mutant 4xUAAA R	TTTATTTATTTATTTATTCTGGTTAATTACCTTTGTCG	
Nrxn2 Mutant 8xUAAA F	AATAAATAAATAAATAAAAATAAATAAATAAATAAACCCACCCACCCACTTCCAAAAC	*Nrxn2* minigene with big mutation
Nrxn2 Mutant 8xUAAA R	TTTATTTATTTATTTATTTTTATTTATTTATTTATTCTGGTTAATTACCTTTGTCG	
Nrxn2 Mutant 4xUAAAA F	AAGGTAATTAACCAGAATAAAATAAAATAAAATAAAACCCACCCACCCACTTC	*Nrxn2* minigene with big mutation
Nrxn2 Mutant 4xUAAAA R	GTGGGTTTTATTTTATTTTATTTTATTCTGGTTAATTACCTTTGTCG	
Nrxn2 Mutant 22xA F	AAGGTAATTAACCAGAAAAAAAAAAAAAAAAAAAAACCCACCCACCCACTTC	*Nrxn2* Mutant with big mutation
Nrxn2 Mutant 22xA R	GTGGGTTTTTTTTTTTTTTTTTTTTCTGGTTAATTACCTTTGTCG	
Nrxn2 Mutant with 4xUAAAA and the rest wild type F	CAAAGGTAATTAACCAGAATAAAATAAAATAAAATAAAACTAACTAACTAACTTTA	Wild type *Nrxn2* minigene
Nrxn2 Mutant with 4xUAAAA and the rest wild type R	TAAAGTTAGTTAGTTAGTTTTATTTTATTTTATTTTATTCTGGTTAATTACCTTTG	
Nrxn2 with sequence replaced to *Tomosyn-2* (1) F	AAGGTAATTAACCAGTACAGTTTAAAATTTGATAAAATTTCGCCCACCCACCCACTTC	*Nrxn2* Mutant with big mutation
Nrxn2 with sequence replaced to *Tomosyn-2* (1) R	GTGGGCGAAATTTTATCAAATTTTAAACTGTACTGGTTAATTACCTTTGTCG	
Nrxn2 with sequence replaced to *Tomosyn-2* (2) F	AAGGTAATTAACCAGTACATTTAAAAGATGATTTAAAAACCCCACCCACCCACTTC	*Nrxn2* Mutant with big mutation

Nrxn2 with sequence replaced to *Tomosyn-2* (2) R	GTGGGGTTTTTAAATCATCTTTTAAATG TACTGGTTAATTACCTTTGTCG	

Appendix II

Publication List

- The tissue-specific RNA binding protein T-STAR controls regional splicing patterns of neurexin pre-mRNAs in the brain.
  Ehrmann I, Dalgliesh C, Liu Y, **Danilenko M**, Crosier M, Overman L, Arthur HM, Lindsay S, Clowry GJ, Venables JP, Fort P, Elliott DJ.
  PLoS Genetics 2013

- Human Tra2 proteins jointly control a CHEK1 splicing switch among alternative and constitutive target exons.
  Best A, James K, Dalgliesh C, Hong E, Kheirolahi-Kouhestani M, Curk T, Xu Y, **Danilenko M**, Hussain R, Keavney B, Wipat A, Klinck R, Cowell IG, Cheong Lee K, Austin CA, Venables JP, Chabot B, Santibanez Koref M, Tyson-Capper A, Elliott DJ.
  Nature Communications 2014

- The splicing landscape is globally reprogrammed during male meiosis.
  Schmid R, Grellscheid SN, Ehrmann I, Dalgliesh C, **Danilenko M**, Paronetto MP, Pedrotti S, Grellscheid D, Dixon RJ, Sette C, Eperon IC, Elliott DJ.
  Nucleic Acids Research 2013

- Murine Joubert syndrome reveals Hedgehog signaling defects as a potential therapeutic target for nephronophthisis.
  Hynes AM, Giles RH, Srivastava S, Eley L, Whitehead J, **Danilenko M**, Raman S, Slaats GG, Colville JG, Ajzenberg H, Kroes HY, Thelwall PE, Simmons NL, Miles CG, Sayer JA.
  PNAS 2014

- Tra2 protein biology and mechanisms of splicing control.
  Best A, Dalgliesh C, Kheirollahi-Kouhestani M, **Danilenko M**, Ehrmann I, Tyson-Capper A, Elliott DJ.
  Biochemical Society Transactions 2014

- CSA-mediated Golgi dysfunction separates Cockayne syndrome from UV sensitive syndrome.
  Brian T. Wilson, Ruth E. Sutton, Jennifer Munkley, Ian J. Wilson, **Marina Danilenko**, Miranda J. Patterson, Lynn Tindale, Heather Fawcett, Yuka Nakazawa, David J. Elliott, Alain Sarasin, Tomoo Ogi, Alan R. Lehmann.
  (Manuscript in revision)