

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

**CHARACTERIZATION OF THE REGULATION OF
CD46 ALTERNATIVE SPLICING**

**TANG SZE JING
SCHOOL OF BIOLOGICAL SCIENCES
2016**

**CHARACTERIZATION OF THE REGULATION OF
CD46 ALTERNATIVE SPLICING**

TANG SZE JING

School of Biological Sciences

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2016

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude to my supervisor, Dr. Xavier Roca, for his constant guidance and patience throughout the course of my project. I am thankful to School of Biological Sciences, Nanyang Technological University for the advance core facilities and financial support. I am also sincerely appreciated my Thesis Advisory Committee, Professor Alex Law and Dr. Eugene Makeyev, for their valuable suggestions on my project.

I would like to thank Dr. Karen Yap Lin Choo and Fursham Hamid for their help in one of the experiments. I would also like to thank Heidi Berger for her help in maintaining Jurkat T cells.

I am indebted to all my present and past lab members especially Luo Shufang and Dr. Liu Jun for their help and guidance throughout these four years.

I am grateful to my friends for their support and accompanying.

Last but not least, I would like to dedicate this thesis to my family who always support me.

TABLE OF CONTENTS	
ACKNOWLEDGEMENTS	1
LIST OF FIGURES	7
LIST OF TABLES	9
ABBREVIATIONS	10
ABSTRACT	16
1. INTRODUCTION	17
1.1 Splicing	17
1.1.1 Exon definition	19
1.2 Alternative splicing	21
1.2.1 Types of alternative splicing	21
1.2.2 Regulation of alternative splicing	22
1.2.2.1 Splice-site strength	22
1.2.2.2 <i>Cis</i> -acting elements and <i>trans</i> -acting factors	25
1.2.2.2.1 SR proteins	30
1.2.2.2.2 SRSF1	30
1.2.2.3 Transcription and chromatin	32
1.2.2.4 RNA secondary structure	33
1.2.3 Alternative splicing and NMD	33
1.2.4 Alternative splicing and diseases	35
1.3 The complement system	38
1.4 Adaptive immune system	41
1.4.1 B cells	41

1.4.2 T cells.....	43
1.5 CD46	47
1.5.1 Gene structure.....	47
1.5.2 CD46 Function	49
1.5.2.1 Role in complement regulation.....	49
1.5.2.2 CD46 and fertilization	49
1.5.2.3 Role of CD46 in epithelial cells	49
1.5.2.4 Role of CD46 on B cells.....	50
1.5.2.5 CD46 in T-cell regulation.....	51
1.5.3 Implications in diseases	54
1.5.4 Alternative splicing of <i>CD46</i>	57
1.6 Alternative splicing in immune responses.....	58
1.7 Objective.....	60
2. MATERIALS AND METHODS	61
2.1 Cell culture.....	61
2.2 Plasmids.....	61
2.2.1 Minigene construction	61
2.2.2 Generation of mutant minigenes	63
2.2.3 Heterologous minigene	70
2.2.3.1 pSXN plasmid.....	70
2.2.3.2 <i>CD46</i> exon 13 in Universal Minigene Vector (UMV).....	71
2.2.4 Overexpression plasmid.....	72

2.3 Bacterial transformation and plasmid extraction	73
2.4 Transfection	74
2.4.1 Adherent cells	74
2.4.1.1 Transfection with minigenes	74
2.4.1.2 siRNA and antisense oligonucleotides (ASOs) transfection ...	74
2.4.2 Suspension cells	77
2.4.2.1 Transfection with minigenes	77
2.4.2.2 siRNA and antisense oligonucleotides (ASO) transfection	77
2.5 Jurkat E6.1 cell activation	77
2.6 RNA extraction and DNase I treatment	78
2.7 Reverse transcription (RT)-PCR and gel electrophoresis	78
2.8 Semi-quantitative (Semi-Q) PCR	80
2.9 Analysis of <i>CD46</i> alternative splicing patterns in human tissues	80
2.10 Western blot	80
2.10.1 Sample preparation	80
2.10.2 SDS-PAGE and transfer	81
2.10.3 Detection	81
2.11 RNA pull-down	82
2.12 Silver staining	83
2.13 Flow cytometry	83
2.13.1 Cell sorting	83
2.13.2 Cell analysis	83

2.14 Inhibition of NMD	83
2.15 Preparation of HeLa nuclear extract.....	84
3. RESULTS	85
3.1 Alternative splicing of <i>CD46</i> is tissue selective	85
3.2 The 5'ss flanking exons 7 and 8 are recognized by U1 using asymmetric loop register	88
3.3 Multiple <i>cis</i> -acting elements regulate the alternative splicing pattern of exon 13.....	96
3.3.1 Exon 13 is rich in ESEs and ESSs.....	96
3.3.2 The intronic sequences near exon 13 have less regulatory effect	102
3.3.3 Validation of <i>cis</i> -acting elements by point mutations	105
3.3.4 Validation of ESEs and ESSs in heterologous context.....	108
3.4 Identification of <i>trans</i> -acting factors as regulators of <i>CD46</i> alternative splicing	110
3.4.1 SRSF1 represses exon 13 inclusion through ESS1	116
3.4.2 PTBP1 is a repressor of exon 13 inclusion	122
3.4.3 TIA1 and TIAL1 promote exon 13 recognition	124
3.5 Manipulation of the alternative splicing pattern of <i>CD46</i> exon 13 by ASOs	128
3.6 Stimulation of Jurkat cells with PMA did not change the inclusion of <i>CD46</i> exon 13	130
3.7 <i>CD46</i> mRNA containing exon 13 is slightly sensitive to NMD.....	132
4. DISCUSSION	134

4.1 Tissue-selective alternative splicing of <i>CD46</i> is likely regulated by <i>trans</i> -acting factors.....	134
4.2 STP cassette exons have different inclusion efficiencies	135
4.3 <i>Cis</i> -acting elements and <i>trans</i> -acting factors regulate <i>CD46</i> exon 13 inclusion	136
4.4 SRSF1 is a splicing repressor for <i>CD46</i> exon 13	142
4.5 Origin of <i>CD46</i> exon 13	144
4.6 ASOs effectively modulate <i>CD46</i> alternative splicing	148
4.7 T-cell signaling and <i>CD46</i> alternative splicing	149
5. CONCLUSIONS AND FUTURE DIRECTIONS	151
6. REFERENCES	154
Appendix I.....	173
Appendix II	177
Appendix III.....	180

LIST OF FIGURES

Figure 1.1 Schematic diagram of splicing and dynamic assembly of the spliceosome	19
Figure 1.2 Recognition of 5'ss and 3'ss by exon definition	20
Figure 1.3 Types of alternative splicing	22
Figure 1.4 U1/5'ss base-pairing registers.....	24
Figure 1.5 Regulatory mechanisms of <i>trans</i> -acting factors.....	26
Figure 1.6 Alternative splicing coupled with NMD	35
Figure 1.7 The complement pathway	40
Figure 1.8 B-cell activation in germinal center.....	43
Figure 1.9 T-cell signaling.....	45
Figure 1.10 Gene structure of <i>CD46</i>	48
Figure 1.11 CD46 functions.....	53
Figure 1.12 Interacting partners of CD46.....	54
Figure 2.1 Schematic diagram of <i>CD46</i> minigene construction.....	62
Figure 2.2 Mutagenesis primers design	64
Figure 2.3 Schematic map of pSXN heterologous plasmid.....	70
Figure 2.4 Schematic map of UMV <i>CD46</i> exon 13 plasmid.....	72
Figure 2.5 Example of design of a siRNA and a dsRNA	75
Figure 3.1 Semi-Q PCR detected <i>CD46</i> alternative splicing patterns in 20 human tissues.....	87
Figure 3.2 Recognition of the 5'ss flanking exon 7 by U1 using asymmetric loop 1 (+3/+4).....	92
Figure 3.3 Recognition of the 5'ss flanking exon 8 by U1 using asymmetric loop 1 (+3/+4).....	94
Figure 3.4 Schematic map of <i>CD46</i> -exon 12-13-14 minigene and serial deletions	97
Figure 3.5 Identification of enhancers and silencers by serial deletions	100
Figure 3.6 Map of enhancers and silencers of exon 13 by deletion assays	104
Figure 3.7 Point mutations validated the presence of ESS1 and ESE1	106
Figure 3.8 Testing ESSs and ESEs in heterologous context.....	109
Figure 3.9 Total proteins bound to ESS1 and ESE1 RNA by pull-down assay ...	113

Figure 3.10 Western blot confirmed protein binding to ESS1 or ESE1	114
Figure 3.11 SRSF1, a common activator, represses <i>CD46</i> exon 13 inclusion	117
Figure 3.12 SRSF1 likely represses inclusion via exon 13 and/or its nearby flanking introns	119
Figure 3.13 SRSF1 and PTBP1 mediate exon 13 skipping through A6 and B9 sequences	121
Figure 3.14 PTBP1 is a weak repressor of endogenous <i>CD46</i> exon 13 inclusion	123
Figure 3.15 TIAL1 and TIA1 enhance exon 13 inclusion	125
Figure 3.16 TIA1 and TIAL1 upregulate exon 13 inclusion via downstream poly-U regions.....	127
Figure 3.17 ASOs are competent to modulate exon 13 inclusion in a dose- dependent manner.....	129
Figure 3.18 T-cell activation does not regulate the alternative splicing pattern of exon 13.....	131
Figure 3.19 <i>CD46</i> mRNA containing exon 13 is slightly sensitive to NMD	133
Figure 4.1 Distal intronic sequences can contain splicing regulatory elements ...	137
Figure 4.2 Deleted intron 12 and 13 middle sequences may contain regulatory information.....	138
Figure 4.3 Regulatory mechanisms of exon 13 ESEs and ESSs.....	140
Figure 4.4 Possible regulatory mechanisms of SRSF1 on <i>CD46</i> exon 13	143
Figure 4.5 Conservation of <i>CD46</i> gene	146

LIST OF TABLES

Table 1.1 List of common SR or SR-like proteins and hnRNPs.....	28
Table 1.2 List of tissue-specific splicing factors.....	29
Table 1.3 Cytokines and transcription factors that induce CD4 ⁺ T-cell differentiation.....	46
Table 1.4 Sizes of <i>CD46</i> exons and introns.....	48
Table 2.1 Primer sequences for generating <i>CD46</i> minigenes.....	63
Table 2.2 Primer design for generating mutant <i>CD46</i> -exon 12-13-14 minigene ...	65
Table 2.3 Primers for mutating the 3'ss and 5'ss flanking exon 7.....	69
Table 2.4 Primers for mutating the 5'ss flanking exon 8.....	69
Table 2.5 Oligonucleotide inserts for pSXN plasmid.....	71
Table 2.6 Primers for subcloning exon 13 and its flanking intronic sequences.....	72
Table 2.7 List of cDNA plasmids used in experiments.....	73
Table 2.8 Primers for cloning splicing factor cDNAs.....	73
Table 2.9 List of siRNAs and dsRNAs.....	76
Table 2.10 List of ASOs.....	76
Table 2.11 List of RT-PCR primers.....	79
Table 2.12 List of antibodies.....	81
Table 2.13 List of RNA oligonucleotides for RNA pull-down.....	82
Table 3.1 Prediction of 5'ss and 3'ss strength using different scoring models.....	91
Table 3.2 Mass spectrometry results for protein bands.....	115
Table 4.1 Alternative splicing patterns of exon 13 homologues in different species	147

ABBREVIATIONS

2' MOE	2'- <u>O</u> - <u>m</u> ethoxy <u>e</u> thyl
3'ss	3' <u>s</u> plice <u>s</u> ite
5'ss	5' splice site
aHUS	<u>A</u> typical <u>h</u> emolytic <u>u</u> remic <u>s</u> ndrome
AMD	<u>A</u> ge-related <u>m</u> acular <u>d</u> egeneration
APC	<u>A</u> ntigen <u>p</u> resenting <u>c</u> ell
ASO	<u>A</u> nti- <u>s</u> ense <u>o</u> ligonucleotides
BCR	<u>B</u> <u>c</u> ell <u>r</u> eceptor
bp	<u>B</u> ase <u>p</u> air
BPS	<u>B</u> ran <u>ch</u> <u>p</u> oint <u>s</u> equence
CCL	<u>C</u> hemokine (<u>C</u> - <u>C</u> motif) <u>l</u> igand
CCP	<u>C</u> omplement <u>c</u> ontrol <u>p</u> rotein
cDNA	<u>C</u> omplementary <u>D</u> N <u>A</u>
CELF	<u>C</u> UG-BP and <u>E</u> TR-3 <u>L</u> ike <u>f</u> actors
CFIm	Pre-messenger RNA <u>c</u> leavage <u>f</u> actor I
CFTR	<u>C</u> ystic <u>f</u> ibrosis <u>t</u> ransmembrane <u>c</u> onductance <u>r</u> egulator
CHX	<u>C</u> yclo <u>h</u> ex <u>i</u> mide
CLC-1	<u>C</u> hloride <u>c</u> hannel 1
CLIP	<u>C</u> ross- <u>l</u> inking <u>i</u> mmunoprecipitation
Clk/Sty	<u>C</u> dc-2 <u>l</u> ike <u>k</u> inase/ <u>s</u> erine, <u>t</u> hreonine and tyrosine kinase
CPSF	<u>C</u> leavage and polyadenylation <u>s</u> pecificity <u>f</u> actor
CTCF	<u>C</u> <u>C</u> <u>C</u> <u>T</u> <u>C</u> -binding <u>f</u> actor
cTNT	<u>C</u> ardiac <u>t</u> roponin <u>T</u>

CUGBP	<u>CUG-binding protein</u>
CYT	<u>Cytoplasmic tail</u>
DAG	<u>Diacylglycerol</u>
DLG4	<u>Disks large homolog 4</u>
DMEM	<u>Dulbecco's modified eagle medium</u>
DMSO	<u>Dimethyl sulfoxide</u>
dsiRNA	<u>Dicer substrate siRNA</u>
DTT	<u>Dithiothreitol</u>
ECL	<u>Enhanced chemiluminescence</u>
EJC	<u>Exon junction complex</u>
Elav	<u>Embryonic lethal abnormal visual</u>
EMSA	<u>Electrophoretic mobility shift assay</u>
ESE/ESS	<u>Exonic splicing enhancer/silencer</u>
FBS	<u>Fetal bovine serum</u>
GOPC	<u>Golgi-associated PDZ and coiled-coil motif containing protein</u>
Grb2	<u>Growth factor receptor-bound protein 2</u>
H3K36me3	Tri-methylation of histone H3 lysine 36
HHV-6	<u>Human Herpes virus 6</u>
hnRNP	<u>Heterogeneous nuclear ribonucleoprotein</u>
ICER/CRE	<u>Inducible cAMP early repressor/cAMP response element</u>
M	<u>modulator</u>
IFN γ	<u>Interferon gamma</u>
Ig	<u>Immunoglobulin</u>
IL	<u>Interleukin</u>

IP3	<u>I</u> nositol 1,4,5- <u>t</u> ris <u>p</u> hosphate
ISE/ISS	<u>I</u> ntronic <u>s</u> plicing <u>e</u> nhancer/ <u>s</u> ilencer
ITAM	<u>I</u> mmunoreceptor <u>t</u> yrosine-based <u>a</u> ctivation <u>m</u> otifs
iTreg	<u>I</u> nduced <u>r</u> egulatory <u>T</u> cell
LAT	<u>L</u> inker of <u>a</u> ctivated <u>T</u> cells
LB	<u>L</u> ysogeny <u>b</u> roth
Lck	<u>L</u> ymphoc <u>y</u> te-specific protein tyrosine <u>k</u> inase
LC-MS/MS	<u>L</u> iquid <u>c</u> hromatography–tandem <u>m</u> ass <u>s</u> pectrometry
LPS	<u>L</u> ipopol <u>y</u> saccharide
MAC	<u>M</u> embrane <u>a</u> ttack <u>c</u> omplex
MASP	<u>M</u> annose-binding lectin <u>a</u> ssociated <u>s</u> erine proteases
MAXENT	<u>M</u> aximum <u>e</u> ntropy model
MBL	<u>M</u> annose- <u>b</u> inding lectin
MBNL	<u>M</u> uscle <u>b</u> lind- <u>l</u> ike
MCS	<u>M</u> ultiple <u>c</u> loning <u>s</u> ite
MDD	<u>M</u> aximum <u>d</u> ependence <u>d</u> ecomposition
MHC	<u>M</u> ajor <u>h</u> istocompatibility <u>c</u> omplex
MM	First-order <u>M</u> arkov <u>m</u> odel
M-MuLV	<u>M</u> oloney <u>M</u> urine <u>L</u> eukemia <u>V</u> irus
mRNA	<u>M</u> essenger RNA
MS	<u>M</u> ultiple <u>s</u> clerosis
Mut	<u>M</u> utation
MV	<u>M</u> easles <u>v</u> irus
NMD	<u>N</u> onsense- <u>m</u> ediated <u>d</u> ecay

NN	<u>N</u> eu <u>r</u> al <u>n</u> et <u>w</u> ork
NO	<u>N</u> itric <u>o</u> xide
NOVA	<u>N</u> eu <u>r</u> o- <u>o</u> ncological <u>v</u> ent <u>r</u> al <u>a</u> ntigen
nt	<u>N</u> ucleo <u>t</u> ide
nTreg	<u>N</u> atural <u>r</u> egulatory <u>T</u> cell
P(Y)	Polyp <u>yr</u> imidine tract
PAMP	<u>P</u> athogen- <u>a</u> ssociated <u>m</u> olecular <u>p</u> attern
PBS	<u>P</u> hosphate- <u>b</u> uffered <u>s</u> aline
PCR	<u>P</u> olymerase <u>c</u> hain <u>r</u> eaction
PCV	<u>P</u> acked <u>c</u> ell <u>v</u> olume
Pen/Strep	<u>P</u> enicillin/ <u>S</u> treptomycin mixture
PIP2	<u>P</u> hosphatidyl <u>i</u> nositol 4,5- <u>b</u> isphosphate
PKC	<u>P</u> rotein <u>k</u> inase <u>C</u>
PLC γ	<u>P</u> hospho <u>l</u> ipase <u>C</u> - γ
PMA	<u>P</u> horbol-12- <u>m</u> yristate-13- <u>a</u> cetate
PNK	<u>P</u> olynucleotide <u>k</u> inase
PNV	<u>P</u> acked <u>n</u> uclear <u>v</u> olume
Pre-mRNA	Pre- <u>m</u> essenger RNA
PSF	<u>P</u> TB-associated <u>s</u> plicing <u>f</u> actor
PTC	<u>P</u> re-mature translation- <u>t</u> ermination <u>c</u> odon
PTK	<u>P</u> rotein tyrosine <u>k</u> inase
PVDF	<u>P</u> olyvinylidene <u>f</u> luoride
P γ S	<u>P</u> resenilin/ γ <u>s</u> ecretase
RA	<u>R</u> heumatoid <u>a</u> rthritis

RISC	<u>R</u> NA- <u>i</u> nduced <u>s</u> ilencing <u>c</u> omplex
RNA-seq	RNA <u>s</u> equencing
RRM	<u>R</u> NA <u>r</u> ecognition <u>m</u> otif
RS	<u>A</u> rginine and <u>s</u> erine dipeptide
RT-PCR	<u>R</u> everse <u>t</u> ranscription PCR
SD	<u>S</u> tandard <u>d</u> eviation
SDS-PAGE	<u>S</u> odium <u>d</u> odecyl <u>s</u> ulfate- <u>P</u> oly <u>a</u> crylamide gel <u>e</u> lectrophoresis
Semi-Q PCR	Semi- <u>q</u> uantitative PCR
SHM	<u>S</u> omatic <u>h</u> yper <u>m</u> utation
siRNA	<u>S</u> mall <u>i</u> nterfering RNA
SLE	<u>S</u> ystemic <u>l</u> upus <u>e</u> rythematosus
SMA	<u>S</u> pinal <u>m</u> uscular <u>a</u> trophy
SMG	<u>S</u> uppressor of <u>m</u> orphogenetic effect on <u>g</u> enitalia
SMN	<u>S</u> urvival of <u>m</u> otor <u>n</u> euron
snRNA	<u>S</u> mall <u>n</u> uclear RNA
snRNP	<u>S</u> mall <u>n</u> uclear <u>r</u> ibon <u>n</u> ucleoprotein
SPAK	<u>S</u> te20/ <u>S</u> PS-1-related <u>k</u> inase
SRPK	<u>S</u> R-specific protein <u>k</u> inase
ss	<u>S</u> plice <u>s</u> ite
STP	<u>S</u> erine, <u>T</u> hreonine, <u>P</u> roline
TCR	<u>T</u> <u>c</u> ell <u>r</u> eceptor
T _H	<u>T</u> <u>h</u> elper
TIA1	<u>T</u> -cell restricted <u>i</u> ntracellular <u>a</u> ntigen 1
TIAL1	<u>TIA1</u> cytotoxic granule-associated RNA binding protein- <u>l</u> ike 1

TLR	<u>T</u> oll <u>l</u> ike <u>r</u> eceptor
Tr1	<u>T</u> <u>r</u> egulatory <u>1</u>
Treg	T <u>r</u> egulatory
U2AF	U2 <u>a</u> uxiliary <u>f</u> actor
UFP	<u>U</u> p- <u>f</u> rameshift <u>p</u> rotein
UTR	<u>U</u> n <u>t</u> ranslated <u>r</u> egion
WMM	<u>W</u> eight <u>m</u> atrix <u>m</u> odel
Ψ	Pseudouridine

ABSTRACT

Alternative splicing gives rise to multiple mRNA and protein isoforms. This process is catalyzed by the spliceosome and regulated by *cis*-acting elements, *trans*-acting factors, transcription and chromatin structure. CD46 is a membrane-bound complement control protein with several extracellular STP domain variants and two mutually exclusive cytoplasmic tails derived from the alternative splicing of cassette exons 7, 8 and 9, or cassette exon 13, respectively. The two cytoplasmic tails have different signaling capacities and regulatory functions in T helper 1 cells and epithelial cells. Besides, the association between aberrant *CD46* alternative splicing and autoimmune diseases proposes a possible novel therapy by amending splicing. To understand the splicing regulatory mechanisms of *CD46*, serial deletion assays in splicing minigenes were performed to identify *cis*-acting elements while the loss- and gain-of-function assays as well as RNA pull-down were carried out to determine the roles of protein regulators. SRSF1 and PTBP1 repress exon 13 inclusion via exonic splicing silencers. SRSF1 is a common activator yet it inhibits exon 13 inclusion in this context and it appears to do so by direct binding within this exon. On the other hand, the splicing activators, TIA1 and TIAL1, strongly enhance exon 13 recognition via the poly-U rich sequences downstream of the 5'ss. The alternative splicing pattern of exon 13 was efficiently modulated by anti-sense oligonucleotides targeting splicing enhancers or silencers in a dose-dependent manner but not by T-cell signaling. From these studies, a better understanding for *CD46* alternative splicing regulation was achieved yet much remains to be elucidated to fully decipher the *CD46* alternative splicing mechanisms. In conclusion, splicing regulation is complicated and context-dependent, and thus intensive studies of alternative splicing events are necessary to improve the current knowledge.

1. INTRODUCTION

Gene expression can be regulated at the points of transcription, post-transcription, translation and post-translational modification. During post-transcriptional processing, pre-messenger RNA (pre-mRNA) splicing generates mature messenger RNAs (mRNAs), yet splicing can also regulate gene expression by producing mRNA isoforms that function differently or that are targeted by mRNA decay. In this thesis, I will discuss about alternative splicing and its regulatory mechanisms using the *CD46* gene as a study model. CD46 has many functions related to immunity and other cellular activities like autophagy and cell-cell adhesion. Its ubiquitous expression and immune regulatory functions are utilized by many pathogens, and thus CD46 is linked to many infectious diseases such as measles and meningitis. Besides, deregulation of CD46 functions related to the complement pathway or T-cell immunity is associated with numerous autoimmune diseases. The CD46 dysfunction is likely caused by aberrant alternative splicing. Many CD46 studies were focused on its functions yet none of them investigated its alternative splicing in detail. This project is the first attempt to study the alternative splicing regulation of *CD46*. This section will introduce general information about splicing and *CD46*, with a focus on its roles in immune cells.

1.1 Splicing

Hybridization of an adenoviral mRNA to different parts of its coding DNA sequence (Berget et al., 1977; Chow et al., 1977) led to the identification of a novel post-transcriptional process called splicing. Pre-mRNA splicing involves a two-step transesterification reaction to remove introns and join exons (Black, 2003). Proper splicing is necessary for gene expression as incorrect splicing may create insertions, deletions, frameshifts and pre-mature stop codons in mRNA leading to production of non-functional or truncated protein. Splicing is also needed for mRNA export to cytoplasm for translation (Brodsky and Silver, 2000; Luo et al., 2001). The importance of splicing and its complex regulatory network leads to intensive studies about basic and regulatory mechanisms of splicing.

Splicing is catalyzed by the spliceosome which is a large RNA-protein complex formed by five small nuclear ribonucleoproteins (snRNPs), which are U1, U2, U4, U5 and U6 snRNP, and more than 150 auxiliary proteins (Wahl et al., 2009). The essential *cis*-acting elements are the 5' splice site (5'ss) located at the 5' end of introns, the 3' splice site (3'ss) located at the 3' end of introns, and the branch point sequence (BPS). The 5'ss and 3'ss are recognized by the U1 small nuclear RNA snRNA and the U2 snRNP auxiliary factor (U2AF), respectively. U2AF helps to recruit U2 snRNP to the BPS to form the A complex together with U1 snRNP. Subsequently, the U4/U5/U6 tri-snRNP is recruited to form the B complex which is still catalytically inactive (Figure 1.1).

After conformational and compositional rearrangements, U1 and U4 snRNP are released and then the spliceosome is activated and able to catalyze the first transesterification reaction. It involves the branch point adenosine-mediated nucleophilic attack of the phosphodiester bond between the last nucleotide in the exon and the first nucleotide in the intron, and formation of the lariat intermediate. During the first catalytic step, the spliceosome undergoes another rearrangement and forms the C complex which subsequently catalyzes the second step. The free 3' hydroxyl group at the 5' exon attacks the last nucleotide in the intron and joins with the 3' exon when the intron lariat is released. After the two catalytic steps, the spliceosome dissociates and is recycled for the next round of splicing (Figure 1.1)(House and Lynch, 2008; Matera and Wang, 2014; Wahl et al., 2009). Throughout the whole splicing event, the essential *cis*-acting elements are recognized by snRNAs and proteins multiple times to ensure proper splicing at the correct site. The intensive remodeling of RNA-RNA interactions throughout splicing is ATP-dependent and carried out by DExD/H-type RNA dependent ATPase/helicases (Cordin and Beggs, 2013; Wahl et al., 2009).

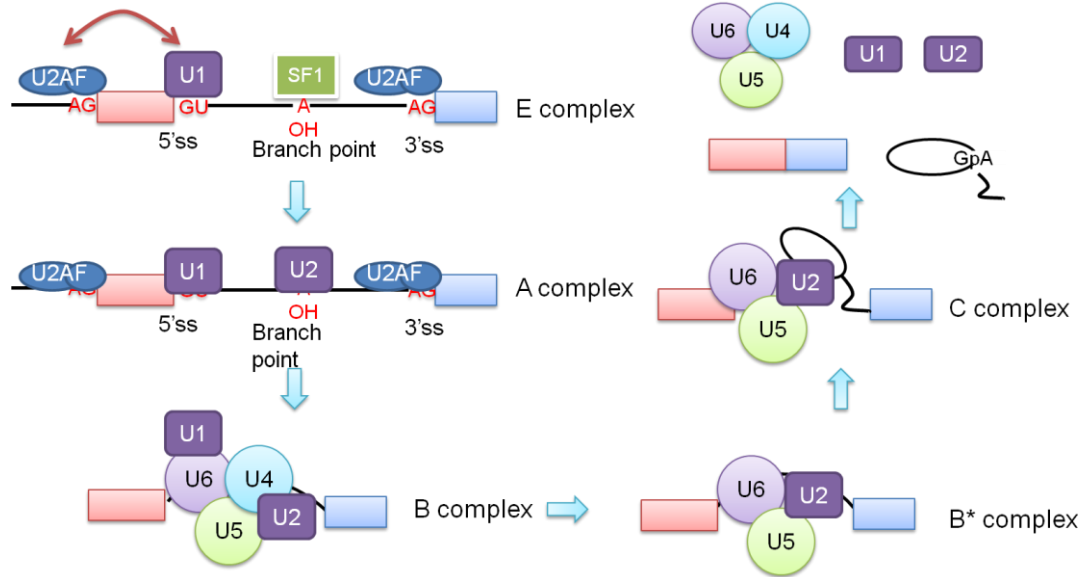


Figure 1.1 Schematic diagram of splicing and dynamic assembly of the spliceosome

U1 snRNP and U2AF bind to the 5'ss and 3'ss, respectively. Their recruitment can be enhanced by each other in the exon definition model (red bidirectional arrow). U2 snRNP is then recruited to the BPS followed by U4/U5/U6 tri-snRNP. Subsequently, U1 and U4 leave and the two catalytic steps occur. Boxes represent exons. Lines represent introns.

1.1.1 Exon definition

Introns generally have the nearly invariant GU and AG dinucleotides at their 5' and 3' ends, respectively (Sheth et al., 2006). Synergistic recruitment of U1 snRNP to the 5'ss and U2 snRNP and U2AF to the downstream BPS and 3'ss helps the recognition of an intron in yeast and lower metazoans which predominantly have small introns in their genomes. In contrast, mammalian genes usually have small exons and large introns. Therefore, the recognition of an exon is facilitated by the exon definition model by which U1 snRNP binding to the 5'ss flanking an exon can help to recruit U2AF to the 3'ss flanking the same exon and vice versa (Figure 1.2). Subsequently, exon juxtaposition occurs to facilitate the spliceosome assembly

across each intron and so excision of the intervening sequence (Berget, 1995; De Conti et al., 2013). Recognition of terminal exons is enhanced by the 5' cap and polyadenylation/ cleavage components such as the pre-messenger RNA cleavage factor I (CFIm) and the cleavage and polyadenylation specificity factor (CPSF), respectively (Figure 1.2) (Kyburz et al., 2006; Lewis et al., 1996; Martinson, 2011). It has been shown that 5' cap is important to stabilize the binding of U1 and U6 snRNP to the 5'ss (Lewis et al., 1996; O'Mullane and Eperon, 1998). In addition, cross-regulation between splicing and polyadenylation at the 3' terminal exon helps to remove the last intron. Mutual enhancement of splicing and polyadenylation can only occur when the polyadenylation/cleavage factors locate downstream of the 3'ss. When the polyadenylation/ cleavage factors locate downstream of the 5'ss, the presence of U1 snRNP at the 5'ss inhibits polyadenylation at intronic region through the interaction between U1 70K and poly A polymerase to prevent the production of truncated mRNA (Berg et al., 2012; Gunderson et al., 1998; Kaida et al., 2010; Martinson, 2011).

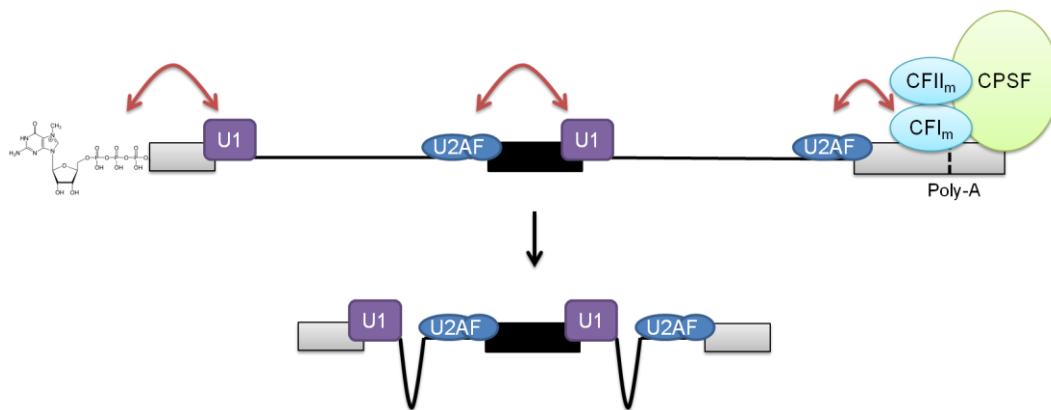


Figure 1.2 Recognition of 5'ss and 3'ss by exon definition

Recruitment of U1 snRNP to the 5'ss enhances the binding of U2AF to the 3'ss flanking the same exon and vice versa. This exon definition model later transforms into exon juxtaposition in order to facilitate the interaction between U1 snRNP and U2AF across intron and so excision of that intron. In addition, recognition of a terminal exon is improved by the 5' cap and polyadenylation machineries.

1.2 Alternative splicing

The proteome diversity in complex organisms is largely generated by alternative splicing, a process by which exons in a pre-mRNA are either retained or excised in various combinations to give rise to various mRNAs and so protein isoforms (Breitbart et al., 1987). Inclusion or exclusion of a region of coding sequences either generates or removes additional protein domains, producing different protein isoforms that may have antagonistic or unrelated functions to regulate cellular processes. Alternative splicing is also coupled to gene expression regulation by inducing nonsense mediated decay (NMD). Skipping of a constitutive exon or inclusion of a poisonous exon or an intron could lead to creation of a premature in-frame stop codon, resulting in mRNA decay (Lewis et al., 2003). In addition, different 3' UTRs derived from alternative splicing have different susceptibility to miRNA regulation. It is known that more than 90% of human genes undergo alternative splicing (Wang et al., 2008). Variations of alternative splicing across tissues are higher than between individuals (Wang et al., 2008). Improper alternative splicing is not only the cause of many genetic diseases such as atypical cystic fibrosis but it is also associated with cancer, implying the importance of accurate splicing (Cartegni et al., 2002; Wang and Cooper, 2007).

1.2.1 Types of alternative splicing

There are seven basic modes of alternative splicing, which are cassette exon, mutually exclusive exons, intron retention, alternative 5'ss or 3'ss, and alternative promoters or poly-A sites (Figure 1.3)(Keren et al., 2010; Nilsen and Graveley, 2010; Wang and Burge, 2008). Cassette exons are alternatively included in mRNAs while only one of the mutually exclusive exons is included in a final splice product. Intron retention refers to the case in which an intron may or may not be excised from the mRNA. The use of alternative 5'ss and 3'ss generates exons with varied lengths and junction sequences. Alternative promoters or poly-A sites associated with different exons generates mRNAs with different terminal exons. The type of alternative splicing I studied is cassette exon which is the most common one. To determine whether a cassette exon is included in or excised from

mRNA, many factors are taken into consideration, which will be discussed in the next section.

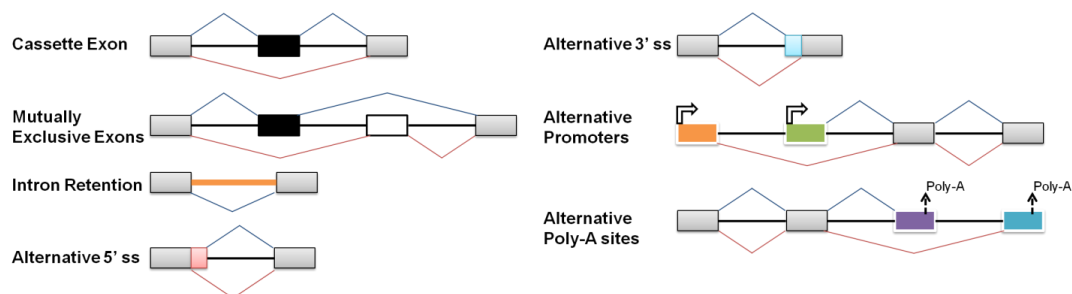


Figure 1.3 Types of alternative splicing

The seven modes of alternative splicing are cassette exon, mutually exclusive exons, intron retention, alternative 5'ss/ 3'ss, alternative promoters or poly-A sites. Image is adapted from Zahler, 2005.

1.2.2 Regulation of alternative splicing

Alternative splicing is regulated by *cis*-acting elements, *trans*-acting factors, transcription, chromatin modifications and RNA secondary structure. The 5'ss, BPS and 3'ss are degenerate in mammalian genomes, therefore additional regulators are required to define an exon. The inclusion of an exon is determined by the balance of positive and negative regulators (Ladd et al., 2005). Generally alternative exons contain weaker splice sites and multiple conserved *cis*-acting elements (Keren et al., 2010).

1.2.2.1 Splice-site strength

The strength of essential *cis*-acting elements which are the 5'ss, BPS and 3'ss largely determines the binding affinity and recruitment efficiency of the spliceosome (Mueller and Hertel, 2012). Recognition of the 5'ss and BPS is dependent on the base-pairing between the pre-mRNA sequence and U1 snRNA and U2 snRNA, respectively. The higher the complementarity between 5'ss and U1

snRNA and between BPS and U2 snRNA, the higher the possibility the exon will be spliced. On the other hand, the length and composition of the polypyrimidine tract are essential for the binding of U2AF proteins. A stretch of 11 uridines was shown to be the strongest polypyrimidine tract for exon inclusion (Coolidge et al., 1997).

The strength of a 5'ss is dependent on its ability to base pair with U1 snRNA. Up to 11 base pairs can be established between U1 snRNA and 5'ss, spanning from the first to the eleventh nucleotide at the 5' end of the snRNA, and the last three exonic nucleotides and first 8 intronic nucleotides at the 5'ss (Figure 1.4A). Only certain mismatches can be tolerated for U1/5'ss base-pairing but the majority of 5'ss are degenerate, suggesting that different mechanisms may be used for 5'ss recognition. Previous works have proved the use of non-canonical base-pairing registers for 5'ss recognition (Roca and Krainer, 2009; Roca et al., 2012). Some atypical 5'ss can complement better with U1 snRNA by shifting the base-pairing 1 nt downstream at the 5'ss side (Figure 1.4B) (Roca and Krainer, 2009). In addition, for certain 5'ss sequences, more stable U1/5'ss base-pairing can be formed when a nucleotide is budged out from the helix at either the 5'ss strand (positions +2 to +5) or the 5' end of U1 snRNA (pseudouridine (Ψ) at position 5 or 6. Ψ is a 5-ribosyl isomer of uridine (Charette and Gray, 2000)) (Figure 1.4C)(Roca et al., 2012). The discovery of non-canonical registers for U1/5'ss base-pairing illustrates the flexibility of U1 snRNA to define different 5'ss and allows a more accurate splicing prediction. Alternative exons are often flanked by a 5'ss which is recognized by the U1 in non-canonical registers. In this thesis, I will show two cassette exons that base pair to U1 in a non-canonical register.

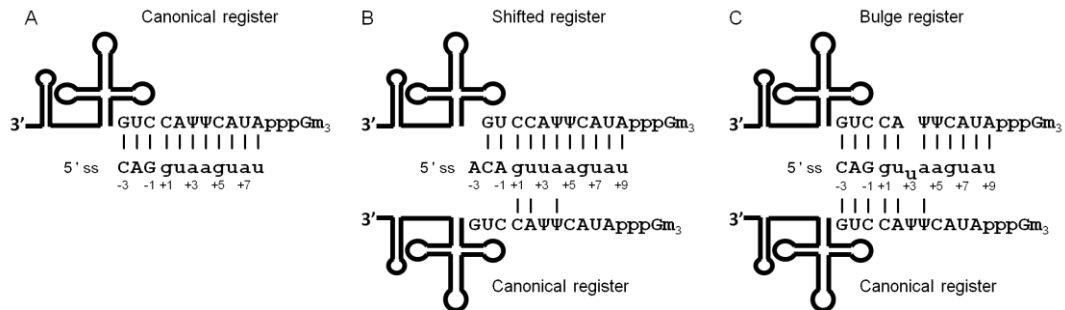


Figure 1.4 U1/5'ss base-pairing registers

(A) U1 snRNA and the consensus 5'ss base pair in the canonical register. (B) The U1/5'ss base-pairing in the shifted register. Only three base pairs can be formed between U1 snRNA and atypical 5'ss in the canonical register. When shifting the base-pairing one nucleotide downstream at the 5'ss strand, more base pairs can be established. (C) The U1/5'ss base-pairing in the bulge register. U1/5'ss base-pairing can be strongly stabilized by bulging out one nucleotide at either the 5'ss or U1 snRNA strand. Uppercase represents exonic sequence and lowercase represents intronic sequence. Exonic position is labeled as +N and intronic position is labeled as -N. Images are adapted from Roca and Krainer, 2009 and Roca et al., 2012.

1.2.2.2 *Cis*-acting elements and *trans*-acting factors

Auxiliary *cis*-acting elements are present within exonic or intronic sequences to regulate exon recognition by either promoting or preventing the binding of snRNPs to the essential *cis*-acting elements. Sequences that increase exon inclusion are enhancers while sequences that increase exon skipping are silencers. Depending on their positions, *cis*-acting elements can be either exonic or intronic splicing enhancers or silencers (ESEs/ESSs, ISEs/ISSs) (Cartegni et al., 2002; Wang and Burge, 2008). *Trans*-acting factors include snRNPs, splicing activators and repressors. They bind to the *cis*-acting elements with certain specificity for functioning (Table 1.1). Splicing activators like SR proteins bind to the ESEs or ISEs and promote exon inclusion by facilitating the recruitment of the spliceosome and antagonizing the negative effects caused by the silencers (Figure 1.5A). On the other hand, splicing repressors like heterogeneous nuclear ribonucleoprotein (hnRNP) inhibit U1 and U2 snRNP recruitment by creating a steric block, inhibiting exon or intron definition and looping-out the exon by binding to two distant ISSs (Figure 1.5B)(Cartegni et al., 2002; Izquierdo et al., 2005; Okunola and Krainer, 2009; Sharma et al., 2005; Wagner and Garcia-Blanco, 2001). Besides, an extended U1/5'ss interaction caused by hnRNP L and hnRNP A1 has been shown to repress the splicing by pausing the spliceosome assembly (Chiou et al., 2013). Cooperative binding of regulators to RNA through direct protein-protein interaction can synergistically enhance the regulatory effects. In contrast, competitive binding of regulators to the same binding site or nearby region tilts the regulatory direction.

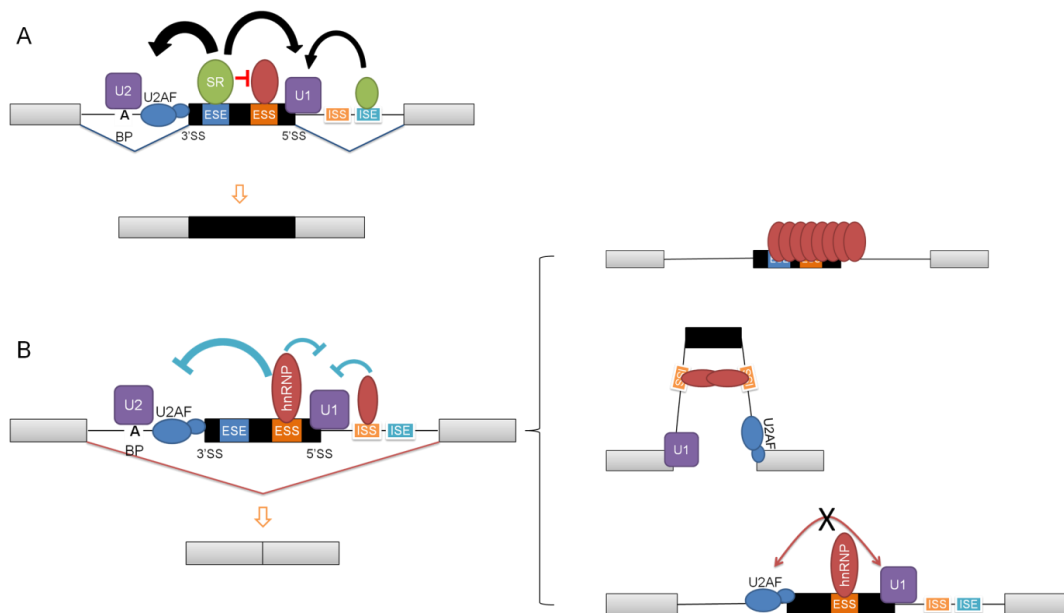


Figure 1.5 Regulatory mechanisms of *trans*-acting factors

(A) ESEs or ISEs are bound by splicing activators like SR proteins which help to recruit the spliceosome and promote exon inclusion. Activators can also inhibit the activity of nearby repressors. (B) ESSs or ISSs provide binding sites for splicing repressors such as hnRNPs to inhibit exon inclusion. The repression can be achieved by different ways. First, recruitment of more repressors by the first bound protein blocks the access of the spliceosome to the splice sites or activator to enhancer sequences. Second, interaction between repressors located at the two flanking introns loops out the alternative exon and causes exon skipping. Third, repressors can prevent exon definition.

In addition to these two major groups of splicing factors, several tissue-specific splicing regulators like neuro-oncological ventral antigen (NOVA), Fox proteins, embryonic lethal abnormal visual (Elav), CUG-BP and ETR-3 Like factors (CELF) and muscleblind-like proteins (MBNL) are present in mammals (Table 1.2). These factors contribute to the tissue-specific splicing events which could be important to establish the physiological functions (Buckanovich and Darnell, 1997; Kino et al., 2009; Ladd et al., 2001; Lisbin et al., 2001; Sun et al., 2012).

The regulatory activities of these RNA binding proteins are context and positional dependent (Fu and Ares, 2014). hnRNPs usually have inhibitory effect on exon inclusion when bound to exons but they can either promote or repress exon inclusion when bound to introns. NOVA is the first protein showing intronic positional effects on splicing (Ule et al., 2006). It usually enhances exon recognition by binding to the downstream region of an exon and inhibits exon inclusion by binding to the upstream region of an exon (Licatalosi et al., 2008). An analogous intron binding positional model also applies to PTBP1 and RBFOX1 (Llorian et al., 2010; Sun et al., 2012). Selection of a splice site involves a competition among several splice sites and hence internal exon inclusion can be improved if the downstream splice site is less favorable for spliceosome recognition due to the binding of repressors, and vice versa. Splicing regulators can auto-regulate their own pre-mRNA splicing or cross-regulate splicing of other factors to adjust the expression level and competency of regulators, and so change the splicing outcomes (Bonomi et al., 2013; Coutinho-Mansfield et al., 2007; Rossbach et al., 2009; Sun et al., 2010). The activities of splicing regulators can be modulated by post-translational modifications like phosphorylation which is usually triggered by signaling pathways (Fu and Ares, 2014; Stamm, 2008). Despite all this information, the detailed mechanisms of how signaling pathways regulate alternative splicing are yet to be elucidated.

Table 1.1 List of common SR or SR-like proteins and hnRNPs

Name	Binding Domains	Binding motif
SRSF1 (ASF/SF2)	RRM, RRMH and RS	RGAAGAAC
SRSF2 (SC35)	RRM and RS	UGCUGUU
SRSF3 (SRp20)	RRM and RS	GCUCCUCUUC
SRSF4 (SRp75)	RRM, RRMH and RS	GAAGGA
SRSF5 (SRp40)	RRM, RRMH and RS	AGGAGAAGGGA
SRSF6 (SRp55)	RRM, RRMH and RS	GGCAGCACCUG
SRSF7 (9G8)	RRM, zinc finger and RS	(GAC) _n
SRSF9 (SRp30c)	RRM, RRMH and RS	CUGGAUU
TRA2 α	RRM and two Arg-rich	GAAARGARR
TRA2 β	RRM and two RS	(GAA) _n
hnRNP A1	RRM, RGG and G	UAGGGA/U
hnRNP A2	RRM, RGG and G	(UUAGGG) _n
hnRNP B1		
hnRNP C1	RRM	U rich
hnRNP C2		
hnRNP F	RRM, RGG and GY	GGGA and G rich
hnRNP G	RRM and SRGY	CC(A/C) and AAGU
hnRNP I (PTBP1)	RRM	UCUU and CUCUCU
hnRNP L	RRM	C and A rich
hnRNP LL	RRM	C and A rich
hnRNP M	RRM and GY	Poly G Or Poly U

Adapted from Chen and Manley, 2009. This table shows an example of binding motifs for different protein regulators. Many novel binding motifs for each protein are also identified in recent studies.

Table 1.2 List of tissue-specific splicing factors

Name	Binding domain	Binding motif	Tissue expression
nPTB	RRM	CUCUCU	Neurons, myoblasts and testes
Nova1	KH	YCA Y	Neurons of the hindbrain and spinal cord
Nova2	KH	YCA Y	Neurons of the cortex, hippocampus and dorsal spinal cord
Fox1	RRM	(U)GCAUG	Muscle, heart and neurons
Fox2	RRM	(U)GCAUG	Muscle, heart and neurons
Rbm35a	RRM	GU rich	Epithelial cells
Rbm35b	RRM	GU rich	Epithelial cells
TIA1	RRM	U rich	Brain, spleen and testes
TIAR	RRM	U rich	Brain, spleen, lung, liver and testes
MBNL	CCCH zinc finger domain	YGCU(U/ G)Y	Muscles, uterus and ovaries
CELF1	RRM	U and G rich	Brain
ETR3	RRM	U and G rich	Heart, skeletal muscle and brain
CELF4	RRM	U and G rich	Muscle
CELF5	RRM	U and G rich	Heart, skeletal muscle and brain
CELF6	RRM	U and G rich	Kidney, brain and testes

Adapted from Chen and Manley, 2009.

1.2.2.2.1 SR proteins

The SR protein family is a group of homologous RNA-binding proteins consisting of an N-terminal RNA recognition motif/s (RRM/s) and a C-terminal arginine and serine dipeptide rich (RS) domain (Howard and Sanford, 2015). SR proteins are capable of complementing *in vitro* splicing and regulate both constitutive and alternative splicing. This protein family has a total of twelve members, that are named SRSF1 to SRSF12, and each SR protein has its own binding specificity (Table 1.1). Though SR proteins generally enhance exon inclusion, they can sometimes repress exon inclusion in a positional and context dependent manner (Erkelenz et al., 2013). Phosphorylation of the RS domain by SR-specific protein kinase (SRPK) or cdc-2 like kinase/serine, threonine and tyrosine kinase (Clk/Sty) regulates activity and localization of SR proteins (Colwill et al., 1996; Misteli et al., 1998; Prasad et al., 1999; Yeakley et al., 1999). Hyperphosphorylation of SR proteins by Clk/Sty releases them from nucleus speckles to participate in splicing (Colwill et al., 1996). During splicing, SR proteins are dephosphorylated and these hypophosphorylated SR proteins remain on the mRNA to facilitate mRNA export (Huang et al., 2004). On the other hand, rephosphorylation of SR proteins by Clk/Sty releases SR proteins from mRNAs to be recycled for the next splicing event. Several SR proteins (SRSF1, SRSF3, SRSF7 and SRSF10) shuttle between cytoplasm and nucleus (Caceres et al., 1998). In the cytoplasm, SR proteins like SRSF1, SRSF3 and SRSF7 can regulate translation (Sanford et al., 2004), and SR proteins can efficiently trigger NMD to degrade premature translation-termination codon (PTC) containing transcripts (Zhang and Krainer, 2004). Cytoplasmic SR proteins are subsequently phosphorylated by SRPK for transportin-mediated nuclear entry (Lai et al., 2001).

1.2.2.2.2 SRSF1

SRSF1 is the founding member of the SR protein family. Its N terminus consists of a canonical RRM followed by a pseudo-RRM while its C terminus comprises an RS domain (Das and Krainer, 2014). SRSF1 was concurrently identified by two groups showing that it is able to promote spliceosome assembly and complement *in*

in vitro splicing with HeLa S100 extract as well as to regulate alternative splicing of the SV40 early pre-mRNA (Ge and Manley, 1990; Krainer et al., 1990a, b). The SRSF1 consensus binding motif identified by different methods like *in vitro* selection, cross-linking immunoprecipitation (CLIP) and RNA sequencing (RNA-seq), is GA rich (Anczukow et al., 2015; Sanford et al., 2008; Smith et al., 2006). SRSF1 regulates both constitutive and alternative exons and it generally promotes inclusion of the exon where it binds. The enhancement is likely mediated through direct recruitment of the spliceosome via protein interaction between SRSF1 RRM8s and U1-70K, which is a component of the U1 snRNP (Cho et al., 2011). On the other hand, when SRSF1 binds to the downstream constitutive exon, it causes skipping of the upstream alternative exon (Ghigna et al., 2005). SRSF1 shuttles between cytoplasm and nucleus to exert different cellular functions (Caceres et al., 1998). The activity and localization of SRSF1 is regulated by phosphorylation, and its expression can be regulated by itself in a negative feedback manner using alternative splicing coupled NMD (Sanford et al., 2005; Sun et al., 2010). Hypophosphorylated SRSF1 can facilitate nuclear export of spliced mRNA by serving as an adaptor of TAP which is a nuclear export factor (Huang et al., 2003; Lai and Tarn, 2004).

SRSF1 regulates translation in splicing-dependent and -independent ways. In the cytoplasm, SRSF1 can directly interact with mTOR to promote phosphorylation of 4EBP1 which is a translational inhibitor, leading to activation of eIF4E followed by initiation of cap-dependent translation (Karni et al., 2008; Michlewski et al., 2008). Alternatively, SRSF1 modulates the alternative splicing of MNK2 and S6 kinase 1 to produce isoforms that enhance translation (Anczukow et al., 2012; Ben-Hur et al., 2013). The ability of SRSF1 to initiate translation contributes to its involvement in triggering NMD (Zhang and Krainer, 2004). In the nucleus, SRSF1 not only regulates splicing but also facilitates Drosha-mediated cleavage of primary miRNA as well as maintains genome stability by preventing formation of R-loop structure during transcription (Li and Manley, 2005; Wu et al., 2010). On the other hand, SRSF1 forms a complex with RPL5/MDM2 upon ribosomal stress, resulting in stabilization of p53 which induces cellular senescence (Fregoso et al., 2013).

SRSF1 is overexpressed in many cancers and its overexpression is regulated by oncogenic transcription factor MYC and splicing regulator Sam68 (Anczukow et al., 2012; Karni et al., 2007). These two regulators increase SRSF1 expression by either enhancing transcription or switching alternative splicing (Das et al., 2012; Valacca et al., 2010). Alternative splicing of many genes which are involved in cell proliferation, apoptosis and cell motility are regulated by SRSF1. Excess SRSF1 causes production of isoforms that favor cellular transformation. In addition, SRSF1 can enhance translation to promote cell proliferation. As mentioned earlier, SRSF1 can induce cellular senescence by stabilizing p53, suggesting that the oncogenic activity of SRSF1 requires disruption of p53. SRSF1 plays many different roles in the cytoplasm and nucleus to maintain cellular homeostasis implying the importance of SRSF1, and hence its misregulation leads to disease.

1.2.2.3 Transcription and chromatin

Transcriptional rate is one of the determinants for exon inclusion. In presence of two competing 5'ss, when the transcription is fast, exons flanked by stronger 5'ss are favored, thus resulting in skipping of cassette exons flanked by weak 5'ss (Nilsen and Graveley, 2010). In contrast, slow RNA polymerase II elongation allows more time for exon recognition by the spliceosome and so increases exon inclusion with some exceptions (Schor et al., 2013). Slow elongation not only provides more interaction time for spliceosome assembly, but it also provides more chance for a repressor to interact with the pre-mRNA leading to more exon skipping (Dujardin et al., 2014). In addition, the carboxy terminal domain of RNA polymerase II provides binding sites for splicing and transcription factors which can regulate splicing efficiency (Schor et al., 2013). Nucleosomes are preferentially positioned over exons and several histone modifications like tri-methylation of histone H3 lysine 36 (H3K36me3) are enriched over exonic region (Luco et al., 2010; Schwartz et al., 2009; Tilgner et al., 2009). The chromatin structure could affect the transcriptional elongation rate which in turn regulates splicing (Luco et al., 2011). Besides, the histone modifications provide a platform for binding of splicing activators or repressors which in turn modulate splicing (Zhou et al., 2014).

Remodeling of histone modifications is time consuming and so it is unlikely the major regulator to modulate splicing upon environmental stimuli. Rather, the histone modifications probably predetermine the basal inclusion level of an exon.

1.2.2.4 RNA secondary structure

RNA tends to form secondary structure by intramolecular base-pairing (Fontana et al., 1993). Such RNA secondary structures can mask the core splicing elements or splicing enhancers, and thus result in exon skipping (Blanchette and Chabot, 1997; Estes et al., 1992; Goguel et al., 1993). On the other hand, the formation of RNA loops can bring two distant splice sites in close proximity and so favor exon inclusion. Binding of some *trans*-acting factors depends on the RNA secondary structure as well as the RNA sequence (Buckanovich and Darnell, 1997; Damgaard et al., 2002), therefore disruption of secondary structure by mutations could affect the protein binding and splicing.

1.2.3 Alternative splicing and NMD

NMD is a translation-coupled mechanism to degrade mRNAs containing an in-frame PTC (Schweiggruber et al., 2013). Translation of mRNAs containing PTCs produces truncated proteins which could have a dominant negative function. Therefore, NMD is important for protein quality control. After splicing, the exon junction complexes (EJC) are placed at 20-24 nt upstream of the exon-exon junctions. At the first, so-called pioneer round of translation, a ribosome displaces the EJCs along the mRNA. When the ribosome's movement is halted at a termination codon, eukaryotic release factor complex eRF1-eRF3 is recruited to ribosome to terminate translation and release the newly synthesized protein. The close distance between the ribosome and the polyadenylation binding protein complex at the 3' UTR prevents interaction between the NMD factor up-frameshit 1 (UPF1) and eRF3. However, when the ribosome stalls at a PTC which is far away from the 3' UTR, UPF1 is then able to interact with eRF3 to trigger NMD. EJC/s which is/are not dissociated from mRNA subsequently facilitate/s the recruitment of UPF2, UPF3 and suppressor of morphogenetic effect on genitalia 1

(SMG1), SMG8, SMG9 and DEAH box polypeptide 34 (DHX34) to assemble an NMD-activating complex together with UPF1. Interaction between UPF1 and UPF2 allows phosphorylation of UPF1 N- and C-terminal domains by SMG1 followed by recruitment of factors that regulate decapping, deadenylation and endonucleolytic cleavage (Brojna and Wen, 2009; Maquat, 2004).

Generally a termination codon located more than 50 nt upstream of the last-exon-exon junction is considered as a PTC, as the EJC at the last exon-exon junction is probably not displaced by the ribosome. PTCs can be generated by DNA mutations (nonsense mutations), genomic rearrangements such as V(D)J recombination of T/B-cell receptor, alternative translation and alternative splicing (Brojna and Wen, 2009; Lykke-Andersen and Jensen, 2015; Maquat, 2004).

Alternative splicing of a poisonous exon or intron retention introduces an in-frame PTC to an mRNA which is then targeted by NMD (Figure 1.6). Coupling alternative splicing with NMD allows for rapid modulation of protein expression level. Many *trans*-acting factors are regulated by this alternative splicing coupled NMD in an auto-regulatory negative feedback mechanism (Hamid and Makeyev, 2014; McGlincy and Smith, 2008; Sun et al., 2010). Excess protein regulators bind to their own mRNAs and promote inclusion or skipping of the poisonous exon to produce PTC-containing mRNAs in order to reduce their expression. In addition, cross-regulation between regulators like PTBP1 and nPTB using alternative splicing coupled NMD can repress expression of one protein in some cell types (Boutz et al., 2007; Coutinho-Mansfield et al., 2007). Generally, alternative splicing coupled NMD is useful to modulate expression of a protein or a protein isoform in response to different cell conditions.

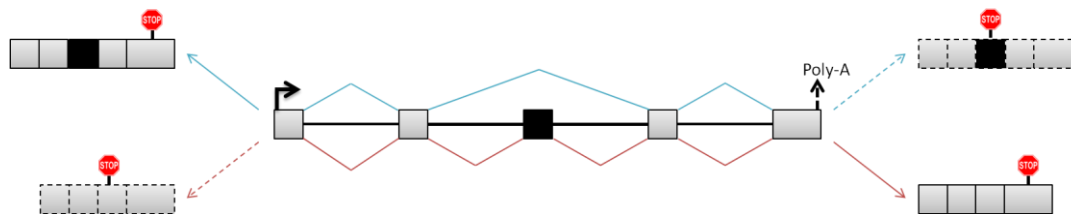


Figure 1.6 Alternative splicing coupled with NMD

Inclusion of an exon containing an in-frame PTC in mRNA results in mRNA decay (right panel). On the other hand, skipping of an exon changes the reading frame and creates an in-frame PTC, leading to mRNA degradation by NMD (left panel). Square boxes depict exons. Lines depict introns. mRNAs outlined with dashed lines are the NMD targets. Image is adapted from Hamid and Makeyev, 2014.

1.2.4 Alternative splicing and diseases

Many genetic diseases are caused by aberrant splicing due to mutations in *cis*-acting elements or in components of the splicing machinery (Cooper et al., 2009). Mutations in *cis*-acting elements only affect the particular genes. The mutations can disrupt the recognition of constitutive exons resulting in production of truncated proteins or triggering NMD. As well, these mutations may favor or disfavor inclusion of an alternative exon, resulting in imbalanced production of protein isoforms which could have antagonistic functions. One disease example for mutations in *cis*-acting elements is cystic fibrosis. It is an autosomal recessive disease caused by the loss of expression of cystic fibrosis transmembrane conductance regulator (CFTR) due to aberrant splicing (Castellani et al., 2008; Cutting, 2015). The severity of cystic fibrosis depends on the expression level of CFTR. Multiple splicing mutations at different exons or introns of the *CFTR* gene have been identified and those mutations cause little or no production of functional CFTR protein. For instance, the polymorphism at the polypyrimidine tract upstream of exon 9 reduces exon inclusion and results in production of non-functional CFTR (Dork et al., 1994).

Mutations in *trans*-acting factors affecting the spliceosome or auxiliary splicing regulators have a bigger and wider impact on splicing and cause a more complex disease compared to *cis*-mutations. MBNL1 regulates splicing events that are required for muscle cell differentiation. A CUG expansion at the 3' UTR of *myotonic dystrophy protein kinase* mRNA forms a RNA hairpin that binds and sequesters MBNL1 (Lee and Cooper, 2009). The loss of MBNL1 caused by sequestration generates non-muscle specific splice isoforms. On the other hand, increased expression of CUG-binding protein (CUGBP) by nuclear accumulation of this MBNL1-bound mRNA further alters the splicing events. MBNL1 and CUGBP regulate many splicing events in opposite directions, and so the decreased MBNL1 and increased CUGBP have a synergistic impact on splicing. Several pre-mRNAs such as *cardiac troponin T (cTNT)*, *insulin receptor* and *chloride channel 1 (CLC-1)* are aberrantly spliced and their missplicing is associated with the symptoms of myotonic dystrophy which are cardiac conduction defect, insulin resistance and myotonia, respectively (Lee and Cooper, 2009).

However, in the case of spinal muscular atrophy (SMA) which is characterized by progressive loss of motor neurons resulting in skeletal muscle atrophy, a decrease in expression of functional survival of motor neuron (SMN) which is required for snRNP assembly, only causes a cell-type specific defect (Lorson et al., 1999; Ruggiu et al., 2012). The reason behind the disease pathogenesis is still not well understood but it is speculated that cell-type specific highly expressed genes are more susceptible to the reduction of snRNP levels.

In cancer cells, many splicing regulators like SRSF1 and hnRNP A1 are upregulated (Anczukow et al., 2012; Wang and Cooper, 2007). Altered signaling pathways that regulate protein modifications like phosphorylation of splicing factors are also identified in cancer (David and Manley, 2010). The overexpression or altered regulatory activities of splicing factors indirectly alters the alternative splicing of apoptotic genes like *Bcl-x*, proto-oncogenes like *cyclin D1* and metastatic genes like *RON* to favor the expression of splice isoforms that promote

cell proliferation, evasion of apoptosis and metastasis (David and Manley, 2010; Faustino and Cooper, 2003).

Understanding splicing codes is useful for prognosis and diagnosis of genetic diseases. However, the prediction accuracy for splicing regulatory elements by the current splicing prediction tools is still low (Douglas and Wood, 2011). Treatment for genetic diseases caused by splicing defects can be achieved by several methods. First, small molecules can block the sequestration of splicing factors by RNA-sequence repeats in the case of myotonic dystrophy. Second, antisense-oligonucleotides (ASOs) can bind to *cis*-acting elements and prevent the recruitment of the spliceosome or splicing factors to modulate splicing. Bifunctional ASOs containing additional protein binding domain, on the other hand, can facilitate the recruitment of splicing factors. Alternatively, inclusion of an exon whose flanking 3'ss or 5'ss is mutated can be rescued by trans-splicing using ASOs containing the 3' or 5' exon/s of the target splice site (Havens et al., 2013).

Cooperative interactions between *cis*-acting elements, protein regulators, RNA secondary structure and histone modifications determine whether an exon is included in an mRNA (Barash et al., 2010; Xiong et al., 2015). The difficulty of deciphering the splicing code due to the complexity of splicing regulation requires thorough studies to improve our current understanding. In this thesis, *CD46* was selected to study splicing regulatory mechanisms. It has two regions of cassette exons that are regulated by unknown mechanisms. *CD46* will be discussed in detail in the next few sections.

1.3 The complement system

The complement system, as one of the oldest immune systems, is part of the innate immunity to eliminate early infection and mediate innate and adaptive immune systems (Sarma and Ward, 2011). The complement system consists of more than 50 membrane-bound and plasma proteins. It can be activated by the classical, alternative and lectin pathway, resulting in removal of pathogens by opsonization, cell lysis and inflammation (Figure 1.7). The activation of complement pathway involves a series of protein cleavages by serine proteases. The three pathways are initiated differently but they all converge at the formation of C3 convertase.

The classical pathway is initiated by the activation of C1 complex which consists of C1q, C1r and C1s. Binding of C1q to the immunoglobulin M or G (IgM or IgG) leads to activation of C1r which is a serine protease followed by cleavage of C1s which is also a serine protease. The activated C1r and C1s are capable of cleaving C4 and C2 to form the C3 convertase (C4b2a) which in turn generates C3b. Subsequently, C4b2a and C3b form the C5 convertase to split C5 into C5b and C5a. C5b then complexes with C6, C7, C8 and C9 to form the membrane attack complex (MAC) which causes cell lysis. The alternative pathway is activated by low level of continuous hydrolysis of C3. C3b generated by hydrolysis forms an alternative C3 convertase (C3bBb) together with factor B. The activation of mannose-binding lectin associated serine proteases (MASP) by its recognition of mannose-binding lectin (MBL) bound to pathogen-associated molecular patterns (PAMPs) leads to the cleavage of C4 and C2, and initiation of the lectin pathway (Figure 1.7)(Dunkelberger and Song, 2010; Sarma and Ward, 2011).

Activation of complement can lead to phagocytosis of C3b-opsonized pathogens, MAC-mediated cell lysis and production of anaphylatoxins like C3a, C4 a, C5a. Complement deficiency caused by dysfunction or absence of complement proteins may result in recurrent infections due to immunodeficiency. However, a lack of activation of the classical pathway (C1q, C4 and C2) predisposes to systemic lupus erythematosus (SLE) which is an autoimmune disease caused by antibodies that target self-antigens resulting in excessive inflammation (Bryan and Wu, 2014;

Truedsson et al., 2007). Complement deficiency impairs the clearance of apoptotic cells which are the main source of autoantigens and inefficiently removes the immune complexes which can cause tissue damage by inflammation. The complement system is also essential for developing B-cell self-tolerance. Hence, complement activation is not only essential for pathogen elimination but also it may be involved in maintaining immune homeostasis.

In addition, deregulation of complement activation is harmful to the host as it can result in autoimmune diseases as characterized by excessive inflammation and tissue damage. Therefore, a proper regulation of complement activation is essential. Several plasma and membrane bound complement control proteins like Factor I, Factor H, CR1, CD55 and CD46 are present to prevent complement activation on host cells by inhibiting the complement activation cascade, disrupting the assembly of MAC and degrading the anaphylatoxins (Dunkelberger and Song, 2010; Oikonomopoulou et al., 2001; Sarma and Ward, 2011; Zipfel and Skerka, 2009). Complement activation can cause inflammation yet its deficiency can also result in autoimmune responses. Therefore, maintaining a moderate complement activation is essential for body wellbeing.

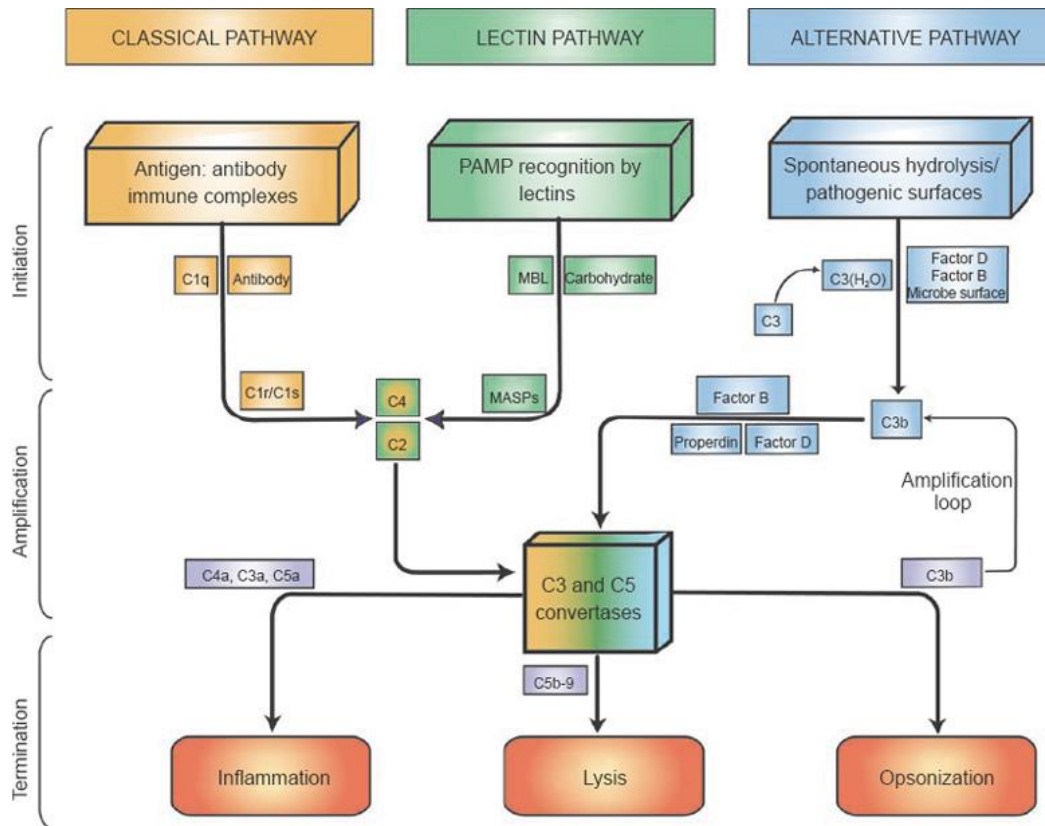


Figure 1.7 The complement pathway

Initiation of the classical, alternative and lectin pathway by recognition of antigen:antibody complex, PAMP:MBL complex and spontaneous hydrolysis of C3, respectively, leads to the formation of C3 and C5 convertase. The consequences of complement activation are inflammation, cell lysis and opsonization. Image is from Dunkelberger and Song, 2010.

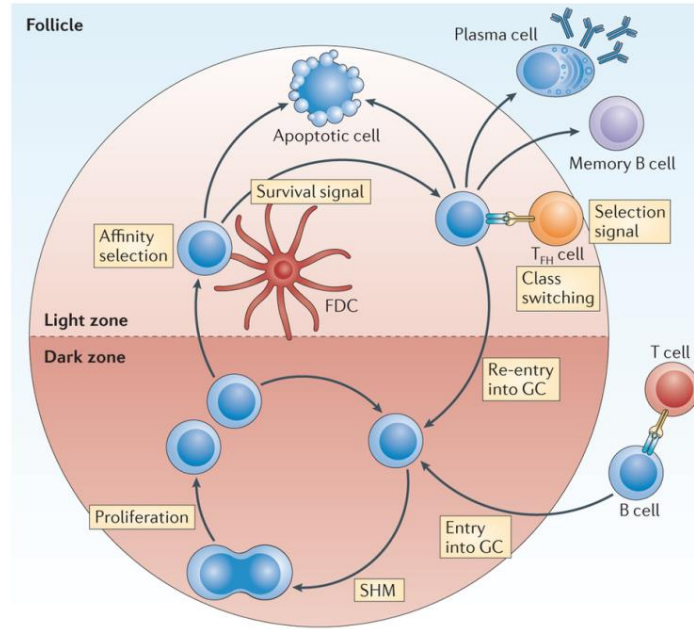
1.4 Adaptive immune system

Adaptive immunity, another part of immune system, is induced by antigen presentation and so it reacts slower than the innate immune response. It is highly specific and able to generate memory cells to protect host from future infections by the same pathogen. The two major players in adaptive immunity are T cells and B cells. The high diversity and specificity of adaptive immunity is generated by random recombination of variable, diversity and joining gene segments (V(D)J recombination) which encode T- and B-cell receptors. Clonal selection of each T cell or B cell bearing different receptors is necessary to remove immune cells carrying receptors that target a self-antigen or are unable to respond to foreign antigens (Rote, 2013).

1.4.1 B cells

The name of B cells came from the bursa of Fabricius which is an organ where B cells mature in birds. B cells fight infections by antibody production and its generation and development occur in bone marrow in humans. Immature B cells leave bone marrow and continue developing into naïve mature B cells in an antigen-independent manner. Mature B cells then migrate to secondary lymphoid organs like spleen and lymph nodes where they activate upon encountering an antigen and differentiate to antibody-producing plasma cells. B-cell activation requires two signals which are from B-cell receptor (BCR)-antigen crosslinking and from either $CD4^+$ T_{helper} (T_H) cell interaction or Toll like receptor (TLR) expressed by B cells. Depending on the second signal, B-cell responses are different in responding duration, class of antibody produced, antibody affinity and ability to become memory cells. T cell-independent B-cell activation is dependent on BCR and TLR signals. Expansion of plasma cells secreting IgM antibodies with low antigen affinity only occurs for a short period. This T cell-independent activated B cells have less competency to form memory cells (Naradikian et al., 2014; Tobon et al., 2013).

On the other hand, B cells acquire, process and present antigens to CD4⁺ T_H cells in the germinal center (Figure 1.8). Activated CD4⁺ T_H1 cells then provide a co-stimulatory signal to B cells via CD40-CD40L and secrete cytokines like IL-4 to induce B-cell proliferation. B cells proliferate in the dark zone of the germinal center where they also undergo somatic hypermutations (SHM) to modify the BCR (antibody) coding gene. Generation of BCR variants through mutations may improve their affinity to antigen. B cells then enter the light zone of germinal center to examine if their BCR have improved affinity. Only those B cells carrying high affinity BCR receive survival signals from follicular dendritic cells and follicular CD4⁺ T_H cells while the rest undergo apoptosis. This process known as affinity maturation lasts for several rounds before B cells differentiate to plasma cells or memory cells. Antibodies function in two different ways to cope with infection. First, antibodies can neutralize either pathogens or toxins to prevent their binding to cell surface. Second, immunoglobulin complexes can trigger the classical complement pathway to remove pathogens via opsonization and phagocytosis (Heesters et al., 2014; Naradikian et al., 2014; Rote, 2013; Tobon et al., 2013).



Nature Reviews | Immunology

Figure 1.8 B-cell activation in germinal center

B cells present antigen to T cells and receive co-stimulatory signal. Activated B cells enter the dark zone of the germinal center for proliferation. During proliferation, B cells undergo somatic hypermutations to modify gene coding for antibody. Subsequently, B cells enter the light zone where they receive survival signals from dendritic cells and T cells if their antibodies have higher affinity to antigen. Finally B cells either differentiate to plasma/memory cells or enter another cycle of SHM and affinity selection. Image is from Heesters et al., 2014.

1.4.2 T cells

The other main player in the adaptive immunity is T cells. Lymphoid progenitor cells from bone marrow migrate to thymus where they differentiate to naïve mature T cells. Two main T cell subsets are $CD4^+$ and $CD8^+$ T cells distinguished by the expression of surface co-receptor CD4 or CD8 and their recognition of different major histocompatibility complex (MHC)-antigen complexes. $CD4^+$ T cells

recognize MHCII-antigen complexes on antigen presenting cells (APCs) like B cells, macrophages and dendritic cells with the T-cell receptor (TCR) and CD4. On the other hand, the TCR and CD8 of CD8⁺ T cells bind to MHCI-antigen complexes expressed on any nucleated cells. Upon activation, T cells become effector T cells with various immune functions to manage infection. Antigen recognition by TCR alone is not sufficient to induce T-cell activation. Additional co-stimulatory signal via CD28 on T cells and CD80/CD86 on APC is necessary for proper T-cell activation. The TCR is non-covalently associated with CD3 which contains immunoreceptor tyrosine-based activation motifs (ITAMs) at its cytoplasmic tail. Upon ligand binding to TCR, phosphorylation of ITAMs by protein tyrosine kinase (PTK) like lymphocyte-specific protein tyrosine kinase (Lck) and Fyn provides a binding site for PTK ZAP-70 which is then activated by Lck-mediated phosphorylation. The activated ZAP-70 then phosphorylates SLP-76 to induce actin cytoskeleton remodeling, and linker of activated T cells (LAT) which is a membrane protein providing binding sites for growth factor receptor-bound protein 2 (Grb2) and phospholipase C- γ (PLC γ) to induce gene expression. Activated PLC γ cleaves phosphatidylinositol 4,5-biphosphate (PIP2) to diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP3) to induce protein kinase C (PKC) and calcium-mediated signaling pathways, respectively. All signaling cascades lead to activation of transcription factors such as NF-AT, ELK-1, Jun and ATF-2 to initiate immune gene expression (Figure 1.9)(Broere et al., 2011; Smith-Garvin et al., 2009).

Activated CD8⁺ T cells recognize MHCI-antigen complexes on infected cells and kill them by two different pathways. T cells can secrete cytolytic granules containing perforin and granzyme B to target cells. Perforin creates pores on target membrane allowing uptake of granzyme B which induces apoptosis in the target cell. Another mechanism is the Fas-mediated apoptosis. Here, the interaction between the Fas ligand on the T cell and the Fas receptor on the target cell triggers a caspase activation cascade and so causes cell death. Therefore, CD8⁺ cytotoxic T cells are in charge of eliminating intracellular pathogens (Broere et al., 2011).

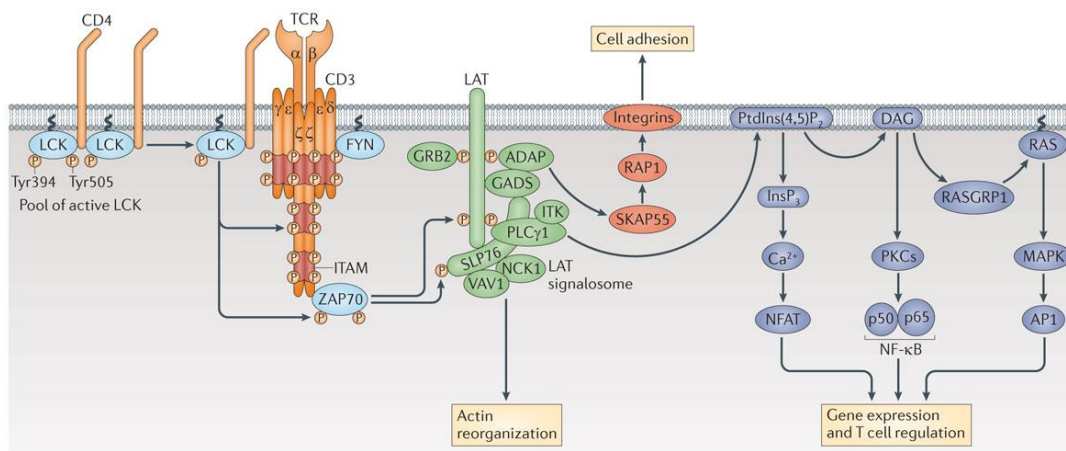


Figure 1.9 T-cell signaling

TCR crosslinking leads to a series of phosphorylations of kinases and triggers downstream pathways, resulting in actin reorganization and gene expression.

Image is from Brownlie and Zamoyska, 2013.

On the other hand, CD4⁺ T cells are not only the helpers to enhance immune responses of CD8⁺ T cells, B cells and macrophages, but they are also involved in immunosuppression. There are several subsets of CD4⁺ T cells established by distinct master transcription factors, secreted cytokine profiles and immune regulatory functions. Differentiation of CD4⁺ T cells to each subset is dependent on environmental cytokines and APCs (Table 1.3). T_H1 cells help to eliminate intracellular pathogens by producing interferon gamma (IFN γ) and IL-2. IFN γ promotes expression of MHC I and MHC II on the APCs as well as enhances phagocytosis of macrophages, while IL-2 is necessary for proliferation of CD8⁺ T cells. In contrast, T_H2 cells are involved in removal of extracellular parasites and allergy. Antibody production by plasma B cells is boosted by the cytokines secreted by T_H2 cells. The main secreted cytokines are IL-4, IL-5, IL-9, IL-10 and IL-13. IL-4 induces IgE class switching and production, and in turn IgE activates mast cells and basophils to release histamine, serotonin and cytotoxic substances to remove parasites. IL-5, IL-9 and IL-13 also activate eosinophils, neutrophils and mast cells to release inflammatory chemicals. Allergen-triggered activation of T_H2

cells and so of granulocytes and mast cells release histamine and inflammatory chemicals which cause allergy symptoms like itchiness and anaphylaxis. On the other hand, IL-17A and IL-17F secreted by T_H17 cells induce production of pro-inflammatory cytokines and chemokines to attract macrophages and neutrophils to the site of infection to remove extracellular bacteria or fungi. Besides, the ability of T_H17 cells to provoke inflammation is associated with autoimmune diseases.

In addition to mounting immune response, there is a subset of CD4⁺ T cells negatively regulating T-cell activity to maintain immune tolerance and cease immune response after pathogen clearance. Such T regulatory (Treg) cells are either derived from the thymus as natural regulatory T cells (nTreg) or induced at the periphery like induced regulatory T cells (iTreg) and T regulatory 1 cells (Tr1). Regulatory T cells produce IL-10 which is an anti-inflammatory cytokine that suppresses T-cell proliferation and activity, and kill activated T cells using granzyme B in a contact-dependent manner. Variations of T-cell effector functions allow efficient removal of both intracellular and extracellular pathogens as well as to maintain homeostasis of the immune system (Broere et al., 2011; Luckheeram et al., 2012).

Table 1.3 Cytokines and transcription factors that induce CD4⁺ T-cell differentiation

CD4 ⁺ Subset	Cytokines	Transcription factors	Inhibitory transcription factors
Th1	IL12, IFN γ	<u>T bet</u> , STAT1, STAT4, Runx 3	GATA3
Th2	IL4, IL2	<u>GATA3</u> , STAT6, STAT5, STAT3	T-bet, Runx3
Th17	IL6, IL 21, IL 23, TGF- β	<u>RORγt</u> , STAT3, ROR α , Runx1	T-bet ⁺ Runx1, Smad3 Runx1 ⁺ FOXP3
iTreg	TGF- β , IL2	<u>FOXP3</u> , Smad2, Smad3, STAT5	
Tr1	IL27, IL10	c-Maf, AhR	

Table is adapted from Luckheeram et al., 2012.

1.5 CD46

CD46 is a type I membrane protein expressed by all nucleated human cells. It not only acts as a complement regulator but also has many functions in adaptive immunity, epithelial cellular activities and fertilization. In addition, CD46 is associated with numerous autoimmune, genetic and infectious diseases. There is much more to learn about CD46 for its functions in different cells and its connections to diseases.

1.5.1 Gene structure

The human *CD46* gene consists of 14 exons, and there are two regions of exons which are alternatively spliced to generate multiple isoforms (Figure 1.10 and Table 1.4)(Purcell et al., 1991). Exons 2-6 encode the N-terminal extracellular domain which consists of four short consensus repeat structures termed complement control proteins (CCPs). CCPs are responsible for the complement regulatory activity (CCP2- 4) and provide binding sites for pathogens (CCP1-3) (Figure 1.12). They are followed by a Serine, Threonine, Proline rich (STP) region encoded by exons 7, 8 and 9. These three exons, also denoted as A, B and C, respectively, are alternatively spliced. The STP region is highly O-glycosylated and the level of glycosylation determines the complement pathway regulated by CD46 (Liszewski and Atkinson, 1996). Exons 11 and 12 code for the transmembrane domain while cassette exon 13 and constitutive exon 14 encode two mutually exclusive cytoplasmic tails (Figure 1.10) (Astier, 2008; Liszewski et al., 2005; Ni Choileain and Astier, 2012; Purcell et al., 1991; Russell, 2004; Russell et al., 1992).

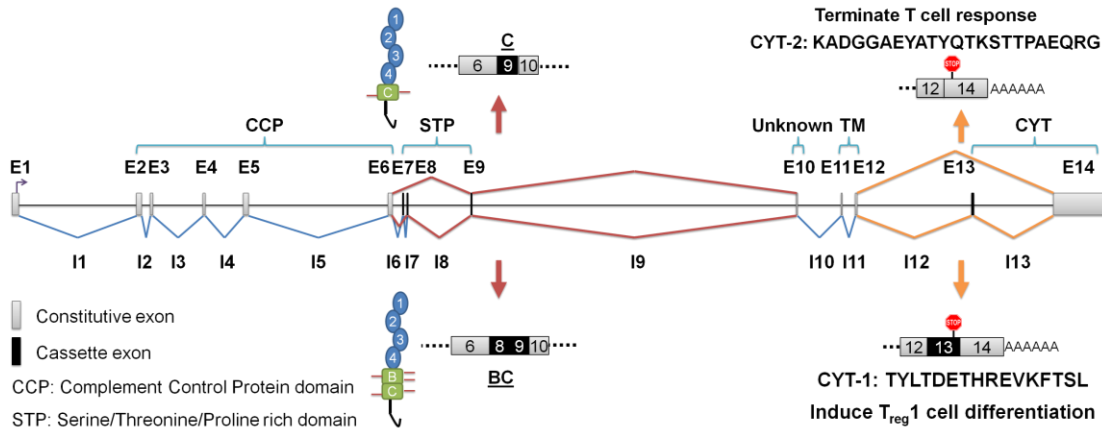


Figure 1.10 Gene structure of *CD46*

With a genomic size of 43,457 base pairs (bp), *CD46* gene consists of 14 exons, with two regions of cassette exons which are exons 7-9 encoding the STP region, and exon 13 encoding CYT1. The two common STP isoforms are BC and C. CYT1 and CYT2 encoded by different exons have different roles in regulating T-cell response. Boxes depict exons and lines depict introns. This schematic is drawn on scale.

Table 1.4 Sizes of *CD46* exons and introns

	Size (bp)		Size (bp)
Exon 1	253	Intron 1	4,704
Exon 2	189	Intron 2	337
Exon 3	103	Intron 3	1,996
Exon 4	86	Intron 4	1,524
Exon 5	198	Intron 5	5,566
Exon 6	183	Intron 6	411
Exon 7	45	Intron 7	127
Exon 8	45	Intron 8	2,497
Exon 9	42	Intron 9	12,929
Exon 10	39	Intron 10	1,740
Exon 11	36	Intron 11	512
Exon 12	64	Intron 12	4,570
Exon 13	93	Intron 13	3,173
Exon 14	1,995		

1.5.2 CD46 Function

As we will see below, CD46 multitasks with many jobs that are related to innate and adaptive immunity, and non-immune functions such as cell adhesion, autophagy and fertilization (Figure 1.11).

1.5.2.1 Role in complement regulation

CD46 was first identified as part of innate immunity, because it can protect host cells from the complement attack by acting as a cofactor for Factor I-mediated proteolytic inactivation of C3b and C4b to prevent the downstream complement pathway (Figure 1.11A)(Barilla-LaBarca et al., 2002; Seya and Atkinson, 1989). The binding of C3b and C4b to CD46 is via the CCP2-4 domains. CD46 with different STP isoforms have a different regulatory ability for the classical pathway. The CD46-BC isoform can bind to the C4b much better than the CD46-C isoform and so the CD46-BC isoform can regulate the classical pathway more efficiently (Figure 1.7 and Figure 1.11A) (Liszewski and Atkinson, 1996).

1.5.2.2 CD46 and fertilization

Expression of CD46 in spermatozoa is located at the inner acrosomal membrane where it regulates the acrosome reaction to facilitate the sperm-egg fusion in humans as well as in mice (Inoue et al., 2003). As such, the loss of or reduced CD46 expression is associated with infertility. The sperm-egg fusion is mediated through the CCP1 domain of CD46, although the interacting partner in the oocyte is not known yet (Riley et al., 2002). CD46 was previously proposed to bind to C3b to regulate the fusion between sperm and oocyte, yet C3 deficient mice did not show any defect in reproduction (Riley et al., 2002). The mechanism of how CD46 mediates fertilization needs further investigation.

1.5.2.3 Role of CD46 in epithelial cells

CD46 expression on epithelial cells not only protects cells from complement activation but also contributes to epithelial polarization and integrity. Interaction of CD46 cytoplasmic tail 1 (CYT1) with PDZ domain disks large homolog 4 (DLG4)

is required for the apical polarization in kidney cells (Ludford-Menting et al., 2002). CD46 is also required for maintaining epithelial barrier integrity in intestinal epithelial cells through its interaction with Ste20/SPS-1-related kinase (SPAK) and E-cadherin. On the other hand, CD46 activation decreases the integrity of tight junctions as characterized by the loss of transepithelial resistance, allowing the transgression of pathogens across intestinal epithelial cells. However, increased epithelial cell proliferation and higher cell mobility due to the loss of cell-cell adhesion by CD46 activation accelerate wound healing (Cardone et al., 2011a). Additionally, CD46 can induce autophagy upon pathogen engagement like measles virus (MV) and *Streptococcus* to eliminate early infection. CD46 CYT1 has been shown to interact with Golgi-associated PDZ and coiled-coil motif containing protein (GOPC) which is a scaffold protein interacting with Beclin1, a component of autophagosome. Interaction between CD46 and GOPC is mediated through tetrapeptide FTSL of CD46 CYT1 and PDZ domain of GOPC (Joubert et al., 2009). CD46 functions in epithelial cells are majorly mediated through CYT1 which is able to interact with PDZ domain of DLG4 and GOPC. Besides, these functions are associated with prevention or elimination of pathogen invasion, suggesting that CD46 is important for the first line immune defense.

1.5.2.4 Role of CD46 on B cells

In addition to its role in innate immunity, CD46 also works as pro- or anti-inflammatory signal for B-cell immunity. CD46-activated T cells have been shown to activate B cells and induce production of IgG and IgM via IL-10 production and cell-cell contact (Fuchs et al., 2009). In contrast, CD46 can suppress allergic response by inhibiting CD40 activation in B cells. CD40 expressed by all mature B cells induces Ig isotype switching to IgE which is an allergic immunoglobulin, and also induces expression of CD23 which is a receptor for IgE-allergen complex and capable of upregulating IgE production. Cross-linking CD46 to CD40 inhibits CD40-mediated upregulation of CD23 and IgE switching (Jabara et al., 2011). Nevertheless, the exact role of CD46 on B cells remains to be clarified.

1.5.2.5 CD46 in T-cell regulation

CD46 possesses two different cytoplasmic tails which are capable of inducing signal transduction. In addition to triggering autophagy and SPAK-mediated signaling, CD46 can function as a co-stimulator for T-cell-receptor signaling by inducing phosphorylation of LAT and p120^{CBL} to promote T-cell proliferation (Astier et al., 2000; Astier, 2008; Cardone et al., 2010; Zaffran et al., 2001). At the early stage of immune response, co-activation of CD46 together with T-cell receptor in the presence of low level of IL-2 induces more IFN γ production to restrain the infection. Activated T cells subsequently secrete IL-2 to maintain their own survival and differentiation. In addition, C3 produced by the activated T cells is converted to C3b for CD46 binding and activation. After a period of T-cell activation and in presence of high IL-2 level, co-activation of CD46 stimulates the differentiation of T_H1 cells to become Tr1 cells with increasing production of IL-10 which is an anti-inflammatory cytokine, and expression of granzyme B which kills effector T cells (Figure 1.11C) (Cardone et al., 2010; Kemper et al., 2005; Kemper et al., 2003). Increased expression of inducible cAMP early repressor/cAMP response element modulator (ICER/CREM) was observed in CD3/CD46 activated CD4⁺ T cells. This resulted in increased nuclear translocation of ICER/CREM which binds to the cAMP responsive elements in IL-2 promoter to repress gene transcription of IL-2 which is essential to maintain T-cell viability (Cardone et al., 2010).

A recent study has shown that the timing of CD46 protein cleavage is the key to its function in T-cell response (Ni Choileain et al., 2011). Upon ligand binding, CD46 is downregulated from the cell surface by metalloproteinases and the cytoplasmic tail is cleaved by Presenilin/ γ secretase (P γ S). Processing of the two mutually exclusive cytoplasmic tails of CD46 (CYT1 and CYT2) by P γ S upon activation is necessary to modulate T-cell activation in a temporal negative feedback mechanism. CYT1 is first downregulated by P γ S to induce production of IL-10, T-cell proliferation as well as Tr1 differentiation, and CYT2 is subsequently processed to terminate the Tr1 cell activation and proliferation. Jagged 1 was

recently identified as a novel CD46 ligand by binding to CCP1 and CCP2 (Figure 1.12)(Le Friec et al., 2012). It is also a ligand for Notch to induce IFN γ production in T cells. The interaction between Jagged 1 and CD46 on resting cells prevents the Notch-Jagged 1 interaction and so inhibits T-cell activation. During normal condition, CD46 acts as a brake to maintain T-cell homeostasis. When there is infection, binding of C3 or pathogens to CD46 releases the Jagged 1 which is now available for Notch (Le Friec et al., 2012). In general, CD46 acts as a negative regulator for T-cell activation as supported by its ability to induce Tr1 cell differentiation and to inhibit spontaneous T-cell activation by interacting with Jagged 1. Nevertheless, CD46 can enhance T-cell activities during infection. Environmental clues like cytokines and chemokines are essential to adjust CD46 functions so that it can properly react to different conditions. Further studies are needed to fully understand the role of CD46 in T cells.

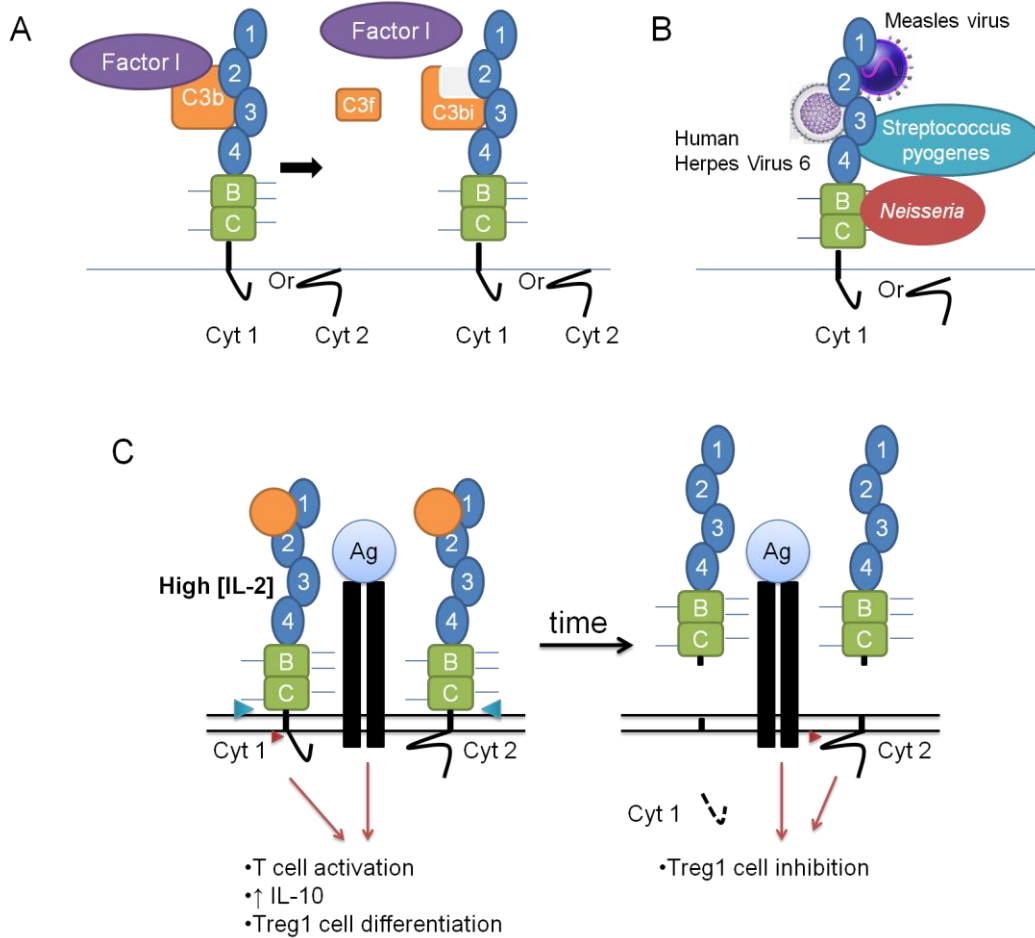


Figure 1.11 CD46 functions

(A) The role of CD46 in innate immunity. CD46 facilitates the Factor I-mediated cleavage of C3b and C4b in the complement system. Image is adapted from Goodship et al., 2004. (B) Pathogens use CD46 as an entry receptor. MV binds to CCP1 and CCP2 for its entry while HHV-6 interacts with CCP2 and CCP3. On the other hand, *Streptococcus* and *Neisseria* bacteria bind to CCP3 and STP region for their entry. (C) The role of CD46 in adaptive immunity. Co-activation together with T-cell receptor results in processing of CYT1 which in turn induces more IFN γ production. In presence of high level of IL-2, CYT1 can subsequently induce the Tr1 differentiation. CYT2 is processed at a much later time which will terminate all T-cell responses and restore immune homeostasis. Ag represents antigen.

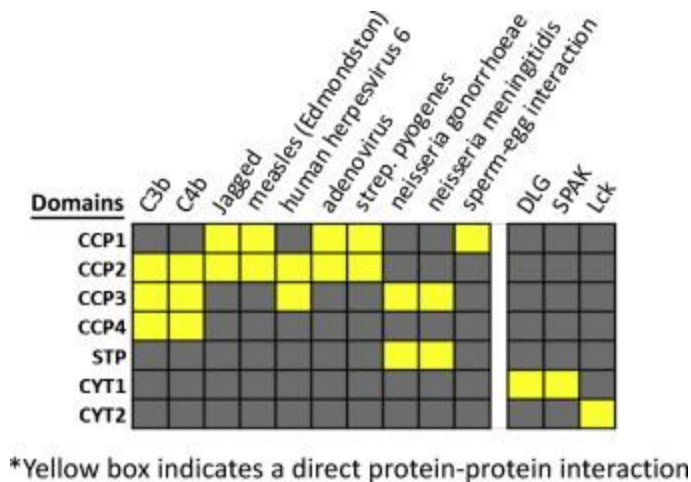


Figure 1.12 Interacting partners of CD46

Complement regulation of CD46 is mediated through CCP2-4. Sequestration of Jagged 1 by CD46 CCP1-2 inhibits T-cell activation. CD46 serves as a pathogen receptor and each pathogen has their preferred binding sites. CYT1 and CYT2 interact with different proteins to modulate different signaling pathways. Image is from Yamamoto et al., 2013.

1.5.3 Implications in diseases

CD46 is involved in different immune responses so its deregulation is associated to a wide diversity of infectious and genetic immune diseases. Individuals who have CD46 deficiency or polymorphisms at the CCP2-4 that affect C3b/C4b binding are predisposed to a rare autoimmune disease known as atypical hemolytic uremic syndrome (aHUS) which is caused by uncontrolled complement activation (Goodship et al., 2004). aHUS has a very low prevalence of 1 in 500,000 people per year. The genetic mutations of *CD46* only account for 12% of total aHUS cases. aHUS is characterized by thrombotic microangiopathy, microangiopathic hemolytic anemia and kidney failure. Prolonged complement activation causes damage to endothelial cells and so blood clots formed by platelets are generated to block the wounds and prevent blood loss. Excess blood clots in blood vessels shatter the red blood cells, resulting in anemia. The blood clots caught in the small

vessels in kidney cause tissue damage, resulting in kidney failure (Noris and Remuzzi, 2009).

Furthermore, excessive immune responses detected in multiple sclerosis (MS) patients could be partly due to the lack of CD46-induced Tr1 cells and production of IL-10 (Astier, 2008; Astier et al., 2006; Ni Choileain and Astier, 2011). Besides, increased IL-23p19, chemokine (C-C motif) ligand 3 (CCL3) and chemokine (C-C motif) ligand 5 (CCL5) production by CD46-engaged myeloid dendritic cells upon lipopolysaccharide (LPS) stimulation were detected in MS patients compared to healthy individuals, suggesting that the CD46 activation pathway is altered in MS patients (Vaknin-Dembinsky et al., 2008). IL-23 is a cytokine required for induction of T_H17 cells which secrete IL-17 and cause inflammation (Table 1.3), while CCL3 and CCL5 are linked to the development of MS lesions. Reduced IL-10 production and granzyme B expression were also observed in CD46-activated CD4⁺ T cells from some asthmatic patients (Tsai et al., 2012; Xu et al., 2010). After stimulation with anti-CD3 and anti-CD46 antibodies, T cells from rheumatoid arthritis (RA) patients showed defective in switching the IFN γ ⁺IL-10⁻ cells to IFN γ ⁻IL-10⁺ cells due to the impairment in shutting down IFN γ production (Cardone et al., 2010; Le Friec et al., 2011). Functional defects of CD46-induced Tr1 cells most likely contribute to the pathogenesis of these autoimmune diseases to some extent. On the other hand, although it is rare, deficiency of CD46 is also associated with recurrent infections as the activation of CD46 is required for T_H1 cell activation and production of IFN γ (Le Friec et al., 2012).

Besides, CD46 is believed to be linked to age-related macular degeneration (AMD), which is caused by dysfunction, detachment and degeneration of retinal pigment epithelium resulting in loss of vision. CD46 was shown to maintain cell adhesion of retinal pigment epithelium together with β 1 integrin (McLaughlin et al., 2003) and its decreased expression was detected in some AMD patients (Ebrahimi et al., 2013). This disease reflects the ability of CD46 to maintain cell-cell adhesion and integrity as discussed earlier.

Ubiquitous expression of CD46 and its immunological functions make it a perfect entry receptor for pathogens. It is known that viruses such as MV and human herpes virus 6 (HHV-6), and bacteria like *Neisseria gonorrhoeae*, *Neisseria meningitidis* and group A *Streptococcus* use CD46 to infect human cells (Figure 1.11B and Figure 1.12)(Riley-Vargas et al., 2004; Russell, 2004; Yamamoto et al., 2013). These pathogens can hijack the immunoregulatory functions of CD46 to prevent their elimination (Cardone et al., 2011b). When binding to CD46, HHV can inhibit macrophage production of IL-12, a cytokine required for T-cell activation and T_H1 cell differentiation (Smith et al., 2003). On the other hand, Streptococcal M5 protein interacts with CD46-BC on CD4⁺ T cells and induces the differentiation of Tr1 cells (Price et al., 2005). MV has different effects on monocyte and macrophage responses upon binding to CD46. In monocytes, CD46 activation by engagement with MV hemagglutinin downregulates IL-12 and causes immunosuppression. In macrophages, activation of CD46 upregulates production of IL-12p40 and synthesis of nitric oxide (NO), resulting in pathogen elimination (Kurita-Taniguchi et al., 2000).

CD46 is over-expressed by some cancer cells to evade the immune recognition and to resist to cancer treatment that triggers complement dependent cytotoxicity (Fishelson et al., 2003; Gorter et al., 1996; Liu et al., 2014; Mamidi et al., 2015; Yan et al., 2008). The expression level of CD46 and other complement regulators like CD55 and CD59 can be used as a diagnostic reference to determine tumor stage. Recently, the ability of MV to induce cell death using CD46 as an entry receptor was utilized for cancer treatment. Virotherapy using MV vaccine strain is in phase I clinical trial to test its efficiency in treating cancer. This therapy targets cancers like non-small cell lung cancer and colorectal cancer which express high level of CD46. Studies have demonstrated effectiveness of this MV vaccine strain to induce apoptosis of cancer cells and reduce tumor size, making it a potential cure for certain cancers (Boisgerault et al., 2013; Deyle et al., 2015; Patel et al., 2014). In a nutshell, diseases associated with CD46 are majorly related to immune defects, implying the importance of CD46 in immune system.

1.5.4 Alternative splicing of *CD46*

Alternative splicing of *CD46* generates multiple isoforms with different STP regions and cytoplasmic tails (Post et al., 1991; Purcell et al., 1991; Russell et al., 1992). Different STP isoforms are derived from alternative inclusion of cassette exons 7, 8 and 9 (also known as A, B and C). Exon 9 has the highest inclusion frequency followed by exons 8 and 7 as shown by the two most common *CD46* isoforms: C isoform containing only exon 9, and BC isoform containing exons 8 and 9. The BC isoform is more O-glycosylated than the C isoform and it preferentially regulates the classical complement pathway as it binds C4b more efficiently (Liszewski and Atkinson, 1996).

Cassette exon 13 can be either included or skipped to generate two mutually exclusive cytoplasmic tails. Inclusion of exon 13 produces a 16 amino acid long CYT1. As an in-frame stop codon is present in exon 13, the CYT1 is solely encoded by exon 13 sequence. In contrast, the 23 amino acid long CYT2 is generated when exon 13 is skipped and exon 14 is translated up to its in-frame stop codon.

CYT1 and CYT2 have different amino acid sequences which can interact with different adaptors or signaling molecules to regulate cellular activities. The two cytoplasmic tails have different processing times such that *CD46* protein with CYT2 has a longer half-life in general (Liszewski et al., 1994), and this observation is consistent with the slower processing of CYT2 in activated T cells (Ni Choileain et al., 2011). As shown in a transgenic mouse expressing human *CD46* with either CYT1 or CYT2 (Marie et al., 2002), CYT1 and CYT2 have antagonistic functions in regulating the T-cell responses including proliferation and inflammation: CYT1 has anti-inflammatory properties while CYT2 has pro-inflammatory properties. Stimulation of *CD46*-CYT1 reduced the cytotoxicity of $CD8^+$ T cells while increased the proliferation of $CD4^+$ T cells. In contrast, activation of *CD46*-CYT2 had opposite effects (Astier, 2008; Marie et al., 2002). However, this result does not reflect the role of *CD46* in human T cells as recently reported (Cardone et al., 2010; Ni Choileain et al., 2011). This discrepancy could be due to the highly

artificial nature of this experiment since mice only express CD46 in male germ cells.

In humans, only processing of CYT1 but not CYT2 can promote the production of IL-10 via the SPAK kinase and so differentiation of Tr1 cells (Cardone et al., 2010; Ni Choileain et al., 2011). Engagement of CD46-CYT1 can trigger autophagy by inducing *de novo* autophagosome formation to eliminate early pathogen infection. This effect is initiated by the interaction between CYT1 and scaffold protein GOPC which then links CYT1 to the autophagosome formation complex. CYT2 does not have the ability to trigger autophagy because it is unable to interact with GOPC (Joubert et al., 2009). All these indicate that the generation of CYT1 by including exon 13 increases CD46 functions. The functional differences between CYT1 and CYT2 allow auto-regulation of cellular activity as in the case of T cells, and also expand the abilities of CD46 in regulating cell signaling.

Exon 13 is only included in some primate *CD46* mRNAs but not in murine or other mammalian *CD46* mRNAs. In addition, expression of murine CD46 is restricted to spermatozoa (Kemper et al., 2001; Tsujimura et al., 1998). These observations suggest that exon 13 and its associated novel functions only newly evolved in primates. Much more about the evolution of exon 13 will be brought about in the Discussion.

1.6 Alternative splicing in immune responses

Recent high-throughput analyses have revealed the importance of alternative splicing in T-cell activation (Ip et al., 2007; Martinez et al., 2012). Numerous T-cell activation-induced alternative splicing events are associated with genes that are involved in effector functions and T-cell viability. These genes include cell surface receptors like *CD44*, *CD45* and *Fas*, transcriptional regulators like *LEF1*, and RNA processing factors like *CELF2* (Martinez and Lynch, 2013). Multiple splicing factors were identified in regulating these splicing events. Alteration in *CD45* alternative splicing upon T-cell activation has been well studied. Three cassette exons (exons 4-6) of *CD45* are predominantly skipped upon antigen engagement to

T cell, giving rise to the smallest isoform, CD45 RO. These three exons encode an extracellular domain that is heavily glycosylated to prevent CD45 homodimerization and so inactivation of CD45 signal transduction. *CD45* alternative splicing is regulated by many proteins using different mechanisms. hnRNP L, hnRNP LL and PTB-associated splicing factor (PSF) bind to the silencers within exon 4 to repress exon inclusion while SRSF1 binds to enhancer within exon 5 to promote its inclusion (Lynch and Weiss, 2000; Melton et al., 2007; Oberdoerffer et al., 2008; Rothrock et al., 2005; Tong et al., 2005). In addition, recruitment of DNA-binding protein CCCTC-binding factor (CTCF) to the *CD45* gene causes RNA polymerase II stalling and favors the inclusion of exon 5 (Shukla et al., 2011).

Another example of alternative splicing in T cells is exon 6 of *Fas* receptor, in which skipping of exon 6 in naive T cells generates soluble Fas and inhibits apoptosis (Izquierdo et al., 2005). Upon T-cell activation, inclusion of exon 6 is strongly enhanced by T-cell restricted intracellular antigen 1 (TIA1) (Izquierdo et al., 2005). The expression of membrane bound Fas receptor on activated T cells upon exon 6 inclusion is used to cease T-cell response by apoptosis. CD44 is a cell surface glycoprotein required for cell adhesion and migration. It consists of 10 alternative exons which are inducibly included by T-cell activation. The alternative splicing change of exon v5 is dependent on Ras-Raf-MEK-ERK signaling pathway which leads to binding of Sam68 and SR-related proteins to one enhancer to promote exon recognition (Cheng et al., 2006; Matter et al., 2002). Sam68 can also recruit SWI/SNF, a nucleosome remodeling complex, through interaction with Brm subunit to pause the transcriptional elongation by chromatin remodeling (Batsche et al., 2006), thus promoting exon inclusion.

Transcriptional alteration during T-cell activation modifies the expression level of splicing regulators and in turn affects splicing events (Ip et al., 2007; Martinez et al., 2012). Furthermore, the T-cell signaling pathways may trigger post-transcriptional modifications of splicing factors to enhance their activities.

Different protein functions are needed for particular cell conditions. Generation of

functionally different protein isoforms by alternative splicing is a faster and sometimes more efficient way than by starting transcription of a gene encoding the required functional protein. Therefore, changes in alternative splicing occur during cell differentiation or activation to produce the needed protein isoforms.

Overall, alternative splicing pattern of an exon is determined by the splice-site strength, protein regulators, transcription and histone modifications. Misregulation of splicing is associated with many genetic diseases. *CD46*, a ubiquitously expressed membrane protein, plays many roles in regulating immune responses and other cellular processes. Inclusion or skipping of *CD46* cassette exon 13, which is the most common type of alternative splicing, gives rise to two cytoplasmic tails with distinct roles in activating, repressing and terminating T-cell activities in a temporal manner (Yamamoto et al., 2013). The proper regulation of *CD46* alternative splicing is required to generate the repertoire of *CD46* with either cytoplasmic tail for each immune response. Association of defective *CD46*-mediated IL-10 production with autoimmune diseases like asthma and RA suggests that inappropriate alternative splicing of *CD46* may disrupt the homeostasis of immune system and result in development of diseases.

Although multiple *CD46* isoforms and its aberrant alternative splicing in disease have been reported previously, no study has been carried out to understand the regulatory mechanism of *CD46* alternative splicing. Therefore, I started this project to gain new insight into the regulation of *CD46* alternative splicing.

1.7 Objective

The objective of my thesis is to study the splicing regulatory mechanisms of *CD46* in order to (i) elucidate the regulatory pathways involved in diseases, and (ii) manipulate its alternative splicing as it could be useful to develop novel treatments for immune diseases. Thus, identification of auxiliary *cis*-acting elements, protein regulators and signaling pathway is the main aim of this project.

2. MATERIALS AND METHODS

2.1 Cell culture

HeLa (*Homo sapiens* cervix adenocarcinoma cell) and HEK293T (*Homo sapiens* embryonic kidney cell) cell lines were maintained in Dulbecco's Modified Eagle Medium (DMEM; Hyclone, USA) supplemented with 10% Fetal Bovine Serum (FBS; Gibco, USA), 100 U/ml Penicillin and 100 µg/ml Streptomycin (Gibco, USA) at 37 °C and 5% CO₂. Jurkat E6.1 (*Homo sapiens* Acute T-cell leukemia) cells were maintained in RPMI-1640 (Hyclone, USA) supplemented with 10% FBS (Gibco, USA), 100 U/ml Penicillin and 100 µg/ml Streptomycin (Gibco, USA).

2.2 Plasmids

2.2.1 Minigene construction

To study the regulatory mechanisms of alternative splicing of cassette exon, several minigenes containing flanking constitutive exons, alternative exon/s and shortened intronic sequences were constructed. For the cassette exon 13, three fragments of *CD46* were amplified from Human Genomic DNA (Promega, USA) using PCR: (i) exon 12 and its downstream 300-nucleotide (nt) intronic sequence using primers Ex12.F.Hind and In12del.R, (ii) exon 13 and 300 nt of its upstream and downstream introns using primers In12del.F and In13del.R and (iii) 425 nt of exon 14 sequence and 300 nt of its upstream intron using In13del.F and Ex14del.R.Xho (Figure 2.1A and Table 2.1). Subsequently, these three fragments were merged into one using overlap PCR. *CD46*-exon 12-13-14 insert was then cloned into pcDNA 3.1+ vector with HindIII and XhoI sites. Similarly, the *CD46*-exon 6-7-8-10 minigene containing the indicated exons was generated to study the alternative splicing of exons 7 and 8. Two insert fragments containing (i) full-length exons 6, 7 and 8, introns 6 and 7, and first 350 nt of intron 8, and (ii) exon 10 and 350 nt of its upstream intron were generated using the same method and ligated into vector using KpnI and XhoI restriction sites (Figure 2.1B and Table 2.1). Mutagenesis

PCR was applied to remove exon 8 to generate *CD46*-exon 6-7-10 minigenes (Figure 2.1C and Table 2.1). All the DNA primers were ordered from IDT, USA.

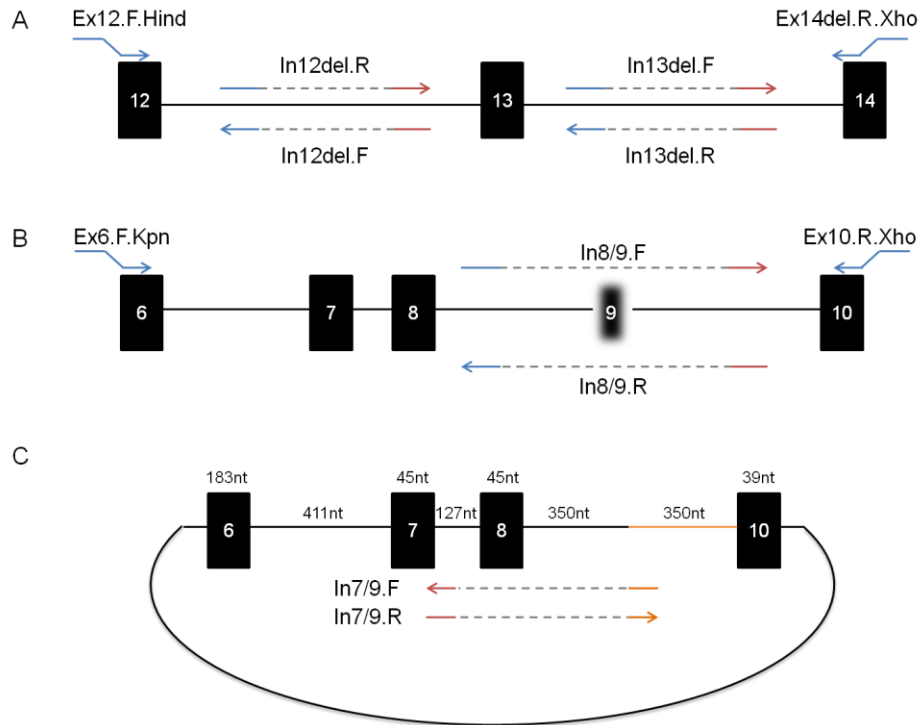


Figure 2.1 Schematic diagram of *CD46* minigene construction

(A) Primer design to generate *CD46*-exon12-13-14 minigene insert. (B) Primer design to generate *CD46*-exon6-7-8-10 minigene insert. (C) Primer design for deleting exon 8 in *CD46*-exon 6-7-8-10 minigene. Arrows represent primers. Dotted lines represent deleted sequences. Primer sequences are shown in Table 2.1.

Table 2.1 Primer sequences for generating *CD46* minigenes

Primers for constructing <i>CD46</i>-exon12-13-14 minigene	
Ex12.F.Hind	AAGCCTGGAAGCTTGGTGGAGTTGCAGTAATTTGTG
Ex14del.R.Xho	GAGAGGACGCTCGAGATTTCAAGCCACTTTCTTTAC
In12del.F	CATGGACCACTCACTATGAGG//AAACATTTTGACATGAATCAC
In12del.R	GTGATTCATGTCAAATGTTT//CCTCATAGTGAGTGGTCCATG
In13del.F	GAATCCAAAAGTAAGAATTTG//AGTCTTTCCATTGGCCCTAC
In13del.R	GTAGGGCCAATGGGAAAGACT//CAAATCTTACTTTTGGATTC
Primers for constructing <i>CD46</i>-exon6-7-8-10 minigene	
Ex6.F.Kpn	AAGCATGAGGTACCTGGTCAAATGTGCGATTTCCAGTAGTC
Ex10.R.Xho	TAAAGGACGCTCGAGCCAAACTGTCAAGTATTCCTTCCTC
In8/9.F	GCACAGTGCCCTGGTGTAGAGAA//TGTGCCCGGCTAATCCCTCTATC
In8/9.R	GATAGAGGGATTAGCCGGGCACA//TTCTCTACACCAGGCACCTGTGC
Primers for deleting exon 8	
In7/9.F	ATGTCTTCTTCCTTATATGTTACAAGATA//TGTGCCCGGCTAATCCCTCTA
In7/9.R	TAGAGGGATTAGCCGGGCACA//TATCTTGTAACATATAAGGAAGAAGACAT

Underline highlights restriction site. // depicts the location of long deleted sequence.

2.2.2 Generation of mutant minigenes

Mutagenesis PCR using KAPA HiFi DNA polymerase (KAPA Biosystems, USA) was applied to introduce deletions or nucleotide substitutions into the wild-type minigenes (Figure 2.2 and Table 2.2-2.4). PCR conditions were as follows: one cycle of denaturation at 95 °C for 2 min, 18 cycles of denaturation at 98 °C for 30 s followed by annealing at 50°C for 1 min and extension at 72 °C for 5 min, and one final extension for 5 min. Several pairs of forward and reverse primers containing only the flanking sequences of each deleted region were used to generate minigenes with deletions. A forward primer containing mutated and flanking sequences and a common reverse primer were used to generate point mutations in the minigenes. Total PCR products were treated with 20 U DpnI (New England Biolabs, USA) to remove methylated PCR templates followed by transformation of chemically competent *E. coli* DH5 α with 4 μ l PCR products. A total of 21 deletions of eight nucleotides were introduced into exon 13 excluding the first two

nucleotides at the 5' end and the last three nucleotides at the 3' end to maintain the integrity of the 3'ss and 5'ss, respectively. Ten 10-nt intronic sequences each for upstream and downstream of exon 13 were separately deleted without disrupting the polypyrimidine tract, the 3'ss and 5'ss. Point mutations in the exonic sequences were designed based on the predictions by *in silico* tools like Human Splicing Finder and SFmap (Akerman et al., 2009; Desmet et al., 2009; Paz et al., 2010) to disrupt the existing ESEs or ESSs without creating new *cis*-acting elements.

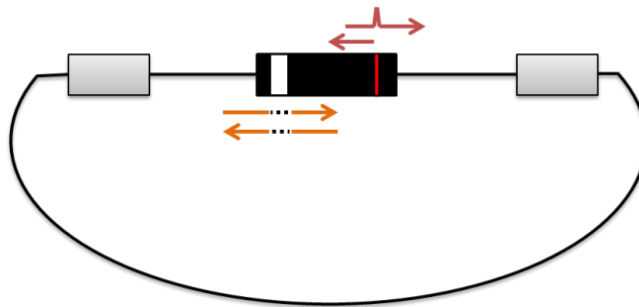


Figure 2.2 Mutagenesis primers design

Orange primers were used to remove the white block sequences. Dotted line represents the sequence to be deleted and that is hence not included in the primers. Red primers were used to introduce point mutations. Arrows depict primers that are shown in Table 2.2-2.4. As multiple sets of point mutations were designed to target one region, a common reverse primer was used to minimize the number of primers to be ordered. Forward primers contain the mutated sequence and so the mutation was introduced into the plasmid by PCR.

Table 2.2 Primer design for generating mutant *CD46*-exon 12-13-14 minigene

Primers for generating first set of serial deletions in exon 13	
CD46.Ex13bin1del.F	CCGTTTTCTCTTCCCTCTGTTCAGCA- ACTGATGAGACCCACAGAGAAGTAA
CD46.Ex13bin1del.R	TTACTTCTCTGTGGGTCTCATCAGT- TGCTGAACAGAGGAAGAGAAAACGG
CD46.Ex13bin2del.F	TTCTCTGTTCAGCACATACCTA- GACCCACAGAGAAGTAAAATTTACTTC
CD46.Ex13bin2del.R	GAAGTAAATTTTACTTCTCTGTGGGTC- TAGGTATGTGCTGAACAGAGGAA
CD46.Ex13bin3del.F	TTCAGCACATACCTAACTGATGA- GAGAAGTAAAATTTACTTCTCTCTGAG
CD46.Ex13bin3del.R	CTCAGAGAGAAGTAAAATTTTACTTCTC- TCATCAGTTAGGTATGTGCTGAA
CD46.Ex13bin4del.F	ATACCTAACTGATGAGACCCACA- AAATTTACTTCTCTCTGAGAAGGAGAG
CD46.Ex13bin4del.R	CTCTCCTTCTCAGAGAGAAGTAAAATTT- TGTTGGGTCTCATCAGTTAGGTAT
CD46.Ex13bin5del.F	AACTGATGAGACCCACAGAGAAGTA- TTCTCTCTGAGAAGGAGAGATGAGA
CD46.Ex13bin5del.R	TCTCATCTCTCCTTCTCAGAGAGAA- TACTTCTCTGTGGGTCTCATCAGTT
CD46.Ex13bin6del.F	GAGACCCACAGAGAAGTAAAATTTAC- GAGAAGGAGAGATGAGAGAAAGGT
CD46.Ex13bin6del.R	ACCTTTCTCTCATCTCTCCTTCTC- GTAAAATTTTACTTCTCTGTGGGTCTC
CD46.Ex13bin7del.F	ACAGAGAAGTAAAATTTACTTCTCTCT- GAGATGAGAGAAAGGTTTGCTTT
CD46.Ex13bin7del.R	AAAGCAAACCTTTCTCTCATCTC- AGAGAGAAGTAAAATTTTACTTCTCTGT
CD46.Ex13bin8del.F	AAATTTACTTCTCTCTGAGAAGGA- AGAAAGGTTTGCTTTTATCATTAATAA
CD46.Ex13bin8del.R	TTTTAATGATAAAAAGCAAACCTTTCT- TCCTTCTCAGAGAGAAGTAAAATTT
CD46.Ex13bin9del.F	CTCTCTGAGAAGGAGAGATGAG- TTGCTTTTATCATTAATAAAGGTATCTGTT
CD46.Ex13bin9del.R	AACAGATACCTTTTAAATGATAAAAAGCAA- CTCATCTCTCCTTCTCAGAGAG
CD46.Ex13bin10del.F	GAAGGAGAGATGAGAGAAAGGT- ATCATTAATAAAGGTATCTGTTTTCTGTTG
CD46.Ex13bin10del.R	CAACAGAAAACAGATACCTTTTAAATGAT- ACCTTTCTCTCATCTCTCCTTC
CD46.Ex13bin11del.F	GAGATGAGAGAAAGGTTTGCTTTT- AAGGTATCTGTTTTCTGTTGTTTATT
CD46.Ex13bin11del.R	AATAACAACAGAAAACAGATACCTT- AAAAGCAAACCTTTCTCTCATCTC

Primers for generating staggered deletions in exon 13	
CD46.Ex13st1del.F	TTCTCTTCCTCTGTTTCAGCACATA- ATGAGACCCACAGAGAAGTAAAATTT
CD46.Ex13st1del.R	AAATTTTACTTCTCTGTGGGTCTCAT- TATGTGCTGAACAGAGGAAGAGAA
CD46.Ex13st2del.F	CTGTTTCAGCACATACCTAACTG- CACAGAGAAGTAAAATTTACTTCTCTCT
CD46.Ex13st2del.R	AGAGAGAAGTAAATTTTACTTCTCTGTG- CAGTTAGGTATGTGCTGAACAG
CD46.Ex13st3del.F	CACATACCTAACTGATGAGACC- AGTAAAATTTACTTCTCTCTGAGAAGGA
CD46.Ex13st3del.R	TCCTTCTCAGAGAGAAGTAAATTTTACT- GGTCTCATCAGTTAGGTATGTG
CD46.Ex13st4del.F	CTAACTGATGAGACCCACAGAGA- TTACTTCTCTCTGAGAAGGAGAGATGA
CD46.Ex13st4del.R	TCATCTCTCCTTCTCAGAGAGAAGTAA- TCTCTGTGGGTCTCATCAGTTAG
CD46.Ex13st5del.F	GATGAGACCCACAGAGAAGTAAAAT- CTCTGAGAAGGAGAGATGAGAGAAA
CD46.Ex13st5del.R	TTTCTCTCATCTCTCCTTCTCAGAG- ATTTTACTTCTCTGTGGGTCTCATC
CD46.Ex13st6del.F	ACCCACAGAGAAGTAAAATTTACTTCT- AGGAGAGATGAGAGAAAGGTTTG
CD46.Ex13st6del.R	CAAACCTTTCTCTCATCTCTCCT- AGAAGTAAATTTTACTTCTCTGTGGGT
CD46.Ex13st7del.F	AGAAGTAAAATTTACTTCTCTCTGAGA- TGAGAGAAAGGTTTGCTTTTATC
CD46.Ex13st7del.R	GATAAAAGCAAACCTTTCTCTCA- TCTCAGAGAGAAGTAAATTTTACTTCT
CD46.Ex13st8del.F	ACTTCTCTCTGAGAAGGAGAGA- AGGTTTGCTTTTATCATTAAAAGGTATC
CD46.Ex13st8del.R	GATACCTTTTAATGATAAAAGCAAACCT- TCTCTCCTTCTCAGAGAGAAGT
CD46.Ex13st9del.F	TGAGAAGGAGAGATGAGAGAA- TTTTATCATTAAAAGGTATCTGTTTCTG
CD46.Ex13st9del.R	CAGAAAACAGATACCTTTTAAATGATAAAA- TTCTCTCATCTCTCCTTCTCA
CD46.Ex13st10del.F	GGAGAGATGAGAGAAAGGTTTGC- TTAAAAGGTATCTGTTTCTGTTGTTT
CD46.Ex13st10del.R	AAACAACAGAAAACAGATACCTTTTAA- GCAAACCTTTCTCTCATCTCTCC
Primers for generating serial deletions in intron 12	
CD46.Ex13us1del.F	GCTACTCGTTTCTTTTTGGTTTG- TTTATTCAGCCGTTTCTCTTCC
CD46.Ex13us1del.R	GGAAGAGAAAACGGCTGAATAAA- CAAACCAAAAAGAAACGAGTAGC
CD46.Ex13us2del.F	CACTTGTTATGCTACTCGTTTCT- AAGTCACTATTTTATTCAGCCGTT

CD46.Ex13us2del.R	AACGGCTGAATAAAAATAGTGA CTT- AGAAACGAGTAGCATAACAAGTG
CD46.Ex13us3del.F	TTTATCCCACCTTGTTATGCT- TTTTGGTTTGAAGTCACTATTTTATTCAG
CD46.Ex13us3del.R	ATAGTGA CTTCAAACCAAAA- AGCATAACAAGTGGGATAAAAATCC
CD46.Ex13us4del.F	TACACTGGATTTTTATCCCAC- ACTCGTTTCTTTTTGGTTTGAAG
CD46.Ex13us4del.R	CAAACCAAAAAGAAACGAGT- GTGGGATAAAAATCCAGTGTAAGAG
CD46.Ex13us5del.F	GCTACTCTCTTACACTGGAT- TTGTTATGCTACTCGTTTCTTTTTGG
CD46.Ex13us5del.R	AAAAGAAACGAGTAGCATAACAA- ATCCAGTGTAAGAGAGTAGCAG
CD46.Ex13us6del.F	CTCAGCTTTCCTGCTACTCTCT- TTTATCCCACCTTGTTATGCTACTCG
CD46.Ex13us6del.R	CGAGTAGCATAACAAGTGGGATAAA- AGAGAGTAGCAGGAAAGCTGAG
CD46.Ex13us7del.F	GAACCTAATTCTCAGCTTTCCT- TACACTGGATTTTTATCCCACCTTG
CD46.Ex13us7del.R	CAAGTGGGATAAAAATCCAGTGTA- AGGAAAGCTGAGAATTAGGTTC
CD46.Ex13us8del.F	TAAGTTATGAACCTAATTCT- GCTACTCTCTTACACTGGATTTTATCC
CD46.Ex13us8del.R	ATCCAGTGTAAGAGAGTAGC- AGAATTAGGTTCATAACTTAAGACTTTCTC
CD46.Ex13usst1del.F	ATGCTACTCG-GTTTGAAGTCACTATTTTATTCAGC
CD46.Ex13usst1del.R	GACTTCAAAC-CGAGTAGCATAACAAGTGGGA
CD46.Ex13usst2del.F	CCCCTTGTT-TTCTTTTTTGGTTTGAAGTCAC
CD46.Ex13usst2del.R	CAAAAAGAAA-AACAAGTGGGATAAAAATCCAGTG
CD46.Ex13usst3del.F	TGGATTTTAT-ATGCTACTCGTTTCTTTTTGGTTTG
CD46.Ex13usst3del.R	CGAGTAGCAT-ATAAAATCCAGTGTAAGAGAGTAGC
Primers for generating serial deletions in intron 13	
CD46.Ex13ds1del.F	GCTTTTATCATTAAAAGGTATCTGT- TTATTTTCAGATGTCCTTTCTTTTG
CD46.Ex13ds1del.R	CAAAAGAAAGGACATCTGAAAATAA- ACAGATACCTTTTAATGATAAAAAGC
CD46.Ex13ds2del.F	AAAAGGTATCTGTTTTCTGTTGT- ATGTCCTTTCTTTTGAAAAATATTCA
CD46.Ex13ds2del.R	TGAATATTTTTCAAAGAAAGGACAT- ACAACAGAAAACAGATACCTTTT
CD46.Ex13ds3del.F	TGTTTTCTGTTGTTTATTTTCAG- TTTTGAAAAATATTCAGTGGATATAT
CD46.Ex13ds3del.R	ATATATCCACTGAATATTTTTCAA- CTGAAAATAAACAGAAAACA
CD46.Ex13ds4del.F	GTTTATTTTCAGATGTCCTTTC- TATTCAGTGGATATATAGATATCAATAA

CD46.Ex13ds4del.R	TTATTGATATCTATATATCCACTGAATA- GAAAGGACATCTGAAAATAAAC
CD46.Ex13ds5del.F	CAGATGTCCTTTCTTTTGAAAAA- ATATATAGATATCAATAACCTGAGCAA
CD46.Ex13ds5del.R	TTGCTCAGGTTATTGATATCTATATAT- TTTTTCAAAGAAAGGACATCTG
CD46.Ex13ds6del.F	CCTTTCTTTTGAAAAATATTCAGTGG- ATCAATAACCTGAGCAAAGAGAT
CD46.Ex13ds6del.R	ATCTCTTTGCTCAGGTTATTGAT- CCACTGAATATTTTTCAAAGAAAGG
CD46.Ex13ds7del.F	TTTTGAAAAATATTCAGTGGATATATAGAT- TGAGCAAAGAGATATTGGCA
CD46.Ex13ds7del.R	TGCCAATATCTCTTTGCTCA- ATCTATATATCCACTGAATATTTTTCAAAA
CD46.Ex13ds8del.F	CAGTGGATATATAGATATCAATAACC- GATATTGGCAGTAAATATCAAAGA
CD46.Ex13ds8del.R	TCTTTGATATTTACTGCCAATATC- GGTTATTGATATCTATATATCCACTG
CD46.Ex13dsst1del.F	GTATCTGTTTTCT- TTCAGATGTCCTTTCTTTTGAAAAATATTCAGTG
CD46.Ex13dsst1del.R	GACATCTGAA-AGAAAACAGATACCTTTTAATGATAAAAAGC
CD46.Ex13dsst2del.F	GTTGTTTATT- CTTTCTTTTGAAAAATATTCAGTGGATATATAG
CD46.Ex13dsst2del.R	CAAAGAAAG- AATAACAACAGAAAACAGATACCTTTTAATG
Primers for generating point mutations in ESS and ESE	
CD46.Ex13ESS1Mut1.F	GAGATGAGAGAAA CGTTAAATT CTATCATTAAAAGGTATCTG TTTTCTGTTG
CD46.Ex13ESS1Mut2.F	GAGATGAGAGAAAG CTTCACTTTAAT CATTAAAAGGTATCTG TTTTCTGTTG
CD46.Ex13ESS1Mut3.F	GAGATGAGAGAAAGGTT CGATT CTATCATTAAAAGGTATCTG TTTTCTGTTG
CD46.Ex13ESS1.R	TTCTCTCATCTCTCCTTCTCAGAG
CD46.Ex13ESE1Mut1.F	CACATACCTAACTG TCGCGACACTCA TACAATTAAAATTTAC TTCTCTCTGAG
CD46.Ex13ESE1Mut2.F	CACATACCTAACTGAT AATTCCCA TATCGGAGTAAAATTTAC TTCTCTCTGAG
CD46.Ex13ESE1Mut3.F	CACATACCTAACTGAT CCGT CCCACAGAGAAGTAAAATTTAC TTCTCTCTGAG
CD46.Ex13ESE1Mut4.F	CACATACCTAACTGATGAGACCCA TAATA AGTAAAATTTAC TTCTCTCTGAG
CD46.Ex13ESE1Mut5.F	CACATACCTAACTGAT CCGT CCCATAGATAAGTAAAATTTAC TTCTCTC
CD46.Ex13ESE1.R	TAGGTATGTGCTGAACAGAGG
CD46.Ex13ESE2Mut1.F	AGTAAAATTTACTTCTCTCTGAGAT TCT AGAGATGAGAGAAAG GTTTGC
CD46.Ex13ESS2Mut1.F	AGTAAAATTTACT CAT ATCTGAGAAGGAGAGATGAGAG

CD46.Ex13ESS2Mut2.F	AGTAAAATTTTACAA CAC ACTGAGAAGGAGAGATGAGAGAAAG G
CD46.Ex13ESS2.R	AAATTTTACTTCTCTGTGGGTC
Primers for generating point mutations in ISE	
CD46.In13ISE1Mut3.F	AA C CAGATGTCCAA A CA A ATGAA T AATATTCAGTGGATATATA GATATC
CD46.In13ISE1Mut3.R	CATCTG TTT G TT G CTT CAG TT AACAGATACCTTTTAATGAT AAAAGC

- depicts the location of the short deleted sequence. Point mutations are labeled in red.

Table 2.3 Primers for mutating the 3'ss and 5'ss flanking exon 7

CD46E7_3'ss +1G F	AAGTGGTTGATCTTCTAACATTATTTTGTTCCTAG GG CTGCCT CCATCTAGTACAAAAC
CD46E7_Com 3'ssR	AATGTTAGAAGATCAACCACTTGGGAG
CD46E7_5'ss m2C F	CCTCCAGCTTTGAGTCATTC CG GTTTAGTAGCTTCTTCCTTATA TGTCTTC
CD46E7_5'ss p6C F	CCTCCAGCTTTGAGTCATTCAGGTTT A CTAGCTTCTTCCTTATA TGTCTTCTTCCTTATA
CD46E7_5'ss p7C F	CCTCCAGCTTTGAGTCATTCAGGTTTAG C AGCTTCTTCCTTATA TGTCTTCTTCCTTATA
CD46E7_Com5'ss m4-m28 R	AATGACTCAAAGCTGGAGGTTTTGT

Point mutations are labeled in red. Cloning and preliminary tests were performed by attachment student Ly Phuong Thao.

Table 2.4 Primers for mutating the 5'ss flanking exon 8

CD46.Ex8.mutag.R	AGGCACTGGACGCTGGAGATTTTGTAGTGGAGAAG
CD46.Ex8.-2C.F	CTCCAGCGTCCAGTGCCTC CG GTTTAGTAATTCCTGCTTATAGTT TTTC
CD46.Ex8.+6C.F	CTCCAGCGTCCAGTGCCTCAGGTTT A CTAATTCCTGCTTATAGTT TTTC
CD46.Ex8.+7C.F	CTCCAGCGTCCAGTGCCTCAGGTTTAG C AATTCCTGCTTATAGTT TTTC

Point mutations are labeled in red. Cloning and preliminary tests were performed by attachment student Goh Eling.

2.2.3 Heterologous minigene

2.2.3.1 pSXN plasmid

pSXN plasmid (gift from Prof. Thomas A. Cooper from Baylor College of Medicine, USA) contains a small alternative exon flanked by a very weak 5'ss and this exon is usually skipped unless a splicing enhancer sequence is inserted in the middle of the exon using the two restriction sites, Sall and BamHI (Figure 2.3 and Table 2.5)(Coulter et al., 1997). A pair of complementary DNA oligonucleotides containing test sequence and corresponding restriction overhangs was first annealed by heating equal molar amounts of the two oligonucleotides at 95 °C for 5 min and gradually cooling to room temperature in 1 h. The annealed oligonucleotides were then 5' end phosphorylated by T4 polynucleotide kinase (T4PNK; New England Biolabs, USA) in 1x ligation buffer [50 mM Tris-HCl, 10 mM MgCl₂, 1 mM ATP and 10 mM DTT] and ligated into the pSXN vector that was cut with Sall and BamHI.

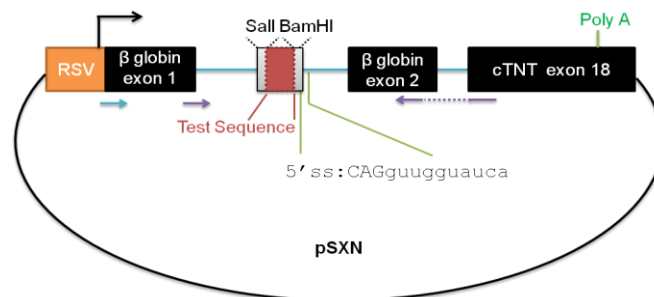


Figure 2.3 Schematic map of pSXN heterologous plasmid

The pSXN plasmid contains a weak exon located between β globin exons 1 and 2. Tested sequences (Table 2.5) were cloned into the weak exon using the Sall and BamHI sites and their effects on exon inclusion were examined by RT-PCR using the purple primers (Table 2.11).

Table 2.5 Oligonucleotide inserts for pSXN plasmid

CD46.Ex13.ESS1.WT.F	<u>TCGACAGAGAAAGGTTTGCTTTTATCATTAAAAG</u>
CD46.Ex13.ESS1.WT.R	<u>GATCC</u> TTTTAATGATAAAAGCAAACCTTTCTCTG
CD46.Ex13.ESS1.Mut 1.F	<u>TCGACAGAGAAAC</u> CGTTAAATT <u>CTATCATTAAAAG</u>
CD46.Ex13.ESS1.Mut 1.R	<u>GATCC</u> TTTTAATGATAGAA TTTAAC <u>GTTTCTCTG</u>
CD46.Ex13.ESE1.WT.F	<u>TCGACTAACTGATGAGACCCACAGAGAAGTAAAG</u>
CD46.Ex13.ESE1.WT.R	<u>GATCC</u> TTTACTTCTCTGTGGGTCTCATCAGTTAG
CD46.Ex13.ESE1.Mut 5.F	<u>TCGACTAACTGAT</u> CCGTCCCA <u>TAGATAAGTAAAG</u>
CD46.Ex13.ESE1.Mut 5.R	<u>GATCC</u> TTTACTT ATCTAT <u>GGGACGGATCAGTTAG</u>
CD46.Ex13.ESS2.WT.F	<u>TCGACATTTACTTCTCTCTGAG</u>
CD46.Ex13.ESS2.WT.R	<u>GATCC</u> TCAGAGAGAAGTAAATG
CD46.Ex13.ESS2.Mut 1.F	<u>TCGACATTTACT</u> CAT <u>ATCTGAG</u>
CD46.Ex13.ESS2.Mut 1.R	<u>GATCC</u> TCAGAT TAT <u>GAGTAAATG</u>
CD46.Ex13.ESE2.WT.F	<u>TCGACCTGAGAAGGAGAGATGG</u>
CD46.Ex13.ESE2.WT.R	<u>GATCC</u> CATCTCTCCTTCTCAGG
CD46.Ex13.ESE2.Mut 1.F	<u>TCGACCTGAGAT</u> CT <u>TAGAGATGG</u>
CD46.Ex13.ESE2.Mut 1.R	<u>GATCC</u> CATCTCT AGA <u>TCTCAGG</u>
pSXN.Linker.F	<u>TCGACATTCGGCTAG</u>
pSXN.Linker.R	<u>GATCC</u> TAGCCGAATG

Insert sequence is in black font, Sall overhang is in blue font and BamHI overhang is in green font. Mutation is labeled in red.

2.2.3.2 CD46 exon 13 in Universal Minigene Vector (UMV)

Exon 13 and its flanking 300 nt intronic sequences were amplified from *CD46*-exon 12-13-14 minigene by PCR using PrimeSTAR Max polymerase (TAKARA Bio, Japan) and primers annealing to introns 12 and 13 (Figure 2.4 and Table 2.6). The PCR fragments were then inserted into UMV plasmid using BamHI and EcoRI sites. UMV plasmid was derived from a *MCAD* minigene plasmid containing *MCAD* exons 8, 9 and 10 with shortened introns in pcDNA3.1+ vector. Dr. Tan Jiazi replaced the *MCAD* exon 9 and its flanking introns with a stretch of multiple cloning site (MCS) so that this UMV can be utilized to study the U1/5'ss recognition and the alternative splicing patterns of cassette exons.

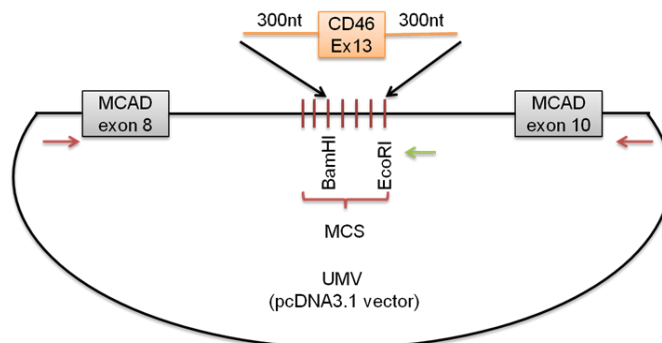


Figure 2.4 Schematic map of UMVCD46exon 13 plasmid

CD46 exon 13 and introns were amplified using the primers shown in Table 2.6 and cloned into UMV using BamHI and EcoRI sites. Red arrows represent RT-PCR primers used to detect splice products (Table 2.11). Green arrow represents sequencing primer.

Table 2.6 Primers for subcloning exon 13 and its flanking intronic sequences

CD46.In12(last 300).F.BamHI	AAAAAAAAGGATC <u>CAAACATTTGACATGAATCACTT</u> CATGC
CD46.In13(1 st 300).R.EcoRI	AAAAAAAAGAAT <u>TCAAATTCTTACTTTTGGATT</u> CACA GCTG

Underline highlights restriction site.

2.2.4 Overexpression plasmid

Protein coding sequences from *trans*-acting factors were amplified from cDNAs using PrimerSTAR Max DNA polymerase (TAKARA Bio, Japan) and cloned into pCGT7 or pcDNA3.1+ vector with respective restriction sites (Table 2.7 and Table 2.8). All restriction enzymes were purchased from New England Biolabs, USA.

Table 2.7 List of cDNA plasmids used in experiments

pCGT7-SF2-WT (SRSF1 cDNA plasmid)	Gifts from Javier F C áceres from MRC Human Genetics Unit at Edinburgh, UK
pCGT7-SF2-ΔRS	
pCGT7-SF2-NRS	
pCGT7-PTBP1	Subcloned from pEM830#48+PTBP (gift from Eugene Makeyev's lab from SBS NTU) using restriction sites HindIII/ BglIII into pCGT7-MCS (HindIII/ BamHI)
pCGT7-TIA1 (transcript variant 2; long isoform)	Cloned from cDNAs with primers annealing to the two ends of the open reading frame
pCGT7-TIAL1 (transcript variant 1; short isoform)	
pcDNA-hnRNP M (M4 isoform)	

Table 2.8 Primers for cloning splicing factor cDNAs

TIA1. ORF. F. XbaI	AAAAAAAAATCTAGAGAGGACGAGATGCCCAAGACTC
TIA1. ORF. R. BamHI	AAAAAAAAAGGATCCTCACTGGGTTTCATACCCTGCCACTC
TIAL1. ORF. F. XbaI	AAAAAAAAATCTAGAATGGAAGACGACGGGCAGC
TIAL1. ORF. R. BamHI	AAAAAAAAAGGATCCTCACTGTGTTTGGTAACTTGCCATAC
hnRNPM. ORF. F. BamHI	AAAAAAAAAGGATCCGAGAAAATGGCGGCAGGGTTCGAAG
hnRNPM. ORF. R. XbaI	AAAAAAAAATCTAGATTAAGCGTTTCTATCAATTCCAACGTC AATCTCTCG

Restriction sites are underlined.

2.3 Bacterial transformation and plasmid extraction

50 µl of chemically competent *E. coli* DH5α was mixed with 4 µl PCR or ligation products and incubated on ice for 30 min followed by a heat shock at 42 °C for 45 s. Subsequently, cells were incubated on ice for another 5 min and then recovered in 1 ml of pure lysogeny broth (LB) at 37 °C for 1 h with agitation. Finally, transformed cells were plated on ampicillin-supplemented LB agar plate (70 µg/ml) and grown at 37°C overnight. Colonies were grown in 5 ml LB with ampicillin (70 µg/ml) overnight and then subjected to miniprep (Omega Bio-Tek, USA). All mutant minigene sequences were confirmed by sequencing (1st Base, Singapore).

For the ligation product, plasmids were digested with the respective restriction enzymes to confirm the insertion of specific fragments before sequencing.

2.4 Transfection

2.4.1 Adherent cells

2.4.1.1 Transfection with minigenes

HEK293T or HeLa cells were seeded at a density of 10^5 cells/ml at 24 h before transfection. 1 μ g DNA containing 83 ng of minigene plasmid and 917 ng of pUC19 mock plasmid which is not expressed in eukaryotic cells was used for a single transfection. DNA was mixed with 100 μ l Opti-MEM (Life Technologies, USA) and XtremeGENE 9 transfection reagent (Roche, Switzerland) in a ratio of 1:3 (DNA:reagent). Transfection mixture was then incubated at room temperature for 20 min before it was added to cells in a drop-wise manner. For U1 suppressor assay, *CD46* minigene, U1 suppressor plasmid and mock plasmid were mixed in a ratio of 1:10:1.

2.4.1.2 siRNA and antisense oligonucleotides (ASOs) transfection

100,000 cells were transfected with 100 pmol of siRNA or ASO and 5 μ l Lipofectamine 2000 (Life Technologies, USA). After 6 h post-transfection, culture media was changed. The list of small interfering RNAs (siRNAs), dicer substrate siRNAs (dsiRNAs) and ASOs is shown in Table 2.9 and Table 2.10. siRNAs are 19 base pairs (bp) long RNAs with 2 nt 3' DNA overhangs at each strand. The 5' end of the antisense strand with weaker Watson-Crick bonding is incorporated into the RNA-induced silencing complex (RISC). In contrast, dsiRNAs are 27-mer RNA duplexes as substrates for Dicer which helps introduce the RNA into RISC (Lee et al., 2004; Pham et al., 2004). The design of dsiRNAs developed by IDT allows production of potent 21-mer RNA duplex with increased RISC incorporation efficiency (Figure 2.5). ASOs generated by Isis Pharmaceuticals, USA are 18 nt long and are chemically modified by 2'-O-methoxyethyl (2'MOE) with phosphorothioate backbone which makes the ASOs more stable and more

RNA-like. 2'MOE ASOs have better binding affinity to the target pre-mRNA as RNA hybridizes more strongly to RNA than to DNA (Novosel and Borkhardt, 2008).

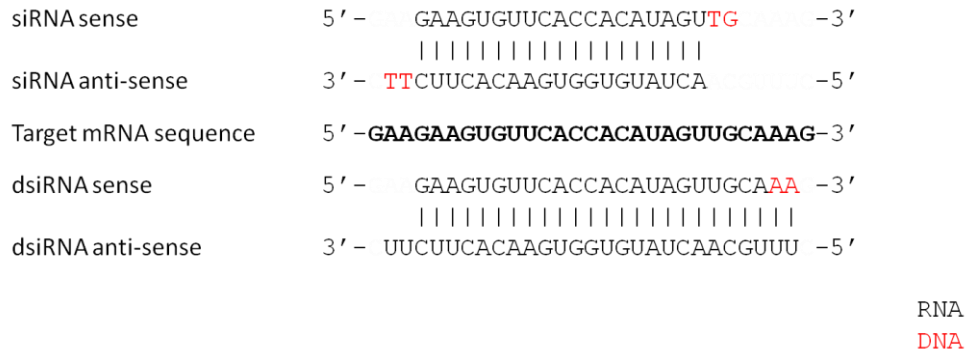


Figure 2.5 Example of design of a siRNA and a dsiRNA

siRNAs have two nt 3' overhang at either of the two strands. On the other hand, dsiRNAs have two nt 3' overhangs only at the antisense strand and an extended base-pairing region towards the 3' end of the sense strand to mimic dicer substrate RNA. Image is adapted from IDT webpage.

Table 2.9 List of siRNAs and dsRNAs

Name	Sequence	Company
Silencer Select Negative control #1 siRNA	—	Ambion, USA
Silencer Select SRSF1 siRNA s12725	Sense: GACCUAUGCAGUUCGAAAA dTdT	Ambion, USA
	Anti-sense: UUUUCGAACUGCAUAGGUC dAdT	
PTBP1 siRNA (Wollerton et al., 2004)	Sense: CUUCCAUCAUCCAGAGAA dTdT	IDT, USA
	Anti-sense: UUCUCUGGAAUGAUGGAAU dTdT	
hnRNP M dsRNA (Venables et al., 2008)	Sense: GAAGUCCUAAACAAGCAUA GUCUdGdA	IDT, USA
	Antisense: UCAGAC UAUGCUUGUUUAGGACUUCG	
TIA1 dsRNA (Izquierdo and Valcarcel, 2007)	Sense: GCUCUAAUUCUGCAACUCU UUAGdCdC	IDT, USA
	Antisense: GGCUAA AGAGUUGCAGAAUUAGAGCUU	
TIAR dsRNA (Izquierdo and Valcarcel, 2007)	Sense: CCAUGGAAUCAACAAGGAU UUGGdAdG	IDT, USA
	Antisense: CUCCAA AUCCUUGUUGAUUCCAUGGUU	

Sequences represent RNA, except that DNA nucleotides are shown as dN (dA, dT, dG, dC). Bold nucleotides represent core sequences of siRNAs and dsRNAs.

Table 2.10 List of ASOs

	Sequence
ASO-CD46-13-ESE1.1	TGTGGGTCTCATCAGTTA
ASO-CD46-13-ESE1.2	TCTCTGTGGGTCTCATCA
ASO-CD46-13-ESS1.1	TAAAAGCAAACCTTTCTC
ASO-CD46-13-ESS1.2	AATGATAAAAAGCAAACCT

2.4.2 Suspension cells

2.4.2.1 Transfection with minigenes

Jurkat E6.1 cells were subcultured one day before transfection so that the cells were in exponential growth phase during transfection. 1 million Jurkat E6.1 cells were transfected with 1 µg DNA (83 ng of minigene plasmid and 917 ng of mock plasmid) using 3 µl XtremeGENE HP transfection reagent (Roche, Switzerland).

2.4.2.2 siRNA and antisense oligonucleotides (ASO) transfection

Half a million of Jurkat E6.1 cells were transfected with 50 pmol of ASOs in 20 µl Nucleofector solution SE, by using the CL-120 program with Amaxa 4D-Nucleofector (Lonza, Switzerland).

2.5 Jurkat E6.1 cell activation

Jurkat E6.1 cells were seeded at a density of 250,000 cells/ml and incubated with phorbol-12-myristate-13-acetate (PMA; Sigma-Aldrich, USA; final conc.: 100 ng/ml) or dimethyl sulfoxide (DMSO; Sigma-Aldrich, USA) for 24 h. Alternatively, 500,000 cells/ml of Jurkat E6.1 were incubated in a 6-well plate coated with 10 µg/ml anti-CD3 (eBioscience, USA) antibodies and supplemented with 10 µg/ml anti-CD28 (eBioscience, USA) or anti-CD46 (BioLegend, USA). For CD3 activation, 1 ml 1x PBS with anti-CD3 antibodies was added to the 6-well plate and incubated at 4 °C overnight. Before adding cells, coated wells were washed with 1x PBS once and then 500 µl fresh media was added to the wells.

2.6 RNA extraction and DNase I treatment

48 h post transfection or at a particular time point, total RNA was extracted using Purelink RNA mini kit (Ambion, USA) and its concentration was measured using Nanodrop 2000 (Thermo Scientific, USA). 4 µg RNA was then treated with 2 U RQ1-RNase free DNase I (Promega, USA) to remove genomic and plasmid DNA. Subsequently, RNA was subjected to ethanol precipitation by first topping up the reaction to 100 µl with RNase-free water, then adding 0.1x volume of 3 M sodium acetate (pH 5.2) and 3x volume of 100% ethanol, and finally incubating at -80 °C overnight. RNA was centrifuged at 13500 rpm for 45 min and then washed with 70% ethanol once before resuspending the pellets in 10µl RNase-free water.

2.7 Reverse transcription (RT)-PCR and gel electrophoresis

1 µg DNase I-treated RNA was used to generate complementary DNA (cDNA) using Moloney Murine Leukemia Virus (M-MuLV) reverse transcriptase (New England Biolabs, USA) and oligo-dT at 42 °C. Subsequently, cDNA was amplified with each respective primer pairs (Table 2.11) using GoTaq polymerase (Promega, USA). The PCR conditions were as follows: 35 cycles of denaturation at 95 °C for 30 s, annealing at respective temperature for 40 s and extension at 72 °C at a polymerization speed of 1 kb/min. PCR products were separated by 1.2% - 2% (w/v) ethidium bromide-stained agarose gel and visualized under UV light. The splice products of minigenes were confirmed by sequencing (1st Base, Singapore). DNA bands on the agarose gel were cut and extracted using QIAquick gel extraction kit (Qiagen, Germany).

Table 2.11 List of RT-PCR primers

pcDNA vector	
pcDNA-F	GAGACCCAAGCTGGCTAGCGTT
pcDNA-R	GAGGCTGATCAGCGGGTTTAAAC
pSXN vector	
pSXN_βglobinEx1_F	AGGTGAACGTGGATGAAGTTGGTGGTG
pSXN_βglobinEx2/cTNTEx18.R	CGTGCAGCCTTTGACCTACTAGTGTG
Primers for detection of CD46 CYT	
CD46.Ex12.F	TTGTTGGAGTTGCAGTAATTTGTGTTGTC
CD46.Ex14a.R	CTTTGAACAAAAAGATTTCAAGCCACTTTCTTTAC
Primers for detection of CD45 alternative splicing	
CD45.Ex3.F	GGCAAAGCCCAACACCTTCCC
CD45.Ex9/10.R	GAAACTTTTCAACCCCTGGTGGCACAT
Primers for detection of SRSF1 isoforms (positive control for NMD)	
SF2.Ex3.F	CACTGGTGTCTGGAGTTTGTACGG
SF2.Ex4.R	TGGGGCAGGAATCCACTCCTATG
Control: GAPDH	
GAPDH.Ex6d.F	CCATGAGAAGTATGACAACAGCCT
GAPDH.Ex8a.R	AGTCTTCTGGGTGGCAGTGA
Primers for determining expression of cDNA in pCGT7 vector	
T7 primer.F	GATGACAGGTGGCCAACAGATG
Primers for determining SRSF1 expression level	
SRSF1.Ex2.F	GAAGACGCGGTGTATGGTCG
SRSF1.Ex2/3.R	CTTGAGGCAGTCCAGAGACAAC
Primers for determining of PTBP1 expression	
PTBP1.Ex2.F	CATTGTCCCAGATATAGCCGTTGGTACAAAG
PTBP1.Ex4.R	CTGGTTTTTCCCCTCAGCATCAGG
Primers for determining T-cell activation	
CD69.Ex2/3.F	CTTATCAGTGGGCCAATACAATTGTC
CD69.Ex4.R	TTCCTCTCTACCTGCGTATCG
Primers for determining hnRNP M expression level	
hnRNPM.Ex2.F	TGAAGGAGAACGACCTGCTCAG
hnRNPM.Ex3/4.R	GCACATCCCCTTGACTTTCCTTC
Primers for determining TIA1 expression level	
TIA1.Ex2.F	ATACGTGGTAACCTTTCCAGAG
TIA1.Ex3/4.R	CTTTGACTTCCTTACCCATTATCTTCC
Primers for determining TIAL1 expression level	
TIAL1.Ex4.F	GAGGTCAAAGTAAACTGGGCAACC
TIAL1.Ex5/6.R	TACCCGGGCATCCGATATTTTACC
TIAL1.Ex7.R	ATTGGTTCGGATTTGACGACCAC

2.8 Semi-quantitative (Semi-Q) PCR

10 pmol forward primer was 5' end labeled using 10U T4 PNK (New England Biolabs, USA) and 10.2 pmol γ -³²P-ATP (60 μ Ci; Perkin Elmer, USA). Labeled primers were then purified by illustra Microspin G-25 column (GE Healthcare Life Sciences, USA) and added into primer mixture containing 100 pmol unlabeled reverse primer and 90 pmol unlabeled forward primer. cDNA was amplified by PCR using Go-Taq polymerase (Promega, USA) with the radiolabeled primer mixture for 23-35 cycles. The PCR products were then separated by 4.5-8% (w/v) native polyacrylamide gel (Acrylamide/Bis 29:1; Biorad, USA) in 1x TBE buffer. Radioactive signals from the gel were subsequently exposed to storage phosphor imaging screen which were read by Typhoon Scanner. The analysis of band intensity was done with ImageQuant TL software (GE healthcare life sciences, USA). For image formation, X-ray films (Kodak, USA) were exposed with the gel for 12-48 h in X-ray cassette at -80 °C and then processed by Kodak Model 2000 X-ray film processor.

2.9 Analysis of *CD46* alternative splicing patterns in human tissues

Total RNAs from 20 human tissues (Ambion, USA) were first used as template to generate cDNAs. These cDNAs were then amplified using semi-quantitative (radioactive) PCR with *CD46* exon 6 and exon 10, *CD46* exon 12 and exon 14 primer sets (Table 2.1 and Table 2.11) to analyze the splice products of STP and cytoplasmic regions.

2.10 Western blot

2.10.1 Sample preparation

Cells were trypsinized and washed with 1x PBS twice before adding lysis buffer [50 mM Tris-HCl, pH8; 150 mM NaCl; 1% Triton X-100; 10% glycerol; 1 mM EDTA; 1x cComplete EDTA-free protease inhibitor cocktail (Roche, Switzerland)]. Protein concentration was measured by Bradford assay and 20 μ g proteins from

each sample were boiled for 10 min with SDS loading buffer before loading into gel.

2.10.2 SDS-PAGE and transfer

Samples were separated by 10-12.5% polyacrylamide gel (Acrylamide/Bis 37.5:1; Biorad, USA) in 1x SDS running buffer [25 mM Tris, 192 mM glycine, 0.1% SDS]. Proteins were transferred to PVDF membrane (Biorad, USA) in transfer buffer [25 mM Tris, 192 mM glycine, 20% methanol] at 100 V for 1 h.

2.10.3 Detection

Membrane was blocked in 5% milk in 1x TBST [144 mM NaCl, 20 mM Tris, 0.05% Tween 20, pH7.6] for 1 h and washed with 1x TBST thrice, each for 5 min. Next, the membrane was incubated with the primary antibody diluted in 1x TBST at 4 °C overnight followed by three washes with 1x TBST and incubation with the secondary anti-mouse/rabbit/goat IgG antibodies (BioLegend, USA/Sigma-Aldrich, USA/Santa cruz biotechnology, USA) diluted in 1x TBST with or without 5% milk (Table 2.12). After three washes, proteins were detected by Western Lightning Plus ECL (Perkin Elmer, USA).

Table 2.12 List of antibodies

Primary Ab		Secondary Ab	
Anti-SRSF1 (SF2)	1:1000	Anti-mouse IgG	1:10000
Anti-hnRNP M (Santa cruz, USA)	1:1000	Anti-mouse IgG	1:10000
Anti-hnRNP A1 (Santa cruz, USA)	1:1000	Anti-mouse IgG	1:10000
Anti-hnRNP F (Santa cruz, USA)	1:1000	Anti-rabbit IgG	1:10000
Anti-hnRNP H (Santa cruz, USA)	1:1000	Anti-rabbit IgG	1:10000
Anti-KSRP (Bethyl Laboratories, USA)	1:2000	Anti-rabbit IgG	1:10000
Anti-PTBP1 (Life Technologies, USA)	1:1000	Anti-mouse IgG	1:10000
Anti-TIA1 (Santa cruz, USA)	1:1000	Anti-goat IgG	1:5000 in 5% milk
Anti-TIAL1 (Santa cruz, USA)	1:1000	Anti-goat IgG	1:5000 in 5% milk
Anti-CD46 (non-reducing sample) (BioLegend, USA)	1:1000	Anti-mouse IgG	1:10000
Anti-Tubulin (Santa cruz, USA)	1:500	Anti-mouse IgG	1:10000
Anti-GAPDH (Santa cruz, USA)	1:1000	Anti-rabbit IgG	1:10000

2.11 RNA pull-down

RNA pull-down was performed using Pierce Magnetic RNA-Protein Pull-down kit (USA). First, 50 pmol synthetic RNA (IDT, USA, Table 2.13) was ligated with desthiobiotinylated cytidine bisphosphate at the 3' end and then purified by ethanol precipitation. Subsequently, labeled RNA was bound to the streptavidin magnetic beads at room temperature followed by incubation with 120 µg HeLa nuclear extract at 4 °C for 2 h with agitation. After incubation, magnetic beads were washed three times with wash buffer [20 mM Tris (pH 7.5), 10 mM NaCl, 0.1% Tween-20] and the bound proteins were eluted using biotin elution buffer. Bound proteins were analyzed by SDS-PAGE and silver staining. For ASO annealing, labeled RNA mixed with ASOs at 1:3 ratio was heated at 90 °C for 2 min and then gradually cooled down to room temperature. RNA:ASO duplexes were subsequently bound to magnetic beads and subjected to pull-down procedure.

Table 2.13 List of RNA oligonucleotides for RNA pull-down

CD46.Ex13ESS1.WT	AUGAG AGAAAGGUUUGCUUUUAUCAUUAA AAGGU
CD46.Ex13ESS1.Mut1	AUGAG AGAAACGUU AAAUCUAUCAUUAA AAGGU
CD46.Ex13ESE1.WT	ACCUA ACUGAUGAGACCCACAGAGAAGUA AAAUU
CD46.Ex13ESE1.Mut3	ACCUA ACUGAUCCGU CCACAGAGAAGUAAAAUU
CD46.Ex13ESE1.Mut5	ACCUA ACUGAUCCGU CCCA UAGAU AAGUAAAAUU
SMN.In7ISS.WT	CCAGCAUUAUGAAAGU
SMN.In7ISS.2A-2C	CC C GCAUUAUGAA C GU

Core sequences of *cis*-acting elements are shown in bold. Mutations are labeled in red. SMN RNA oligonucleotide sequences were obtained from Hua et al., 2008.

2.12 Silver staining

Protein gels were first incubated in 50% methanol twice, then in 5% methanol and in water with 60 mM DTT once for fixation. Each incubation was 15 min long with constant shaking. Subsequently, the gel was incubated in 0.1% silver nitrate solution (Sigma-Aldrich, USA) for 15 min followed by incubation with developing solution (3% sodium carbonate (Sigma-Aldrich, USA) and 0.05% formaldehyde). The protein bands which were absent or less intense in the mutant control samples were analyzed by LC-MS/MS, a service provided by NTU mass spectrometry core facilities, and confirmed by western blot.

2.13 Flow cytometry

2.13.1 Cell sorting

Activated Jurkat T cells were collected by centrifugation and washed once with 1x PBS. Cells were resuspended in 100 μ l 1x PBS and incubated with 2 μ l goat serum (Sigma-Aldrich, USA) for 2 min followed by 15 min-incubation with FITC-conjugated CD69 (BioLegend, USA). After washing with 1x PBS, stained cells were resuspended in culture media and sorted using FACSAria Cell Sorter (Becton Dickinson, USA).

2.13.2 Cell analysis

Stained cells were resuspended in 400 μ l 1x PBS and analyzed using LSRFortessa X-20 (Becton Dickinson, USA). The data were processed using FlowJo software.

2.14 Inhibition of NMD

HEK293T cells at 80% confluency were treated with 100 μ g/ml cycloheximide (CHX; Sigma-Aldrich, USA) for 5 h before RNA extraction. The effect of CHX on NMD was examined using RT-PCR. SRSF1 was served as positive control as it has two NMD-sensitive mRNA isoforms, whose expression levels were increased by inhibition of translation by CHX.

2.15 Preparation of HeLa nuclear extract

HeLa cells were grown in suspension and harvested at the density of 5×10^5 cells/ml. Cells were washed with 1x PBS twice and then the packed cell volume (PCV) was determined. Cell pellet was resuspended in 2x PCV of Buffer A [Hypotonic, 10 mM HEPES-KOH pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT] and incubated on ice for 10 min before homogenization using glass dounce homogenizer. The cell lysate was then centrifuged at 2000 x g for 15 min. The supernatant corresponded to cytoplasmic S100 and the pellet corresponded to nuclei. 0.5x packed nuclear volume (PNV) of low salt Buffer C [25% glycerol, 20 mM HEPES-KOH pH7.9, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 200 mM NaCl, 0.5 mM DTT, 0.5 mM PMSF] was drop-wisely added to the nuclei with constant mixing followed by addition of 0.5x PNV of high salt Buffer C [25% glycerol, 20 mM HEPES-KOH pH7.9, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 1.2 M NaCl, 0.5 mM DTT, 0.5 mM PMSF]. Nuclei lysate was homogenized before incubating at 4 °C for 30 min with constant agitation followed by centrifugation at high speed. The nuclear extract in supernatant was transferred to dialysis tubing (nominal molecular weight cut off: 3500 kD; Fisherbrand, USA) and dialyzed against 1 L Buffer D [20 mM HEPES-KOH pH 7.9, 20% glycerol, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT] at 4 °C for 2 h. Dialyzed nuclear extract was aliquoted and snap frozen using liquid nitrogen.

3. RESULTS

3.1 Alternative splicing of *CD46* is tissue selective

Variations of alternative splicing across tissues are more common than variations among individuals (Wang et al., 2008). Previous studies have illustrated that *CD46* splice isoforms are differentially expressed across tissues (Johnstone et al., 1993). To further verify these findings, total RNA from 20 human tissues purchased from Ambion, Life Technologies was analyzed by semi-qRT-PCR with two primer sets annealing to the constitutive exons flanking the cassette exon/s to determine the prevalence of STP and cytoplasmic isoforms in different tissues. The total RNA is comprised of RNA from all cell types of that tissue and is from at least 3 tissue donors, and so the tissue-selective alternative splicing pattern of *CD46* is a collection of *CD46* isoforms of different cell types in that tissue.

Both *CD46* cytoplasmic tails were detected in all tissue RNAs with a predominance of CYT2 (exon 13 skipping), suggesting that the inclusion of exon 13 is generally repressed (Figure 3.1A). The exon 13 inclusion level in different tissues varied, ranging from as low as 6% (kidney, lane 8 and small intestine, lane 15) to as high as 47% (lung, lane 10).

At least two STP isoforms were observed in all tissue RNAs and the predominant two isoforms are inclusion of single exon and two exons (Figure 3.1B). The identities of the included exons in mRNA were determined by PCR with exons 7, 8 and 9 specific primers (data not show). Results showed that the single included exon was exon 9 giving rise to the STP-C isoform, while exons 8 and 9 were included together in mRNA giving rise to the STP-BC isoform. Most tissues predominantly express either STP-C or STP-BC isoforms, while nearly equivalent expression level of both STP-C and STP-BC isoform was seen in prostate, thyroid and trachea (lanes 13,19-20). Few tissues, such as bladder and brain (lanes 2,3), have a little amount of mRNAs with all three STP exon skipping, and a very low level of *CD46* STP-ABC isoform was only detected in lung (lane 10). Similar to the findings in the Johnstone et al.1993, brain and heart mainly express STP-C

isoform (Figure 3.1B, lanes 3,7), while kidney and lung predominantly express STP-BC isoform (lanes 8,10). Here, prostate and testes are shown to express equivalent STP-BC and STP-C isoforms or predominant STP-C isoform, respectively (lanes 13,17). However, in the Johnstone et al. 1993, prostate mainly expresses STP-C isoform while testes express both STP-BC and STP-C isoforms. In conclusion, *CD46* isoform expression changes across different tissues and the two most predominant isoforms are STP-C isoform with CYT2 (C2) and STP-BC isoform with CYT2 (BC2). There is no clear correlation between the alternative splicing pattern of STP exons especially exon 8, and inclusion of exon 13 when I compared exon 8 inclusion against exon 13 inclusion. My result is largely consistent with the previous data, showing that CYT2 is the predominant cytoplasmic tail and exon 7 is mostly skipped.

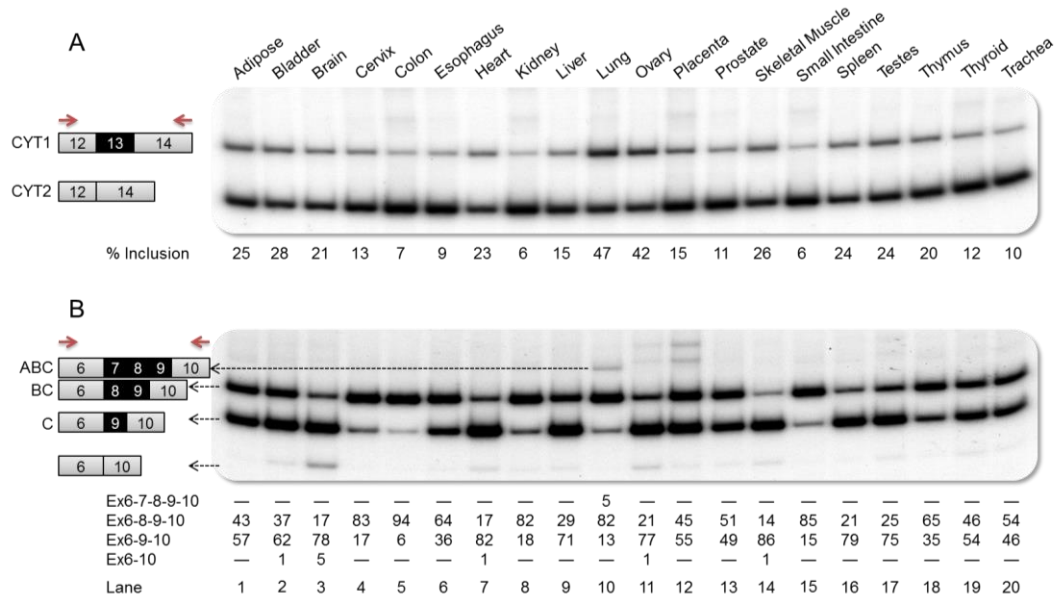


Figure 3.1 Semi-Q PCR detected *CD46* alternative splicing patterns in 20 human tissues

Radioactive RT-PCR was performed to examine alternative splicing patterns of STP exons and exon 13. Primers (denoted as red arrows) annealing to exons 12 and 14 were used to determine inclusion of exon 13 while primers annealing to exons 6 and 10 were used to examine the alternative splicing patterns of exons 7, 8 and 9. (A) CYT2 is predominantly expressed in human tissues. Skipping of exon 13 gives rise to CYT2 while inclusion of exon 13 gives rise to CYT1. (B) STP-BC and STP-C are the two most common STP isoforms. For the STP region, inclusion of a single STP exon, usually exon 9, generates the STP-C isoform, inclusion of two STP exons, usually exons 8 and 9, generates the STP-BC isoform, and inclusion of all three STP exons produces the STP-ABC isoform. The splice isoform for each DNA band is indicated at the side of the gel images.

3.2 The 5'ss flanking exons 7 and 8 are recognized by U1 using asymmetric loop register

The 5'ss motifs flanking exons 7 and 8 substantially deviate from the consensus sequence and so they would exhibit a very weak 5'ss strength compared to exon 9 as predicted by different 5'ss scoring models (Table 3.1) (Brunak et al., 1991; Reese et al., 1997; Yeo and Burge, 2004). If the 5'ss flanking exons 7 and 8 base pair with U1 snRNA in the canonical register, only 5 base pairs would form between U1 snRNA and 5'ss (Roca et al., 2013; Roca et al., 2012). However, if two nucleotides at 5'ss positions +3 and +4 and one nucleotide at U1 snRNA position 6 form an asymmetric loop in the helix, this allows formation of 4 additional base pairs thus stabilizing the U1/5'ss helix despite the disruption by the unpaired nucleotides (Figure 3.2B and Figure 3.3B). The asymmetric loop 1 (+3/+4) is defined as a loop in the 5'ss/U1 duplex that has uneven number of unpaired nucleotides, in which the number differs by 1. The +3/+4 represents the positions of the looped nucleotides at the 5'ss.

Here, I hypothesized that the 5'ss flanking exons 7 and 8 use this non-canonical register for U1 base-pairing. To examine this hypothesis, mutational analysis and U1 suppressor experiment were performed. First, *CD46*-exon 6-7-10 minigene containing exons 6, 7 and 10 with full-length intron 6, first 50 nt of intron 7 and last 350 nt of intron 9 was used as the wild-type control for studying the test 5'ss recognition by U1 (Figure 3.2A; see Appendix I for details). Exons 8 and 9 were removed from the minigene in order to simplify data interpretation as exon 7 is very lowly included in endogenous mRNA. The minigene transcripts were detected by RT-PCR using vector specific primers, allowing specific detection of minigene mRNA but not endogenous *CD46* mRNA.

In the context of this minigene, exon 7 was poorly recognized and included in the mature transcript (Figure 3.2C, lane 1), and this is consistent with endogenous *CD46* mRNA in which exon 7 was also rarely detected (Figure 3.1B). In order to improve exon 7 inclusion, the position +1 of the 3'ss flanking exon 7 (first nucleotide in exon 7) was mutated from uridine (non-consensus) to guanosine

(consensus) in the context of *CD46*-exon 6-7-10 minigene (Table 2.3 and Figure 3.2B) (Sheth et al., 2006). Inclusion of exon 7 was largely improved by this mutation (Figure 3.2C, lane 1 vs. 2). Therefore, this minigene with an improved 3'ss was used in the subsequent mutational assay. When the adenosine at position -2 of the 5'ss flanking exon 7 was mutated to cytosine, it reduced the number of base pairs in either the canonical or asymmetric loop register, and inclusion level of exon 7 was largely decreased by 46% (Figure 3.2B, green colored nucleotide and Figure 3.2C, lane 2 vs.3). Nucleotides at positions +6 and +7 of the 5'ss form base-pairing with U1 in the asymmetric loop register but not in the canonical register (Figure 3.2B, red and dark blue nucleotides). Hence, nucleotide substitutions at these two positions were introduced to examine whether recognition of exon 7 is affected. If the U1/5'ss base-pairing is in the canonical register, inclusion of exon 7 will not be affected by these mutations. Conversely, if the U1/5'ss recognition is using the asymmetric loop register, inclusion of exon 7 will be reduced. Inclusion of exon 7 was completely abolished by these two mutations (Figure 3.2C, lane 2 vs. 4/7), suggesting that the 5'ss is recognized by U1 in such non-canonical register.

Subsequently, rescue experiments were carried out to further confirm the base-pairing register used by the 5'ss flanking exon 7. A U1 snRNA suppressor is a plasmid-encoded U1 that carries mutations at specific positions to base pair to the corresponding mutated 5'ss. U1 snRNA suppressor was co-introduced into HEK293T together with mutant *CD46* minigenes to restore the recognition of the mutated 5'ss and so exon inclusion.

In the +6C mutant minigene, position +6 of the 5'ss was mutated to cytosine. In order to create a base pair at this position in the canonical register, a guanosine at position 3 of U1 snRNA (G3) was introduced (Figure 3.2B, sky blue nucleotide). However, this U1 suppressor G3 failed to improve exon inclusion (Figure 3.2C, lane 4 vs. 5). Similarly, U1 suppressor G2 (Figure 3.2B, dark green nucleotide) is required to form base-pairing with cytosine at position +7 of the 5'ss in the canonical register in the context of +7C mutant minigene. Again, no inclusion of exon 7 was detected in samples supplemented with U1 suppressor G2 (lane 7 vs. 8).

These results suggest that the 5'ss flanking exon 7 does not use the canonical register for U1 recognition.

In contrast, in the asymmetric loop register, U1 suppressors G4 and G3 (Figure 3.2B, purple and pink nucleotides, respectively) were used to restore the base-pairing with the cytosine at positions +6 and +7 of the 5'ss, respectively. When +6C mutant minigene was co-expressed together with U1 suppressor G4, a 3% increase in exon 7 containing minigene transcript was detected (Figure 3.2C, lane 4 vs. 6). When U1 suppressor G3 was applied to rescue recognition of the 5'ss with +7C mutation, exon 7 inclusion was restored back to similar level as the original construct (wild-type with improved 3'ss) (lane 2/7 vs. 9). This result confirms that the 5'ss flanking exon 7 base pairs with U1 in the asymmetric loop register.

Analogous experiments were performed to examine the U1/5'ss base-pairing for the 5'ss flanking exon 8. Although exon 9 is the only exon identified in endogenous transcript with single included STP exon, it is possible that exon 8 is singly included in mRNA without exon 9 in the context of minigene. To simplify data interpretation, exon 9 was removed from the minigene. This minigene contains exons 6, 7, 8 and 10 with full length of introns 6 and 7, first 350 nt of intron 8 and last 350 nt of intron 9 (Figure 3.3A; see Appendix I for details). Our *CD46*-exon 6-7-8-10 minigene thus showed two spliced products, one with exons 6, 10 and another with exons 6, 8 and 10 (Figure 3.3C, lane 1).

Like in exon 7, mutation at position -2 of the 5'ss abolished inclusion of exon 8 (Figure 3.3C, lane 1 vs. 2). Additionally, mutations at positions +6 and +7 of the 5'ss disrupting the putative base-pairings in the asymmetric loop register strongly reduced exon 8 inclusion ($\approx 50\%$ reduction) (Figure 3.3C, lane 1 vs. 3/6). For +6C mutation, both U1 suppressors G3 and G4, which form base-pairing with the 5'ss +6C in the canonical or non-canonical register, respectively, barely rescued the inclusion (lane 3 vs. 4/5). On the other hand, U1 suppressor G3 restored inclusion of exon 8 with the +7C mutation using the asymmetric loop register (lane 6 vs. 8) while U1 suppressor G2 which rescues the canonical U1/5'ss base-pairing failed to (lane 6 vs. 7). These results imply the use of asymmetric loop register for the

recognition of the 5'ss flanking exon 8 by U1. In summary, both mutational analysis and U1 suppressor experiment strongly support my hypothesis that the 5'ss flanking both exons 7 and 8 are recognized by U1 via the asymmetric loop 1 (+3/+4) register.

Table 3.1 Prediction of 5'ss and 3'ss strength using different scoring models

3'ss	Sequence	NN	MAXENT	MDD	MM	WMM
Exon6	agguuucucuaauuuuccagUGG	0.41	6.59	–	6.47	7.62
Exon7	aacauuauuuuguuuccuagUGC	0.92	7.08	–	7.71	8.24
Exon8	auuauuauuuuguuuccagUGU	0.95	8.12	–	9.72	10.52
Exon9	aaucaaacuauuuuucuagGUC	0.98	7.93	–	8.38	7.64
Exon12	uguuuccugguuucuuauagUUG	0.93	7.75	–	9.65	11.09
Exon13	uuuucucuuccucuguucagCAC	0.89	9.4	–	10.86	12.88
Exon14	cuuuauaccuugguuugcagGAA	0.85	8.18	–	9.15	8.41
<hr/>						
5'ss	Sequence	NN	MAXENT	MDD	MM	WMM
Exon6	AAGguacaaag	0.89	6.82	12.18	6.32	4.92
Exon7	CAGguuuagua	0.02	3.14	5.88	3.33	0.79
Exon8	CAGguuuagua	0.01	3.14	5.88	3.33	0.79
Exon9	CAGguugguua	0.69	8.08	11.68	7.20	5.75
Exon12	AGGguaauua	0.93	5.83	10.08	5.53	6.97
Exon13	AAGguaucugu	0.74	7.56	10.48	5.81	5.93
Exon14	UCAugaugaug	–	-15.02	-12.27	-12.82	-14.92

Footnote: The 5'ss flanking constitutive exon 6 and alternative exon 9 have much higher score than the 5'ss flanking exons 7 and 8. These numbers help to explain the low inclusion level of both exons 7 and 8. Constitutive and cassette exons are flanked by 3'ss with similar strength except the 3'ss flanking exon 6, whose score is lower. The prediction suggests that 5'ss play a more important role than 3'ss in determining inclusion of STP exons and exon 13. Lowercase sequences depict introns; Uppercase sequences depict exons; NN, Neural Network; MAXENT, Maximum Entropy Model; MDD, Maximum Dependence Decomposition Model, MM, First-order Markov Model; WMM, Weight Matrix Model (Brunak et al., 1991; Reese et al., 1997; Yeo and Burge, 2004).

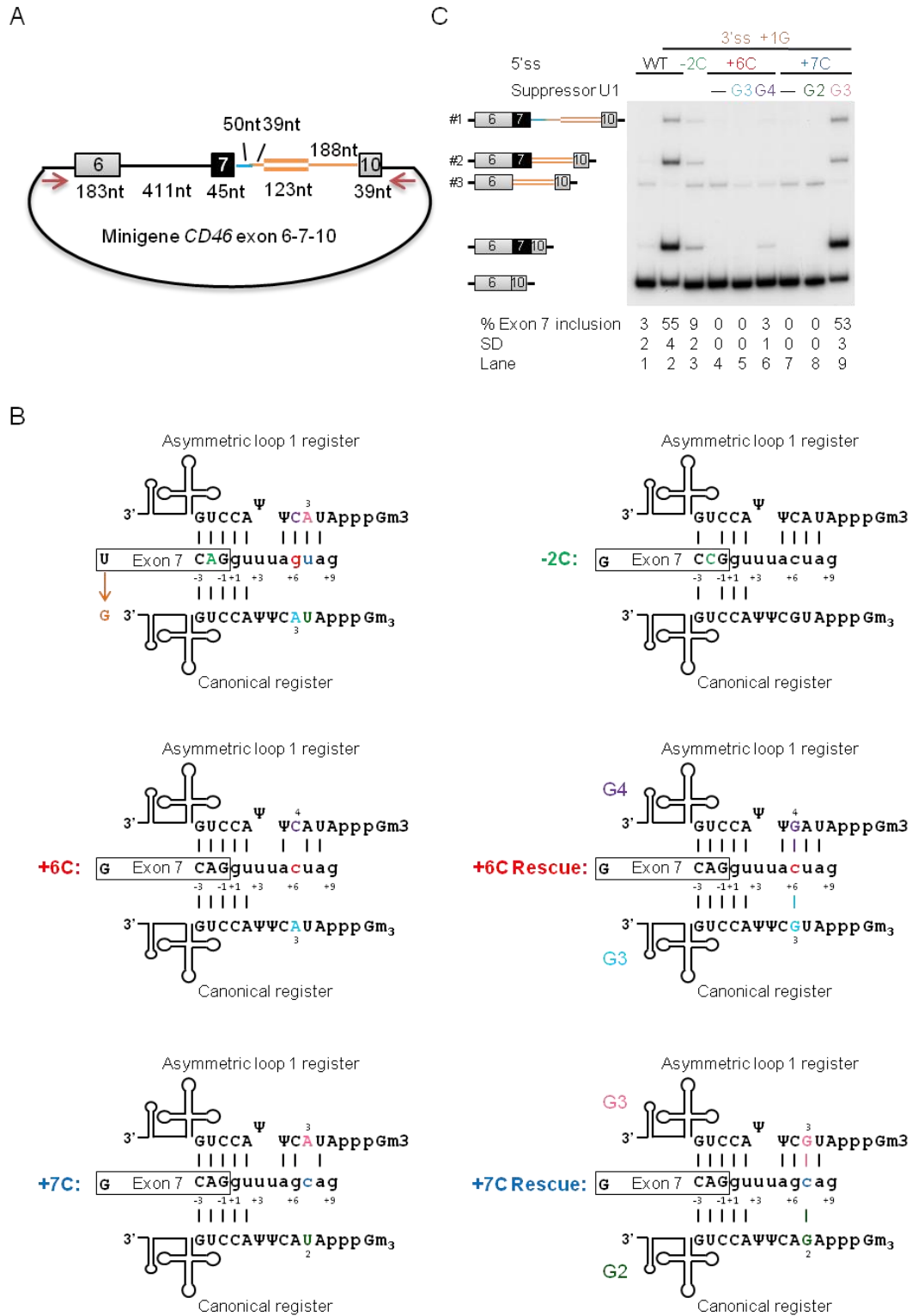


Figure 3.2 Recognition of the 5'ss flanking exon 7 by U1 using asymmetric loop 1 (+3/+4)

(Continued on the next page)

(Continued from the previous page)

(A) Schematic of *CD46*-exon 6-7-10 minigene used to study the 5'ss/U1 snRNA interaction. Red arrows depict RT-PCR primers for detecting minigene transcripts.

(B) Schematic of U1/5'ss base-pairing for the 5'ss flanking exon 7 in the canonical or asymmetric loop register. The number of base pairs between the 5'ss flanking exon 7 and U1 snRNA is increased by forming an asymmetric loop involving two nucleotides at 5'ss positions +3 and +4, and one nucleotide at U1 snRNA.

Uppercase in 5'ss sequence indicates the exonic sequence while lowercase indicates intronic sequence. The position +1 of 5'ss is conventionally assigned to the first intronic nucleotide while the position -1 is given to the last exonic nucleotide.

Position numbering in U1 snRNA starts from the 5' end of snRNA. Font colors label the point mutation at the 3'ss and the different point mutations at the 5'ss, and the corresponding rescue nucleotides at suppressor U1 snRNAs. (C) Mutational analysis and suppressor U1 experiments demonstrate asymmetric loop register for the 5'ss flanking *CD46* exon 7. Inclusion of exon 7 is negligible in wild type and it can be improved by increasing 3'ss strength (lanes 1 vs. 2). Mutations disrupted the U1 snRNA binding (lanes 3,4,7) and the effects can only be rescued by U1

suppressor that restores base-pairing using the asymmetric loop register (lane 9). A cryptic 5'ss (AAU | gucaguuu) which is 212 nt downstream of exon 7 was activated and resulted in partial intron retention (Band #1). Besides, a 123-nt long cryptic exon within intron 9 sequence was activated (Bands #2 and #3). This cryptic exon is 89 nt away from exon 7 and uses the same cryptic 5'ss and a novel cryptic 3'ss (uaauuag | AAU). Presence of these cryptic splice products could be due to the artificially shortened introns in the minigenes. The location of this cryptic exon is close to the junction of intron deletion, and so it is likely that the deleted intronic sequence possesses repressing effect on the inclusion of this cryptic exon (see Appendix I for details). Nevertheless, this cryptic exon does not change the conclusions of this experiment. Each combination of mutant minigene and suppressor was analyzed at least in triplicate. SD: standard deviation.

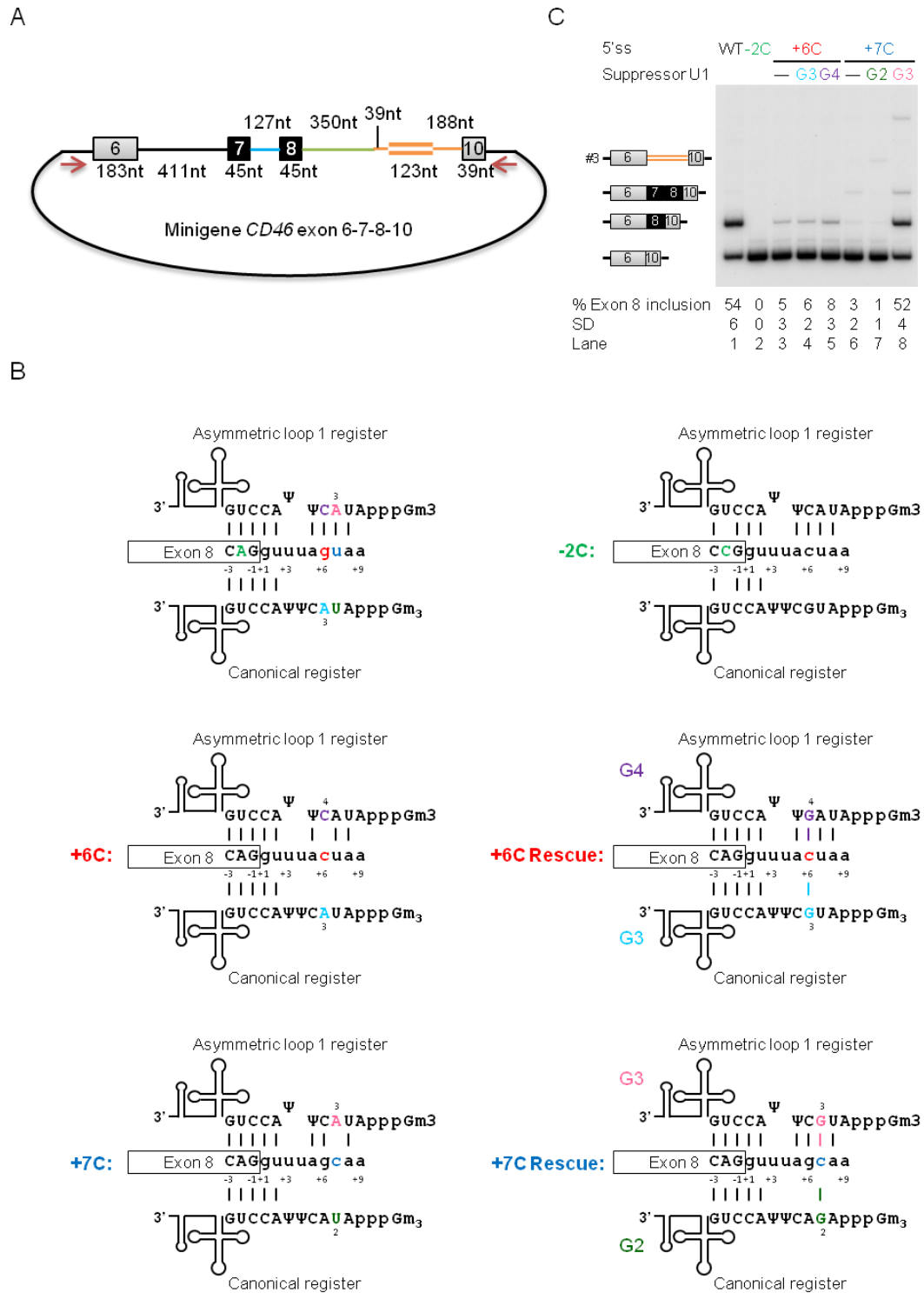


Figure 3.3 Recognition of the 5'ss flanking exon 8 by U1 using asymmetric loop 1 (+3/+4)

(Continued on the next page)

(Continued from the previous page)

(A) Schematic of *CD46*-exon 6-7-8-10 minigene used to study the 5'ss/U1 snRNA interaction. Red arrows depict RT-PCR primers for detecting minigene transcripts. (B) Schematic of U1/5'ss base-pairing for the 5'ss flanking exon 8 in the canonical or asymmetric loop register. The number of base pairs between the 5'ss flanking exon 8 and U1 snRNA is increased by forming an asymmetric loop involving two nucleotides at 5'ss positions +3 and +4, and one nucleotide at U1 snRNA. Font colors label the different point mutations at the 5'ss, and the corresponding rescue nucleotides at suppressor U1 snRNAs. (C) Mutational analysis and suppressor U1 experiments demonstrate asymmetric loop register for the 5'ss flanking *CD46* exon 8. Exon 8 is 50% included in the context of wild-type minigene while little exon 7 inclusion was detected (lane 1). Mutations disrupted the U1 snRNA binding (lane 1 vs. 2,3,6), and the effects were only rescued by U1 suppressor that restores base-pairing using the asymmetric loop register (lane 1 vs.6 vs. 8). The same 123-nt long cryptic exon as above (Figure 3.2C) was also activated in the context of *CD46*-exon 6-7-8-10 minigene pre-mRNA (Band #3). Each combination of mutant minigene and suppressor was analyzed at least in triplicate. SD: standard deviation.

3.3 Multiple *cis*-acting elements regulate the alternative splicing pattern of exon 13

From here on, characterization of the alternative splicing pattern of cassette exon 13 was the focus of my thesis because the different biological functions of the two cytoplasmic tails are important and well studied, and its misregulation is associated with autoimmune diseases.

3.3.1 Exon 13 is rich in ESEs and ESSs

In order to understand the alternative splicing regulation of exon 13, auxiliary *cis*-acting elements in the exon 13 were first identified. Serial deletions of exon 13 sequence were introduced into the *CD46* minigene containing exons 12, 13 and 14 with shortened introns 12 and 13 to determine the importance of these short sequences in splicing. *Cis*-acting elements generally have a length of 6-8 nucleotides (Cartegni et al., 2002), therefore the 93-nt long exon 13 sequence excluding the portions at the 5'ss and 3'ss, leaving 2 nt and 3 nt at the 5' and 3' end of exon 13 intact, was divided into 11 fragments of 8 nucleotides (Figure 3.4).

Three possible outcomes of this deletion assay are expected: (i) the deletion causes no effect on the alternative splicing pattern, suggesting that the deleted sequence contains no *cis*-acting elements, (ii) the deletion increases inclusion, suggesting that the deleted sequence contains an ESS, and (iii) the deletion results in more exon skipping, suggesting that the deleted sequence contains an ESE. A potential drawback for this deletion assay is that the new junctions formed by the deletions may create new *cis*-acting elements and so affect the splicing outcome. In order to limit such false positive results, a set of staggered deletions with different junctions was generated to further support the results from the first set of deletions. If results from the two sets of deletion assays were not consistent, the regulatory activity of the corresponding region was considered as inconclusive.

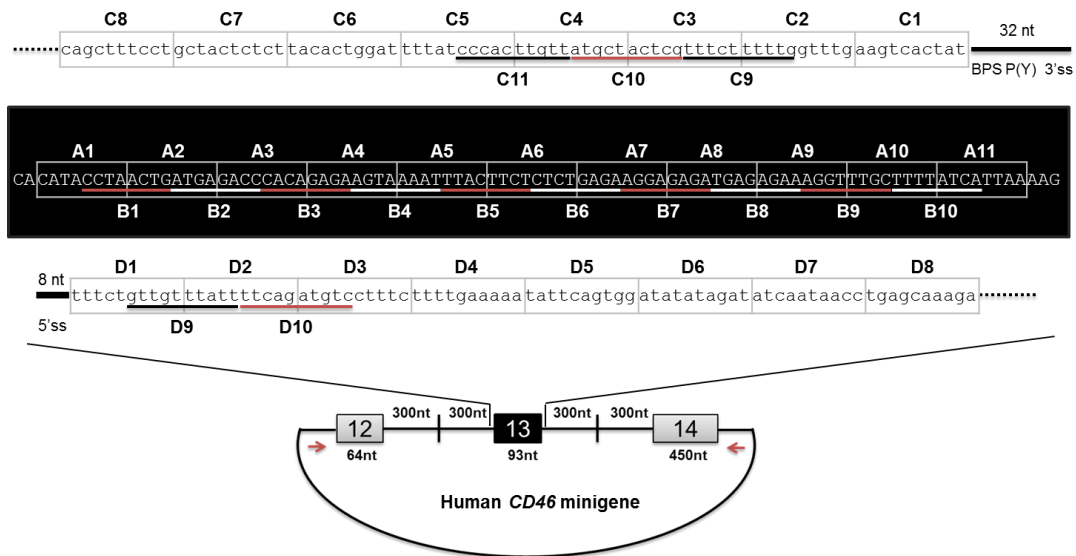


Figure 3.4 Schematic map of CD46-exon 12-13-14 minigene and serial deletions

This expression plasmid, schematically shown in bottom, contains exons 12, 13 and 14 with 300 nt of terminal intronic sequences of introns 12 and 13. The design of serial deletions in exonic and intronic sequences is shown above the plasmid schematic. Exon 13 shown in black box is 93 nt long. The sequences flanking the black box are introns 12 and 13. The first 2 nt and the last 3 nt of exonic sequences which are part of the 3'ss and 5'ss, respectively, were excluded from the deletions, not to affect the splice sites. The rest of the exonic sequence was split into 11 fragments, and each is 8 nt long. These 11 fragments are serially deleted (A1-A11, marked in light gray bordered boxes) in minigenes to examine their regulatory roles in exon 13 inclusion. Likewise, 8 deletions of 10-nt intronic sequence were introduced into introns 12 and 13, respectively (C1-C8, D1-D8, marked in light gray bordered boxes). The first deletion in intron 12 starts 32 nt away from the intron-exon junction, leaving the 3'ss, the polypyrimidine tract (P(Y)) and the predicted branch point sequence intact. On the other hand, the first deletion in intron 13 starts 8 nt away from the exon-intron junction to avoid disruption of the 5'ss. In order to further validate the results derived from the first set deletions, staggered deletions which span half of the sequences from two adjacent deletions were generated (marked by underlines; B1-B10 for exon 13; C9-C11 for intron 12;

D9-D10 for intron 13). Dark red solid arrows depict vector specific primers used for RT-PCR. See Appendix II for sequence details about this minigene.

To identify *cis*-acting elements that regulate exon 13 inclusion, wild-type and mutant minigenes (Figure 3.4, A1-A11) were separately introduced into HeLa, HEK293T and Jurkat cell lines by transfection. RNA extractions were performed after 48 h post-transfection and subsequently the alternative splicing patterns of minigene transcripts were determined by RT-PCR with vector specific primers. Only those deletions that cause significant increase or decrease in exon inclusion level as compared to the wild type are identified as enhancers or silencers. The strength of each *cis*-acting element is determined by the absolute difference in exon inclusion level. The stronger the increase or decrease in exon inclusion, the stronger the silencer or enhancer is.

The majority of wild-type *CD46* minigene splice products contain exon 13 in all three different cell lines (Figure 3.5, lane 1), and so this alternative splicing pattern is different from that of endogenous *CD46* mRNA (see Discussion). All 11 deletions affected the alternative splicing pattern of *CD46* exon 13 to different extents (Figure 3.5A). Deletion of A2, A3, A4 or A7 sequences resulted in nearly complete skipping of exon 13 (Figure 3.5A, lanes 3-5,8). This result strongly suggests that these sequences contain ESE/s which promote the recognition and inclusion of exon 13 since their absence caused virtually no inclusion of exon 13. In contrast, A5, A6, A8, A9, A10 and A11 sequences contain ESS/s as exon 13 skipping was significantly decreased when either of these sequences was removed (Figure 3.5A, lanes 6-7,9-12).

Staggered deletions of exon 13 sequence were then generated to further map the regions of ESEs and ESSs (Figure 3.5B). Deletion of B1, B2 or B3 sequence caused exon skipping (Figure 3.5B, lanes 2-4) as their overlapping deletions A2, A3 and A4 did. Besides, removal of B5, B9 or B10 fragment resulted in more exon inclusion (Figure 3.5B, lanes 6,10-11) which is consistent with the results of A5, A6, A9, A10 and A11. The consistency of these two deletion assays very strongly suggests the existence of multiple ESEs and ESSs within exon 13 and their essential regulatory roles in determining exon 13 inclusion. Consistent results were seen in all three cell lines, suggesting that these *cis*-acting elements are similarly regulated in different cell types.

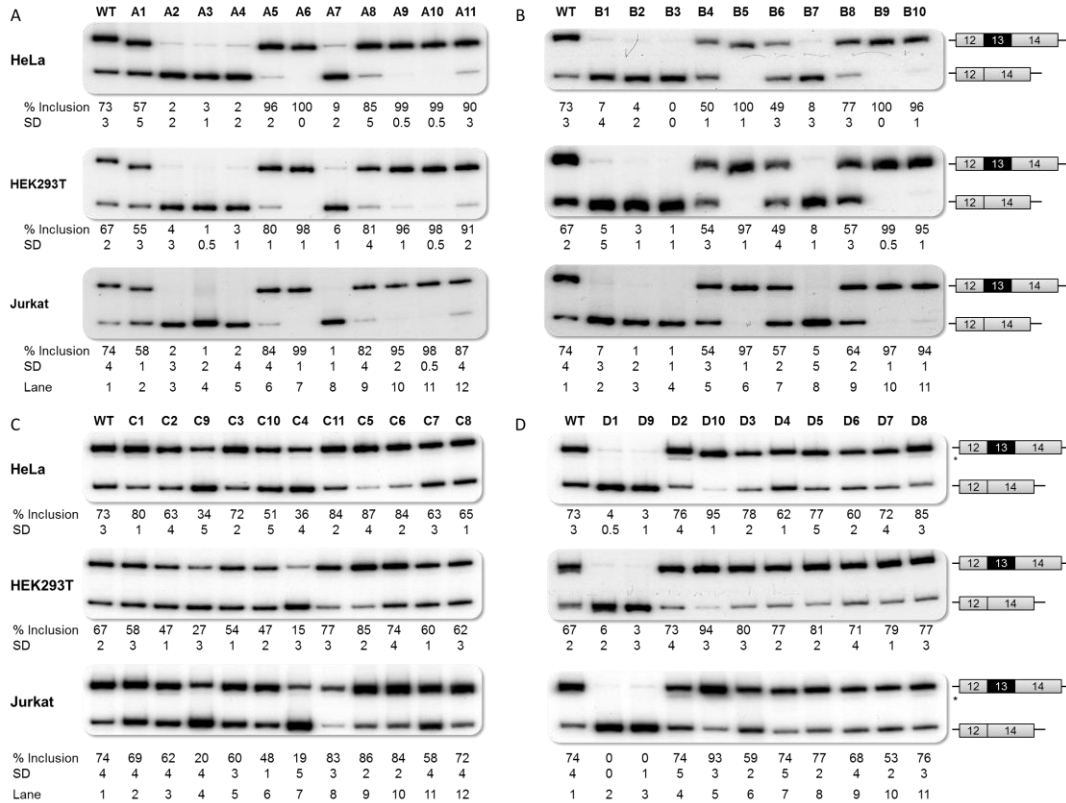


Figure 3.5 Identification of enhancers and silencers by serial deletions

(A) First set deletions of exon 13 sequence (A series, A1-A11) revealed two regions of enhancers and silencers, respectively. Deletions of A2, A3, A4 or A7 sequence (lanes 3-5,8) caused more skipping while removal of A5, A6, A8, A9, A10 or A11 sequence (lanes 6-7,9-12) resulted in more inclusion. (B) Results from the staggered deletions (B series, B1-B10) agreed with the observations from the first assay. Absence of B1, B2 and B3 led to exon skipping (lanes 2-4) just like their overlapping A1-A4, while removal of B5, B9 and B10 had a similar effect on exon inclusion (lanes 10-11) as their overlapping A5, A6, and A9, A10, A11. (C) Deletion analysis of intronic sequence upstream of exon 13 (C series, C1-C11) showed that proximal intron 12 does not contain any strong ISE or ISS. Most of the deletions had little impact on exon 13 inclusion except C4 (lane 7) as this deletion caused a rather strong repression on exon inclusion. (D) Deletion analysis of intronic sequence downstream of exon 13 (D series, D1-D10) identified a strong ISE located immediately downstream of the 5'ss flanking exon 13 (lanes 2-3). * is likely a heterohybridization duplex of exon 13 skipped and included PCR products.

Mean exon inclusion values are derived from experimental triplicate (different transfections) and the standard deviation (SD) is also indicated.

3.3.2 The intronic sequences near exon 13 have less regulatory effect

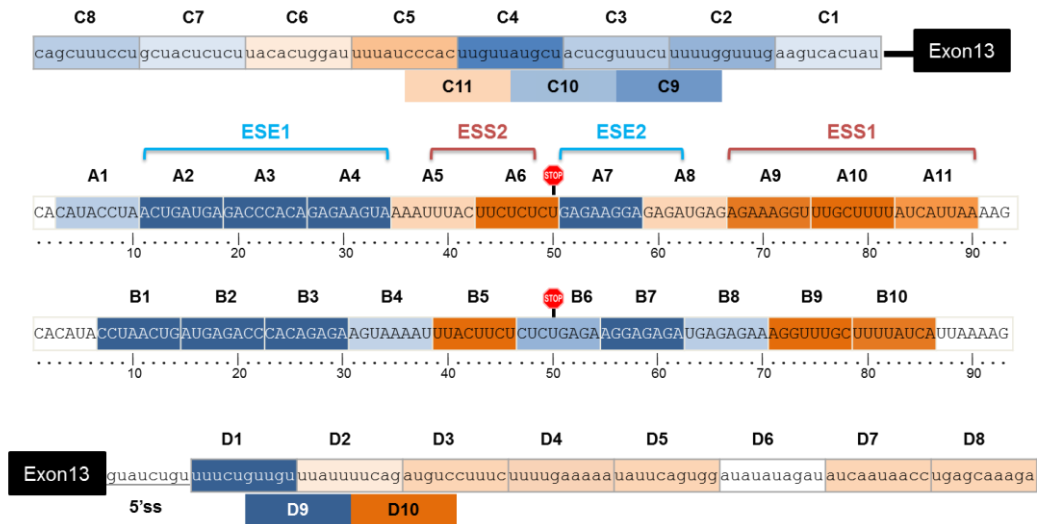
In addition to the exonic *cis*-acting elements, introns usually contain regulatory information to aid in proper splicing (Damgaard et al., 2002; Del Gatto-Konczak et al., 2000; Hovhannisyan and Carstens, 2005; Hua et al., 2008). Analogous to the previous experiment, eight 10-nucleotide deletions were introduced into introns 12 and 13 near exon 13 without disturbing the 3'ss, the BPS and the 5'ss, thus leaving intact the last 32 nt of intron 12 and the first 8 nt of intron 13 (Figure 3.4, C1-C8 and D1-D8).

The majority of deletions in both introns did not trigger a significant change in exon 13 inclusion (Figure 3.5C and D). Only removal of few regions like C2, C4, C5 and D1 changed exon 13 inclusion, and staggered deletions for these regions (Figure 3.4, C9-C11 and D9-D10) validated this result. Intron 12 has several weak and moderate ISEs (C2, C4 and C7), and one weak ISS at C5 and C6 region (Figure 3.5C). On the other hand, intron 13 has fewer *cis*-acting elements. Nevertheless, a strong ISE is present in intron 13 immediately downstream of the 5'ss, as its removal led to a strong reduction in exon inclusion confirmed by D1 and D9 (Figure 3.5D, lanes 2-3). The D10 potentially revealed a strong ISS but its staggered sequences D2 and D3 had similar exon 13 inclusion efficiency as the wild-type minigene, suggesting that a new silencer was likely created by deleting D10 sequence (Figure 3.5D, lanes 4-6). Few discrepancies were also observed in intron 13 deletion assays among the three cell lines. Two weak ISSs were detected in HeLa and Jurkat E6.1 but in different regions. However, in HEK293T, nearly the entire 80-nucleotide sequence downstream of the 5'ss flanking exon 13 contains weak ISSs (Figure 3.6). Altogether, these results suggest that intronic sequences near exon 13 may have a relatively minor effect on regulating inclusion of this exon compared to its exonic sequence.

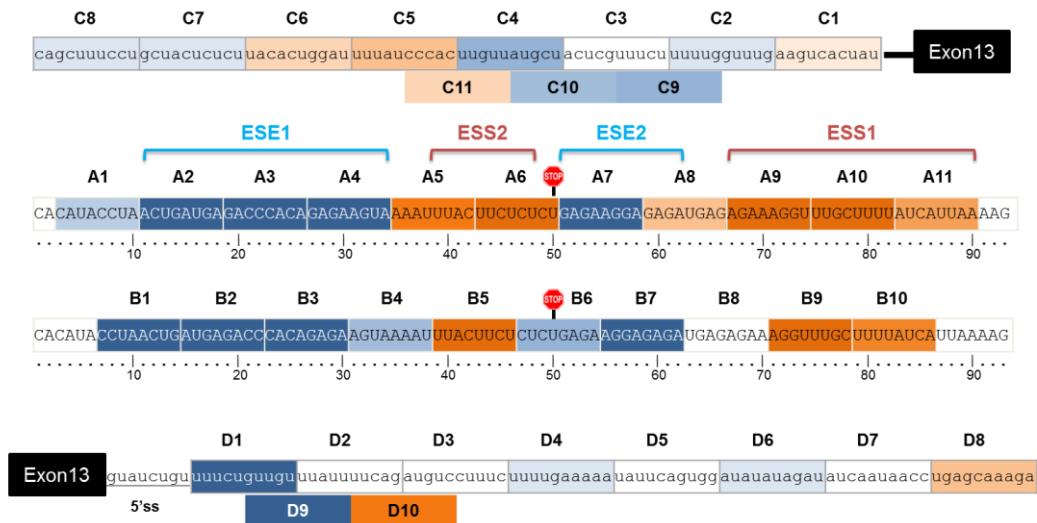
To summarize this series of deletion experiments, exon 13 consists of four regions of *cis*-acting elements with two strong ESEs and two strong ESSs which are denoted as ESE1, 2 and ESS1, 2, respectively. These ESEs and ESSs are alternatively arranged in the exon. A strong ISE is located downstream of the 5'ss

flanking exon 13 in addition to multiple weak ISEs and ISSs within the proximal intron 12 and 13 sequences (Figure 3.6).

A HEK293T



B HeLa



C Jurkat E6.1

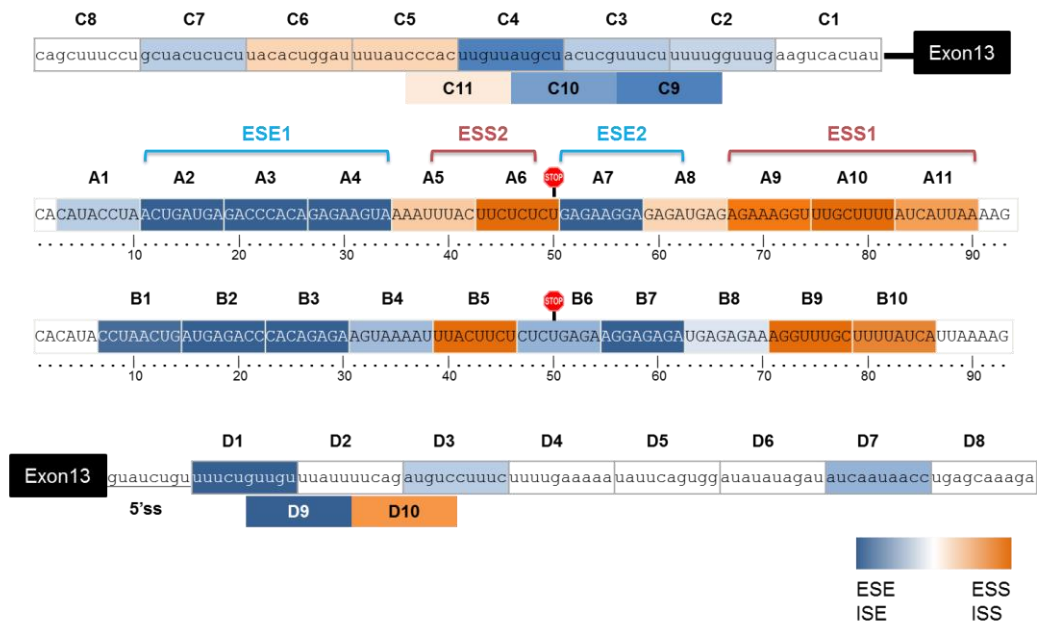


Figure 3.6 Map of enhancers and silencers of exon 13 by deletion assays

Strong ESEs and ESSs are present along exon 13 sequence. The long ESE and ESS are annotated as ESE1 and ESS1 while the short ESE and ESS region are annotated as ESE2 and ESS2. A strong ISE was identified in intron 13 while many weak ISSs and ISEs also exist in introns 12 and 13. The blue color highlighted sequence depicts enhancer while the orange color highlighted sequence depicts silencer. Color intensity correlates with the strength of *cis*-acting elements as illustrated in the color gradient rectangle. The strength of each *cis*-acting element is determined by the absolute difference in exon inclusion level as compared to the wild type. HeLa and Jurkat E6.1 have a similar map of *cis*-acting elements as HEK293T.

3.3.3 Validation of *cis*-acting elements by point mutations

Two ESEs and two ESSs have been identified through deletion assays using splicing minigenes. Subsequently, sets of nucleotide substitutions in ESE1 and ESS1 (Figure 3.7A) were generated to further confirm the regulatory role of these *cis*-acting elements and to be used as a negative control for RNA pull-down. The point mutations were designed such that they disrupt the predicted protein binding without creating a novel *cis*-acting element by using the prediction tool Human splicing finder (Desmet et al., 2009). Based on the prediction, I first mutated the potential protein binding sequences and then analyzed the mutant sequence as compared to the wild type using the prediction tool, to determine whether the mutations can disrupt the predicted binding sites without generating new *cis*-acting elements. The alternative splicing pattern of mutant exon 13 in the context of minigene was analyzed by RT-PCR. The point mutations in ESE1 abolished exon inclusion while the point mutations in ESS1 caused 100% exon inclusion, suggesting the complete disruption of this enhancer or silencer by nucleotide substitutions (Figure 3.7B, lanes 4-6 and Figure 3.7C, lanes 4-8). This result further validated the enhancer and silencer found in the deletion assays.

Since the point mutations had such strong effects, it was possible that the mutations created novel enhancers or silencers, instead of or in addition to removing the existing ones. Combination of A2 and B9 deletions restored exon inclusion back to the wild-type level (Figure 3.7B, lane 7), so ESS1 and ESE1 point mutations were combined with A2 or B9 deletion respectively to determine their competency as a negative control (Figure 3.7B, lanes 8-10 and Figure 3.7C, lanes 10-14). In the case of ESS1, mutations (mut) 2 and 3 were too strong to be reversed by removing the enhancer (Figure 3.7B, lanes 9-10). The enhancing effect brought by ESS1 mut 1, on the other hand, was neutralized by removing A2 sequence (Figure 3.7B, lane 8), so this mut 1 was chosen as a negative control for ESS1 RNA pull-down. In the case of ESE1, repression of exon inclusion caused by mut 3 was rescued by the B9 deletion (Figure 3.7C, lane 12), but such rescue did not occur for the other mutations, so ESE1 mut 3 was used as control. ESE1 mut 5 which had additional

two nucleotide substitutions as compared to mut 3 was also chosen as an additional negative control. Details about RNA pull-down experiment will be discussed in section 3.4.



Figure 3.7 Point mutations validated the presence of ESS1 and ESE1

(A) Several sets of point mutations in ESE1 and ESS1 were designed based on the prediction tools. The mutations are shown aligned by dotted vertical lines to the wild-type exon 13 sequence below. The blue or orange exon sequences are denoted as ESE1 and ESS1, respectively. (B) Point mutation analysis of ESS1. Mut 1-3 caused complete exon inclusion (lanes 4-6). This result was similar to the deletion analysis and thus further confirms the presence of silencer at the 3' end of exon 13. The A2 enhancer sequence was deleted in the ESS1 mutant minigenes to determine if the enhancing effect caused by those mutations was due to removal of silencer but not creation of novel *cis*-acting elements (lanes 8-10). The effect caused by

ESS1 mut 1 can be neutralized by removal of A2 (lane 8), and hence mut 1 was chosen as a negative control for pull-down assay. (C) Point mutation analysis of ESE1. All point mutations except mut 3 caused complete exon skipping (lanes 4-8), confirming the presence of this enhancer. However, only the repressive effect of mut 3 was reversed by B9 (silencer) deletion (lane 12). Mut 5 was similar to mut 3 with two additional nucleotide substitutions, so mut 3 and 5 were both chosen as negative controls. Both experiments were done at least in triplicate with HEK293T cells.

3.3.4 Validation of ESEs and ESSs in heterologous context

The previous deletion and point mutation assays have identified and confirmed the ESEs and ESSs in exon 13. To examine if these ESEs and ESSs are functional in a heterologous context (in a different exon), these sequences were cloned into the pSXN plasmid (Figure 3.8A and B). This plasmid was previously used to identify enhancers as it contains a very weak and small alternative exon whose inclusion can only be achieved by insertion of an enhancer using the two restriction sites within this weak exon (Coulter et al., 1997). The alternative splicing pattern of this heterologous minigene was examined by RT-PCR with beta globin primers specific for this vector.

ESE1 was shown to be a very strong enhancer as insertion of this sequence into the weak exon caused almost complete exon inclusion which was abolished by mut 5 as shown in the previous mutational analysis (Figure 3.8C, lanes 7-8). However, ESE2 appeared to include a weak enhancer as only a very faint inclusion band was detected after inserting this sequence into the alternative exon (Figure 3.8C, lanes 9-10). The inclusion of this ESE2-containing alternative exon was improved by changing the 5'ss sequence for better base-pairing to U1 snRNA (Figure 3.8C, lane 11). Similarly, the mut 1 in ESE2 in the context of such improved 5'ss also resulted in strong exon skipping (Figure 3.8C, lane 12).

Although the pSXN plasmid was originally applied to identify enhancer sequences, the *CD46* exon 13 ESS sequences were also tested in this context. The insertion of silencer sequences did not promote exon inclusion (Figure 3.8C, lanes 3,5), suggesting that the inclusion of ESE1/2 containing exons was not due to the increased exon size but rather to the presence of promoting elements, and then the ESS sequences probably possess silencing ability. The ESS2 mutation had similar band pattern as its wild type (Figure 3.8C, lane 6), suggesting that the mutation did not create novel *cis*-acting elements. However, a 15% inclusion level of mutated ESS1-containing exon was observed (Figure 3.8C, lane 4), and this implies the creation of an enhancer by the mut1. To summarize, the enhancers and silencers

found in the deletion assays were further validated by this heterologous minigene experiment, proving the regulatory roles of ESEs and ESSs in exon 13.

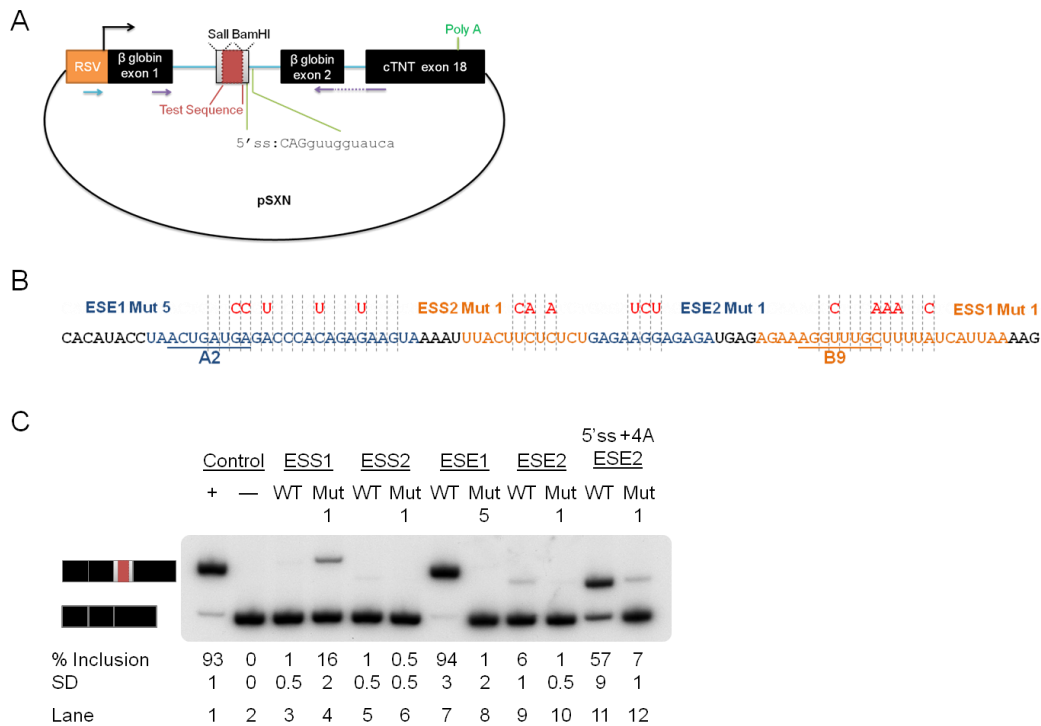


Figure 3.8 Testing ESSs and ESEs in heterologous context

(A) Schematic diagram of the pSXN plasmid. Test sequences were inserted into the gray exon using Sall and BamHI sites. Purple arrows depict RT-PCR primers for detecting minigene transcripts. (B) The blue or orange exonic sequences are denoted as ESE1/2 and ESS1/2, respectively. These wild-type or mutant sequences are inserted into pSXN plasmid to examine their regulatory roles in exon inclusion. (C) The pSXN 13 containing enhancer sequence and the pSXN-linker containing short neutral sequence were the positive and negative control. Exon inclusion suggests the enhancing effect of the inserted sequences (lanes 1,7,11), while virtually absence of exon inclusion implies the silencing or neutral effect of the inserted sequence (lanes 2-3,5-6,8,10). Experiment was done at least in triplicate with HEK293T.

3.4 Identification of *trans*-acting factors as regulators of *CD46* alternative splicing

Cis-acting elements provide binding sites for *trans*-acting factors to exert their regulatory functions on splicing by either promoting or repressing exon recognition by the spliceosome. Since multiple enhancers and silencers have been verified in the previous deletion, mutation and heterologous context assays, I sought to identify *trans*-acting factors that regulate exon 13 inclusion via these *cis*-acting elements. Both candidate and systematic approaches were carried out to search for the potential regulators. For the candidate approach, exon 13 sequence was analyzed by splicing prediction tools like Human Splicing Finder, SFmap and Splicing Rainbow (Akerman et al., 2009; Desmet et al., 2009; Paz et al., 2010). These prediction tools gave a list of potential splicing proteins. I picked up SRSF1 and PTBP1 from the list and examine their roles in alternative splicing of exon 13 by loss- and gain-of-function assays (see next section).

Concurrently, RNA pull-downs were performed to identify *trans*-acting factors bound to the ESS1 and ESE1. Synthetic RNA oligonucleotides consisting of the wild-type or mutant ESS1 or ESE1 sequence and additional 5 flanking nucleotides at each end were purchased from IDT (Figure 3.8A, blue or orange colored sequence). These RNA oligonucleotides were attached to streptavidin magnetic beads and then incubated with HeLa nuclear extract. Bound proteins eluted from RNA were separated by SDS-PAGE and detected by silver staining or western blot. As these RNA oligonucleotides were incubated with high concentration of proteins, it is likely that some of the proteins bound to the RNA in a non-sequence specific manner. In order to isolate the sequence-specific binders, mutant RNA oligonucleotides were also used for RNA pull-down. In silver stained gel, those protein bands that have reduced intensity or are absent in the mutant as compared to the wild-type sample were further analyzed by mass spectrometry.

Numerous proteins with varied molecular weights were pulled down together with wild-type ESS1 RNA. Similar band pattern was present in the mutant ESS1 lane (Mut1; Figure 3.9A) in which some of the protein bands had the same band intensity as in the wild type while some had reduced intensity. The protein band #1

which has reduced intensity in mutant sample was analyzed by mass spectrometry. The mass spectrometry result suggests that this band corresponds to hnRNP M, a splicing repressor, and this was confirmed by western blot (Table 3.2, #1 and Figure 3.10A). However, functional studies of hnRNP M imply that this protein does not regulate exon 13 inclusion (data not shown). In addition, I have assessed this ESS1 pull-down by western blot with available splicing factor antibodies like hnRNP A1, hnRNP H, KSRP, PTBP1 and SRSF1 (data not shown). Most of the proteins were either not detected in the eluate or showed similar intensity in both wild-type and mutant eluates. Only PTBP1 and SRSF1 showed reduced binding in the mutant sample (Figure 3.10A).

In ESE1 RNA pull-down, protein binding in mut 3 was similar to wild type and one protein band at around 65kD was more intense than the one in wild type (Figure 3.9B). Additionally, mut 5 had more protein bands than wild type. These data suggest that these two mutants created novel silencers instead of removing the enhancers. An alternative control for RNA pull-down was applied, and this was achieved by antisense oligonucleotide (ASO)-mediated blocking of protein binding. The ASO targeting ESE1 region was annealed to the wild-type RNA oligonucleotides before incubation with proteins. Several protein bands disappeared in the RNA:ASO lane (Figure 3.9B, # labeled), suggesting the effectiveness of ASO-mediated blocking of protein binding. These proteins were identified by mass spectrometry, showing that most of them are hnRNPs and spliceosome components which could not be properly tested because of time constraints (Table 3.2). Similarly, western blot was performed to detect if certain candidate proteins bind to the ESE1 (Figure 3.10B).

Although binding of SRSF1 and PTBP1 was confirmed by western blot, we did not detect these proteins by mass spectrometry. PTBP1 has a molecular weight at around 57 kD and the protein band at this molecular weight did not show strong reduction in intensity in mutant ESS1. Hence, this band is not selected for mass spectrometry. Furthermore, there was only weak reduction in PTBP1 binding as shown by the western blot. On the other hand, SRSF1 was not detected by mass

spectrometry and this could be due to the lower abundance of SRSF1 relative to other proteins in the silver stained gel.

In summary, several *trans*-acting factors bound to the ESS1 and ESE1 were identified by RNA pull-down. Binding of SRSF1, PTBP1 and hnRNP M to ESS1 sequence was detected by western blot and the binding efficiency was diminished by the ESS1 mut1 (Figure 3.10A). SRSF1 also interacted with ESE1 and the binding was reduced by ASO (Figure 3.10B). The details about ASOs are discussed in section 3.5. In addition, hnRNP A1 also appeared to bind with more affinity to wild-type ESE1 than to ESE1 plus ASO. The binding of SRSF1 to ESE1 and ESS1 as well as the binding of PTBP1 to ESS1 agree with the predictions by the previously mentioned prediction tools. In order to examine the roles of SRSF1 and PTBP1 in regulating exon 13, knockdown and overexpression experiments were subsequently carried out.

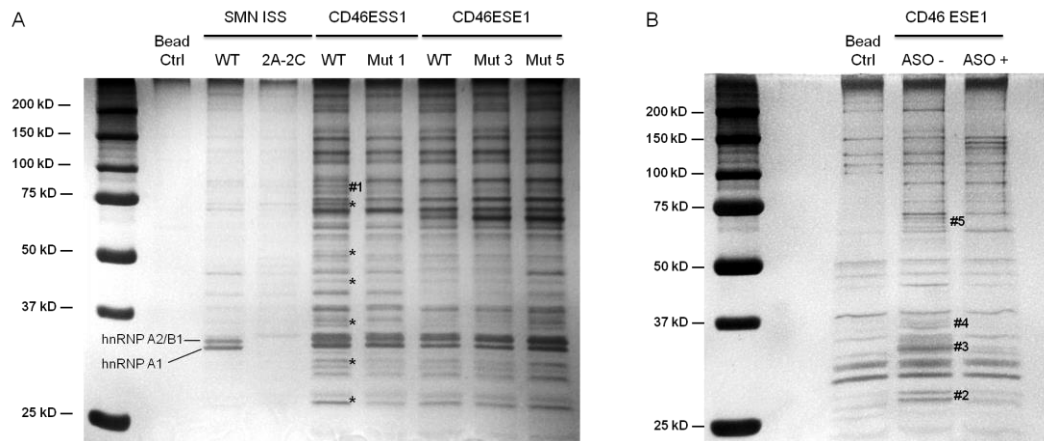


Figure 3.9 Total proteins bound to ESS1 and ESE1 RNA by pull-down assay

(A) SMN ISS wild-type (intron 7 positions +10 to +25) and 2A-2C mutant RNA were used as positive controls, as the 2A-2C mutant in which +12A and +23A are changed to C is known to disrupt the binding of hnRNP A1/A2 (Hua et al., 2008). Some protein bands (*/#) had reduced binding in the mutant ESS1 RNA compared to wild-type ESS1, suggesting that these proteins specifically interact with ESS1 sequence and they are likely the regulators. In contrast, the ESE1 mutant had even more bound proteins as compared to the wild type. (B) Binding of proteins was reduced on ASO-annealed ESE1 RNA sequence, as shown by the number sign. The identities of protein bands (#2-#5) with reduced intensity in control were determined by mass spectrometry and they are hnRNPs and spliceosome components as shown in Table 3.2. RNA pull-downs for both ESS1 and ESE1 were repeated once.

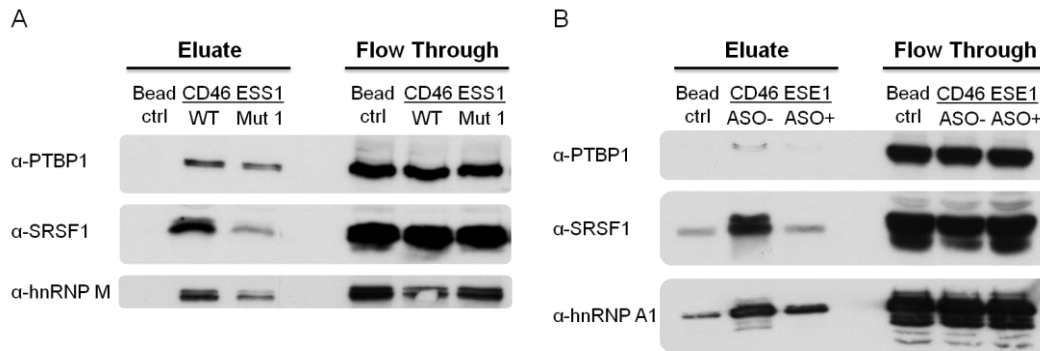


Figure 3.10 Western blot confirmed protein binding to ESS1 or ESE1

(A) SRSF1 specifically bound to the ESS1 and this interaction was reduced in mutant ESS1. PTBP1 had more similar binding affinity to wild-type and mutant ESS1, although the band in ESS1 WT was slightly stronger than that on the ESS1 mut1. Reduced hnRNP M binding on mutated ESS1 was also detected. (B) SRSF1 and hnRNP A1 bound to the ESE1 WT, and ASO annealing blocked the access of these proteins to ESE1. PTBP1 was barely detected in wild-type ESE1, suggesting that ESE1 is not a PTBP1 binding site. This observation also provides specificity to the preferential detection of the other proteins in the pull-downs with wild-type sequence.

Table 3.2 Mass spectrometry results for protein bands

	Mass	Score	emPAI
Band #1			
Heterogeneous nuclear ribonucleoprotein M (<i>HNRNPM</i>)	77749	8140	5.43
Band #2			
Heterogeneous nuclear ribonucleoprotein A1 (<i>HNRNPA1</i>)	33249	80	0.10
U2 small nuclear ribonucleoprotein B" (<i>SNRPB2</i>)	25470	37	0.28
Small nuclear ribonucleoprotein-associated proteins B and B' (<i>SNRPB</i>)	24765	33	0.14
Band #3			
Heterogeneous nuclear ribonucleoproteins A2/B1 (<i>HNRNPA2B1</i>)	37464	843	4.91
Heterogeneous nuclear ribonucleoprotein A1 (<i>HNRNPA1</i>)	38837	625	2.70
DnaJ homolog subfamily C member 8 (<i>DNAJC8</i>)	29823	87	0.37
ELAV-like protein (<i>ELAVL1</i>)	39200	57	0.08
Serine/arginine-rich splicing factor 5 (<i>SRSF5</i>)	31359	51	0.11
Serine/arginine-rich splicing factor 6 (<i>SRSF6</i>)	39677	51	0.08
Band #4			
Heterogeneous nuclear ribonucleoprotein A1 (<i>HNRNPA1</i>)	38837	305	1.09
Heterogeneous nuclear ribonucleoproteins A2/B1 (<i>HNRNPA2B1</i>)	37464	233	1.14
Heterogeneous nuclear ribonucleoprotein H3 (<i>HNRNPH3</i>)	24925	100	0.13
Heterogeneous nuclear ribonucleoprotein A3 (<i>HNRNPA3</i>)	39799	69	0.38
Band #5			
Heterogeneous nuclear ribonucleoprotein K (<i>HNRNPK</i>)	51230	294	0.98
Heterogeneous nuclear ribonucleoprotein L (<i>HNRNPL</i>)	64720	225	0.22
Heterogeneous nuclear ribonucleoprotein R (<i>HNRNPR</i>)	60144	77	0.05
Heterogeneous nuclear ribonucleoprotein Q, Isoform 4 (<i>SYNCRIP</i>)	58927	77	0.06
Cleavage stimulation factor subunit 2 (<i>CSTF2</i>)	63017	65	0.11
Probable ATP-dependent RNA helicase DDX5 (<i>DDX5</i>)	69557	61	0.26

Footnote: Selected protein IDs that are human and related to splicing and with higher score among the result list are presented in this table. Gene name is shown in parenthesis. The exponentially modified protein abundance index (emPAI) shows an approximate and relative quantitation of the proteins in a mixture (MatrixScience). The score is determined by the match of query peptides to references.

3.4.1 SRSF1 represses exon 13 inclusion through ESS1

As stated earlier, analysis of exon 13 sequence using *in silico* prediction tools like Human Splicing Finder, SFmap and Splicing Rainbow (Akerman et al., 2009; Desmet et al., 2009; Paz et al., 2010) predicted the regulation of exon 13 inclusion by SRSF1 by binding to the ESE1 region. Interaction between SRSF1 and ESS1 or ESE1 was also observed in the previous pull-down assays. Therefore, loss- and gain-of-function assays were performed to test the predicted role of SRSF1 as an activator of exon 13 inclusion. Control or SRSF1 siRNA was introduced into cells to knockdown expression of SRSF1 while empty vector or T7 tagged cDNA plasmid was used for overexpression assay. Endogenous *CD46* transcripts were then examined by RT-PCR using exon 12- and 14-specific primers.

Unexpectedly, exon 13 was included at higher levels in SRSF1-depleted cells (Figure 3.11, lane 1 vs.2, 30% increase), which is contradictory to the prediction. This observation suggests that SRSF1 instead of being an activator is repressing exon 13 inclusion. The enhancement of exon 13 inclusion by depletion of SRSF1 was partially reversed by providing exogenous SRSF1 protein to cells by introducing plasmids expressing siRNA resistant SRSF1 cDNAs (Figure 3.11, lanes 3-4). However, exon 13 inclusion did not change in SRSF1-overexpressing cells (Figure 3.11, lanes 5-6). This experiment suggests that SRSF1 represses exon 13 inclusion.

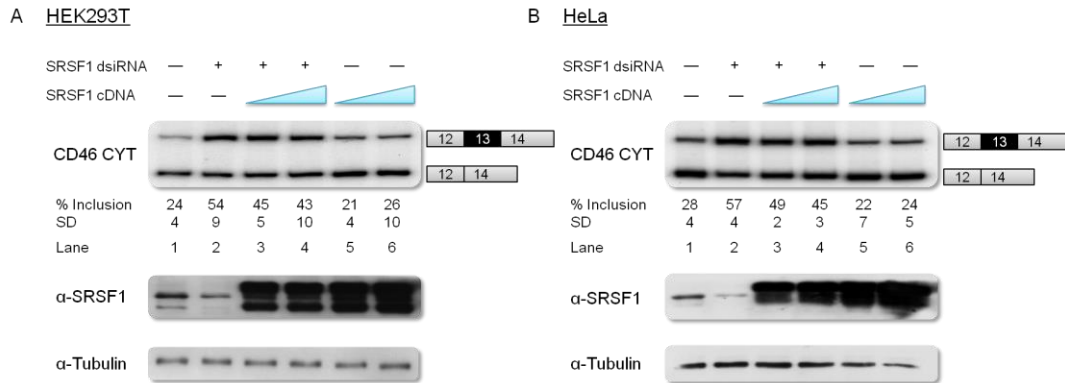


Figure 3.11 SRSF1, a common activator, represses *CD46* exon 13 inclusion

Exon 13 inclusion increased when SRSF1 was depleted (lane 2) and this enhancement could be partially reversed by replenishing SRSF1 using cDNA plasmid (lanes 3-4). The SRSF1 mediated repression was consistent between (A) HEK293T and (B) HeLa. Empty vector was used as a control for overexpression cDNA plasmid. Expression level of SRSF1 was examined by western blot with whole-cell extract (bottom panels). Each transfection was done at least in triplicate.

It is surprising to find out that SRSF1 acts as a repressor for exon 13 inclusion since SRSF1 is a well-known splicing activator. A previous study has reported that binding of SRSF1 on *RON* exon 12 favors the inclusion of exon 12 and so indirectly inhibits the recognition of cassette exon 11 (Ghigna et al., 2005). However, there is so far no study showing direct repression of SRSF1 on its binding exon. To test that SRSF1-mediated exon 13 repression occurs via this exon and its proximal flanking introns, this region was cloned into a heterologous minigene containing *MCAD* constitutive exons 8 and 10 (Figure 3.12A). This hybrid minigene was introduced into HEK293T cells together with either SRSF1 siRNA or SRSF1 cDNA plasmid. Hybrid minigene transcripts were analyzed by RT-PCR with vector specific primers after 48 h incubation. SRSF1 knockdown had a small effect on exon 13 inclusion in the context of this hybrid minigene but exon 13 was largely skipped when SRSF1 was overexpressed (Figure 3.12B, lanes 1-4). In this heterologous context, the depletion or enrichment of SRSF1 affected the inclusion of exon 13 in the same direction as that of endogenous *CD46* and that of *CD46*-exon 12-13-14 minigene (Figure 3.11, Figure 3.12B and Figure 3.13). This result strongly suggests that SRSF1 regulates this alternative splicing event directly through exon 13 and/or its nearby flanking introns.

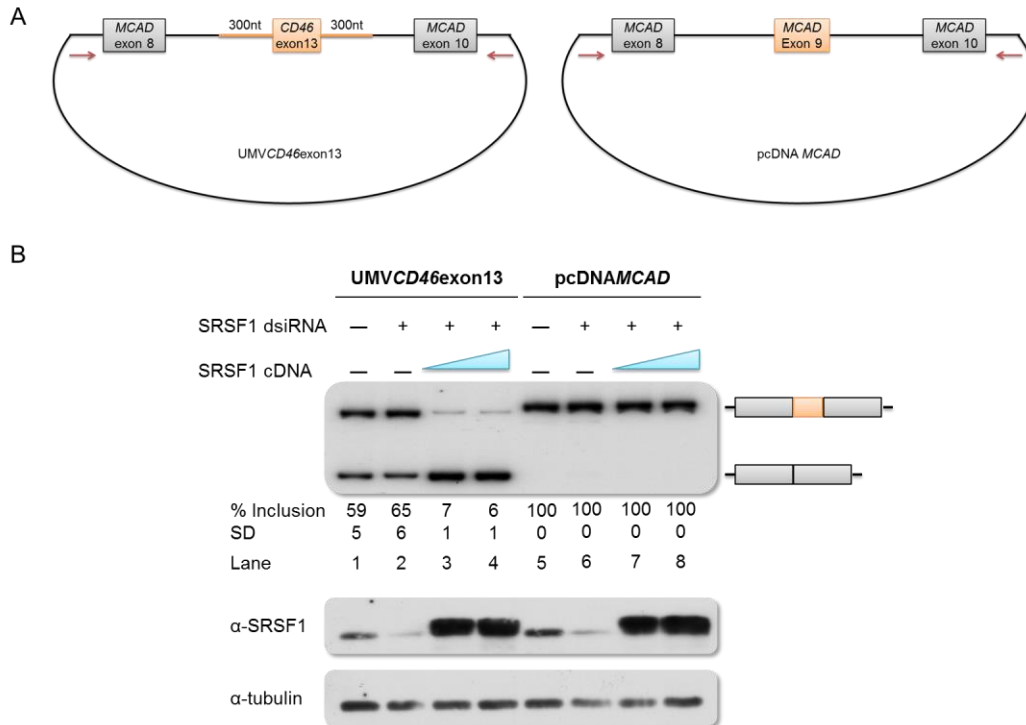


Figure 3.12 SRSF1 likely represses inclusion via exon 13 and/or its nearby flanking introns

(A) A hybrid minigene containing constitutive *MCAD* exons and *CD46* exon 13 was generated to test if SRSF1 regulates splicing via exon 13 and/or its flanking intronic sequences (left). The pcDNAMCAD minigene containing *MCAD* exons 8, 9 and 10 was used as a control (right). (B) As an unspecific control, the original pcDNAMCAD minigene did not respond to either knockdown or overexpression of SRSF1 (lanes 5-8) while the hybrid minigene pre-mRNA was susceptible to overexpression of SRSF1 (lanes 3-4), implying that the site that mediates SRSF1 regulation is located within exon 13 and/or its nearby flanking intronic sequences. Control siRNA and empty vector are used as control for knockdown and overexpression assays, respectively. Expression level of SRSF1 was examined by western blot with whole-cell extract (bottom panels). Experiment was done at least in triplicate with HEK293T cells.

As shown previously, less SRSF1 binding to the mutant ESS1 in comparison to wild-type ESS1 was detected in the RNA pull-down (Figure 3.10A), suggesting that SRSF1 indeed interacts through the ESS1 region to repress exon 13 inclusion. The effect of SRSF1 on exon 13 was examined using the *CD46* minigene together with SRSF1 overexpression (Figure 3.13, lanes 10-18). Co-transfections of wild-type or mutant *CD46*-exon-12-13-14 minigene and SRSF1 cDNA plasmid or empty vector were performed. Subsequently, the alternative splicing pattern of minigene transcripts was studied. Exon 13 inclusion was mostly repressed by exogenous SRSF1 in the context of wild-type minigene (Figure 3.13, lane 1 vs. 10). On the other hand, the mutant exon 13 with deleted A6 or B9 sequence but not others like A11 was strongly resistant to SRSF1 overexpression (Figure 3.13, lane 2 vs. 11, lane 7 vs. 16, lane 5 vs. 14), suggesting that these two regions could contain the actual SRSF1 binding sites. Other deletions that severely reduced the effects of SRSF1 overexpression were A9, A10 and B5 (Figure 3.13, lane 3 vs. 12, lane 4 vs. 13, lane 6 vs. 15), albeit to a lesser extent than the previous two. Altogether, this experiment suggests that the A6 and B9 and perhaps some flanking nucleotides contain the actual SRSF1 binding sites within *CD46* exon 13, resulting in skipping of this exon.

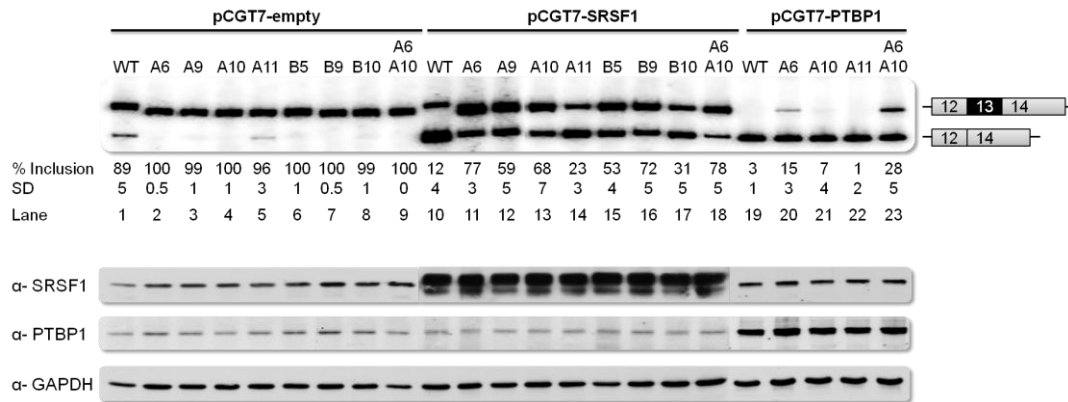


Figure 3.13 SRSF1 and PTBP1 mediate exon 13 skipping through A6 and B9 sequences

When wild-type *CD46* exon 12-13-14 minigene was co-expressed with exogenous SRSF1 or PTBP1 proteins, the inclusion of exon 13 was largely repressed (lane 1 vs. 10/19). Exon 13 with deleted A6, A9, A10, B5 or B9 sequence became largely resistant to SRSF1 overexpression (lanes 11-13,15-16). On the other hand, only exon 13 without A6 sequence was weakly resistant to the repression caused by PTBP1 (lane 20). pCGT7-empty plasmid is a control plasmid for overexpression assay. Expression levels of SRSF1 and PTBP1 were examined by western blot with whole-cell extract (bottom panels). Experiment was done at least in triplicate with HEK293T.

3.4.2 PTBP1 is a repressor of exon 13 inclusion

Next, I focused on the role of PTBP1 in exon 13 inclusion as PTBP1 was predicted to be a repressor. There are two PTBP1 consensus binding sequences within ESS1 and ESS2 and its binding to ESS1 was confirmed by RNA pull-down (Figure 3.10A). Likewise, the role of PTBP1 in regulating the endogenous alternative splicing pattern of exon 13 was investigated by knockdown and overexpression assays. A 15% increase in exon 13 inclusion was observed in PTBP1-knockdown HEK293T cells, yet the repressive effect of PTBP1 was weaker than that of SRSF1 (Figure 3.14A, lane 2). The PTBP1-mediated repression was reversed by adding back exogenous PTBP1 (Figure 3.14A, lanes 3-4). Besides, overexpression of PTBP1 did not have an effect on exon 13 inclusion (Figure 3.14A, lanes 5-6). These data suggest that PTBP1 may not be the major repressor of endogenous exon 13 inclusion.

Subsequently, co-expression of *CD46* wild-type or mutant minigenes and PTBP1 cDNA plasmid in cells was done to examine the PTBP1 regulatory site within exon 13. The inclusion of exon 13 in the context of *CD46* minigene was highly susceptible to PTBP1 overexpression as there was nearly no exon 13 inclusion detected in PTBP1-enriched cells (Figure 3.13, lane 19). Removal of either PTBP1 consensus sequences, A6 and A10, barely rescued exon 13 inclusion from PTBP1 overexpression although exon 13 with A6 deletion had a small improvement in inclusion level (12% increase). In addition, deletion of both putative PTBP1 binding sites had synergistic enhancing effects on exon inclusion upon PTBP1 overexpression (Figure 3.13, lanes 20-23). Besides, PTBP1 may inhibit exon inclusion through introns since removal of the exonic putative binding motifs did not completely rescue the phenotype. Overall, these data strongly suggest that PTBP1 acts as a splicing repressor for exon 13 inclusion mediated through A6 and probably other intronic regions.

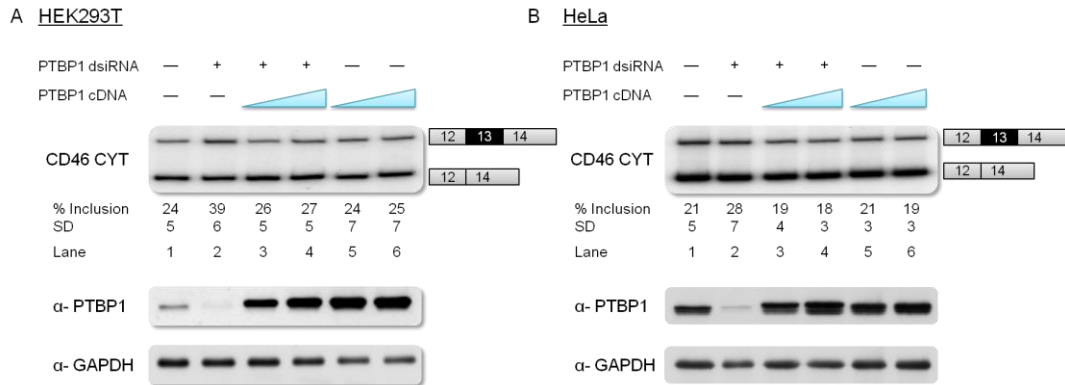


Figure 3.14 PTBP1 is a weak repressor of endogenous *CD46* exon 13 inclusion

(A) A small increase in exon inclusion level was observed in PTBP1-depleted HEK293T cells (lane 1 vs. 2). Inclusion of exon 13 was rescued by replenishing PTBP1 (lanes 3-4). Control siRNA and empty vector was used as a control for knockdown and overexpression assays, respectively. (B) PTBP1 had a negligible effect on exon 13 inclusion in HeLa as neither knockdown nor overexpression of PTBP1 significantly changed the alternative splicing pattern of exon 13. Expression level of PTBP1 was examined by western blot with whole-cell extract (bottom panels). Experiment was done at least in triplicate.

3.4.3 TIA1 and TIAL1 promote exon 13 recognition

After identifying two repressors that inhibit exon 13 inclusion through the ESSs, I searched for activators of inclusion of this exon. TIA1 and TIAL1 have been shown to regulate alternative splicing patterns of exons flanked with weak 5'ss by binding to the U-rich sequences downstream of the 5'ss and recruiting U1 snRNP via direct interaction with the U1 snRNP specific polypeptide U1-C (Forch et al., 2002; Izquierdo et al., 2005; Le Guiner et al., 2001; Singh et al., 2011). *CD46* exon 13 is flanked by a relatively weak 5'ss (Table 3.1) followed by a stretch of U-rich sequence, and removal of this U-rich sequence strongly suppressed exon inclusion as shown in the deletion assays (Figure 3.5D), making it a potential binding site for TIA1/TIAL1.

Depletion or overexpression of both protein regulators had an impact on the inclusion of exon 13 in endogenous *CD46* mRNA. Overexpression of TIA1 and TIAL1 both improved exon 13 inclusion by 25% or more increase but TIAL1 seemed to have a stronger enhancing effect (Figure 3.15, lane 1 vs. 2/3). Consistently, removal of TIAL1 caused a 12% decrease in exon 13 inclusion, confirming the role of TIAL1 in regulating exon 13 (lane 1 vs. 7). In addition, the repression caused by TIAL1 knockdown was partially reversed by TIA1 overexpression and TIAL1 reconstitution (lanes 8-9). In contrast, exon 13 inclusion level in TIA1-depleted cells was similar to the control (lane 1 vs. 4). It could be that the presence of TIAL1 is enough to support the exon 13 inclusion. Double knockdown of TIA1 and TIAL1 had more impact on the alternative splicing pattern than single knockdown of either one, and exon inclusion could be just weakly rescued by individual reconstitution of TIA1 or TIAL1 (lanes 10-12), indicating that TIA1 and TIAL1 both are involved in regulating exon 13 with TIAL1 as the major activator.

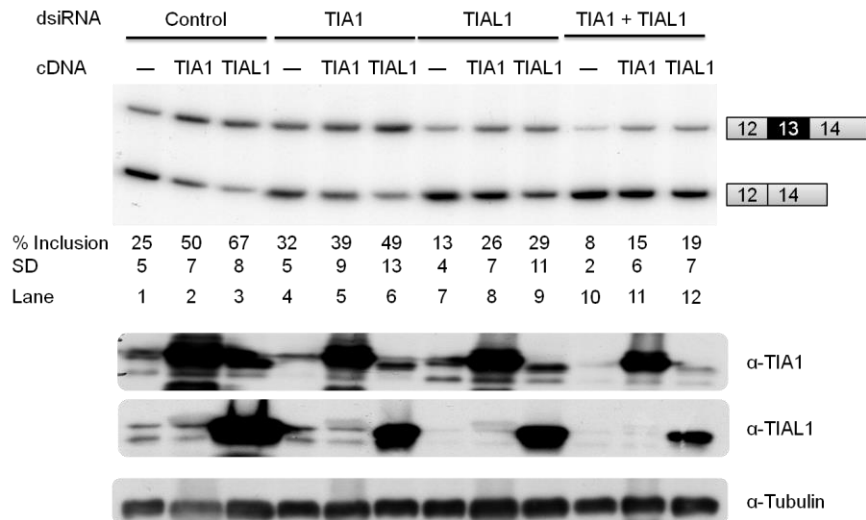


Figure 3.15 TIAL1 and TIA1 enhance exon 13 inclusion

Overexpression of TIA1 or TIAL1 strongly improved exon inclusion in endogenous transcripts (lanes 2-3). However, only depletion of TIAL1 caused a large reduction in exon inclusion (lane 7) which could be further enhanced by simultaneous removal of TIA1 (lane 10). Control siRNA and empty vector was used as a control for knockdown and overexpression assays, respectively. Expression levels of TIA1 and TIAL1 were examined by western blot with whole-cell extract (bottom panels). Experiment was done at least in triplicate with HEK293T.

To examine if TIA1 and TIAL1 indeed activate exon 13 inclusion via the downstream U-rich region, mutant minigenes with deletion of D1 or poly-U substitution downstream of the 5'ss were co-expressed in HEK293T together with exogenous TIA1 or TIAL1 protein. Inclusion of exon 13 in minigene transcripts was then assessed by RT-PCR. Exon 13 inclusion decreased in the wild-type minigene in presence of additional TIA1 or TIAL1 (Figure 3.16, lane 1 vs. 6/11). This observation is contradictory to the previous data with endogenous *CD46* mRNA (Figure 3.15) and will be further commented in the Discussion. On the other hand, removal of D1 region did not completely abolish the enhancing effects of TIA1 and TIAL1 as exon 13 inclusion was still largely promoted by overexpression of these factors (Figure 3.16, lane 4 vs. 9/14, 20%-30% increase). This suggests that the U-rich sequence further downstream in intron 13 can also mediate the TIA1 and TIAL1 activation of exon 13 inclusion. Although these might only constitute weak binding sites in normal condition, the activation by these U-rich sequences can be stronger with excess of TIA1 and TIAL1. Removal of all adjacent poly-U regions made exon 13 completely unresponsive to TIA1 or TIAL1 overexpression (Figure 3.16, lane 5 vs. 10/15), suggesting that these poly-U sequences downstream of the 5'ss can be bound by TIA1 or TIAL1 to promote exon recognition. All in all, TIA1 and TIAL1 strongly enhance exon 13 inclusion via its downstream poly-U rich region and that TIAL1 appears to be the most competent of the two in promoting exon 13 inclusion.

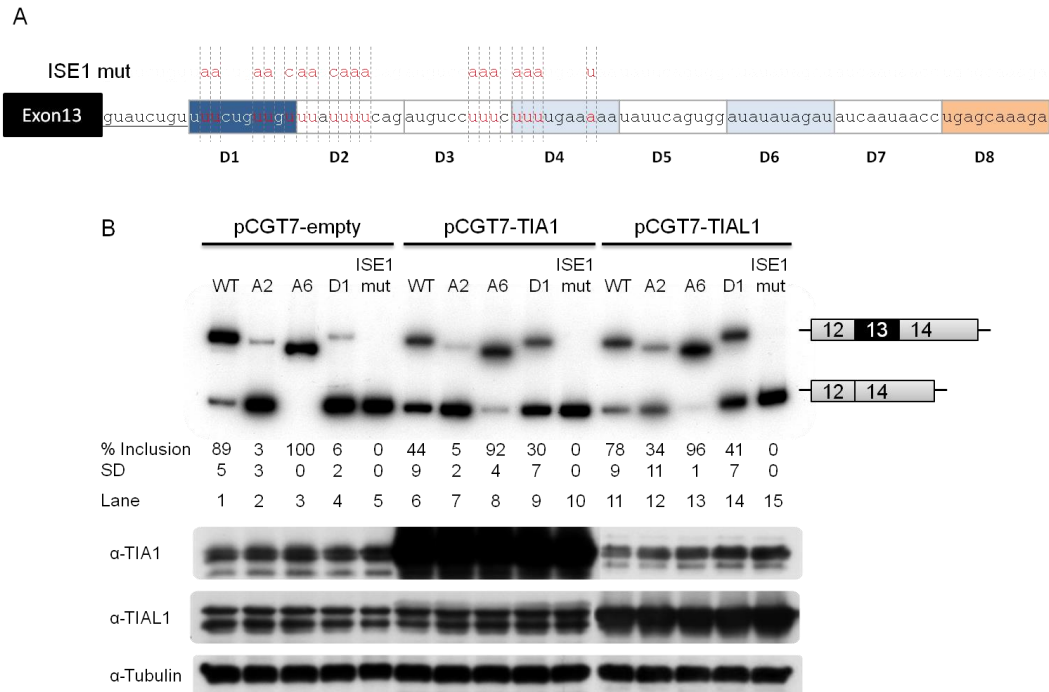


Figure 3.16 TIA1 and TIAL1 upregulate exon 13 inclusion via downstream poly-U regions

(A) Design of ISE1 point mutations. Uridines were mostly replaced with adenosines and few other nucleotides were changed to cytosine or uridine to avoid creation of ISEs according to the prediction by Human Splicing Finder. (B) TIA1 and TIAL1 are able to improve inclusion of exon 13 when the D1 region is removed, suggesting that the extended poly-U regions further downstream are bound by these two factors. The enhancing effects of TIA1 and TIAL1 on exon 13 inclusion were abolished by removal of poly-U regions (lanes 10,15). A2 mutant minigene is a negative control for this experiment. Empty vector is a control for overexpression. Expression levels of TIA1 and TIAL1 were examined by western blot with whole-cell extract (bottom panels). Experiment was done at least in triplicate with HEK293T. Migration of exon 13 in A2 mutant minigene and A6 was faster due to the 8 nt-deletion of exonic sequence.

3.5 Manipulation of the alternative splicing pattern of *CD46* exon 13 by ASOs

With the understanding of basic splicing regulation of *CD46* exon 13, supported by identification of *cis*-acting elements and *trans*-acting factors, I attempted to manipulate the alternative splicing pattern of exon 13 to alter its inclusion level. As defective expression of *CYT1* is associated with autoimmune diseases, manipulation of exon 13 inclusion could have therapeutic applications. One way to switch the alternative splicing patterns to increase either exon inclusion or skipping is to use ASOs which can anneal to pre-mRNAs and block the recruitment of either the spliceosome, or splicing activators or repressors.

Our ASOs have a length of 18 nt while the ESE1 and ESS1 are more than 20 nt long. In order to cover the whole ESE1 and ESS1, two ASOs with 2' O-methoxyethyl and phosphorothioate backbones were designed for each ESE1 and ESS1 to modulate exon 13 inclusion, and the annealing positions of the two ASOs differed by 4 or 5 nt (Figure 3.17A). Cells were transfected with respective control or ESE1/ESS1 targeted ASOs, and subsequently endogenous *CD46* transcripts were examined. Both ESE1 ASOs efficiently blocked exon 13 recognition, causing nearly complete exon skipping (Figure 3.17B, lanes 2-3). On the other hand, only one of the ESS1 ASOs weakly improved exon inclusion while the other one further repressed exon inclusion (Figure 3.17B, lanes 4-5)(see Discussion for details). Consistent effects of ASOs on exon 13 inclusion were observed in all three cell lines.

The enhancing or repressing effects of ASOs on exon 13 are dose-dependent as shown in Jurkat cells treated with increasing amount of ESE1 and ESS1 ASOs (Figure 3.17C). 25 pmol of ESE1.2 ASO was sufficient for maximal repression of exon inclusion leaving only 1% of exon 13- included mRNA. Exon inclusion was completely eliminated when cells were treated with 100 pmol ESE1.2 ASOs (Figure 3.17C, lanes 4-6). ESS1.2 ASO, on the other hand, weakly promoted exon inclusion (5% increase) at 25 pmol and the enhancing effects reached a maximum at 50 pmol with an inclusion frequency of 25% (Figure 3.17C, lanes 7-9). This analysis indicates the feasibility of artificial manipulation of *CD46* exon 13

inclusion levels by ASOs and future functional analysis of each cytoplasmic tail in epithelial or T cells. Besides, this ASO-mediated alternative splicing manipulation of exon 13 suggests a potential novel therapeutics for autoimmune diseases although optimization of ASO design and further studies in disease models are needed.

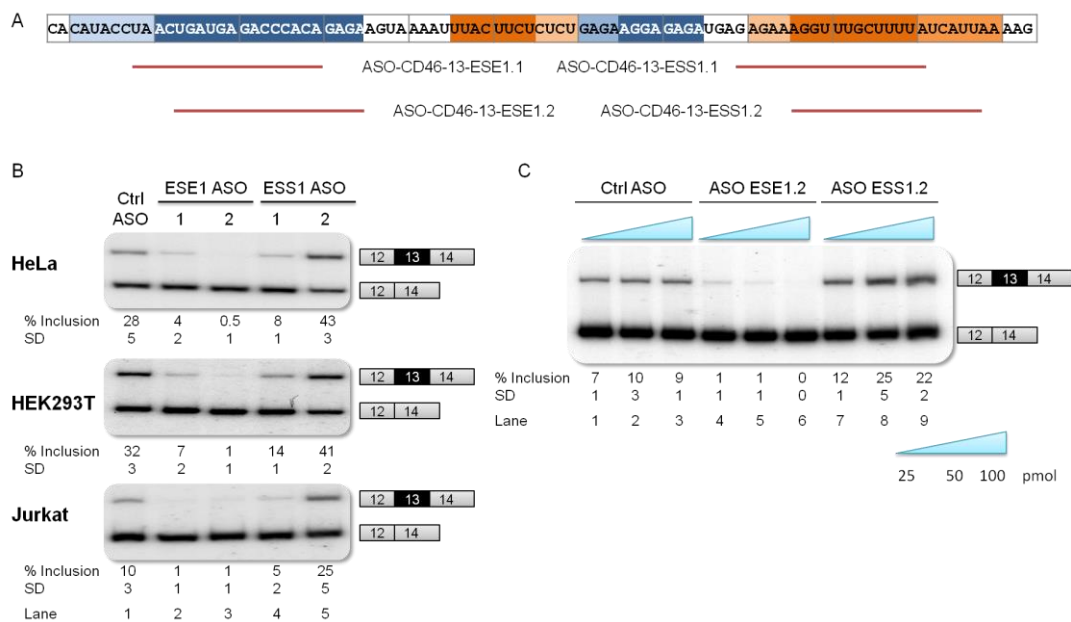


Figure 3.17 ASOs are competent to modulate exon 13 inclusion in a dose-dependent manner

(A) A total of four ASOs were designed to target ESE1 and ESS1. (B) ESE1 ASOs strongly repressed exon inclusion (lanes 2-3) while only ESS1.2 ASO weakly enhanced exon inclusion (lane 5). (C) As little as 25 pmol ESE1.2 ASOs was sufficient to repress exon inclusion by blocking the binding of activators (lane 4). ESS1.2 ASO, in contrast, could weakly improve exon inclusion by about 15% at most (lanes 7-9). Experiments were done at least in triplicate.

3.6 Stimulation of Jurkat cells with PMA did not change the inclusion of *CD46* exon 13

It has been known that the global pattern of alternative splicing changes during T-cell activation (Martinez et al., 2012). Additionally, a previous study has reported that alternative splicing of *CD45*, which regulates T-cell activation by removing an inhibitory phosphate group at Lck, in JSL1 cells (derived from Jurkat cells), changed upon T-cell activation. This study proposed that alternative splicing of *CD45* is regulated by PKC signaling pathway upon T-cell activation (Lynch and Weiss, 2000). Since *CD46* is involved in regulating T-cell activation, it is likely that this process could trigger a change in *CD46* alternative splicing which in turn regulates the T-cell response. To examine this hypothesis, Jurkat T cells were treated with either PMA which is a phorbol ester that can trigger the PKC signaling pathway (Pino et al., 2008), or anti-CD3/CD28 antibodies. During T-cell activation, many genes related to proliferation and cytokine production are upregulated, and these genes can be used as markers to determine the activation status of T cells (Shipkova and Wieland, 2012). *CD69*, which is a transmembrane lectin protein involved in T-cell proliferation, is largely upregulated upon T-cell activation so it is frequently used as an early activation marker (Simms and Ellis, 1996; Testi et al., 1989). After 24 h post activation, *CD69*-positive cells were sorted and RNAs were extracted. Endogenous *CD46* mRNA isoforms were examined by semi-qRT-PCR with primers annealing to exons 12 and 14. No significant change in exon 13 inclusion was detected in Jurkat T cells activated by any of the two stimuli (Figure 3.18B), indicating that alternative splicing pattern of exon 13 was not affected by T-cell signaling pathways in the Jurkat E6.1 cell line.

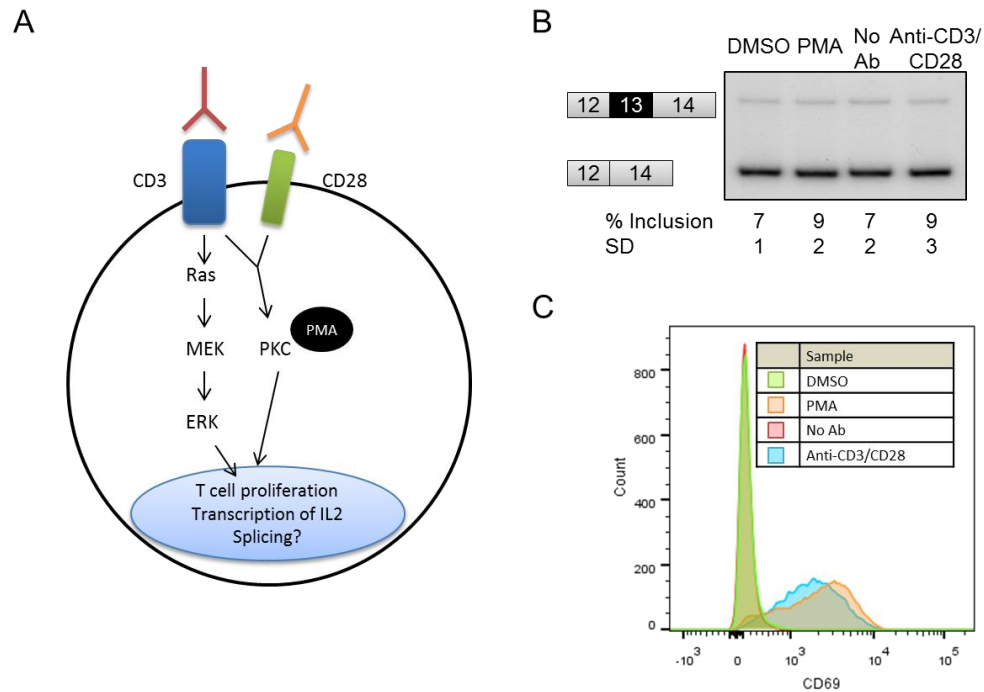


Figure 3.18 T-cell activation does not regulate the alternative splicing pattern of exon 13

(A) T-cell signaling pathways triggered by anti-CD3/CD28 antibodies and PMA. (B) Jurkat T cells were stimulated with either PMA or anti-CD3/CD28 antibodies but none of these stimulations significantly changed exon 13 inclusion level. (C) Assessment of CD69 expression level by FACS was used to determine the expected activation status of Jurkat T cells. Experiments were done at least in triplicate.

3.7 *CD46* mRNA containing exon 13 is slightly sensitive to NMD

The low inclusion level of exon 13 in endogenous transcripts could also be due to other post-transcriptional processes. One possibility is NMD, which is typically triggered by an in-frame PTC that locates 50 nt or further upstream from the last exon-exon junction (Brognia and Wen, 2009), albeit this 50-nt boundary is flexible. The stop codon within exon 13 is 41 nt away from the last exon-exon junction, which is not far from the 50 nt boundary. Therefore, endogenous *CD46* mRNA containing exon 13 may be degraded by NMD resulting in less exon 13 inclusion being detected. A test was then carried out to examine this hypothesis. Cells were treated for 5 h, with cycloheximide (CHX), which is a translation inhibitor, to prevent the occurrence of NMD and so degradation of NMD-targeted mRNA. A small but reproducible increase in exon 13 included mRNA ($\approx 10\%$ increase) was detected in CHX-treated HEK293T cells and HeLa cells (Figure 3.19). This result strongly suggests the *CD46* mRNA containing exon 13 is slightly targeted by NMD.

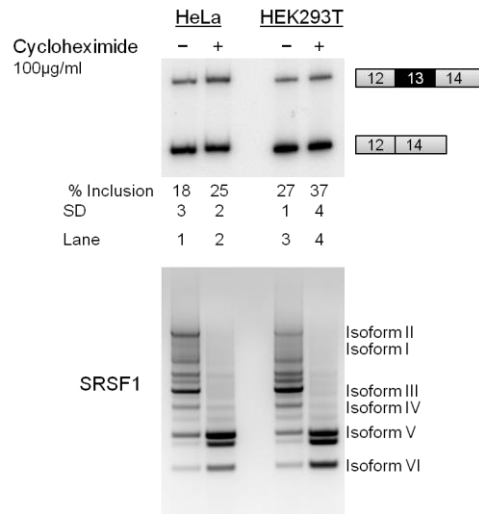


Figure 3.19 *CD46* mRNA containing exon 13 is slightly sensitive to NMD

Translational inhibition caused by CHX treatment resulted in more exon13-included mRNA, suggesting that some of these mRNAs are degraded in normal condition. As a positive control, SRSF1 has splice isoforms V and VI that are subjected to NMD and these isoforms greatly increase upon CHX treatment (Sun et al., 2010). Experiments were done at least in triplicate.

4. DISCUSSION

Multiple *CD46* isoforms with tissue-selectivity were previously detected due to alternative splicing. Altered *CD46* alternative splicing is associated with numerous autoimmune diseases such as asthma, MS and RA (Astier et al., 2006; Le Fricc et al., 2011; Tsai et al., 2012; Xu et al., 2010). This project is the first attempt to understand the regulatory mechanisms of *CD46* alternative splicing. Multiple strong ESEs and ESSs were identified along exon 13 and a strong ISE is located downstream of the 5'ss flanking exon 13. These *cis*-acting elements are bound by splicing activators (TIA1/TIAL1) or repressors (PTBP1) to either promote or repress exon 13 inclusion. Surprisingly, here it was found that a common splicing activator like SRSF1 represses exon 13 inclusion, and this regulation appears to be direct. A better understanding of *CD46* alternative splicing regulatory mechanisms not only expands the current splicing knowledge but may also be helpful to develop therapy for autoimmune diseases.

4.1 Tissue-selective alternative splicing of *CD46* is likely regulated by *trans*-acting factors

In order to understand the trend of *CD46* alternative splicing in tissues, the alternative splicing patterns of *CD46* in twenty human tissues were examined by semi-qRT-PCR using primers which target the immediately flanking constitutive exons. The results suggest that *CD46* alternative splicing is tissue-selective but with a preference for BC2 and C2 isoform (Figure 3.1.) The variations across tissues could be caused by tissue-specific *trans*-acting factors or ubiquitous *trans*-acting factors with different expression levels in various tissues (Buckanovich and Darnell, 1997; Coutinho-Mansfield et al., 2007). Nevertheless, each tissue sample comprises different cell types and so the alternative splicing pattern of *CD46* in a tissue is a collection of heterogeneous cells. Cell-type selective expression of *CD46* isoforms may slightly deviate from what observed in the tissue sample but it does not change the conclusion that exon 13 is predominantly skipped while exons 8 and 9 are predominantly included.

The analysis of STP isoforms in twenty tissues suggests that the STP region is essential for certain CD46 functions because the splice product containing no STP cassette exons was barely detected in few tissues (Figure 3.1B). Besides, the differential expression of STP variants suggests that each STP isoform may have different binding capacities for ligands or pathogens.

In addition to the expression levels of *trans*-acting factors, modifications of *trans*-acting factors may influence their subcellular location as well as their binding affinity to RNA, and eventually cause a change in splicing (Stamm, 2008). For instance, if the modification causes translocation of a splicing activator from nucleus to cytoplasm or decreases its binding affinity, the alternative splicing pattern of its regulated exons would become skipping. Future works should elucidate the identities of the *trans*-acting factors responsible for the tissue selective differences in *CD46* alternative splicing.

4.2 STP cassette exons have different inclusion efficiencies

Exon 9 is included more efficiently in mRNA followed by exons 8 and 7 as virtually all endogenous *CD46* isoforms contain exon 9. Exon 8 inclusion level ranged from 15% to 90% in different tissues and inclusion of exon 7 was only barely detected in very few tissues (Figure 3.1B). One reason for the imperfect inclusion of exons 7 and 8 might be that their flanking 5'ss severely deviate from the consensus 5'ss sequence. Mutational analysis in this thesis has demonstrated that the recognition of the 5'ss flanking exons 7 and 8 by U1 snRNA occurs via the asymmetric loop register. Breaking any of the base pairs specific to the asymmetric loop register resulted in loss of exon inclusion (Figure 3.2C and Figure 3.3C). The use of non-canonical register is often associated with alternative exons (Roca et al., 2012) and the base-pairing between U1 and 5'ss might be affected by *trans*-acting factors. Exons 7 and 8 are flanked by almost the same 5'ss sequence but exon 8 is included much more efficiently than exon 7 (Figure 3.1-3.3). It is likely that exon 7 and its downstream intron have more negative regulators to inhibit 5'ss recognition, based on the splicing prediction. On the other hand, inclusion of exon 8 may be dependent on exon 9 inclusion. The recruitment of the spliceosome at exon 9 may

help promote the binding of U1 snRNP to the 5'ss flanking exon 8 through intron definition. On the other hand, the steric hindrance caused by the very short distances between exon 7 and its neighboring exons (intron 6 is 411 nt long and intron 7 is 127 nt long) may disrupt the recruitment of the spliceosome, resulting in extremely low exon 7 inclusion. The STP domain is the binding site for *Neisseria* bacteria and the BC isoform has higher bacterial adherence than the C isoform (Källström et al., 1997; Källström et al., 2001), suggesting that exons 7 and 8 are included less to prevent bacterial infection.

4.3 *Cis*-acting elements and *trans*-acting factors regulate *CD46* exon 13 inclusion

Endogenous exon 13 is predominantly skipped in all twenty tissues (Figure 3.1A). Serial deletion and point mutation assays using splicing minigenes were carried out to identify the silencers and enhancers regulating the alternative splicing pattern of exon 13. Results showed that virtually the entire exon 13 sequence is covered in ESEs and ESSs with two of each (Figure 3.6). In contrast, the adjacent flanking introns contain only one strong ISE and several weak ISSs and ISEs. Recently it is becoming clear that distal intronic sequences can regulate alternative splicing (Lovci et al., 2013; Thomas et al., 2010), and hence it is possible that the distal intron 12 and 13 sequences instead of the proximal sequences contain strong ISSs or ISEs for regulating exon 13 inclusion (Figure 4.1).

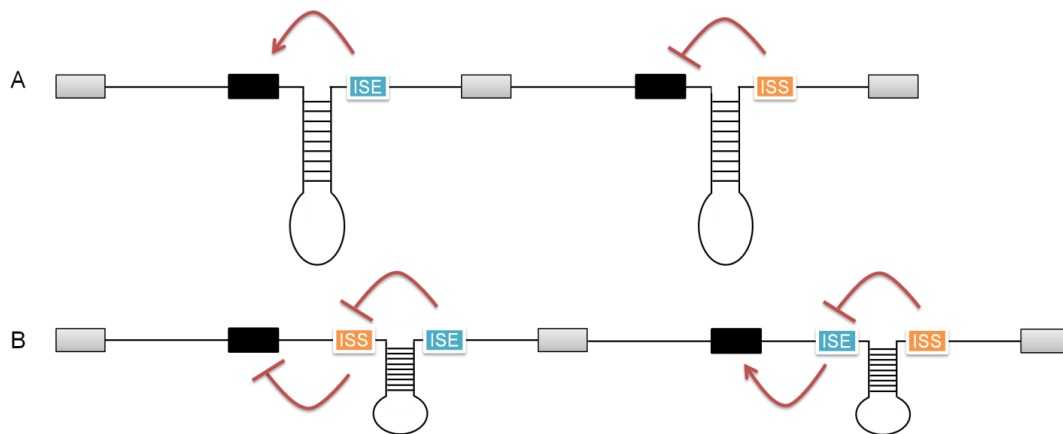


Figure 4.1 Distal intronic sequences can contain splicing regulatory elements

(A) RNA loops form within an intron allowing a distal ISE or ISS to promote or repress exon recognition. (B) A distal ISE counteracts the repressing effect of a more proximal ISS to promote exon inclusion. Vice versa, a distal ISS inhibits a proximal ISE to promote exon skipping. Images are adapted from Buratti and Baralle, 2004 and Lovci et al., 2013.

Each deletion affected the alternative splicing pattern of *CD46* exon 13 in a similar manner in three distinct cell lines: HEK293T, HeLa and Jurkat T cells. This result implies that the *cis*-acting elements have very similar roles across cell types. It also further supports that the varied preferences of *CD46* splice isoforms in different tissues could be regulated by tissue-specific *trans*-acting factors. However, two discrepancies were observed between the expression ratio of two cytoplasmic tails in endogenous *CD46* mRNAs and in minigene mRNAs. First, mRNAs derived from wild-type minigene mostly contained exon 13 (Figure 3.5) while endogenous mRNAs predominantly exhibited exon 13 skipping (Figure 3.1A). Second, no clear cell type-selective alternative splicing pattern of wild-type *CD46* minigene was observed in three different cell lines. These discrepancies suggest that the minigene may not completely reconstitute the *in vivo* situation, yet it allowed the identification of many general *cis*-acting elements.

One possible cause would be the RNA secondary structure formed by the intronic sequences or the 3' end of exon 14 which are not included in the minigene. The

RNA secondary structure would originally bury the enhancer sequences or the essential *cis*-acting elements like the 5'ss and 3'ss from *trans*-acting factors, leading to exon skipping. These inhibitory effects would be impaired when the structure is altered due to the deletions of some sequences, resulting in more exon inclusion. According to the mfold results (Zuker, 2003), there is no significant change in the secondary structure formed between exon 13 and its surrounding introns after removing the middle intronic sequences (data not shown). This observation suggests that the deleted sequences may not mask the exon 13 from the spliceosome. Besides, the deleted intronic sequences may contain *cis*-acting elements that are bound by tissue specific *trans*-acting factors to generate tissue specific alternative splicing (Figure 4.2). As seen in Figure 3.13 and Figure 3.14, PTBP1 has an incredibly strong repressive effect on minigene exon 13 as compared to the endogenous one. This could be due to the absence of middle intronic sequences in the minigenes since these sequences might contain regulatory elements that can counteract the PTBP1-mediated repression.

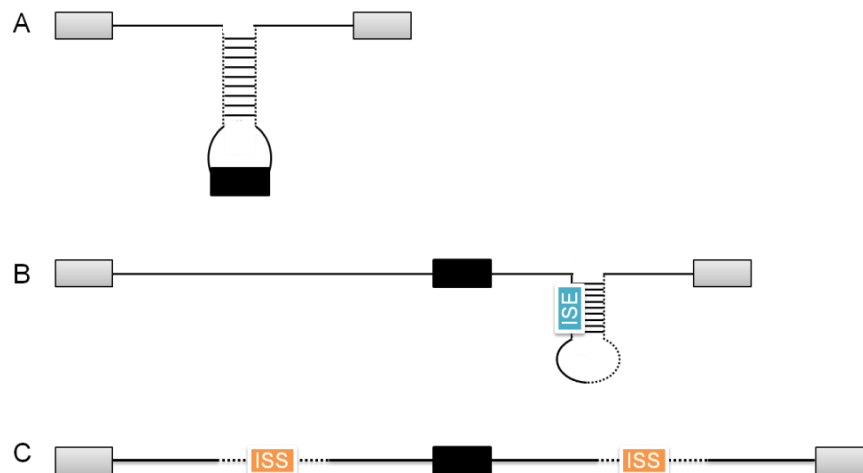


Figure 4.2 Deleted intron 12 and 13 middle sequences may contain regulatory information

(A) Hybridization between the deleted intronic sequences may mask the exon 13 from inclusion. (B) The internal loop forms within an intron may bury enhancer from protein binding. (C) The deleted intron may contain silencers to repress exon inclusion. Dotted line represents deleted sequences.

In addition, the different chromatin structures and histone modifications between plasmid DNA and endogenous DNA probably favor the recognition of exon 13, leading to more exon 13 inclusion in minigene splice products. The less complex chromatin structure which forms around minigene DNA may favor exon 13 inclusion due to increased transcriptional rate or other mechanisms (Dujardin et al., 2014; Kornblihtt et al., 2004; Kornblihtt et al., 2013). Slow transcriptional elongation rate caused inhibition of endogenous exon 13 inclusion (preliminary data not shown), suggesting that slow transcription may promote recruitment of a repressor resulting in exon skipping. Besides, differences in histone modifications between endogenous and minigene DNA could cause the alternative splicing change since different modifications provide a platform for different protein bindings (Luco et al., 2011; Luco et al., 2010). On the other hand, the minigene uses the strong CMV promoter for its transcription. Although very little minigene plasmid was used for the transfection, the amount of transcripts generated might be still enough to exhaust the splicing repressors resulting in inclusion of exon 13 in most of the transcripts.

The translational inhibition experiment showed that exon 13 containing endogenous *CD46* mRNA is targeted by NMD although the effect is very mild (Figure 3.19). The *CD46* minigene mRNAs which have no translation start codon are not degraded by NMD. Thus, this NMD-insensitivity of the minigene mRNA could just partially explain the higher exon 13 inclusion in minigene mRNA.

The alternative arrangement of ESEs and ESSs in exon 13 suggests the competitive binding of activators and repressors to exon 13 caused by steric blocking (Figure 4.3) (Fu and Ares, 2014). The ESE1 may directly interact with U2AF35 and hence stabilize the binding of U2AF to the 3'ss (Fu and Ares, 2014; Wu and Maniatis, 1993; Zhu and Krainer, 2000). The ESS1 most likely prevents the recruitment of U1 snRNP to the 5'ss by steric blocking or inhibits the subsequent spliceosome assembly (Chiou et al., 2013; House and Lynch, 2008). The ESS2 and ESE2 located in the middle region of exon 13 may regulate splicing by either promoting or inhibiting exon definition (Izquierdo et al., 2005). The ISE downstream of the

5'ss is bound by TIA1/TIAL1 thus probably interacting with U1C to help stabilize the U1-5'ss interaction (Le Guiner et al., 2001). *CD46* further exemplifies that the splicing outcome is dependent on the balance of positive and negative regulators (Ladd et al., 2005). Exon 13 is flanked by a relatively weaker 5'ss which deviates a lot from the consensus 5'ss sequence. Hence, enhancers are needed for exon inclusion. The number of enhancers and silencers with similar strength found through our deletion assays is almost even, yet the endogenous exon 13 inclusion level is still low. This suggests that the expression of protein regulators is the determinant for exon 13 inclusion. It is likely that the expression of activators for exon 13 is low in normal condition, resulting in less exon inclusion.

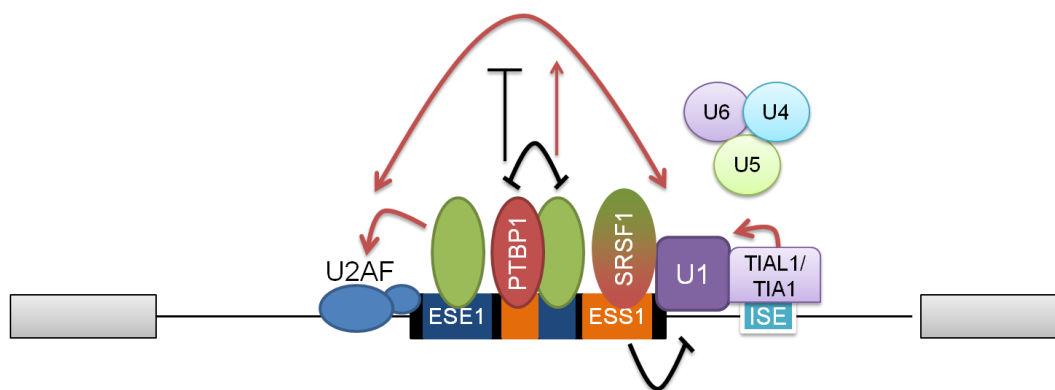


Figure 4.3 Regulatory mechanisms of exon 13 ESEs and ESSs

Competitive binding of activators or repressors to ESEs or ESSs and cross-inhibition between protein regulators determine the alternative splicing pattern. The activators bound to ESE1 may help to recruit U2AF35 to the 3'ss. The repressors bound to ESS1 may inhibit the U1 recruitment. Nevertheless, the extended interaction between the repressors and U1 snRNP may prevent the recruitment of tri-snRNP (U4/U6, U5) to the 5'ss. In addition, these regulators may either promote or prevent the exon definition.

Neither single knockdown of one repressor (SRSF1) nor of one activator (TIAL1) completely included or excluded exon 13 from mRNA (Figure 3.11, Figure 3.14 and Figure 3.15), suggesting that the splicing regulation is a combination of different *trans*-acting factors. However, depletion of both SRSF1 and PTBP1 did not show any synergistic enhancing effect on inclusion of exon 13 (data not shown). SRSF1 and PTBP1 may have the same binding site on exon 13, and so only one of these two proteins can repress the exon inclusion at any given time. On the other hand, overexpression of repressors has a negligible effect on exon 13 inclusion. It could be that the endogenous SRSF1 and/or PTBP1 are already saturated for regulating *CD46* pre-mRNA alternative splicing.

Trans-acting factors are cross-regulated by each other (Bonomi et al., 2013; Boutz et al., 2007) and this makes the interpretation of splicing codes more challenging. TIA1 or TIAL1 overexpression caused a decrease in exon 13 inclusion in the context of the *CD46* minigene (Figure 3.16). One possible reason could be that the expression or activity of one or several repressors is regulated by TIA1 and TIAL1. Therefore, increased functional repressors caused by exogenous TIA1 and TIAL1 resulted in exon 13 skipping. Meanwhile, exon 13 was less repressed in TIAL1-overexpressed cells. TIAL1 probably has better enhancing strength than TIA1 or it has less impact on regulating other *trans*-acting factors. On the other hand, TIA1 and TIAL1 have been shown to regulate alternative splicing of each other (Izquierdo and Valcarcel, 2007). Here, TIAL1 appeared to regulate expression of TIA1 as there was a small increase in TIA1 expression in cells supplemented with TIAL1 cDNA plasmid (Figure 3.15, lanes 3,6,9). This might help to explain the stronger enhancing effect of TIAL1.

4.4 SRSF1 is a splicing repressor for *CD46* exon 13

SRSF1 is a well known activator (Rooke et al., 2003; Sun et al., 1993; Zhu and Krainer, 2000), yet in this thesis it was found to be a repressor of *CD46* exon 13. Knockdown and overexpression assays showed that depletion of SRSF1 led to increased exon 13 inclusion while exon 13 with removal of ESS1 or ESS2 resisted to SRSF1-mediated repression (Figure 3.11 and Figure 3.13). In addition, binding of SRSF1 to ESS1 was detected by RNA pull-down (Figure 3.10A). The exact regulatory mechanism of SRSF1 on exon 13 is still unknown but it is unexpected to observe that SRSF1 represses exon inclusion by likely interacting with the exonic sequence that it regulates. A previous study has shown that SRSF1 depletion caused either increase or reduction in exon inclusion in different mRNAs in mouse embryonic fibroblasts (Pandit et al., 2013), yet this study did not show the correlation between SRSF1 binding site and its direction of regulation. So far, SRSF1 is only known to repress exon inclusion through interaction with a downstream constitutive exon like the case of *RON* exon 11 (Ghigna et al., 2005).

It is known that SR proteins and hnRNPs regulate splicing in a positional dependent manner (Erkelenz et al., 2013). SR proteins enhance exon inclusion when bound to the exon but they repress exon inclusion when recruited to the region downstream of the 5'ss. In contrast, hnRNPs recruited to the exon or the upstream intron generally repress exon recognition and only promote exon inclusion when bound to the downstream intron. This study also showed that binding of these regulators did not impair the U1 snRNP recruitment to the 5'ss. Rather, these proteins affect the transition of E complex to A complex. All these data blurs the definition of SR proteins as activators and hnRNPs as repressors. Especially my result strongly suggests that SRSF1 acts on ESS/s to repress exon inclusion, though further experiment like *in vitro* splicing is required to confirm the direct regulation. Thus, SRSF1 might likely interfere with the exon definition or with the progression of spliceosome assembly. The interaction between the RS domain of SRSF1 and U1 snRNP helps in the recruitment of U1, but an extended interaction may inhibit the subsequent spliceosome rearrangements (Chiou et al.,

2013). Besides, SRSF1 may also interact with other proteins to inhibit exon inclusion (Figure 4.4).

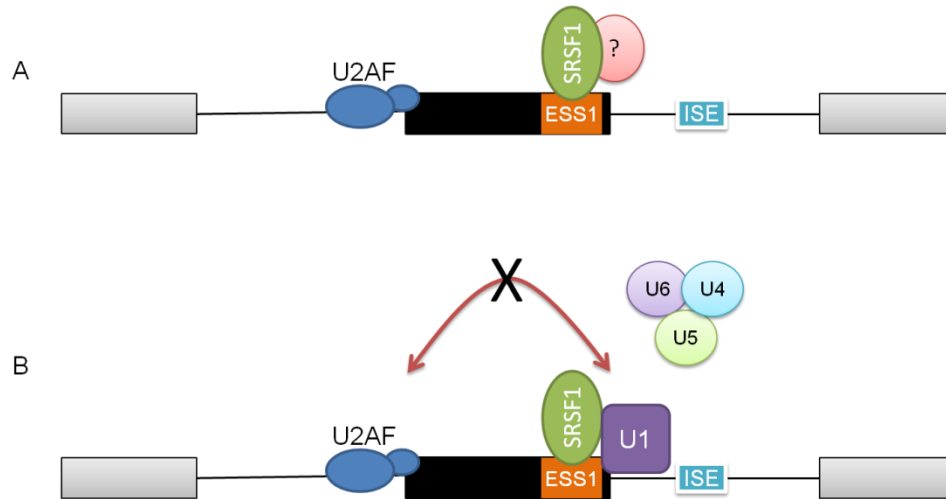


Figure 4.4 Possible regulatory mechanisms of SRSF1 on *CD46* exon 13

(A) SRSF1 may interact with other proteins and inhibit the recruitment of U1 snRNP to 5'ss. (B) SRSF1 may interfere with the exon definition and its prolonged interaction with U1 snRNP may prevent the spliceosome assembly.

Another implication of my result is that it is difficult to predict the role of a *trans*-acting factor in splicing just based on its binding position and protein family. The length of exon and intron, and the flanking protein binding sequences may also influence the binding affinity and function of a regulator (Pandit et al., 2013). Other types of regulatory mechanisms may be present in addition to looping out exon, masking splice sites, nucleation of proteins, blocking exon definition and direct recruitment of U1 snRNP. Thus, there might be many unknowns needed to be solved to fully understand all splicing regulatory mechanisms.

4.5 Origin of *CD46* exon 13

The *CD46* exon 13 sequence exists in the primate genomes and in some mammalian genomes like in dog, elephant and dolphin but not in mice and some other metazoans like birds (Figure 4.5A). Although exon 13 sequence is present in the *CD46* genes of primates and other mammals, this sequence is not recognized as an exon but rather it is part of an intron or UTR in most of the animals except human, gorilla, gibbon, macaque, marmoset, baboon and green monkey (Table 4.1). This observation suggests that exon 13 is newly evolved through exonization by which a sequence gains novel splice sites by mutations (Keren et al., 2010), and as such its inclusion efficiency may not be optimal as illustrated by the slightly weaker 5'ss (Table 3.1). Despite the high conservation of exon 13 sequence and flanking intronic sequences among primate *CD46* genes (Figure 4.5B), the alternative splicing patterns of exon 13 sequence vary in these primates (Table 4.1). The exon 13 sequence is annotated as part of an intron in chimpanzee and orangutan, while it is an alternative terminal exon in macaque and marmoset. Exon 13 appears to be constitutively included in baboon, gibbon and green monkey *CD46* mRNAs but it is clearly a cassette exon in gorilla and human. Few nucleotide differences in exon 13 and intronic sequence may affect auxiliary *cis*-acting elements resulting in variations in alternative splicing patterns. Mutations that arise in the genome throughout evolution could either create a novel exon to expand protein function or weaken an existing exon that is non-essential or harmful. Besides, genetic variations like single nucleotide polymorphism (SNP) can possibly perturb splicing and contribute to disease progression (Lu et al., 2012).

The CYT1 derived from this new exon adds new *CD46* functions which may not be essential but expand cellular activity and functional plasticity. It has been known that the two cytoplasmic tails have different roles in regulating T-cell immune response (Ni Choileain and Astier, 2012; Yamamoto et al., 2013). CYT1 is required to induce Tr1 cell differentiation in order to suppress T-cell response at a post-infection stage. It is possible that low CYT1 expression is sufficient for the

function and high expression of CYT1 may cause immunodeficiency and therefore the inclusion of exon 13 is maintained at a low level.

In addition, CD46 maintains the integrity of epithelial cell-cell junctions and CD46 activation accelerates wound healing (Cardone et al., 2011a; McLaughlin et al., 2003). Although most of the epithelial functions are mediated through CYT1, the presence of these two cytoplasmic tails with different signaling capacities allows more functional roles for CD46 in different cell types. On the other hand, CYT1 can induce the autophagosome which is the first line of pathogen elimination. This function is coincident with its novel regulation in T_H1 cells, suggesting that immunological roles of CD46 are further expanded by alternative splicing of exon 13. However, pathogens like MV use CD46 as an entry receptor and hijack its function to evade elimination (Kurita-Taniguchi et al., 2000; Smith et al., 2003). Upon binding to CD46, pathogens induce CD46-CYT1 mediated autophagy to get into the cytoplasm where they escape from the autophagosome and start the infection (Joubert et al., 2009). Therefore, the low level of exon inclusion is probably a defense strategy to reduce MV invasion.

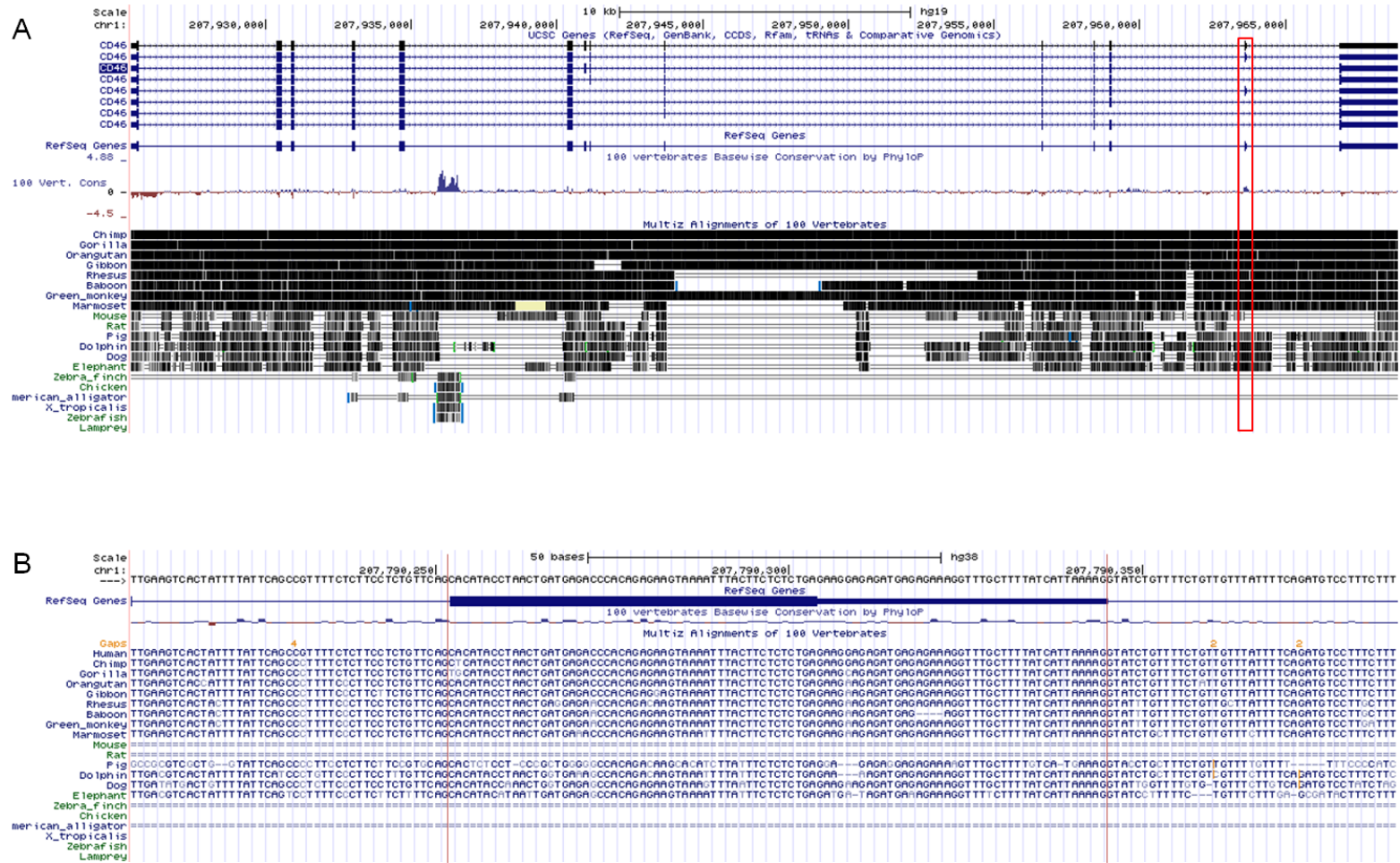


Figure 4.5 Conservation of *CD46* gene

(Continued on the next page)

(Continued from the previous page)

(A) The top part of diagram shows the human *CD46* reference gene. Square boxes represent exons while lines represent introns. The bottom part shows the gene conservation among different species as a grayscale density plot. The darker values in conservation histogram indicate the higher levels of overall conservation. Double horizontal line represents a long gap in aligned sequence between species. *CD46* gene is highly conserved among primates while it is less conserved in other mammals. In chicken, alligator, western clawed frog and zebrafish genomes, only a small fragment of gene sequence aligns to human *CD46* gene. Red square highlights the exon 13 region among different species. (B) Sequence alignment of human *CD46* exon 13 to different species. Conserved nucleotide is labeled in dark blue while non-conserved nucleotide is labeled in light blue. Single line represents missing nucleotide while double line represents sequence gap. Red vertical lines mark the start and end of exon 13 sequence. These sequence alignments (A,B) are obtained from UCSC genome browser. A longer sequence alignment of exon 13 with 100 nt flanking introns between human and other species is shown in Appendix III.

Table 4.1 Alternative splicing patterns of exon 13 homologues in different species

Intron/UTR	Constitutive exon	Cassette exon	Alternative terminal exon
Chimpanzee	Baboon	Human	Macaque
Orangutan	Gibbon	Gorilla	Marmoset
Dog	Green monkey		
Dolphin			
Elephant			

Footnote: This table is generated based on the annotated transcripts obtained from Ensembl genome browser.

4.6 ASOs effectively modulate *CD46* alternative splicing

ASOs are able to anneal to single stranded RNAs and cause mRNA degradation, translational inhibition, prevention of protein binding and splicing switch (Bennett and Swayze, 2010; Dias and Stein, 2002; Kole et al., 2012). They were first designed to reduce gene expression via the RNase-H mediated mRNA cleavage. ASOs are nowadays also applied to modulate splicing by either promote exon inclusion or skipping to produce functional or partially functional proteins. Duchenne muscular dystrophy (DMD) is caused by the absence of functional dystrophin protein. The protein expression can be restored by modulating exon skipping in mRNA using ASOs to generate a shorter but functional protein (Arechavala-Gomez et al., 2012; Koo and Wood, 2013).

ASOs targeting ESE1 completely inhibited exon recognition (Figure 3.17B), and this can be achieved by blocking the binding of activators to enhancers and the recruitment of the spliceosome to the 3' splice site due to steric hindrance. On the other hand, ESS1.1 ASO instead of promoting exon inclusion caused further repression on exon recognition. It is likely that the ESS1.1 ASO blocked the binding of activator to ESE2 due to the close proximity leading to more exon repression. ESS1.2 can only weakly improve exon 13 inclusion and this could be due to the presence of another strong silencer in exon 13 sequence which could cancel out the enhancing effect. Besides, the binding of ASOs at the 3' end of exon could interrupt the recruitment of U1 snRNP due to the steric hindrance. Albeit the effectiveness of ASOs on exon repression, more considerations are needed for ASO design such as binding position to improve exon inclusion. As shown in this thesis, the alternative splicing pattern of *CD46* exon 13 can now be artificially modulated using ASOs. This opens up research opportunities to study the role of each cytoplasmic tail in cellular activities.

4.7 T-cell signaling and *CD46* alternative splicing

Many alternative splicing events like *CD45* and *Fas* change upon T-cell activation (Martinez et al., 2012). *CD46* as a T-cell co-activator (Astier et al., 2000) might also be regulated at the point of splicing during T-cell activation to facilitate the T-cell response. However, no alteration in *CD46* exon 13 inclusion was detected in Jurkat T cells stimulated with PMA (Figure 3.18B). In addition, there was no change in the alternative splicing pattern of STP exons (data not shown). These findings suggest that the signaling pathways triggered by PMA do not regulate *CD46* alternative splicing in this cell line. In addition, it has been reported that PMA can trigger signaling pathways that are either similar to or distinct from those in primary T cells (Martinez et al., 2012). Therefore, signal cascades activated by PMA in Jurkat cells may not mimic the canonical pathways that induce a change in *CD46* alternative splicing in T cells.

However, activation of Jurkat T cells by cross-linking CD3 and CD28 did not change the alternative splicing pattern either. Nevertheless, a recent study showed that engagement of CD46 with locally produced C3b upon T-cell activation (Cardone et al., 2010) leads to MMP-mediated cleavage of CD46 ectodomain and subsequently the cytoplasmic tail is cleaved by P γ S. Temporal cleavage of CYT1 and CYT2 is sufficient to induce Tr1 differentiation and then terminate T-cell response (Ni Choileain et al., 2011). Hence, change in *CD46* alternative splicing might not be necessary to switch the expression of two cytoplasmic tails to trigger different signaling pathways. According to this model, alternative splicing of exon 13 might only be needed to generate a diverse cytoplasmic tail repertoire which is then temporarily regulated by proteolytic cleavage.

Splicing regulation of *CD46* is dependent on a regulatory network of *cis*-acting elements, protein regulators, transcription, NMD and probably cell signaling. The complexity and flexibility of alternative splicing may allow cells to quickly respond to environmental stimuli by simply changing expression of few regulators which can either repress or enhance inclusion of different exons, to produce isoforms needed or degrade proteins through NMD. This study is only the

beginning of the characterization of *CD46* alternative splicing. Further experiments are required to elucidate the regulatory mechanisms and networks of protein regulators, and other processes like transcription. Comprehensive studies about *CD46* alternative splicing could be useful for medical application since its aberrant alternative splicing is associated with autoimmune diseases.

5. CONCLUSIONS AND FUTURE DIRECTIONS

Altogether, the presented data shows that *CD46* alternative splicing is regulated by multiple factors and its alternative splicing can be modulated by ASOs. A weak 5'ss flanking the cassette exons is one of the reasons causing low exon inclusion. STP exons 7 and 8 are flanked by a 5'ss that weakly complements to U1 snRNA in the canonical register, yet this study showed that these two 5'ss can form additional base pairs with U1 snRNA in the asymmetric loop register to stabilize the U1 snRNP binding. On the other hand, exon 13 is flanked by a relatively stronger 5'ss compared to that of exons 7 and 8 yet its inclusion remains low in most tissues, implying that additional factors are present to modulate the alternative splicing pattern of exon 13. Multiple ESEs, ESSs, ISEs and ISSs have been identified along exon 13 and flanking introns through the deletion assays. Knockdown and overexpression assays suggest that PTBP1 weakly represses exon 13 inclusion through the ESS2 region. Furthermore, exon 13 inclusion is strongly repressed by SRSF1 and the regulatory site of SRSF1 is likely the ESS1 region as suggested by the pull-down and overexpression assays. On the other hand, recognition of exon 13 is largely promoted by TIAL1 via the U rich sequences immediately downstream of the 5'ss. Besides, a low level of exon 13 containing *CD46* mRNA in cells is partially due to NMD. *CD46* isoforms are important for T-cell signaling, yet no feedback mechanism appears to regulate *CD46* alternative splicing. Based on the results from deletion assays, ASOs were designed to target ESE1 and ESS1 to modulate exon 13 inclusion for more *CD46*-CYT1 production. Though further studies are required, switching expression of *CD46* isoforms could be helpful to promote immunosuppression in autoimmune disease. Besides, manipulation of *CD46* alternative splicing could recapitulate pathogenesis of diseases, allowing identification of defective regulatory pathways. Consistent with previous mechanistic characterization of other alternative splicing events, this study further shows that splicing regulation is complex and context dependent.

Despite the identification of *cis*-acting elements and *trans*-acting factors, few more questions remain to be addressed to fully understand the alternative splicing

regulation of *CD46* exon 13 in the future. The first issue would be the regulatory mechanism and actual binding site of SRSF1 to repress exon 13 inclusion. The answer can be achieved by *in vitro* splicing and spliceosome assembly assay to examine which stage of splicing is inhibited by SRSF1, and by electrophoretic mobility shift assay (EMSA) to determine the binding site. Several proteins bound to ESE1 have been identified through RNA pull-down and mass spectrometry. The second future direction would be use of loss- and gain-of-function assays to investigate whether these proteins are involved in the regulation of exon 13 alternative splicing.

ASO-mediated artificial modulation of exon 13 inclusion allows the study of respective functions of each cytoplasmic tail in T-cell response as well as in epithelial cells or other cell types. Combination of ASO treatment and CD46 activation would be applied to separately examine the CD46 signaling mediated by the two cytoplasmic tails in different cells. Aberrant *CD46* alternative splicing or defective CD46-induced Tr1 generation is associated with many autoimmune diseases such as RA, asthma and MS. Using ASO to manipulate the expression of CD46 isoforms in primary T cells from healthy donor or patients might be able to recapitulate the disease pathogenesis in order to identify causes of disease and to assess usefulness of ASO as a therapy. Besides, RNA-seq could be performed with samples from patients and healthy individuals to compare the differences in alternative splicing events. Through the RNA-seq, a novel protein regulator for exon 13 might be identified.

Though the alternative splicing regulation of exon 13 was the focus of this thesis, further study could be carried out to elucidate the alternative splicing regulation of the STP exons. Analogous to the studies of exon 13, identification of *cis*-acting elements could be done with sequence substitutions while identification of *trans*-acting factors could be achieved by RNA pull-down, knockdown and overexpression experiments. It would be interesting to examine if the low inclusion of exon 7 is due to the exonic sequence or the flanking short introns. One way to examine would be replacing the exon 8 with exon 7 and vice versa. If exon 7 still

has a low inclusion efficiency, it would suggest that the different nucleotides between the exons 7 and 8 affect the *cis*-acting elements. In contrast, it would suggest that exon 7 surroundings are the cause for its low inclusion. All in all, the importance of CD46 in immune cells and pathologies warrants further studies of alternative splicing using this model substrate.

6. REFERENCES

- Akerman, M., David-Eden, H., Pinter, R.Y., Mandel-Gutfreund, Y., 2009. A computational approach for genome-wide mapping of splicing factor binding sites. *Genome Biology* 10, R30.
- Anczukow, O., Akerman, M., Clery, A., Wu, J., Shen, C., Shirole, N.H., Raimer, A., Sun, S., Jensen, M.A., Hua, Y., Allain, F.H., Krainer, A.R., 2015. SRSF1-Regulated Alternative Splicing in Breast Cancer. *Mol Cell* 60, 105-117.
- Anczukow, O., Rosenberg, A.Z., Akerman, M., Das, S., Zhan, L., Karni, R., Muthuswamy, S.K., Krainer, A.R., 2012. The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nature Structural & Molecular Biology* 19, 220-228.
- Arechavala-Gomez, V., Anthony, K., Morgan, J., Muntoni, F., 2012. Antisense oligonucleotide-mediated exon skipping for Duchenne muscular dystrophy: progress and challenges. *Current Gene Therapy* 12, 152-160.
- Astier, A., Trescol-Biemont, M.C., Azocar, O., Lamouille, B., Rabourdin-Combe, C., 2000. Cutting edge: CD46, a new costimulatory molecule for T cells, that induces p120CBL and LAT phosphorylation. *The Journal of Immunology* 164, 6091-6095.
- Astier, A.L., 2008. T-cell regulation by CD46 and its relevance in multiple sclerosis. *Immunology* 124, 149-154.
- Astier, A.L., Meiffren, G., Freeman, S., Hafler, D.A., 2006. Alterations in CD46-mediated Tr1 regulatory T cells in patients with multiple sclerosis. *The Journal of Clinical Investigation* 116, 3252-3257.
- Barash, Y., Calarco, J.A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencowe, B.J., Frey, B.J., 2010. Deciphering the splicing code. *Nature* 465, 53-59.
- Barilla-LaBarca, M.L., Liszewski, M.K., Lambris, J.D., Hourcade, D., Atkinson, J.P., 2002. Role of membrane cofactor protein (CD46) in regulation of C4b and C3b deposited on cells. *The Journal of Immunology* 168, 6298-6304.
- Batsche, E., Yaniv, M., Muchardt, C., 2006. The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nature Structural & Molecular Biology* 13, 22-29.
- Ben-Hur, V., Denichenko, P., Siegfried, Z., Maimon, A., Krainer, A., Davidson, B., Karni, R., 2013. S6K1 alternative splicing modulates its oncogenic activity and regulates mTORC1. *Cell reports* 3, 103-115.
- Bennett, C.F., Swayze, E.E., 2010. RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annual Review of Pharmacology and Toxicology* 50, 259-293.

- Berg, M.G., Singh, L.N., Younis, I., Liu, Q., Pinto, A.M., Kaida, D., Zhang, Z., Cho, S., Sherrill-Mix, S., Wan, L., Dreyfuss, G., 2012. U1 snRNP determines mRNA length and regulates isoform expression. *Cell* 150, 53-64.
- Berget, S.M., 1995. Exon recognition in vertebrate splicing. *The Journal of Biological Chemistry* 270, 2411-2414.
- Berget, S.M., Moore, C., Sharp, P.A., 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America* 74, 3171-3175.
- Black, D.L., 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry* 72, 291-336.
- Blanchette, M., Chabot, B., 1997. A highly stable duplex structure sequesters the 5' splice site region of hnRNP A1 alternative exon 7B. *RNA* 3, 405-419.
- Boisgerault, N., Guillerme, J.B., Pouliquen, D., Mesel-Lemoine, M., Achard, C., Combredet, C., Fonteneau, J.F., Tangy, F., Gregoire, M., 2013. Natural oncolytic activity of live-attenuated measles virus against human lung and colorectal adenocarcinomas. *BioMed Research International* 2013, 387362.
- Bonomi, S., di Matteo, A., Buratti, E., Cabianca, D.S., Baralle, F.E., Ghigna, C., Biamonti, G., 2013. HnRNP A1 controls a splicing regulatory circuit promoting mesenchymal-to-epithelial transition. *Nucleic Acids Research* 41, 8665-8679.
- Boutz, P.L., Stoilov, P., Li, Q., Lin, C.H., Chawla, G., Ostrow, K., Shiue, L., Ares, M., Jr., Black, D.L., 2007. A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes & Development* 21, 1636-1652.
- Breitbart, R.E., Andreadis, A., Nadal-Ginard, B., 1987. Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Annual Review of Biochemistry* 56, 467-495.
- Brodsky, A.S., Silver, P.A., 2000. Pre-mRNA processing factors are required for nuclear export. *RNA* 6, 1737-1749.
- Broere, F., Apasov, S., Sitkovsky, M., van Eden, W., 2011. A2 T cell subsets and T cell-mediated immunity. In: Nijkamp, F.P., Parnham, M.J. (Eds.), *Principles of Immunopharmacology*. Birkhäuser Basel, pp. 15-27.
- Brogna, S., Wen, J., 2009. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nature Structural & Molecular Biology* 16, 107-113.
- Brownlie, R.J., Zamoyska, R., 2013. T cell receptor signalling networks: branched, diversified and bounded. *Nature Reviews. Immunology* 13, 257-269.

- Brunak, S., Engelbrecht, J., Knudsen, S., 1991. Prediction of human mRNA donor and acceptor sites from the DNA sequence. *Journal of Molecular Biology* 220, 49-65.
- Bryan, A.R., Wu, E.Y., 2014. Complement deficiencies in systemic lupus erythematosus. *Current Allergy and Asthma Reports* 14, 448.
- Buckanovich, R.J., Darnell, R.B., 1997. The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. *Molecular and Cellular Biology* 17, 3194-3201.
- Buratti, E., Baralle, F.E., 2004. Influence of RNA secondary structure on the pre-mRNA splicing process. *Molecular and Cellular Biology* 24, 10505-10514.
- Caceres, J.F., Sreaton, G.R., Krainer, A.R., 1998. A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes & development* 12, 55-66.
- Cardone, J., Al-Shouli, S., Kemper, C., 2011a. A novel role for CD46 in wound repair. *Frontiers in Immunology* 2, 28.
- Cardone, J., Le Friec, G., Kemper, C., 2011b. CD46 in innate and adaptive immunity: an update. *Clinical and Experimental Immunology* 164, 301-311.
- Cardone, J., Le Friec, G., Vantourout, P., Roberts, A., Fuchs, A., Jackson, I., Suddason, T., Lord, G., Atkinson, J.P., Cope, A., Hayday, A., Kemper, C., 2010. Complement regulator CD46 temporally regulates cytokine production by conventional and unconventional T cells. *Nature Immunology* 11, 862-871.
- Cartegni, L., Chew, S.L., Krainer, A.R., 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature Reviews. Genetics* 3, 285-298.
- Castellani, C., Cuppens, H., Macek, M., Cassiman, J.J., Kerem, E., Durie, P., Tullis, E., Assael, B.M., Bombieri, C., Brown, A., Casals, T., Claustres, M., Cutting, G.R., Dequeker, E., Dodge, J., Doull, I., Farrell, P., Ferec, C., Girodon, E., Johannesson, M., Kerem, B., Knowles, M., Munck, A., Pignatti, P.F., Radojkovic, D., Rizzotti, P., Schwarz, M., Stuhmann, M., Tzetis, M., Zielenski, J., Elborn, J.S., 2008. Consensus on the use and interpretation of cystic fibrosis mutation analysis in clinical practice. *Journal of Cystic Fibrosis* 7, 179-196.
- Charette, M., Gray, M.W., 2000. Pseudouridine in RNA: What, Where, How, and Why. *IUBMB Life* 49, 341-351.
- Chen, M., Manley, J.L., 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature Reviews. Molecular Cell Biology* 10, 741-754.
- Cheng, C., Yaffe, M.B., Sharp, P.A., 2006. A positive feedback loop couples Ras activation and CD44 alternative splicing. *Genes & Development* 20, 1715-1720.

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

Cho, S.Y., Hoang, A., Sinha, R., Zhong, X.Y., Fu, X.D., Krainer, A.R., Ghosh, G., 2011. Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proceedings of the National Academy of Sciences of the United States of America* 108, 8233-8238.

Chow, L.T., Gelinas, R.E., Broker, T.R., Roberts, R.J., 1977. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 12, 1-8.

Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J.L., Bell, J.C., Duncan, P.I., 1996. The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *The EMBO journal* 15, 265-275.

Coolidge, C.J., Seely, R.J., Patton, J.G., 1997. Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Research* 25, 888-896.

Cooper, T.A., Wan, L., Dreyfuss, G., 2009. RNA and Disease. *Cell* 136, 777-793.

Cordin, O., Beggs, J.D., 2013. RNA helicases in splicing. *RNA Biology* 10, 83-95.

Coulter, L.R., Landree, M.A., Cooper, T.A., 1997. Identification of a new class of exonic splicing enhancers by in vivo selection. *Molecular and Cellular Biology* 17, 2143-2150.

Coutinho-Mansfield, G.C., Xue, Y., Zhang, Y., Fu, X.D., 2007. PTB/nPTB switch: a post-transcriptional mechanism for programming neuronal differentiation. *Genes & Development* 21, 1573-1577.

Cutting, G.R., 2015. Cystic fibrosis genetics: from molecular understanding to clinical application. *Nature Reviews. Genetics* 16, 45-56.

Damgaard, C.K., Tange, T.O., Kjems, J., 2002. hnRNP A1 controls HIV-1 mRNA splicing through cooperative binding to intron and exon splicing silencers in the context of a conserved secondary structure. *RNA* 8, 1401-1415.

Das, S., Anczukow, O., Akerman, M., Krainer, A.R., 2012. Oncogenic Splicing Factor SRSF1 Is a Critical Transcriptional Target of MYC. *Cell reports* 1, 110-117.

Das, S., Krainer, A.R., 2014. Emerging Functions of SRSF1, Splicing Factor and Oncoprotein, in RNA Metabolism and Cancer. *Molecular Cancer Research* 12, 1195-1204.

David, C.J., Manley, J.L., 2010. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes & Development* 24, 2343-2364.

- De Conti, L., Baralle, M., Buratti, E., 2013. Exon and intron definition in pre-mRNA splicing. *Wiley Interdisciplinary Reviews. RNA* 4, 49-60.
- Del Gatto-Konczak, F., Bourgeois, C.F., Le Guiner, C., Kister, L., Gesnel, M.C., Stevenin, J., Breathnach, R., 2000. The RNA-binding protein TIA-1 is a novel mammalian splicing regulator acting through intron sequences adjacent to a 5' splice site. *Molecular and Cellular Biology* 20, 6287-6299.
- Desmet, F.O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M., Beroud, C., 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Research* 37, e67.
- Deyle, D.R., Escobar, D.Z., Peng, K.W., Babovic-Vuksanovic, D., 2015. Oncolytic measles virus as a novel therapy for malignant peripheral nerve sheath tumors. *Gene* 565, 140-145.
- Dias, N., Stein, C.A., 2002. Antisense oligonucleotides: basic concepts and mechanisms. *Molecular Cancer Therapeutics* 1, 347-355.
- Dork, T., Fislage, R., Neumann, T., Wulf, B., Tummeler, B., 1994. Exon 9 of the CFTR gene: splice site haplotypes and cystic fibrosis mutations. *Human Genetics* 93, 67-73.
- Douglas, A.G., Wood, M.J., 2011. RNA splicing: disease and therapy. *Briefings in functional genomics* 10, 151-164.
- Dujardin, G., Lafaille, C., de la Mata, M., Marasco, L.E., Munoz, M.J., Le Jossic-Corcus, C., Corcos, L., Kornblihtt, A.R., 2014. How slow RNA polymerase II elongation favors alternative exon skipping. *Molecular Cell* 54, 683-690.
- Dunkelberger, J.R., Song, W.C., 2010. Complement and its role in innate and adaptive immune responses. *Cell Research* 20, 34-50.
- Ebrahimi, K.B., Fijalkowski, N., Cano, M., Handa, J.T., 2013. Decreased membrane complement regulators in the retinal pigmented epithelium contributes to age-related macular degeneration. *The Journal of Pathology* 229, 729-742.
- Erkelenz, S., Mueller, W.F., Evans, M.S., Busch, A., Schoneweis, K., Hertel, K.J., Schaal, H., 2013. Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. *RNA* 19, 96-102.
- Estes, P.A., Cooke, N.E., Liebhaber, S.A., 1992. A native RNA secondary structure controls alternative splice-site selection and generates two human growth hormone isoforms. *The Journal of Biological Chemistry* 267, 14902-14908.
- Faustino, N.A., Cooper, T.A., 2003. Pre-mRNA splicing and human disease. *Genes & Development* 17, 419-437.

- Fishelson, Z., Donin, N., Zell, S., Schultz, S., Kirschfink, M., 2003. Obstacles to cancer immunotherapy: expression of membrane complement regulatory proteins (mCRPs) in tumors. *Molecular Immunology* 40, 109-123.
- Fontana, W., Konings, D.A., Stadler, P.F., Schuster, P., 1993. Statistics of RNA secondary structures. *Biopolymers* 33, 1389-1404.
- Forch, P., Puig, O., Martinez, C., Seraphin, B., Valcarcel, J., 2002. The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites. *The EMBO Journal* 21, 6882-6892.
- Fregoso, O.I., Das, S., Akerman, M., Krainer, A.R., 2013. Splicing-factor oncoprotein SRSF1 stabilizes p53 via RPL5 and induces cellular senescence. *Molecular cell* 50, 56-66.
- Fu, X.D., Ares, M., Jr., 2014. Context-dependent control of alternative splicing by RNA-binding proteins. *Nature Reviews. Genetics* 15, 689-701.
- Fuchs, A., Atkinson, J.P., Fremeaux-Bacchi, V., Kemper, C., 2009. CD46-induced human Treg enhance B-cell responses. *European Journal of Immunology* 39, 3097-3109.
- Ge, H., Manley, J.L., 1990. A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. *Cell* 62, 25-34.
- Ghigna, C., Giordano, S., Shen, H.H., Benvenuto, F., Castiglioni, F., Comoglio, P.M., Green, M.R., Riva, S., Biamonti, G., 2005. Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene. *Molecular Cell* 20, 881-890.
- Goguel, V., Wang, Y., Rosbash, M., 1993. Short artificial hairpins sequester splicing signals and inhibit yeast pre-mRNA splicing. *Molecular and Cellular Biology* 13, 6841-6848.
- Goodship, T.H., Liszewski, M.K., Kemp, E.J., Richards, A., Atkinson, J.P., 2004. Mutations in CD46, a complement regulatory protein, predispose to atypical HUS. *Trends in Molecular Medicine* 10, 226-231.
- Gorter, A., Blok, V.T., Haasnoot, W.H., Ensink, N.G., Daha, M.R., Fleuren, G.J., 1996. Expression of CD46, CD55, and CD59 on renal tumor cell lines and their role in preventing complement-mediated tumor cell lysis. *Laboratory Investigation* 74, 1039-1049.
- Gunderson, S.I., Polycarpou-Schwarz, M., Mattaj, I.W., 1998. U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Molecular Cell* 1, 255-264.
- Hamid, F.M., Makeyev, E.V., 2014. Emerging functions of alternative splicing coupled with nonsense-mediated decay. *Biochemical Society Transactions* 42, 1168-1173.

- Havens, M.A., Duelli, D.M., Hastings, M.L., 2013. Targeting RNA splicing for disease therapy. *Wiley interdisciplinary reviews. RNA* 4, 247-266.
- Heesters, B.A., Myers, R.C., Carroll, M.C., 2014. Follicular dendritic cells: dynamic antigen libraries. *Nature Reviews. Immunology* 14, 495-504.
- House, A.E., Lynch, K.W., 2008. Regulation of alternative splicing: more than just the ABCs. *The Journal of Biological Chemistry* 283, 1217-1221.
- Hovhannisyanyan, R.H., Carstens, R.P., 2005. A novel intronic cis element, ISE/ISS-3, regulates rat fibroblast growth factor receptor 2 splicing through activation of an upstream exon and repression of a downstream exon containing a noncanonical branch point sequence. *Molecular and Cellular Biology* 25, 250-263.
- Howard, J.M., Sanford, J.R., 2015. The RNAissance family: SR proteins as multifaceted regulators of gene expression. *Wiley Interdisciplinary Reviews-Rna* 6, 93-110.
- Hua, Y., Vickers, T.A., Okunola, H.L., Bennett, C.F., Krainer, A.R., 2008. Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice. *American Journal of Human Genetics* 82, 834-848.
- Huang, Y.Q., Gattoni, R., Stevenin, J., Steitz, J.A., 2003. SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Molecular cell* 11, 837-843.
- Huang, Y.Q., Yario, T.A., Steitz, J.A., 2004. A molecular link between SR protein dephosphorylation and mRNA export. *Proceedings of the National Academy of Sciences of the United States of America* 101, 9666-9670.
- Inoue, N., Ikawa, M., Nakanishi, T., Matsumoto, M., Nomura, M., Seya, T., Okabe, M., 2003. Disruption of mouse CD46 causes an accelerated spontaneous acrosome reaction in sperm. *Molecular and Cellular Biology* 23, 2614-2622.
- Ip, J.Y., Tong, A., Pan, Q., Topp, J.D., Blencowe, B.J., Lynch, K.W., 2007. Global analysis of alternative splicing during T-cell activation. *RNA* 13, 563-572.
- Izquierdo, J.M., Majos, N., Bonnal, S., Martinez, C., Castelo, R., Guigo, R., Bilbao, D., Valcarcel, J., 2005. Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. *Molecular Cell* 19, 475-484.
- Izquierdo, J.M., Valcarcel, J., 2007. Two isoforms of the T-cell intracellular antigen 1 (TIA-1) splicing factor display distinct splicing regulation activities. Control of TIA-1 isoform ratio by TIA-1-related protein. *The Journal of Biological Chemistry* 282, 19410-19417.
- Jabara, H.H., Angelini, F., Brodeur, S.R., Geha, R.S., 2011. Ligation of CD46 to CD40 inhibits CD40 signaling in B cells. *International Immunology* 23, 215-221.

- Johnstone, R.W., Russell, S.M., Loveland, B.E., McKenzie, I.F., 1993. Polymorphic expression of CD46 protein isoforms due to tissue-specific RNA splicing. *Molecular Immunology* 30, 1231-1241.
- Joubert, P.E., Meiffren, G., Gregoire, I.P., Pontini, G., Richetta, C., Flacher, M., Azocar, O., Vidalain, P.O., Vidal, M., Lotteau, V., Codogno, P., Roubourdin-Combe, C., Faure, M., 2009. Autophagy induction by the pathogen receptor CD46. *Cell Host & Microbe* 6, 354-366.
- Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., Dreyfuss, G., 2010. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468, 664-668.
- Källström, H., Blackmer Gill, D., Albiger, B., Liszewski, M.K., Atkinson, J.P., Jonsson, A.B., 2001. Attachment of *Neisseria gonorrhoeae* to the cellular pilus receptor CD46: identification of domains important for bacterial adherence. *Cellular Microbiology* 3, 133-143.
- Källström, H., Liszewski, M.K., Atkinson, J.P., Jonsson, A.-B., 1997. Membrane cofactor protein (MCP or CD46) is a cellular pilus receptor for pathogenic *Neisseria*. *Molecular Microbiology* 25, 639-647.
- Karni, R., de Stanchina, E., Lowe, S.W., Sinha, R., Mu, D., Krainer, A.R., 2007. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nature Structural & Molecular Biology* 14, 185-193.
- Karni, R., Hippo, Y., Lowe, S.W., Krainer, A.R., 2008. The splicing-factor oncoprotein SF2/ASF activates mTORC1. *Proceedings of the National Academy of Sciences of the United States of America* 105, 15323-15327.
- Kemper, C., Chan, A.C., Green, J.M., Brett, K.A., Murphy, K.M., Atkinson, J.P., 2003. Activation of human CD4⁺ cells with CD3 and CD46 induces a T-regulatory cell 1 phenotype. *Nature* 421, 388-392.
- Kemper, C., Leung, M., Stephensen, C.B., Pinkert, C.A., Liszewski, M.K., Cattaneo, R., Atkinson, J.P., 2001. Membrane cofactor protein (MCP; CD46) expression in transgenic mice. *Clinical and Experimental Immunology* 124, 180-189.
- Kemper, C., Verbsky, J.W., Price, J.D., Atkinson, J.P., 2005. T-cell stimulation and regulation: with complements from CD46. *Immunologic Research* 32, 31-43.
- Keren, H., Lev-Maor, G., Ast, G., 2010. Alternative splicing and evolution: diversification, exon definition and function. *Nature Reviews. Genetics* 11, 345-355.
- Kino, Y., Washizu, C., Oma, Y., Onishi, H., Nezu, Y., Sasagawa, N., Nukina, N., Ishiura, S., 2009. MBNL and CELF proteins regulate alternative splicing of the skeletal muscle chloride channel CLCN1. *Nucleic Acids Research* 37, 6477-6490.

- Kole, R., Krainer, A.R., Altman, S., 2012. RNA therapeutics: beyond RNA interference and antisense oligonucleotides. *Nature Reviews. Drug Discovery* 11, 125-140.
- Koo, T., Wood, M.J., 2013. Clinical trials using antisense oligonucleotides in duchenne muscular dystrophy. *Human Gene Therapy* 24, 479-488.
- Kornblihtt, A.R., De la Mata, M., Fededa, J.P., Munoz, M.J., Nogues, G., 2004. Multiple links between transcription and splicing. *RNA* 10, 1489-1498.
- Kornblihtt, A.R., Schor, I.E., Allo, M., Dujardin, G., Petrillo, E., Munoz, M.J., 2013. Alternative splicing: a pivotal step between eukaryotic transcription and translation (vol 14, pg 153, 2013). *Nature Reviews. Molecular Cell Biology* 14.
- Krainer, A.R., Conway, G.C., Kozak, D., 1990a. The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. *Cell* 62, 35-42.
- Krainer, A.R., Conway, G.C., Kozak, D., 1990b. Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes & development* 4, 1158-1171.
- Kurita-Taniguchi, M., Fukui, A., Hazeki, K., Hirano, A., Tsuji, S., Matsumoto, M., Watanabe, M., Ueda, S., Seya, T., 2000. Functional modulation of human macrophages through CD46 (measles virus receptor): Production of IL-12 p40 and nitric oxide in association with recruitment of protein-tyrosine phosphatase SHP-1 to CD46. *The Journal of Immunology* 165, 5143-5152.
- Kyburz, A., Friedlein, A., Langen, H., Keller, W., 2006. Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing. *Molecular Cell* 23, 195-205.
- Ladd, A.N., Charlet-B, N., Cooper, T.A., 2001. The CELF Family of RNA Binding Proteins Is Implicated in Cell-Specific and Developmentally Regulated Alternative Splicing. *Molecular and Cellular Biology* 21, 1285-1296.
- Ladd, A.N., Stenberg, M.G., Swanson, M.S., Cooper, T.A., 2005. Dynamic balance between activation and repression regulates pre-mRNA alternative splicing during heart development. *Developmental Dynamics* 233, 783-793.
- Lai, M.C., Lin, R.I., Tarn, W.Y., 2001. Transportin-SR2 mediates nuclear import of phosphorylated SR proteins. *Proceedings of the National Academy of Sciences of the United States of America* 98, 10154-10159.
- Lai, M.C., Tarn, W.Y., 2004. Hypophosphorylated ASF/SF2 binds TAP and is present in messenger ribonucleoproteins. *The Journal of biological chemistry* 279, 31745-31749.

Le Friec, G., Cardone, J., Cope, A., Kemper, C., 2011. CD3/CD46-mediated generation of IL-10-secreting T cells is defective in rheumatoid arthritis. *Annals of the Rheumatic Diseases* 70, A48.

Le Friec, G., Sheppard, D., Whiteman, P., Karsten, C.M., Shamoun, S.A., Laing, A., Bugeon, L., Dallman, M.J., Melchionna, T., Chillakuri, C., Smith, R.A., Drouet, C., Couzi, L., Fremeaux-Bacchi, V., Kohl, J., Waddington, S.N., McDonnell, J.M., Baker, A., Handford, P.A., Lea, S.M., Kemper, C., 2012. The CD46-Jagged1 interaction is critical for human TH1 immunity. *Nature Immunology* 13, 1213-1221.

Le Guiner, C., Lejeune, F., Galiana, D., Kister, L., Breathnach, R., Stevenin, J., Del Gatto-Konczak, F., 2001. TIA-1 and TIAR activate splicing of alternative exons with weak 5' splice sites followed by a U-rich stretch on their own pre-mRNAs. *The Journal of Biological Chemistry* 276, 40638-40646.

Lee, J.E., Cooper, T.A., 2009. Pathogenic mechanisms of myotonic dystrophy. *Biochemical Society Transactions* 37, 1281-1286.

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., Carthew, R.W., 2004. Distinct Roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA Silencing Pathways. *Cell* 117, 69-81.

Lewis, B.P., Green, R.E., 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 of the United States of America* 100, 189-192.

Lewis, J.D., Izaurralde, E., Jarmolowski, A., McGuigan, C., Mattaj, I.W., 1996. A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes & Development* 10, 1683-1698.

Li, X., Manley, J.L., 2005. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* 122, 365-378.

Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., Darnell, J.C., Darnell, R.B., 2008. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 456, 464-469.

Lisbin, M.J., Qiu, J., White, K., 2001. The neuron-specific RNA-binding protein ELAV regulates neuroglial alternative splicing in neurons and binds directly to its pre-mRNA. *Genes & Development* 15, 2546-2561.

Liszewski, M.K., Atkinson, J.P., 1996. Membrane cofactor protein (MCP; CD46). Isoforms differ in protection against the classical pathway of complement. *The Journal of Immunology* 156, 4415-4421.

Liszewski, M.K., Kemper, C., Price, J.D., Atkinson, J.P., 2005. Emerging roles and new functions of CD46. *Springer Seminars in Immunopathology* 27, 345-358.

Liszewski, M.K., Tedja, I., Atkinson, J.P., 1994. Membrane cofactor protein (CD46) of complement. Processing differences related to alternatively spliced cytoplasmic domains. *The Journal of Biological Chemistry* 269, 10776-10779.

Liu, M., Yang, Y.J., Zheng, H., Zhong, X.R., Wang, Y., Wang, Z., Wang, Y.G., Wang, Y.P., 2014. Membrane-bound complement regulatory proteins are prognostic factors of operable breast cancer treated with adjuvant trastuzumab: a retrospective study. *Oncology Reports* 32, 2619-2627.

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., Smith, C.W., 2010. Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB. *Nature Structural & Molecular Biology* 17, 1114-1123.

Lorson, C.L., Hahnen, E., Androphy, E.J., Wirth, B., 1999. A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proceedings of the National Academy of Sciences of the United States of America* 96, 6307-6311.

Lovci, M.T., Ghanem, D., Marr, H., Arnold, J., Gee, S., Parra, M., Liang, T.Y., Stark, T.J., Gehman, L.T., Hoon, S., Massirer, K.B., Pratt, G.A., Black, D.L., Gray, J.W., Conboy, J.G., Yeo, G.W., 2013. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nature Structural & Molecular Biology* 20, 1434-1442.

Lu, Z.X., Jiang, P., Xing, Y., 2012. Genetic variation of pre-mRNA alternative splicing in human populations. *Wiley Interdisciplinary Reviews. RNA* 3, 581-592.

Luckheeram, R.V., Zhou, R., Verma, A.D., Xia, B., 2012. CD4(+)T cells: differentiation and functions. *Clinical & Developmental Immunology* 2012, 925135.

Luco, R.F., Allo, M., Schor, I.E., Kornblihtt, A.R., Misteli, T., 2011. Epigenetics in alternative pre-mRNA splicing. *Cell* 144, 16-26.

Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., Misteli, T., 2010. Regulation of alternative splicing by histone modifications. *Science* 327, 996-1000.

Ludford-Menting, M.J., Thomas, S.J., Crimeen, B., Harris, L.J., Loveland, B.E., Bills, M., Ellis, S., Russell, S.M., 2002. A functional interaction between CD46 and DLG4 - A role for DLG4 in epithelial polarization. *The Journal of Biological Chemistry* 277, 4477-4484.

- Luo, M.L., Zhou, Z., Magni, K., Christoforides, C., Rappsilber, J., Mann, M., Reed, R., 2001. Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature* 413, 644-647.
- Lykke-Andersen, S., Jensen, T.H., 2015. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nature reviews. Molecular cell biology* 16, 665-677.
- Lynch, K.W., Weiss, A., 2000. A model system for activation-induced alternative splicing of CD45 pre-mRNA in T cells implicates protein kinase C and Ras. *Molecular and Cellular Biology* 20, 70-80.
- Mamidi, S., Hone, S., Teufel, C., Sellner, L., Zenz, T., Kirschfink, M., 2015. Neutralization of membrane complement regulators improves complement-dependent effector functions of therapeutic anticancer antibodies targeting leukemic cells. *Oncoimmunology* 4, e979688.
- Maquat, L.E., 2004. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nature reviews. Molecular cell biology* 5, 89-99.
- Marie, J.C., Astier, A.L., Rivaller, P., Rabourdin-Combe, C., Wild, T.F., Horvat, B., 2002. Linking innate and acquired immunity: divergent role of CD46 cytoplasmic domains in T cell induced inflammation. *Nature Immunology* 3, 659-666.
- Martinez, N.M., Lynch, K.W., 2013. Control of alternative splicing in immune responses: many regulators, many predictions, much still to learn. *Immunological Reviews* 253, 216-236.
- Martinez, N.M., Pan, Q., Cole, B.S., Yarosh, C.A., Babcock, G.A., Heyd, F., Zhu, W., Ajith, S., Blencowe, B.J., Lynch, K.W., 2012. Alternative splicing networks regulated by signaling in human T cells. *RNA* 18, 1029-1040.
- Martinson, H.G., 2011. An active role for splicing in 3'-end formation. *Wiley Interdisciplinary Reviews. RNA* 2, 459-470.
- Matera, A.G., Wang, Z., 2014. A day in the life of the spliceosome. *Nature Reviews. Molecular Cell Biology* 15, 108-121.
- Matter, N., Herrlich, P., Konig, H., 2002. Signal-dependent regulation of splicing via phosphorylation of Sam68. *Nature* 420, 691-695.
- McGlinchy, N.J., Smith, C.W., 2008. Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends in Biochemical Sciences* 33, 385-393.
- McLaughlin, B.J., Fan, W., Zheng, J.J., Cai, H., Del Priore, L.V., Bora, N.S., Kaplan, H.J., 2003. Novel role for a complement regulatory protein (CD46) in retinal pigment epithelial adhesion. *Investigative Ophthalmology & Visual Science* 44, 3669-3674.

- Melton, A.A., Jackson, J., Wang, J., Lynch, K.W., 2007. Combinatorial control of signal-induced exon repression by hnRNP L and PSF. *Molecular and Cellular Biology* 27, 6972-6984.
- Michlewski, G., Sanford, J.R., Caceres, J.F., 2008. The splicing factor SF2/ASF regulates translation initiation by enhancing phosphorylation of 4E-BP1. *Molecular cell* 30, 179-189.
- Misteli, T., Caceres, J.F., Clement, J.Q., Krainer, A.R., Wilkinson, M.F., Spector, D.L., 1998. Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. *The Journal of cell biology* 143, 297-307.
- Mueller, W.F., Hertel, K.J., 2012. RNA Elements Involved in Splicing. In: Stamm, S., Smith, C.W.J., Lüthmann, R. (Eds.), *Alternative pre-mRNA Splicing: Theory and Protocols*. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 21-31.
- Naradikian, M., Scholz, J., Oropallo, M., Cancro, M., 2014. Understanding B Cell Biology. In: Bosch, X., Ramos-Casals, M., Khamashta, M.A. (Eds.), *Drugs Targeting B-Cells in Autoimmune Diseases*. Springer Basel, pp. 11-35.
- Ni Choileain, S., Astier, A.L., 2011. CD46 plasticity and its inflammatory bias in multiple sclerosis. *Archivum Immunologiae et Therapiae Experimentalis* 59, 49-59.
- Ni Choileain, S., Astier, A.L., 2012. CD46 processing: a means of expression. *Immunobiology* 217, 169-175.
- Ni Choileain, S., Weyand, N.J., Neumann, C., Thomas, J., So, M., Astier, A.L., 2011. The dynamic processing of CD46 intracellular domains provides a molecular rheostat for T cell activation. *PLOS One* 6, e16287.
- Nilsen, T.W., Graveley, B.R., 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463, 457-463.
- Noris, M., Remuzzi, G., 2009. Atypical hemolytic-uremic syndrome. *The New England Journal of Medicine* 361, 1676-1687.
- Novosel, A., Borkhardt, A., 2008. Antisense Oligonucleotides. In: Offermanns, S., Rosenthal, W. (Eds.), *Encyclopedia of Molecular Pharmacology*. Springer Berlin Heidelberg, pp. 185-189.
- Oberdoerffer, S., Moita, L.F., Neems, D., Freitas, R.P., Hacohen, N., Rao, A., 2008. Regulation of CD45 alternative splicing by heterogeneous ribonucleoprotein, hnRNPLL. *Science* 321, 686-691.
- Oikonomopoulou, K., Reis, E.S., Lambris, J.D., 2001. *Complement System and Its Role in Immune Responses*. eLS. John Wiley & Sons, Ltd.
- Okunola, H.L., Krainer, A.R., 2009. Cooperative-Binding and Splicing-Repressive Properties of hnRNP A1. *Molecular and Cellular Biology* 29, 5620-5631.

- O'Mullane, L., Eperon, I.C., 1998. The pre-mRNA 5' cap determines whether U6 small nuclear RNA succeeds U1 small nuclear ribonucleoprotein particle at 5' splice sites. *Molecular and Cellular Biology* 18, 7510-7520.
- Pandit, S., Zhou, Y., Shiue, L., Coutinho-Mansfield, G., Li, H., Qiu, J., Huang, J., Yeo, G.W., Ares, M., Jr., Fu, X.D., 2013. Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Molecular Cell* 50, 223-235.
- Patel, M.R., Jacobson, B.A., Belgum, H., Raza, A., Sadiq, A., Drees, J., Wang, H., Jay-Dixon, J., Etchison, R., Federspiel, M.J., Russell, S.J., Kratzke, R.A., 2014. Measles vaccine strains for virotherapy of non-small-cell lung carcinoma. *Journal of Thoracic Oncology* 9, 1101-1110.
- Paz, I., Akerman, M., Dror, I., Kosti, I., Mandel-Gutfreund, Y., 2010. SFmap: a web server for motif analysis and prediction of splicing factor binding sites. *Nucleic Acids Research* 38, W281-285.
- Pham, J.W., Pellino, J.L., Lee, Y.S., Carthew, R.W., Sontheimer, E.J., 2004. A Dicer-2-Dependent 80S Complex Cleaves Targeted mRNAs during RNAi in *Drosophila*. *Cell* 117, 83-94.
- Post, T.W., Liszewski, M.K., Adams, E.M., Tedja, I., Miller, E.A., Atkinson, J.P., 1991. Membrane cofactor protein of the complement system: alternative splicing of serine/threonine/proline-rich exons and cytoplasmic tails produces multiple isoforms that correlate with protein phenotype. *The Journal of Experimental Medicine* 174, 93-102.
- Prasad, J., Colwill, K., Pawson, T., Manley, J.L., 1999. The protein kinase Clk/Sty directly modulates SR protein activity: Both hyper- and hypophosphorylation inhibit splicing. *Molecular and cellular biology* 19, 6991-7000.
- Price, J.D., Schaumburg, J., Sandin, C., Atkinson, J.P., Lindahl, G., Kemper, C., 2005. Induction of a regulatory phenotype in human CD4+ T cells by streptococcal M protein. *The Journal of Immunology* 175, 677-684.
- Purcell, D.F., Russell, S.M., Deacon, N.J., Brown, M.A., Hooker, D.J., McKenzie, I.F., 1991. Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a regulator of complement activation. *Immunogenetics* 33, 335-344.
- Reese, M.G., Eeckman, F.H., Kulp, D., Haussler, D., 1997. Improved splice site detection in Genie. *Journal of Computational Biology* 4, 311-323.
- Riley, R.C., Kemper, C., Leung, M., Atkinson, J.P., 2002. Characterization of human membrane cofactor protein (MCP; CD46) on spermatozoa. *Molecular Reproduction and Development* 62, 534-546.

- Riley-Vargas, R.C., Gill, D.B., Kemper, C., Liszewski, M.K., Atkinson, J.P., 2004. CD46: expanding beyond complement regulation. *Trends in Immunology* 25, 496-503.
- Roca, X., Akerman, M., Gaus, H., Berdeja, A., Bennett, C.F., Krainer, A.R., 2012. Widespread recognition of 5' splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. *Genes & Development* 26, 1098-1109.
- Roca, X., Krainer, A.R., 2009. Recognition of atypical 5' splice sites by shifted base-pairing to U1 snRNA. *Nature Structural & Molecular Biology* 16, 176-182.
- Roca, X., Krainer, A.R., Eperon, I.C., 2013. Pick one, but be quick: 5' splice sites and the problems of too many choices. *Genes & Development* 27, 129-144.
- Rooke, N., Markovtsov, V., Cagavi, E., Black, D.L., 2003. Roles for SR proteins and hnRNP A1 in the regulation of c-src exon N1. *Molecular and Cellular Biology* 23, 1874-1884.
- Roszbach, O., Hung, L.H., Schreiner, S., Grishina, I., Heiner, M., Hui, J., Bindereif, A., 2009. Auto- and cross-regulation of the hnRNP L proteins by alternative splicing. *Molecular and Cellular Biology* 29, 1442-1451.
- Rote, N.S., 2013. Adaptive Immunity. In: Huether, S.E.M., Kathryn L. (Ed.), *Understanding Pathophysiology*, 5th Edition, pp. 143-165.
- Rothrock, C.R., House, A.E., Lynch, K.W., 2005. HnRNP L represses exon splicing via a regulated exonic splicing silencer. *The EMBO Journal* 24, 2792-2802.
- Ruggiu, M., McGovern, V.L., Lotti, F., Saieva, L., Li, D.K., Kariya, S., Monani, U.R., Burghes, A.H., Pellizzoni, L., 2012. A role for SMN exon 7 splicing in the selective vulnerability of motor neurons in spinal muscular atrophy. *Molecular and Cellular Biology* 32, 126-138.
- Russell, S., 2004. CD46: a complement regulator and pathogen receptor that mediates links between innate and acquired immune function. *Tissue Antigens* 64, 111-118.
- Russell, S.M., Sparrow, R.L., Mckenzie, I.F.C., Purcell, D.F.J., 1992. Tissue-Specific and Allelic Expression of the Complement Regulator Cd46 Is Controlled by Alternative Splicing. *European Journal of Immunology* 22, 1513-1518.
- Sanford, J.R., Coutinho, P., Hackett, J.A., Wang, X., Ranahan, W., Caceres, J.F., 2008. Identification of nuclear and cytoplasmic mRNA targets for the shuttling protein SF2/ASF. *PloS one* 3, e3369.
- Sanford, J.R., Ellis, J.D., Cazalla, D., Caceres, J.F., 2005. Reversible phosphorylation differentially affects nuclear and cytoplasmic functions of splicing factor 2/alternative splicing factor. *Proceedings of the National Academy of Sciences of the United States of America* 102, 15042-15047.

- Sanford, J.R., Gray, N.K., Beckmann, K., Caceres, J.F., 2004. A novel role for shuttling SR proteins in mRNA translation. *Genes & development* 18, 755-768.
- Sarma, J.V., Ward, P.A., 2011. The complement system. *Cell and Tissue Research* 343, 227-235.
- Schor, I., Gómez Acuña, L., Kornblihtt, A., 2013. Coupling Between Transcription and Alternative Splicing. In: Wu, J.Y. (Ed.), *RNA and Cancer*. Springer Berlin Heidelberg, pp. 1-24.
- Schwartz, S., Meshorer, E., Ast, G., 2009. Chromatin organization marks exon-intron structure. *Nature Structural & Molecular Biology* 16, 990-995.
- Schweingruber, C., Rufener, S.C., Zund, D., Yamashita, A., Muhlemann, O., 2013. Nonsense-mediated mRNA decay - mechanisms of substrate mRNA recognition and degradation in mammalian cells. *Biochimica et Biophysica Acta* 1829, 612-623.
- Seya, T., Atkinson, J.P., 1989. Functional properties of membrane cofactor protein of complement. *Biochemical Journal* 264, 581-588.
- Sharma, S., Falick, A.M., Black, D.L., 2005. Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E complex. *Molecular Cell* 19, 485-496.
- Sheth, N., Roca, X., Hastings, M.L., Roeder, T., Krainer, A.R., Sachidanandam, R., 2006. Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Research* 34, 3955-3967.
- Shipkova, M., Wieland, E., 2012. Surface markers of lymphocyte activation and markers of cell proliferation. *Clinica Chimica Acta* 413, 1338-1349.
- Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., Oberdoerffer, S., 2011. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479, 74-79.
- Simms, P.E., Ellis, T.M., 1996. Utility of flow cytometric detection of CD69 expression as a rapid method for determining poly- and oligoclonal lymphocyte activation. *Clinical and Diagnostic Laboratory Immunology* 3, 301-304.
- Singh, N.N., Seo, J., Ottesen, E.W., Shishimorova, M., Bhattacharya, D., Singh, R.N., 2011. TIA1 prevents skipping of a critical exon associated with spinal muscular atrophy. *Molecular and Cellular Biology* 31, 935-954.
- Smith, A., Santoro, F., Di Lullo, G., Dagna, L., Verani, A., Lusso, P., 2003. Selective suppression of IL-12 production by human herpesvirus 6. *Blood* 102, 2877-2884.

- Smith, P.J., Zhang, C., Wang, J., Chew, S.L., Zhang, M.Q., Krainer, A.R., 2006. An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Human molecular genetics* 15, 2490-2508.
- Smith-Garvin, J.E., Koretzky, G.A., Jordan, M.S., 2009. T cell activation. *Annual Review of Immunology* 27, 591-619.
- Stamm, S., 2008. Regulation of alternative splicing by reversible protein phosphorylation. *The Journal of Biological Chemistry* 283, 1223-1227.
- Sun, Q., Mayeda, A., Hampson, R.K., Krainer, A.R., Rottman, F.M., 1993. General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. *Genes & Development* 7, 2598-2608.
- Sun, S., Zhang, Z., Fregoso, O., Krainer, A.R., 2012. Mechanisms of activation and repression by the alternative splicing factors RBFOX1/2. *RNA* 18, 274-283.
- Sun, S., Zhang, Z., Sinha, R., Karni, R., Krainer, A.R., 2010. SF2/ASF autoregulation involves multiple layers of post-transcriptional and translational control. *Nature Structural & Molecular Biology* 17, 306-312.
- Testi, R., Phillips, J.H., Lanier, L.L., 1989. T cell activation via Leu-23 (CD69). *The Journal of Immunology* 143, 1123-1128.
- Thomas, C.P., Raikwar, N.S., Kelley, E.A., Liu, K.Z., 2010. Alternate processing of Flt1 transcripts is directed by conserved cis-elements within an intronic region of FLT1 that reciprocally regulates splicing and polyadenylation. *Nucleic Acids Research* 38, 5130-5140.
- Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcarcel, J., Guigo, R., 2009. Nucleosome positioning as a determinant of exon recognition. *Nature Structural & Molecular Biology* 16, 996-1001.
- Tobon, G.J., Izquierdo, J.H., Canas, C.A., 2013. B lymphocytes: development, tolerance, and their role in autoimmunity-focus on systemic lupus erythematosus. *Autoimmune Diseases* 2013, 827254.
- Tong, A., Nguyen, J., Lynch, K.W., 2005. Differential expression of CD45 isoforms is controlled by the combined activity of basal and inducible splicing-regulatory elements in each of the variable exons. *The Journal of Biological Chemistry* 280, 38297-38304.
- Truedsson, L., Bengtsson, A.A., Sturfelt, G., 2007. Complement deficiencies and systemic lupus erythematosus. *Autoimmunity* 40, 560-566.
- Tsai, Y.G., Niu, D.M., Yang, K.D., Hung, C.H., Yeh, Y.J., Lee, C.Y., Lin, C.Y., 2012. Functional defects of CD46-induced regulatory T cells to suppress airway inflammation in mite allergic asthma. *Laboratory Investigation* 92, 1260-1269.

- Tsujimura, A., Shida, K., Kitamura, M., Nomura, M., Takeda, J., Tanaka, H., Matsumoto, M., Matsumiya, K., Okuyama, A., Nishimune, Y., Okabe, M., Seya, T., 1998. Molecular cloning of a murine homologue of membrane cofactor protein (CD46): preferential expression in testicular germ cells. *Biochemical Journal* 330 (Pt 1), 163-168.
- Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B.J., Darnell, R.B., 2006. An RNA map predicting Nova-dependent splicing regulation. *Nature* 444, 580-586.
- Vaknin-Dembinsky, A., Murugaiyan, G., Hafler, D.A., Astier, A.L., Weiner, H.L., 2008. Increased IL-23 secretion and altered chemokine production by dendritic cells upon CD46 activation in patients with multiple sclerosis. *Journal of Neuroimmunology* 195, 140-145.
- Valacca, C., Bonomi, S., Buratti, E., Pedrotti, S., Baralle, F.E., Sette, C., Ghigna, C., Biamonti, G., 2010. Sam68 regulates EMT through alternative splicing-activated nonsense-mediated mRNA decay of the SF2/ASF proto-oncogene. *Journal of Cell Biology* 191, 87-99.
- Wagner, E.J., Garcia-Blanco, M.A., 2001. Polypyrimidine tract binding protein antagonizes exon definition. *Molecular and Cellular Biology* 21, 3281-3288.
- Wahl, M.C., Will, C.L., Luhrmann, R., 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701-718.
- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., Burge, C.B., 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470-476.
- Wang, G.S., Cooper, T.A., 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nature Reviews. Genetics* 8, 749-761.
- Wang, Z., Burge, C.B., 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* 14, 802-813.
- Wu, H., Sun, S., Tu, K., Gao, Y., Xie, B., Krainer, A.R., Zhu, J., 2010. A splicing-independent function of SF2/ASF in microRNA processing. *Molecular cell* 38, 67-77.
- Wu, J.Y., Maniatis, T., 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* 75, 1061-1070.
- Xiong, H.Y., Alipanahi, B., Lee, L.J., Bretschneider, H., Merico, D., Yuen, R.K., Hua, Y., Gueroussov, S., Najafabadi, H.S., Hughes, T.R., Morris, Q., Barash, Y., Krainer, A.R., Jovic, N., Scherer, S.W., Blencowe, B.J., Frey, B.J., 2015. RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science* 347, 1254806.

Xu, Y.Q., Gao, Y.D., Yang, J., Guo, W., 2010. A defect of CD4+CD25+ regulatory T cells in inducing interleukin-10 production from CD4+ T cells under CD46 costimulation in asthma patients. *The Journal of Asthma* 47, 367-373.

Yamamoto, H., Fara, A.F., Dasgupta, P., Kemper, C., 2013. CD46: The 'multitasker' of complement proteins. *The International Journal of Biochemistry & Cell Biology* 45, 2808-2820.

Yan, J., Allendorf, D.J., Li, B., Yan, R., Hansen, R., Donev, R., 2008. The role of membrane complement regulatory proteins in cancer immunotherapy. *Advances in Experimental Medicine and Biology* 632, 159-174.

Yeakley, J.M., Tronchere, H., Olesen, J., Dyck, J.A., Wang, H.Y., Fu, X.D., 1999. Phosphorylation regulates in vivo interaction and molecular targeting of serine/arginine-rich pre-mRNA splicing factors. *The Journal of cell biology* 145, 447-455.

Yeo, G., Burge, C.B., 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Journal of Computational Biology* 11, 377-394.

Zaffran, Y., Destaing, O., Roux, A., Ory, S., Nheu, T., Jurdic, P., Roubourdin-Combe, C., Astier, A.L., 2001. CD46/CD3 costimulation induces morphological changes of human T cells and activation of Vav, Rac, and extracellular signal-regulated kinase mitogen-activated protein kinase. *The Journal of Immunology* 167, 6780-6785.

Zahler, A.M., 2005. Alternative splicing in *C. elegans*. *WormBook : The Online Review of C. elegans Biology*, 1-13.

Zhang, Z., Krainer, A.R., 2004. Involvement of SR proteins in mRNA surveillance. *Molecular cell* 16, 597-607.

Zhou, H.L., Luo, G., Wise, J.A., Lou, H., 2014. Regulation of alternative splicing by local histone modifications: potential roles for RNA-guided mechanisms. *Nucleic Acids Research* 42, 701-713.

Zhu, J., Krainer, A.R., 2000. Pre-mRNA splicing in the absence of an SR protein RS domain. *Genes & Development* 14, 3166-3178.

Zipfel, P.F., Skerka, C., 2009. Complement regulators and inhibitory proteins. *Nature Reviews. Immunology* 9, 729-740.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* 31, 3406-3415.

Appendix I

CD46 exon 6→ exon 10 genomic sequence

Uppercase: Exon

Lowercase: Intron

Exon 6 Exon 7 Exon 8 Exon 9 Exon 10

Gray colored sequence is not included in minigenes

Dotted underlined sequence is not included in minigene *CD46* exon 6-7-10

Cryptic 3'ss

Cryptic exon

Cryptic 5'ss

```

TGGTCAAATGTCGATTCCAGTAGTCGAAAATGGAAAACAGATATCAGGATTGGAAAAAATTTACTACAAAGCAACAGTT
ATGTTTGAATGCGATAAGGGTTTTTACCTCGATGGCAGCGACACAATTTGCTGTGACAGTAACAGTACTTGGGATCCCCAGT
TCCAAGTGCTTAaaggtacaaaggttatccttttctgtccttgggtttgtattggttggctgttcattttagactttatt
tctttgatattaactatcagtcatacaaaaataactgaaaagaacaatttttagtatttaactctgtcctgtattcatttctat
gccagatgaatgacacgaaattcacataaaaattctgctgtgtgtattttttgtgctttccagggttcttagcacgttatgta
cattgcatgggtatagcttttaataatTTTTATGTATAAAAAGTGAATTACAACAACCTTTGGAAATGAAACATGGGCATT
Ttactctaagtaagtcacaatggcataaattcataataaataaagaaatgagagcaataactccaagtggttgatccttaacatta
tttggtttctagTGGTGCCTCCATCTAGTACAAAACCTCCAGCTTTGAGTCATTTCAGgtttagtagcttcttctctatatgt
cttcttcttatatgttacaagatataaggaattcctgggaaattggcagcaataactccaagtggttgggtccaatctaca
tattatttttggtttccagTGTGACTTCTTCCACTACAAAATCTCCAGCGTCCAGTGCCTCAGgtttagtaatttctctgctt
atagtttttcaaaaatcctttaaattcctgggtgattttttataaaaatccttcaaatcttgggtgagttcacttatttttagt
aatgaaaggaggaggacatttgtatagccatttttagttgtatacatagtgattctaaaatatttggcataatctgatata
aacaggaatatgacctgggaaagttgtataaaatttgaatggtttatatatatatggttacattatatatagagaaattga
tttagtaccttcttatgggtattagtaaggtataaagtggtgacatttaaagcacttagcacagtgccgtggttagagaaa
gctcttagagctagccaatatgacttatcctacacttaaatatgttagggaaggggaaggttcatgtagtttgttagtat
gtgtgtgtgtaaatgtaatatgtttgtttgtaaatatgtagtattttgtgtcatcaactttgctatttttcttccaga
agtgatctgttgacaccttactgagtttgttcttattgacagactcttctatttctggatgcaactagtttagtctctt
gatgtaaaacatttaaacagggaaatttctgctgtcctcagaacaagatctgtatttctgctcttccctaccaccctctt
ccacacctcataatgttatttttttctcttagtgggcagttttatctggcaatagcaactcaattttatagcaactg
aaagcaggaaaagtccttatttactgaaaataaaatagaagacttttgacgggactcctggtcaaccactgtcttagctg
gagagattattttgatgatggaaggaattatcctgatctcagcactaccatctctggtttcactttgctccagcagaagc
aagagacaatagtaattgtgggtagcactcagctctgggagactgtgcttagttaacatcatttctgtggtctcagtttcc
cttttagtataatgtgagtggtgaactgaaatctcaatcactggcttcaatttgggtatatatttcttactcctaatattt
taaggtctgaagagaattatgcatgttaactagtaactctttaggctgggtgggtgggtgagcagcagcagcagcagcagc
ataccagaactttgggaggccgaggtgggtggatcacgaggtcagaagtcagaaccagcctgggtaagtgggtgaaaccca
tctctactaaaaatcaaaaaatagcggggtgctgtggcaggaacttgtaatccagctgctcgggaggccgagcagagaa
tcgcttgaaccagctgggcagaggttgcagtgagccaagatcatgccactgcactccagcctgggtgacagagtgagactctg
tctcaaaaacaaaaccttctataaattcctgtaagagcttggtaactggtagcctcctgagtgagtgagtgagtgagtgag
ggcaggaagagatgagtttttagtgactgatacatgagcaaaacaagattgcatgtgaaattatactcagtgaggcatcc
gggttaactcacatgctcctagtggttatagcactaactacgtttgttgatcacagctctttgaaagagttgatgataag
catttcataaagttattataaaactgggtatagcttagtggtgagagcagaagagagaattgctgggattgtcaatgat
tcttatcctgctagaggaagactaaaatacaaacatcatgagtttaaggtagatcctaaatggtttaacaagaagaagttg
aggaagctcaagtggaagagaagatgtaaatatgtagtatcactacttacagtttttaaaaagatacatatgttta
ccccagcacatcccttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
tcaactgcaacctccgctcccggttcaagcgggttctcctgccacctcagactccaagtagctgggtattacaggcaccacc
actatgcccggctaattttttgtatttttagtagagacggggtttctccatgtaggctggctagctcgaactcctgacctca
agtgatccactgctcggcctcccacagtgctgggattacaggcgtgagccaccacaccatcctcacattactttcatgat
ggttatcttctaagatttccattataaaaatcacagagtggtgagcattctgtacatacatcttacatactgttgcaaatat
atatgtagggttaatttcaagatgtggaattgcaaaattgttaaggtgtgagttttcatttgataaaactaaaaaatcaata
caaaaataaagctcaaattttatattgataaagccctgtgtaatttataaaaatcaaaacttattttctagTGGTGCCTCC
ATCTAGTACAAAACCTCCAGCTTTGAGTCATTTCAGgttgggttaactcttatacctactgatattggttaagaattttatttttt
tattatactttaaagcttagggatgtgtgacagcgtgaggtttgttacataggtatagctgtgccatggtgtgtgtgtgc
accatcaactcatcatttatattaggtatttctcctaagtctatccctccccagccctaaccctgacaggccccagctgt
gtgatgttccccaccctgtgtccatgtgttctcgtttgtaactcctacatagagtgagaacatgaggtgtgtgtgtgtgtgt
tcttgtgatagtttgattagaatgatggtttccagcttccatgctcctgcaaggacatgaactcatcctttttttatgg
ctgcatagttatccatggtgtatattgtgccacatttcttaatccagctctatcattgatggacattcgggtgtgttccaagtc
ttgtctattgtgaatagtgccacaataaacatacgtgtgcatgtgtctttatagtagcatgatttataatcctttgggtat
accagtgtaagggattactgggtcaaatggtatttctagttgtagactcctcaggaatcaccacactgcttcccaaatggtt
aactaatttacagctcccaccaacagtgtaaaaggttctcgttttctccacatcctcctcagcatctgtgttctcgtactttt
aataatcgcaatcctaactggcgtgagtggtatctcattgtggttttgattgcatcttctctgatgaccagtgatgatgaga

```

tttttcatatgtctgttggtgtataaaatgtctcttttggagaagtgctgttcatatcctttgccactttttgatgggtg
tgttttttcttgaatattttaaagttctttgtagattctggatattagccctttgtcagatgggcagatgcaaaagattt
tccccattctgacgttgctgtcactctgatgatagttcttttgcctgtgcagaagctctttagttcattagatcccat
ttgtctattttggcttttgggtgcttttgggttttagttatgaagtatttggcctgacctatgtcctgaaagttat
gcctaggttttctgttggtttttaggtgttaggtcttacattaaagcttccaatccatcttgagttaaattttgtataag
gtgtgaggaagggatccagttcagctttctacatcaggttagtcagttttcccagcaccgtttatataatagggaaatccct
ccccattctgtttttgtcaggtttgtcaaaagatcagatgggtttagatgtgagttgttatttcaagcctctattctgtt
ccattgatctatatactgttttgggtaccagttaccatgtctgttttgggtcactgtagcctttagtatgatttgaagttcagga
gctttagatgtctccagctttgttcttttggcttaggttgtcttggctatataggactctttttgggttccatatagaactttaa
gtagtttttccaattctgtgaagaaattcagtggttagctttagatgtgtagcattgaaatctataaaattaccttgggttagtat
ggcattttccaatattgattcttccatccatgagcatggaatgttcttccattgtttgtgtaactctttatttctgttga
gcagtgggttgcacttgtcttgaagatgcttccacatccatccttgaagttggattcctaggtattttattctcttgt
agcaattgtgaaatgggaatcactcattatttggctctctgtttgtctattattgggtgataggaatgcttgtgatttttga
cattgattttgtatctgagacttctgctgaagttgtttatcagcttgtggagatttgggtgagacgattgggttttctaaa
tatgcaatcatgtcactctgcaaacagagacaattgacttcccttttccaaatgaataaccctttatttcttttcttgcct
gatccactggccagaactccaacacttgaatagagtggtgagagagggcatcctgtctgtgcccgttttccaaag
ggaatgtccagttttgcccattcgggtatgatatggctgtgggtttgtcttaaatagctcttattattttgagatataatcc
atcaatacctagtttattgagagtttttagcatgaagggctgttgaattttgtcaaggccttttctgcatctattgaaatag
tcatgtggttttgtttagttctgttttgtgatggatgatttttagattgttaattgtgaaccagccttgatccca
gggtagaacctgaacttgtcttgaagatgcttccacatccatccttgaagttggattcctaggtattttattctcttgt
tgcactgagttctcagggatattggcctaaaattatctcttttgggtgcacctctaccaggttttgggtatcaggatgatgc
tagcttctcaaaatgaattagggaggttccctcttttctgtttagtgaatagtttcaagggaatggtaaccagctcctct
tgtagatctggttgaattcggctgtgaatccatctggctggacttttttgggttggtaggctattaatattgcccctcaat
tcagaaactgttatttgggtctatcagaggttcaactttttagctgtttagctgtgaggttgggttgggtgaggaatctac
attgcttctagattttctagcttatttgtgtagaggtgtttagtattctctgaggttagttgtatttctgtgggtcag
gggtgatccctttatcattttttatgctgtctattgattattctctcttactcttatttagtcttctgtagcgggtcatt
tttgttgatctttcaaaaaaccagctcctggaatcattgagtttttgaagggtttttatcttcttaccctcctcagttctg
ctctgactcttagttattttcttgcctctctgctagcttttgaatttggcttttgcctcttctctagttctttaaagggtgatgt
aggggttggatttttagatcttctctgcttctcttgggtcatttagtgctataaaatttccctctacacactgctttaaagt
ttcccagagatctcgttacgtttagatctttgttctcgttgggttttaaagaacaactttatttccagctctcattcatttta
cccagtagtcatccaagagcaggttgttcagtttccatgattttgtgtggtttttagtgaatttataatcctgatttgcact
gtggtctgagagacagtttgttgtgatttctgttcttttaccatttgcctgaggttcttctaccaatattgaggtcaatttt
agaataagtgcatgtggtgctgagaagaatgtatattctattcatttgggtgtgaggttctgtagatgatttaggtctg
cttgggtccagagctagttcaagctcctggatctccttggtaacttctctgtctcattgatctgtccaatatgacagttgggtgt
taaaggctgccattatattgtgtgggtctatgtctcttggtaggtctcaaggccttggctttatgaaatctgggtgtct
tgtattgggtcctgatatatttaggatagcttcttcttggtagaattgactcctttaccattatgtaggtgccccttctt
ctcttttgatctttgttggtttaagtctgttttatcagaactagaattgcaacccttttttttggcttccatttgcct
ggtagatctcctctgtccctttattttgagcctatgtgcatcttgcactgtgagataggtctcctgaaatagacacacagat
gggtctgactctttatccaatggcagctctgtgtttttaaattggggcatttagcccattttcattaaaggctaaatagttg
tatttggtaattttagtctctgtcattatgagtttagctgttttaccatttggcctgaggttcttctaccaatattgaggtcaatt
gtctttaccatttggtagtttttgcagttggctggtagccttcttccatgttttagtcttctcaggagctcttgt
aaggcaggtcctgggtggcgaaaaaatctgtcagcatttggctgtctataaaatatttttcttctcacttacaaggttag
tttggctggatagagatttgggttgaataatttttctttaaagaatgttgaatattggccccattcttttaaaggctcga
gggttctgtctgagacatccgctgtatgttggataggttcccttcttgggtgaaacctgacctttctctgtgctccctaac
attttttcttctcgtttcaaccttggtagactgagattatgtgtcttgggttgcacttttcaaggatatttttgggtg
tctctgtatttctgaaattgaaatgttggcctgtcttctgtagtttagggaagctcctctgagacaatcctgaaaggtgtt
tctaactgttccattctccccgtcacttccaggtacaccaatcaaacgttagatttgggttttccatagctcccatattt
ttggaggtcttggtttcttcttcttctaaacttctctcttcttcttcttcttcttcttcttcttcttcttcttcttcttct
tagtttcttccactttagcgaatcagctattgaaagctttagcactgctgacgaagttcttgtgcccgttcttccat
catgtcatttgaaggtcttctctacactgtttattctagttagccattcatctaaacttttttcaagggtttttaccctcctcga
gatgggttagaacaatcttagcttggagaagtttgttattactgacctctgaaacctacttctgtttaaactcatcaaacat
ctctgtccagtttttggcttccctgtcggcaaggagctgtgactcttggaggagaagaggcagctgttttttggaaatttccag
ctttctgtctgttttctccatcttgggttttctcactcttcttcttcttcttcttcttcttcttcttcttcttcttcttct
ttgatgtttagtctattccttctgttttgggtgttttcttccaacagtcagaccctcagctgcaggtgtgttggagtttgc
tggaggtccactccagacctgttttctttagtatcaccagcagaggtgcagaaccacaaatatttctgtcctgactcctcct
ctggaagctctcgtcccaggtggggcaccaccacctgtttgaggtgtttgtcggcctactgggaggtcttcccagtcagg
ctacatgggtgtcagggaccacctgagagggcagctgttccaggtgcaagctgcaagctgcaagctgcaagctgcaagctgca
tcttccagagctgtcagacagggacgtttaaagctgtcagaagctgtcggcttttggcttactatgcccctgccccagaggtgga
atctatagagggcagtagcccttactgagctgcgggtgggtcctccagtttgggttccagttctcttggtttaccactgtgagc
tactcaagcctcagcaatggcagacatctctccctcatcaggtcagctgtcagaggtcaatctcagactgctgcgctagca
gtgacaaggttctgtgagcgtgggacctgccaagcctgacacagggaggtatctcctgtgctgtctgttgaagcttgg
gaaaagtgagatttgggtcaggtgtactgtttctccaggtacagctgtcagcgttcccttggctaggaaagggaaatc
ccccagctccttgcctcctcgggtgagggcagctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct
aacaggtatctcagttgaaatgagaaatcactgtcttctgcatctatctcctgagctgacaccagagctgttctcta
ttcagccatcttggacaagatttttttaccataaaattttttaaattaaaaacttaagacatactaaaaaaatgtaaga
atttttaaattgatgggtgaaaaataacatcttgttttaatttgcatttcttgggtcttgggtttttagttgttttgggtt
ttttttagaccgggtctcactctgttggccaggtcggatgaggtggtgcaaccacggcacactgcagcctcaacctcctgg
gctcaggtgattctcccactcagcctcccaggtagctggaactacaggcacctgccaccacaccagctatttttttgggtg
gaaattttttagtagagacaggggttttggcaatttggcaagcttcttccaaactcctgggttaagcagctcctcctgctcag
cttcccaaggtgctaggattacaggtgtgaccaccatgcccggcaatttaccatttctttaaattatcagtgaaatctgaa
tttttgatctacattatttattcatgtgcttggctatttttctgttaggtttctaaaaatatttttcttagtgattttttag

agctcattatgaattaaagcccttttaatatgaagtgaacagttttttccagggtctctttgtttgtcagttttaagcttggt
atactactggtatagattactttttataacttttaattacatagtagtgaataatactttttagaatttcaaacatt
acaaaagatataagagtaaaagatagtttttttttaactcataccaggtaaatcaactgtaaatagtttagggtttcttt
ttaaactatttttttgtacatacaacatcttttttttaaaaagttttttttagatttgggggtacaggtgaagggtttgtacata
agtaaacgtgtcacagggatttgtgtacataaacttttcatcaccaggtattaagcctagtacccaatagttacctttct
gctcttctctactcccaccctcccgggtcaagatatacccaggtgctgtgttctctctttgtgttcatagttcttacc
atttagccccacttataagtgagaacatcggttatttgggagcttctgctgaagacatgctgctggttttttttaggcta
cgtgatttccacgggtgatatgtaccacattttctgttccagctctgcattgatgggcatttagggttattctgtctt
gctattgtgagcagtgctgacgtgaacatctatgtgcatatcttttatggcagaatgatttatattctctgggtatatacc
cagtaatgggatgtctgggttgaatggtagttctgcttttaggtctttgaggaatcgccactgctcccacaatggggaa
ctaaattacactcccacgaacagcatataagtggtcccttttctctacaacctcaccagcatctattatttttgactttta
ataatagccattctgactggtgtgaaatgggtaccactgctgcttttgatttgcatttctcctaagatcaggtgactgagct
ttcttctgtatgtttttggccacatatatgtcttctttgaaaagtgtctgttcacgtgctttgcccacttttaatggag
ttgtttttctctgttaaaattttttaagttcctgttaaatgtcaaatattaggcctttgtcagatgcatagtttgcaaaaatgt
tctcccattctgtaagttgtctgttactctggttagatagcttctttgtctgtgcagaagcttccaatttactccaatcgta
ttgtcaattttgtggttgcgattgttcaattttgtgcatgaaatctttgtctgttatattagccagagtgatt
gctaggttcttccagggttttatagttttgggttttacctgcaaatctttaatccatcttgagttgattttgtgtatg
gtgtaaggaagggatccagtttcaatctctgcatatggctagccaggttatcccagcaccatttttaatagggaaatctttt
tcccattgctgtttttgtcagttctgtggaagatcagatgggtcatagatgtgacgcttattctaggtctctctatctgtt
ccattgtgtatctctgtttttgtaccagtagcatttccactctctgtttttgggttagcgtgacctgttagtagagat
atgtgatgctccagctttgttcttttgcgtaagatttcttgactatccaggctctttttggctccatagaaattttaaaa
tagttttttctagttctgtgaagaatgtctgttggtagtttgataggaatagcattgaaatctgtaaatagctttgggcagata
gccattttaatgatattgattctccatccatgacatgggagttttttccattgtttgtgcttctctgattctcttgagc
ttgtttttataattctcatgtcagagatgttccactctctgcttttagctgtatctctaggtattttttttagtagta
tttgactgggatgctctctgattgactctcagtttggctgtgtgtgtgtaggaatgctagtgattttgtatattga
ttttgtatctctgcaactttggtgaaatgtttatcatctggagcagcttttaggccaagactgtggggtttctcaaatataga
atcatgtgtctgtggacagagatagttgacttctctctctctatttggatgcccctgtttcttctctctctctctctctg
ttgctcaggaccatcaacatacttttatgtgtaattatatacattagttttgggtgttttctctttcttcaaaaatgatattag
ctgtacataattttctgcaactttttccacttacctcttagatgttagttgtgttttagatttattctcacttctcaatga
tgacttaagatataaccataccttatttaactattctcctatccagaggtcatttatatgttcagacgttttatataataggt
agttggatgtattgatctttccctaatggctatgactttgggtgtcttcttactaaaacttttaggtctctctgtatgctgaa
ttgtatgaagatttcaaccatgtgttctgactttggctctttgactttaaactcattagtaaatctgaaatctgaaatctggtgat
ataatgagatataatgggtccacctcgccccaccagcaaccaccatccatagctaaatagcaccatttttgaataagttg
tcttctcccactgatattgaaatgtcaattttgtatcatacattgcattcataggtgtccatcctctttctcctttgact
ttgtccattctctctttctgtagcaaaaagcaagcttttttttttttttttttactgttagtactatatttttaataatag
ccggaaactctttctcagaaaatttctaaaggaaactctatgttt
attttaagtttacaattcagtggtcattcagtagcactgacattgttatgttaacctcagtagccatccatctctagaatattata
atcatcccactgaaatctctactcattaaacagtaactctccatctcctccatccccatcacctgatgataaccactatct
acttttgtctcatgaatttgctattctaggtaaatttatataattataataagttggagttccacttatatggcatcccacta
ttgtcctttttgtgtgtggcttattttacttagcataatgttttttaaggttccatctagatctatctgtgacatgaacagaa
cttcatctcttttttaggtgaataatagttcgttatatgtatataactatatttcatatttattcattcatctgttctgtagcac
ttgggtttgtttgtttttaccttttggctatttttagtaatgctgtttataaacaataggtgtacacatatctgtttgagttccc
gtattcagttcttttaggtatatacctgggagtagaattgctagatcataagggaaattctatttttaataatttttaggaactg
gcactgtattcccaggtggctgcaccattttacattcccaccagcagtgaggaaacttttaacaaaggttccaattttttccac
atctccaccaacttattaaattgttttttagtaatagccaagtgtcttagtctgtttttgtatgtgataacagaataccac
agactgagtaacttataatgaacagaaatttatttcttagagttctggaggctgagaagtctaatacaaggcatgggcatct
gtgagggccctctgtcttgcataatgatggaaggtcaccatgggtgaaaggaaaagagacgggtgagagagagagtaag
gtggacaaaaccatgccctaaatacaaacacaccctgcataaacgacattagtccatctatgaggacagagccataatgac
ccaaacacctctaaaggccaccaccaataacattacattgacaattaaatttcatcatgagttttgaaagggacattcaaa
ccatagaccgtcctaatagatataagtagatatacactagatagcatttccctaaaaattagaatgtgttgagaatct
tttcatgtgcttattgggtcattgtatattctcttttagaagaacatcttttttaagttcttttgccattttcaagttgggttga
catttgttactgagttttaaaggttatttatatgttctggatattacttcttatacaaatatagatttacaataatttctct
ctattctgtgggttgccttttcaactctgtttagtagtgcattttgtagcaaaaattgtcttttaattgtgataaataatttt
ttcttgtcttgcctgggttttagagtcacatccaataaatcatagccagatcaaatgtcaagaagcatttctctgttttct
ttctgagagttttatagttttacctcttatgttttaggtctctgactatttttagtttgtgtttgtatagataaggggtccag
cttcaacttttttgcatgtggatataccagttttaccagcaccattgtgagaagagactccctttcccatttgactgttctct
gtcagcgttttgaaaaccgtttgaccatataatgcaggggtttatttccgaggtctctgttctattccattgttctgtatgct
tgctttttgtcataccattgtttttgatattatagtttttggaaatcagaaggtgtgagccttctcaactttgctctttgttt
ctaggattgttttggctattttgtggtcccttgagattccatagttattttggaaatggatttttctatctcagaaaaaaaac
caaaaaataaccatcatcaggattttcataggggttgcactaaatctgaagattgctttgtatagattgacagtttcaaaaaa
ctaaatttttccaattccataaacacaaggtcttctccattttatattatctttaaatttttttagcagttttgttagtttca
gtgtacaagcttttccctcttagtttatttcttaagtattttattctttttgtatgctattcaaaatgatattgttttcttaa
tttccctttcagattatctattgttagtgatagaaatacaactgagttttgtatgttgattttgtaacctgccacttaactg
aattcatttatagtttcaagagatttttgcagtggtgtgatatgtgggactcttagggttttcaataaaaagatcattgtc
atctgcaaacagagcgaattttatttctctcttccatttgaattttcttttcttaacctaattgtctattagaacttca
gtcctatgtcagatagaagtggaacacatggacacctgtctgttctctgatcttagggaaaaagccttcatcttaaacat
tgagtatgatgttagctgtgggttttaatatatggccttattatgttgaggaagtttcttccatttcaggtcaactctaa
agcatttattttataatccctcttt
tgacagagagctctgctgtgtttgccagactggag
tgacgtgtgtgtctcggctcactgcaacctcccggttcaagcagattctcctgctcagcctcccagtagctgg
gattacaggcagcactaccacaccagctaaatttttgttttttttttttttttagtagagacgggtgtcccatgtggccag
gatggtctcgaactcctgactcgtgatccgctcctcggcctcccaaagtgctgggattacaggcgtgagccactgtgccc

ggctaaccctctatcttttaattttaattagaaatcacattggatttatagactaattttgagaaaatgtacagataaatt
gcatgttgaaccacagaatgttgtcacagaaaatgtgagtggggaggaagaagaagattatgacattaatgtaagtttata
aattttatgatactactttgagaaacattaatatagtttataaggaaaattatattttaaggattttctacaaagtgaaa
aaaaatcacctatgagtttaaaggattttaagctttatattaattctttccttttattaatttaaatctatatttctc
tttttctagGATATCCTAACCTGAGGAAGGAATACTTGACAGTTGG

Appendix II

CD46 exon 12→ exon 14 genomic sequence

Uppercase: Exon

Lowercase: Intron

Exon 12 Exon 13 Exon 14

Gray colored sequence is not included in minigene

SNP rs63193962 (T/-)

SNP rs41317997 (C/T)

Predicted branch point sequencePredicted branch point adenosineIntronic sequence for serial deletionsIn-frame stop codonPolyadenylation signal sequence

TTGTTGGAGTTGCAGTAATTTGTGTTGTCCCGTACAGATATCTTCAAAGGAGGAAGAAGAAAGCgtaaattaaagcatgtttc
 ttttaacttcttggctccttcttataacttaacatgcttttggcagcttcagtttgaatctgtattgcatgctatcctttttt
 tttttt[aaaaaaatgcctgcattagtttgcctaccaattgcatttcttggctaaaaatatttccagctacttctaacaatca
 ctgcaacataaaaaacatctgtgcagagttcctacagctattatgactttaagaaaaatttcttttggatttgagggttaaagtt
 atttagcaaatcatggaccactcactatgaggccttctctaccctattccctatttccaaaattgcactaatggccaaaatta
 aatttaaagtatctcactgtggagactttgacagttgtattcactggaaat[tttataatgaaagttccactgtctatataag
 cttagctagttgcaagaagctatgtaggcatatcatttaaataccaatttgggtgtctttcatttaaagttgtattttatac
 atgttaaactgggggattc[aaacacacacatc]ccgatttccctcttatgttctctgaaattaccagattgcagtaaaactgat
 agggttaaagtt[aaaaat]gtctgagaggggttagatctataaggtacctcaat[taactgtgtcttt]taataataaaaaatgtaa
 aaagtgagtggttagaataagaagaaaaatgtacaagtcacatacttcttaaagaaaataggggttt[aaacacacacac]ca
 attttagttaaagagcttacagtttaacaaaagacttagaatgaatcctggaaacacagactaagattaggtgctttgaaattt
 ctgaaagataataacatttactgatttttaaaaactgttttccctatagttttagaatagctttgaaatcaaac
 ctaataattttaaagacttttggtaactcatggttttatgatataatccttttatactgatcatcaggtttgggctaagtcac
 aaaagtaacaagcttactcataaggaactcaaac[tttt]catataacttgggtatctttgaaataccataatcccttcac
 ttgaaatc[taaat]ctgttataatggaacagatagcattttagttatctatcagttgtatattc[taaa]agttacatc
 tacaacgtgagtacaaaatattatcaatggaatatttttacaatccataaggtttgctgtagaacaagttc[taaa]gaaaaca
 aaacaaagtcgcatctgtcttatttccctctcttggtt[taaa]atatacactgcatcagaatttt[taaa]gggtcatt[aa]caagat
 taatcaaacctctctagctagcagtt[aa]tgggttttctgtgtgtgtttttgctttttgtgtagggagcttgcctgtc
 ccagctggagtgcaatggcgcaatctggctcgctgcaacttggcctcctgggttcaagggattctcgtcctcagcctcac
 gagttagctgggattacagctgctgaccaccaccacgggctaatttgggtgactcctgtagagatgggggttccaccatggtgg
 tcaggtggtcttgaactcctgacctcaggtaatccgccaactcggcctcccaaggtgctaggattacaggtatgagccacc
 ggcgccggccttaatgggtcttt[aa]agctcaacagttatctgactgttacactcccaatgaaaggttcaatagcagtagtg
 aactgagttccaaaat[taatt]gtccagtgatttttt[aa]agcccatattttacagcataggggtccactcagactgtcc
 attat[taaa]ttt[taaa]atatttgggaaggtatgatattttagccaatgcaacataat[ta]atcctgcaaatagtagcaca
 cttaggaaaaatagtcagaaaaatgggtgtgagagacgaaactcctaattgaccttgaactt[taaa]atataaacctgcgtggga
 atcttttagaatagatgttaccctttttatgacaggaacatactcgattggaatcagcatatgtaatggctgaaatc[aa]ag
 ctaggctctctttatggatctattat[tt]gtagttccatgttattttgtagctgaaactgaaatgtaggggtggttatcttcc
 ttctcagagcttagaggaatgtttctctgacttt[aa]tgtgacatttagtcacctggaaatctgttaaagtagcagattctgat
 tcattgggtttggggtggagcctgagaccctgcctgtctagtgagctccagtgctgatgctgctgctgctgctgctgggctcac
 tgaccaggggtgtgcccactactgatgtgctatattgtgaaagctgaagaaacactatgaagcttatggcaaat[tt]tct
 aatatt[aa]gataaattgaaagtttt[aa]agattcctcatccatggtgatacct[taaa]aaaaggctatggggt[aa]ataaatagctt
 ggcttttctaccttctt[aa]actct[taaa]aaaacaagatttccagccggcctggtggctcacgctgtaatccagcacttt
 gggaggccgagggcggcagatcagaggtcaggagatcaagaccatcctggctaacacggtgaaaccccgctctact[taaa]aa
 taca[aa]aaaa[aa]aaaa[aa]aaaa[aa]tagtggcgggctgtagctccagctactctgggaggctgagggcaggagaatggcgtgaa
 ccccgaggcggagctgcaagtgcagagattgcccagctgcaactccagcctgggtgacggagtgagactccatccataaac
 aaaaaacacaaaaaaccccaagatttccattatttctt[aa]tgcactgtt[aa]atgtccataatctt[aa]tccccagattggct
 gttgtggccttagaaacttctataaagactataaaatgtctggaggcacttctccacttccgggggtaactggaagcaggct
 agcaggcaca[aa]taatctcaaatgtgt[aa]agcctggcaca[aa]atgtatttctgtaaacctggtattt[aa]tgggtatttggg
 aacgtaattactctctgagcctc[aa]tctcaaaaatacaaaaatagtgtaaggagtcattt[aa]gatcatccaaatcattt
 cactactgataaaaagttgggcaaaaggt[aa]agattctgocacttttctgttcatagctctcagctgagaacagtagtagg
 ataccagttgggtccgaaacaggtcccatccacatgctccctcctttccctccat[aa]tcttcttactaattcttctacc
 atacttttagtcttggaaagccacagccctgtctt[aa]tgcactgctcagctcctgtattgcatggcatgactgtgcttctgaa
 aacagatgctttgtttacataaatatt[aa]agtaagtttttt[aa]aatgcatgatttttt[aa]actcagtagt[aa]agc
 cctggataagtgacagtttggctaagttagtgaaagttattt[aa]cataat[aa]tataat[aa]taccctgtgacaaatagttttc
 ctttttccatt[aa]tgggtattt[aa]tattt[aa]tataacctaactgacttt[aa]tattacatt[aa]tattgactgaccaat[aa]c

TTTCAAATTTAATATTTATATTTAGAGATCTATATATGTATAAAATATGTATTTTGTCAAATTTGTTACTTAAATATATAGAGAC
CAGTTTTCTCTGGAAAGTTTGTTTAAATGACAGAAGCGTATATGAATTCAGAAAAATTAAGCTGCAAAAATGTATTTGCTATA
AAATGAGAAGTCTCACTGATAGAGGTTCTTTATTGCTCATTTTTAAAAAATGGACTCTTGAAATCTGTTAAAAATAAAATGT
ACATTTGGAGATGTTTCATGA

Appendix III

Sequence alignment of exon 13 and flanking 100 nt of each introns 12 and 13 between species.

Uppercase: Exon 13, as referenced by human genome

Lowercase: Intron

Analysis by MView:

Reference sequence: Human

Identities normalized by aligned length.

Colored by: identity + property

		161	.	.	.	2	240	
1 Human	100.0%	TGAGAAGGAGAGATGAGAGAAAGGTTTGCTTTTATCATTAAAAGGtatctgttt	tctgt	tg	ttt	at	tt	tt	tc	o	---	gat	
2 Chimp	98.3%	TGAGAAGGAGAGATGAGAGAAAGGTTTGCTTTTATCATTAAAAGGtatctgttt	tctgt	tg	ttt	at	tt	tt	tc	o	---	gat	
3 Gorilla	97.6%	TGAGAAGCAGAGATGAGAGAAAGGTTTGCTTTTATCATTAAAAGGtatctgttt	tctgt	tg	ttt	at	tt	tt	tc	o	---	gat	
4 Orangutan	97.0%	TGAGAAGAAGAGATGAGAGAAAGGTTTGCTTTTATCATTAAAAGGtatctgttt	tctat	tg	ttt	at	tt	tt	tc	o	---	gat	
5 Gibbon	96.6%	TGAGAAGAAGAGATGAGAGAAAGGTTTGCTTTTATCATTAAAAGGtatctgttt	tctgt	tg	ttt	at	tt	tt	tc	o	---	gat	
6 Rhesus	94.9%	TGAGAAGAAGAGATGAGAGAAAGGTTTGCTTTTATCATTAAAAGGtatctgttt	tctgt	tg	ttt	at	tt	tt	tc	o	---	gat	
7 Baboon	94.2%	TGAGAAGAAGAGATGAG-----AGGTTTGCTTTTATCATTAAAAGGtatctgttt	tctgt	tg	ttt	at	tt	tt	tc	o	---	gat	
8 Greenmonkey	95.9%	TGAGAAGAAGAGATGAGAGAAAGGTTTGCTTTTATCATTAAAAGGtatctgttt	tctgt	tg	ttt	at	tt	tt	tc	o	---	gat	
9 Marmoset	92.3%	TGAGAAGAAGAGATGAGAGAAAGGTTTGCTTTTATCATTAAAAGGtatctgttt	tctgt	tg	ttt	at	tt	tt	tc	o	---	gat	
10 Pig	64.2%	TGAGGA---GAGAGGAGAGAAAAGTTTGCTTTTGTCA-TGAAAGGtatctgttt	tctgttt	tg	ttt	at	tt	tt	tc	o	---	---	
11 Dolphin	82.4%	TGAGAA---AAGATGAGAGAAAGGTTTGCTTTTATCATTAAAAGGtatctgttt	tctgttg	tg	ttt	at	tt	tt	tc	o	---	gagat	
12 Dog	80.4%	TGAGAAGAAGAGATGAGAGAAAGGTTTGCTTTTATCATTAAAAGGtatctgttt	tg	tg	---	tg	ttt	at	tt	tt	tc	o	gagat
13 Elephant	73.9%	TGAGATGA-TAGATGAAAGAAAGGTTTCTTTTATCATTAAAAGGtatctgttt	tc	---	tg	ttt	at	tt	tt	tc	o	gagat	
consensus/100%		TGAGr...AGA.GAr...ArGTTTCTTTTrTCA-TGAAAGGTAYy.YTTT...YGYTT...T...G...											
consensus/90%		TGAGAA...rAGATGAGAGAAAGGTTTGCTTTTATCATTAAAAGGTATyyGyTTTCTr...TGTTT.TTTTy...G...											
consensus/80%		TGAGAAG...GAGATGAGAGAAAGGTTTGCTTTTATCATTAAAAGGTATyTGyTTTCTGTGTGTTT.TTTTCA GAT											
consensus/70%		TGAGAAGrAGAGATGAGAGAAAGGTTTGCTTTTATCATTAAAAGGTATyTGTTTCTGTGTGTTT.TTTTCA GAT											

		241	:	3	.	.	.	320																	
1 Human	100.0%	gtcctttccttt	tg	aaaa	at	tt	cag	tg	g	a	t	a	t	a	t	a	g	at	at	ca	aa	ac	---	tgagca	a	a	g	a	t	at	tg	g	ca
2 Chimp	98.3%	gtcctttccttt	tg	aaaa	at	tt	cag	tg	g	a	t	a	t	a	t	a	g	at	at	ca	aa	ac	---	tgagca	a	a	g	a	t	at	tg	g	ca
3 Gorilla	97.6%	gtcctttccttt	tg	aaaa	at	tt	cag	tg	g	a	t	a	t	a	t	a	g	at	at	ca	aa	ac	---	tgagca	a	a	g	a	t	at	tg	g	ca
4 Orangutan	97.0%	gtcctttccttt	tg	aaaa	at	tt	cag	tg	g	a	t	a	t	a	t	a	g	at	at	ca	aa	ac	---	tgagca	a	a	g	a	t	at	tg	g	ca
5 Gibbon	96.6%	gtcctttccttt	tg	aaaa	at	tt	cag	tg	g	a	t	a	t	a	t	a	g	at	at	ca	aa	ac	---	tgagca	a	a	g	a	t	at	tg	g	ca
6 Rhesus	94.9%	gtccttgcttt	tg	aaaa	at	tt	cag	tg	g	a	t	a	t	a	t	a	g	at	at	ca	aa	ac	---	tgagca	a	a	g	a	t	at	tg	g	ca
7 Baboon	94.2%	gtccttgcttt	tg	aaaa	at	tt	cag	tg	g	a	t	a	t	a	t	a	g	at	at	ca	aa	ac	---	tgagca	a	a	g	a	t	at	tg	g	ca
8 Greenmonkey	95.9%	gtccttgcttt	tg	aaaa	at	tt	cag	tg	g	a	t	a	t	a	t	a	g	at	at	ca	aa	ac	---	tgagca	a	a	g	a	t	at	tg	g	ca
9 Marmoset	92.3%	gtcctttccttt	tg	aaaa	at	tt	cag	tg	g	a	t	a	t	a	t	a	g	at	at	ca	aa	ac	---	ggagca	a	a	g	a	t	at	tg	g	ca
10 Pig	64.2%	-tttcccatc	tg	aaaa	at	tt	ccat	g	acc	at	g	ta	g	cc	g	tg	gg	ta	ac	---	---	---	---	tgagca	a	a	g	a	t	at	tg	g	ca
11 Dolphin	82.4%	gtcctttccttc	tg	aaaa	at	tt	caat	g	aat	at	tt	ta	g	at	at	tg	ata	ac	---	---	---	---	tgagca	a	a	g	a	t	at	tg	g	ca	
12 Dog	80.4%	gtcctatctagaacag	aaaa	at	tt	caat	g	aat	at	tt	ta	g	at	at	ct	g	ta	cc	---	---	---	---	tgagca	a	a	g	a	t	at	tg	g	ca	
13 Elephant	73.9%	atcctttccttt	tc	aaaa	at	ct	at	tg	at	at	tg	at	at	tc	aa	ac	tg	ag	ta	aga	---	---	tgagca	a	a	g	a	t	at	tg	g	ca	
consensus/100%		.T.yy.....y.rAAAr.TC.rT.r.yAT.TAG.yryy.rTA..y.....ArC.AA...r.ry.Gr.A																															
consensus/90%		rTyCTy.CTTy TrAAAAATyTC.rTGATAT.TAGATyCrrTAAYC TGArC.AAGA.rTrT.GGCA																															
consensus/80%		GTCCTT.CTTy TGAAAAATyTCArTCrATAT.TAGATATCrATAACC TGArC.AAGA.rATrTTGGCA																															
consensus/70%		GTCCTTyCTTT TGAAAAATATTCArTGATATrTAGATATCAATAACC TGAGC.AAGA.rATATTGGCA																															

		321]	322
1	Human	100.0%	gt	
2	Chimp	98.3%	gt	
3	Gorilla	97.6%	gt	
4	Orangutan	97.0%	gt	
5	Gibbon	96.6%	gt	
6	Rhesus	94.9%	gt	
7	Baboon	94.2%	gt	
8	Greenmonkey	95.9%	gt	
9	Marmoset	92.3%	gt	
10	Pig	64.2%	gg	
11	Dolphin	82.4%	gt	
12	Dog	80.4%	gt	
13	Elephant	73.9%	gt	
	consensus/100%		G	
	consensus/90%		GT	
	consensus/80%		GT	
	consensus/70%		GT	