Regulation of splicing Integrin α6 during development and differentiation

Mohammed Ali Alshehri

Institute of Genetic Medicine

Thesis submitted for the degree of Doctor of Philosophy

November 2017

Declaration

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

…………………………..

Mohammed ALShehri

Acknowledgements

First, I would like to thank my pervious supervisors Dr. Julian Venables and Professor John Burn for their support and guidance for the first two years for my project. I would thank my supervisor Professor David Elliott for his help, guidance and support to keep the project progress during my PhD. I would like thank you for your helping and supervision, guide me and encouraging my research to complete this project.

I would like to thank all members of David's lab, both past and present including Mrs Caroline Dalgliesh, Doctor Ingrid Ehrmann and Doctor Jennifer Munkley for helping, support and assisting me with techniques. I would also thank Dr. Andrew Best and Dr Marina Danilenko for their helping in techniques and valuable advice. I would particularly like to thank Dr Abd Al-Hasan for sharing his valuable expertise in lab and techniques, also for his supporting and advisor all time. I would like to thank members of the Elliott and Jackson labs for their help and advice

I would also thank including Doctor Philippe Fort (centre Nationale de la Recherche Scientifique, CRBM-UMR5237, Université de Montpellier, Montpellier, France), Professor Chris Smith (Department of Biochemistry, University of Cambridge),

My thanks also go to the Saudi Arabian culture bureau in London and the Najran University who supported the studentship at Newcastle University.

I would like to extend thank you to all my friends and family. In particular, thank you to my mum -Aisha for her unconditional love, support and kindness you have shown through all my life. Thank you to my father -Ali for his support and kindness. Thank you for always believing in me. This thesis is for you. A special thanks to my wife – Nawal- for her love, support and care. Thank you to all my brothers and sisters for their support over the past 4 years. I am truly grateful to all of you.

Table of contents

Declaration .... **II**

Acknowledgements .... **III**

Table of contents .... **IV**

List of Figure .... **VII**

List of Table .... **XI**

List of Abbreviations .... **XII**

Abstract .... **XIV**

Chapter 1: Introduction **1**

1.1 Pre-mRNA **1**

1.1.1 One gene multiple protein 1

1.1.2 Overview of mRNA 2

1.2 Splicing **3**

1.2.1 Splicing mechanism 5

1.2.3 The minor spliceosome 7

1.2.4 Alternative splicing 8

1.3 RNA-protein interaction: **10**

1.4 The regulation of splicing **11**

1.5Cis elements: splicing enhancers and silencers **11**

1.5.1 Strength of splice site and position 11

1.5.2 Cis- elements activators and repressors 12

1.5.3 Trans-acting RBPs 12

1.6 Other factors that affect splicing **13**

1.6.1 Intron/exon structure 13

1.6.2 RNA secondary structure 14

1.6.3 Other factors effect on splicing 14

1.7 Alternative splicing in ITGA6 Integrins **17**

1.8 Integrins **18**

1.8.1 Integrin structure 21

1.8.1.1Extracellular domains 21

1.8.3 Integrin cytoplasmic domains and cell signalling 24
1.8.4 α subunits in the cytoplasmic domain 26
1.9 Integrin α6 (ITGA6) 26
1.10 Research aims and Hypothesis 29
1.10.1 Aims: 29
Chapter 2: Investigation of orthologous of stem cell regulated exons in zebrafish 30
2.1 Introduction 30
2.2 Aims: 32
2.3 Materials and Methods 33
2.3.1 Zebrafish dissection method: 33
2.3.2 RNA extraction: 34
2.3.4 Geneious program 34
2.3.5 Primer design 34
2.3.6 PCR 35
2.3.7 Gel electrophoresis 37
2.4 Results: 38
2.4.1 Alternative splicing in 10 genes in zebrafish head and tail. 38
2.4.2 Investigation of ITGA6 alternative splicing in vertebrates 41
2.4.3 Investigation of paralogues and orthologues of ITGA6 in Zebrafish Head and Tail after Post- Fertilization. 44
2.5 Discussion 46
2.5.1 Validation of a novel alternative splicing event from Zebrafish development: 46
2.5.2 ITGA6 alternative splicing pattern in zebrafish 48
2.6 Chapter Summary 49
Chapter 3: Identification of PTBP as a novel regulator of the ITGA6 splicing switch 50
3.1 Introduction 50
3.2 Aims 51
3.3 Materials and Methods 52
3.3.1 Quantitative real time PCR (qPCR) 52
3.3.2 Cell culture 54
3.3.3 Calculation of percentage splicing inclusion (PSI %) 56
3.3.4 Gel electrophoresis 56
3.3.5 Western immunoblotting 56
3.3.6 siRNA Transfection: 57
3.4 Results 58

3.4.1 Splicing pattern of ITGA6 in different cell lines (HeLa, MCF7 and MDA-MB231), stem cells and fibroblasts 58

3.4.3 SiRNA knockdown by MBNL1, RBFOX2, and PTBPs in HeLa, MCF7 and MDA-MB231 cell lines........ 60

3.4.2 Measurement of the relative levels of factors that regulate ITGA6 alternative splicing 70

3.5Discussion 73

3.5.1Investigation of ITGA6 alternative splicing in different cancer cell lines........ 73

3.5.2 How is ITGA6 alternative splicing regulated? 74

3.5.3 Identification of PTBP as a novel regulator........ 75

Chapter 4: Establishment of a minigene system to study regulation of the ITGA6 alternative exon 78

4.1Introduction 78

4.2 Aims........ 80

4.3 Methodology........ 81

4.3.1 ITGA6 Minigene essay:........ 81

4.3.2 siRNA transfection 88

4.3.3 Western blot 88

4.3.4 Transfection 89

4.4 Results 90

4.3.1 ITGA6 alternative exon minigenes respond to splicing control similarly to the endogenous ITGA6 alternative exon........ 90

4.4.2 The ITGA6 alternative exon encoded with a minigene response to changes in regulation factors similarly to the endogenous ITGA6 alternative exon 97

4.4.3 Effects of PTBP overexpression/KD in different cell lines on splicing pattern from the minigene 103

4.4.4 The downstream binding site for PTB represses splicing of the ‘a’ exon through RNA protein interactions 106

4.5 Discussion 113

Chapter 5 General discussion and conclusion........ 120

Bibliography 125

List of Figures

Figure1.1: Multiple mRNAs from single gene. Post transcriptional mechanism including a) alternative splicing b) alternative polyadenylation c) rna editing(siomi and dreyfuss, 1997) **2**

Figure1.2: Cis elements that control pre-mrna splicing. A) the four components of the splice site in an intron are the 3' splice site, polypyrimidine tract, branch point and 5' splice site. B) splicing complex formation by exon definition. Spliceosomal complexes form along the pre-mrna during splice site recognition and enable commitment to splicing. U1snrnp binds at the 5'ss by base pairing, whereas U2AF binds at the 3'ss and polypyrimidine tract. The branch point is bound by u2 snrna. Other rna-binding proteins including 'sr' proteins bind to cis-acting splicing enhancers to further help define the correct splice sites. Image redrawn from (wahl et al., 2009) 4

Figure 1.3: The two steps of the splicing mechanism. Both reactions are nucleophilic attacks. In step one an oh group at the branch site adenosine attacks the ('splice donor') 5' splice site. In step two the 5' splice site attacks the 3' ('acceptor') splice site and thus the intron is removed and the two exons are ligated. **5**

Figure 1.4 :Schematic diagram showing the splicing reaction (wahl et al., 2009).................................. **7**

Figure 1.5:The different kinds of alternative splicing event (see text for details). Image redrawn from (srebrow & kornblihtt, 2006).. **9**

Figure1. 6: RNA-binding protein processing in various post-transcriptional processes at different locations in eukaryotic cells. RBPS are involved in 5 major processes that are represented by red circles. RBPS play a major role in splicing of mrna in the nucleus, and after that mrna export from nucleus to the cytoplasm by various RBPS. In cytoplasm, rbps contribute to mrna localization in compartments such as the mitochondria. RBPS are also responsible for rna stability and lead to degradation. RBPS are also involved in translational control. Image adapted from (janga and mittal, 2011).. **10**

Figure 1.7: Sequence elements that affect splicing include both splicing enhancer (ese/ise) and splicing silencers (ess/iss). Image modified from (srebrow & kornblihtt, 2006) **12**

Figure1.8: Schematic representing a classification of integrin family of heterodimers, image adapted from (barczyk et al., 2010) **19**

Figure 1.9: Integrin structure. A) primary structure of integrin α and β subunits. Domains 1-7 represent the homologous repeats in the α subunit. Stars represent divalent cation-binding sites (mg2 as orange star, and ca2 as grey stars). Green represents the conserved domain in the β subunit. The grey bar represents the plasma membrane. B) integrin secondary structure. My project concerns alternative splicing of the cytoplasmic portion of the alpha subunit that starts with the sequence gffkr (lower case). Figure from (barczyk, carracedo et al. 2010).. **22**

Figure 1.10: Domain structure of α and β subunits in cytoplasmic tail of integrins. B subunit binding sites for multiple integrin binding proteins including talin are shown in red. Image adapted from (morgan et al., 2007). **23**

Figure 1.11. Modes of integrin signalling: a, b and c shows inside-out signalling. D and e show outside-in signalling. Image adopted from (askari et al., 2009) **25**

Figure 1.12: The two alternatively spliced forms of ITGA6 genes initialise showing the last three exons and the peptide sequences of the alternative c termini produced a) structure of the two for itga6 isoform b) the two alternatively spliced forms of ITGA6 , peptide in black is common between the two isoforms.. **28**

Figure 2.1: Aconserved regulated program of alternative splicing in germ layer specification (venables, lapasset, et al., 2013).. **31**

Figure 2.2: The steps of head and tail dissection from zebrafish embryos after 24 hours. A zebrafish embryo is covered by shield. B. The embryonic shield is pulled away. C zebrafish embryo at 24

hours. D. Head and tail after dissection. Photos taken from a previous dissection by alicia madgwick.. **33**

Figure 2.3: Developmentally regulated alternative splicing in zebrafish. Agarose gel illustrating changes in ases in zebrafish embryogenesis: a) alternative splicing in 10 genes in zebrafish head and tail (see materials and methods) 24 hours post-fertilization. KIF13A and itga6, in the third column, out of 10 genes show alternative splicing events between head and tail. The lower band represents the short form (exon exclusion) and upper form represents the long form (exon inclusion) **39**

Figure 2.4: Developmentally regulated alternative splicing in zebrafish. Agarose gel illustrating changes in ases in zebrafish embryogenesis. Agarose gel showing the two ases that shift in their inclusion ratios during embryogenic development from 2 to 48 hours. **40**

Figure 2.5: Paralogues and orthologous of ITGA6. Genetic tree for ITGA3, ITGA6 and ITGA7 is shown for both the 'a' and 'b' alternatively spliced isoform in different species. **42**

Figure 2.6: Alignment of the sequence paralogues and orthologous of ITGA6. C-termini of various vertebrate species shown... **43**

Figure 2.7: Alternative splicing of 5 zebrafish paralogues of ITGA6 in head and tail, 24 post-fertilization. Upper band represents long form (exon inclusion) and lower band represents short form (exon exclusion). Itga3a, itg3b and itga7 did not produce any rt-pcr product... **45**

Figure 2.8 : Alternative splicing for ITGA6A and ITGA6B in development stages from 2 hours to 48 hours post-fertilization in zebrafish. A) agarose gel illustrating alternative splicing for itga6a and ITGA6B in zebrafish development. The long form (292bp) clearly appears at 15 hours whereas the short form (195bp) is visible at all stages from 11 to 48 hours in ITGA6A. In ITGA6B, the long form (186bp) is visible just at 28 and 48 hours, whereas the short form (137bp) clearly shows in all stages. B) chart shows different of alternative splicing event between ITGA6A and ITGA6B during zebrafish development stages. **45**

Figure 2.9: Stages of embryonic development of the zebrafish. Image adopted from (kimmel, ballard, kimmel, ullmann, & schilling, 1995) .... ERROR! BOOKMARK NOT DEFINED.

Figure 3.1: Investigation of itga6 alternative splicing control in different cancer cell lines. Rt-pcr was performed using specific primers across the alternative exon of itga6 using rna from different cancer cells. A) graphs show the percentage of splicing inclusion (psi%) in different cell lines , stem cells and fibroblasts. B) agarose gel illustrating changes in ases of ITGA6 in different cancer cell lines. Exclusion of the cassette exon is represented by the lower band (247bp), while the upper band represents exon inclusion (377bp). This experiment showed that exon 25 is skipped in hela and stem cells, but included in mda-mb231, mcf7 and fibroblasts.. **59**

Figure 3.2. ITGA6 alternative splicing pattern after knocking down splicing regulation factors in the mda-mb231, mcf7 and hela cell lines. Agarose gel shows itga6 splicing after knocking down ptbp1/2, mbnl1 and rbfox2, respectively, in the mda-mb-231 , mcf7 and hela cell lines. Exclusion of cassette exon appears as the lower band, while the upper band represents the inclusion. **61**

Figure 3.3: The percentage splicing inclusion (psi) of ITGA6 after knockdown of splicing regulation factors in the mcf7 cell line. Graphs show the percentage of splicing inclusion of the ITGA6 alternative exon which was compared between controls (without sirna) and cells with knockdown of splicing regulation factors (ptbp, mbnl1, rbfox2 and esrp) (green). Data is represented as the mean of three biological replicates in each case. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. Capillary gel electrophoresis image shows ITGA6 alternative splicing after knocking down regulation factors in the mcf7 cell line. Exclusion of the cassette exon appears in the lower band (246bp), while the upper band represents the exon inclusion (377). **63**

Figure 3.4 : Knockdown of mbnl1 in mcf7 tested by western blotting and real time pcr. A) depletion of endogenous mbnl1 was confirmed in trplicate in mcf7 cells by western blot. Expression of gapdh was monitored as a loading control. B) qpcr detected the knockdown of mbnl1 in mcf7 cell line (represented in red) compared with the control (untreated cell) represented in blue. Data for the three biological replicates and statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001 65

Figure 3.5:The percentage splicing inclusion (psi) of ITGA6 with knockdown of splicing regulation factors in the hela cell line. Graphs show the percentage of splicing inclusion of ITGA6 alternative exon compared between controls (without sirna) with gfp (dark brown) and knockdown of regulation factors (ptbps, mbnl1, rbfos2 and esrps) green. Data represents the mean of three biological replicated in each case. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. Capillary gel electrophoresis image shows ITGA6 alternative splicing after knocking down splicing regulation factors in the hela cell line. Exclusion of cassette exon appears in the lower band (246bp), while the upper band represents exon inclusion (377). **67**

Figure 3.6: Knockdown of PTBP1 in hela cell line confirmed by western blotting and real time pcr. A) depletion of endogenous ptbp1 was confirmed in triplicate in hela cells by western blot. Expression of gapdh was monitored as a loading control. B) qpcr detected the expression of kncodown of PTBP1 in hela cell line (shown in red) compared with the control (untreated cell) in blue. Data for the mean three biological replicates and statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001 69

Figure 3.7: Regulation factor expression in different cell lines, stem cells and fibroblasts. Qpcr was used to quantify expression of ptbp (ptbp1 and ptbp2), mbnl1, rbfox2 and esrp (esrp1&esrp2) in mcf7, stem cell, fibroblast, mda-mb231 and hela cells. Data for the three biological replicates and statistical significance were calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. .... **72**

Figure 3-8: Schematic diagram explaining the RNA map. This image was adopted and modified from (ule et al., 2006). **75**

Figure 3-9: Schematic illustration of the vegf signaling pathway in cancer stem cells **76**

Figure 4.1: Schematic diagram to explain mutagenesis procedure using overlap pcr. The first step of the reaction created mutants of fragment by using specific primers designed with specific mutations in order to produce two overlapping dna fragments with complementary mutations. These were joined by overlap pcr and then cloned into the pxj41 vector. **86**

Figure 4.2: Sequence of the human alternative exon of ITGA6 and surrounding intron sequences. The alternative exon is shown in red font. Rbfox binding sites are shown in red highlight (strong site) and pink (weak site). Mbnl sites (ygcy) are highlighted blue (strong site) green (medium) and yellow (weak site). **91**

Figure 4.3. Establishment of a minigene system to study regulation of the ITGA6 alternative exon to test if the ITGA6 alternative exon is repressed by ptb, i made a minigene construct for the itga6 alternative exon along with the surrounding introns and transfected this into different cancer cell lines. After rt-pcr and gel electrophoresis, the lower band represented short mrna isoform (excluded exon) whereas, upper band represented the long isoform (inclusion exon). **92**

Figure 4.5: Analysis of ITGA6 minigene constructs. The agarose gel shows pcr products made from of cDNA of hela, mda-mb231 and mcf7 cells transfected with itga6 4kb , ITGA6 1.3kb and itga6short minigenes. The upper band represents the inclusion ITGA6 alternative exon (318 bp), which appears only in itga61.3kb and itga64kb. The lower band represents the exclusion ITGA6 alternative exon (vector without insertion) (188 bp). **96**

Figure 4.6: Effect of knock down of splicing regulation factors on expression of the ITGA61.3kb minigene construct in the mcf7 cell line. A) the endogenous itga6 alternative splicing patterns after knocking down (ptbp1/2, mbnl1, rbfox2 and esrp1 in mcf7 cells. The upper band (377 bp) represents the long form (with inclusion of the alternative exon), whereas the lower band (249 bp) represents exclusion of the alternative exon. B) agarose gel of rt-pcr using minigene primers around the itga6 alternative exon. A pcr product (itga61.3kn minigene) including the alternative exon is shown in the upper band (inclusion exon) and the lower band (exclusion exon). **99**

F**igure 4.7: The ITGA6 1.3kb minigene construct responds to knock down of pbtbp1 and ptbp2 similarly to the endogenous gene.** Agarose gel shows. A) the endogenous itga6 alternative exon knockdown ptb1 and ptb2. The upper band represents the exon inclusion (377 bp), whereas the lower band shows the exclusion of the alternative exon (247 bp). B) rt-pcr using minigene primers

around the alternative itga6 exon. The upper band represents exon inclusion and the lower band exon skipping. Double knockdown of ptb1 and ptb2 in itga6 causes a shift from the lower band to the upper band for both the minigene and endogenous itga6. .... **100**

Figure 4.8: Overexpression of splicing regulation factors with ITGA61.3KB minigene construct in the mcf7 cell line. Bar chart shows the percentage splicing inclusion (psi) of the itga6 minigene with overexpression constructs encoding ptbs & mbnl (blue) compared with itga6 wild type gfp (red). Mbnl has higher psi compared to gfp, whereas ptbp1 has less. Data represents the mean of three biological replicates in each case. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. Capillary gel electrophoresis shows mbnl drives more inclusion ('a' form) whereas ptb1 represses the 'a' exon, comparing with coexpression of gfp. **102**

Figure 4.9: Effects of PTBP over-expression and knockdown in different cell lines on minigene. Bar chart shows the percentage splicing inclusion (psi) from the itga6 minigene after knockdown (orange) and overexpression(bright blue) of ptbp regulator factors comparing with wild type gfp (green) in a) mda-mb-231, b) mcf7 & c) hela cell lines. Data represent the mean of three biological replicates in each case. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. .... **105**

Figure 4.10: Effect of mutating the ptb binding sites downstream of the ITGA6 exon. A) sequence of the human alternative exon of ITGA6 and surrounding intron sequences with deletion for candidate ptb sites (del a & del b) in downstream intron. B) bar chart shows the psi level of the itga6 minigene with del a (red) &del b (pink) giving less inclusion for the alternative exon compared with wild type ITGA61.3 gfp (green). This means that even though ptb expression represses exon 'a', these binding sites are activating. **108**

Figure 4.11: PTB protein structure. A) schematic illustration of ptb rna binding sites including rrma1 (red) rrm2 (grey) rrm3 (yellow) and rrm4 (green). Each rrm motif has different binding affinity for pyrimidine-rich sequences on mrna. B) model of the interaction ptb four motifes with rna. Image adapted from(sawicka et al., 2008a) **110**

Figur4.12: Effect of different rrm mutation upon to PTB inhibitor a) schematic illustrate mutant sitat at 4 domain ptb binding sites, and table shows the location of the mutant.b) bar chart shows transfecting itga6 minigenes with ptb different mutant. Data represented the mean of three biological replicated in each case after calculate. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. B) western blotting for ptbp different mutant expression with flag. It dos not show any expression compare with wt. ... **112**

Figure 4.13. Illustration of the possible binding site for mbnl1 in ITGA6. The sequence represents the ITGA6 alternative exon (in red) with surrounding introns. Mbnl1 candidate bind sites are in yellow, blue and green boxes. The diagram underneath represents where mbnl1 binds in order to activate itga6 alternative exon. **115**

Figure 4.14 Schematic illustrate the possibility binding sites regulation factors regulate ITGA6 exon. The sequence represents the itga6 alternative exon (in red) with around intron in black, mbnl1 candidate bind sites represent in yellow, blue and green, ptb candatit binding site represent in dark green. Diagram shows the regulations binding sites and how might they regulate itga6 alternative exon. **118**

Figure5.1Schematic illustrate the two hypothesis of how ITGA6 alterntive exon regulated . A) possibility binding sites regulation factors regulate itga6 exon. **124**

List of Tables

Table1.1: Classification of integrins. From (reddy and mangale, 2003) **20**
Table 2.1: Primers for orthologous of 10 genes **35**
Table 2.2: ITGA6 paralogues and orthologue zebrafish primers **35**
Table 2.3 : 5xbuffer stock (1ml) **36**
Table 2.4: PCR component for 50µlx reaction **36**
Table 2.5: PCR procedure **36**
Table 2.6: Electrophoresis gel protocol (1.5%) **37**
Table 2.7: 10x tbe (1litre) components **37**
Table 2.8: 0.5m EDTA (at ph8) **37**
Table 3.1: Components of reverse transcription **52**
Table 3.2: Temperature sequence of rt reaction **52**
Table 3.3: Components of qPCR **53**
Table 3.4 :Steps of qPCR reaction **53**
Table 3.5: Primers designed to targeted regulatory proteins for qPCR **53**
Table 3.6: Transfection protocol **57**
Table 4.1: Primers designed to clone different size of itag6 regulatory **81**
Table 4.2: Components of cloning pcr **81**
Table 4.3:Steps of PCR reaction for cloning **82**
Table 4.4: Constructs restriction digest **83**
Table 4.5: Vector restriction digest **83**
Table 4.6: Component for ligation **83**
Table 4.7: Pxj41 vector primers for colony screening and for sequencing **84**
Table 4.8: Steps for cloning screening PCR **84**
Table 4.9: PCR1 .... **87**
Table 4.10: PCR2 .... **87**
Table 4.11:PCR3 .... **87**
Table 4.12: Primers was used for overlapping pcr (mutagenesis) del PTBPA **88**
Table 4.13: Primers was used for overlapping pcr (mutagenesis) del PTBPB **88**
Table 4.14: Transfection protocol for ITGA6 minigene: first we mixed (mix1) and (mix2) together and incubated for 10-20 min. After that, we spin it for short time and then transfected to cell. **89**

List of Abbreviations

3’ss	3’ prime splice site
5’ss	5’ prime splice site
AS	Alternative splicing
ASE	Alternative splice events
BP	Bbranch pointe
CLIP	UV Cross-Linking and ImmunoPrecipitation
C-terminus	Carboxy-terminus
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPS	deoxynucleotide triphosphates
EDTA	Ethylenediaminetetraacetic acid
ESRP	Epithelial splicing regulatory protein
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
ECM	Extracellular matrix
FBS	Fetal Bovine Serum
GFP	Green Fluorescent Protein
HeLa	human cervical carcinoma cell line
iCLIP	individual nucleotide resolution CLIP
hnRNP	Heterogeneous nuclear ribonucleoprotein
ISE	Intronic splicing enhancer
ISS	Intronic splicing silencer
KH	K homology
mRNA	Messenger ribonucleic acid
MBNL	Muscleblind-like
MIDAS	Metal ion-dependent adhesion site
MCF-7	Michigan Cancer Foundation-7 human breast cancer cell line
MBNL1	Muscle-blind like protein1
MDA-MB231	MD Anderson human invasive breast cancer cell line
NMR	Nuclear Magnetic Resonance

miRNA	micro RNA
PBS	Phophate Buffered Saline
PCR	Polymerase chain reaction
RGD	Arginine-glycine-aspartic acid
PPT	Polypyrimidine Tract
PSI	Percentage Splicing Inclusion
PTB	Polypyrimidine Tract Binding protein
PPT	Polypyrimidine tract
qPCR	quantitative real-time PCR
QKI	Quaking
RBFOX2	RNA binding protein, fox-1 homolog
RGD	Arginine-glycine-aspartic acid
RNA	Ribonucleic acid
RNAi	Ribonucleic acid interference
RRM	Ribonucleic acid recognition motif
RT-PCR	Reverse transcription polymerase chain reaction
RBP	RNA-binding protein
RRM	RNA recognition motif
RNAi	RNA interference
RNA-seq	RNA sequencing
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
siRNA	Small interfering ribonucleic acid
snRNP	Small nuclear ribonucleoprotein
TBST	Tris-Buffered Saline and Tween-20
TE	Tris/EDTA
UTR	Untranslated region
UCSC	University of California, Santa Cruz

Definition

Paralogues   a pair of genes that derives from the same ancestral gene.

Orthologues   genes in different species that evolved from a common ancestral gene by speciation.

Abstract

Alternative splicing is an important mechanism for creating protein diversity. Integrins are significant in many aspects of cell biology, including cell signalling and interaction with the cell matrix. ITGA6 has two different cytoplasmic C-termini (a6A and a6B) that shift 100% between stem cells and fibroblasts. The primary aim in this thesis was to monitor splicing patterns during development and differentiation integrin subunit alpha 6 (ITGA6) to see which alternative splicing events are similarly regulated in fish and humans using early zebrafish development. The a6A and a6B integrins had been differentially implicated in the expression in the function of breast cancer and cancer stem cells. Therefore, the second aim was to monitor splicing patterns for ITGA6 in different cancer cell lines and to compare them with stem cell patterns, fibroblast, and zebrafish, determining which splicing regulator protein regulates the ITGA6 alternative exon. It was confirmed that the ITGA6 alternative exon 25 was activated by MBNL1, RBFOX2 and ESRP in cancer cell lines, and PTBP was discovered as a novel regulator for ITGA6 splicing that inhibited the exon of ITGA6 in cancer cell line. The third aim for this project was to identify the mechanism of splicing of this ITGA6 alternative exon, including identifying the PTB binding site that regulates ITGA6. A minigene system was established to study the regulation of the ITGA6 alternative exon. The ITGA6 1.3 minigene positively responded to siRNA mediated depletion of splicing factors in the same way as the endogenous gene, indicating this minigene was a good model. The alternative exon of ITGA6 was activated by MBNL1 and was inhibited by PTBP, leading to more production of ITGA6B. Using this minigene plasmid it was confirmed that PTBP inhibited alternative splicing of ITGA6. The last aim of this chapter was to discover the PTB binding sites. Through a series of in silico analyses, a binding site for PTB was identified downstream of the regulated exon. Surprisingly, loss of this PTB binding site actually repressed this splicing event. These data suggest that PTB both activates this alternative splicing event through direct RNA-protein interactions, but also more strongly represses this exon, possibly through protein interactions with other regulatory factors. .

Chapter 1: Introduction

1.1 Pre-mRNA

1.1.1 One gene multiple protein

There are approximately 23,000- 30,000 protein coding genes in the human genome which produce a huge variety of proteins (International Human Genome Sequencing, 2004) (Gerstein et al., 2007). Protein-coding genes numbers are significant when we take into an account that the human genome is estimated to produce over one million different protein species (Nilsen and Graveley, 2010). This means there must be some method to amplify information. This occurs via a series of post-transcriptional mechanisms. After DNA is transcribed to precursor messenger RNA (pre-mRNA), it is modified by splicing, RNA editing, 3'polyadenylation, 5' capping, transcription and termination. These processing steps explain the protein complexity from our relatively modest number genes figure1.1 (Lander, 2011).

In individual cells, gene expression is regulated at the RNA level by modification of the primary transcript (pre-mRNA) which produces multiple mRNA isoforms from a single gene (Kalsotra and Cooper, 2011). This can lead a single gene to encode multiple protein isoforms. mRNA variants can encode functionally and structurally different proteins. Proteins can also be modified via post-transcriptional modifications including phosphorylation, methylation, ubiquitination, glycosylation, etc.) (Wilhelm et al., 2014). Taken together post-transcriptional modifications with subsequent post-translational modifications to proteins leads to the massive surplus of proteins over protein-coding genes.

It has been reported that roughly 95% of human genes are subjected to alternative splicing. Alternative splicing allows each human gene to generate two to three different mRNA isoforms (Djebali et al., 2012). While not all alternative splicing events are of functional importance, alternative spicing is a significant mechanism and key for posttranscriptional gene expression.

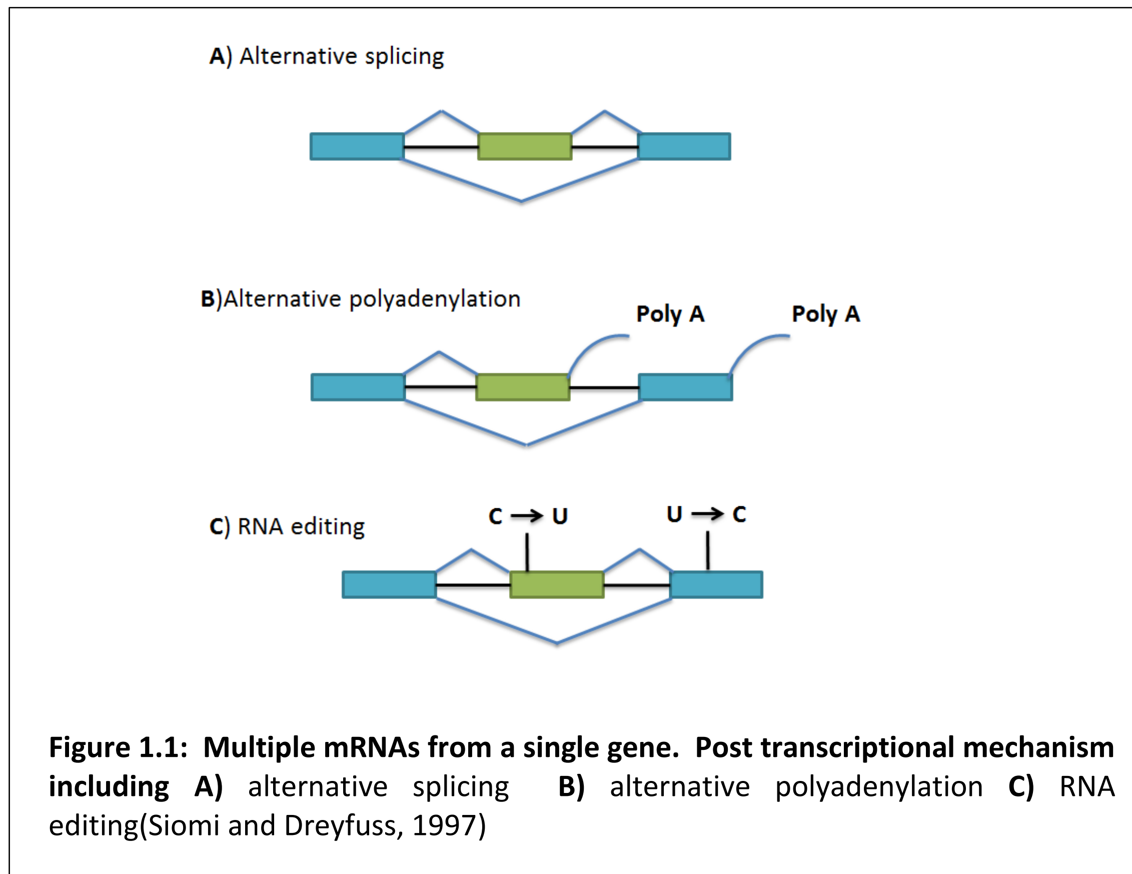


Figure 1.1: Multiple mRNAs from a single gene. Post transcriptional mechanism including A) alternative splicing **B)** alternative polyadenylation **C)** RNA editing(Siomi and Dreyfuss, 1997)

1.1.2 Overview of mRNA

DNA-encoded information in eukaryotic under goes several steps of gene expression. The Initial stage of eukaryotic gene expression is the process of transcription in which the DNA is transcribed to mRNA by RNA polymerase II (Darnell, 2013). All pre-mRNA molecules are subject to a series of processing events, which all known as post-transcriptional modifications, before they leave the nucleus. Post-transcriptional processing steps include capping, splicing and polyadenylation (Mandal et al., 2004). mRNA capping takes place after nascent RNA is produced by RNA polymerase. A methylated Guanine base is added to the 5’ end of the transcript. The function of the cap at the 5’ end is to distinguish between mRNAs and other RNA species (e.g. RNA pol I and III produce uncapped RNAs). Capping also protects RNA molecules from degradation (Mandal et al., 2004).

Eukaryotic genes consist of introns and exons. Splicing removes introns and joins exons in order to produce functional mRNAs. Spliceosomes contain small nuclear

ribonucleoproteins (snRNPs), heterogeneous nuclear ribonucleoproteins (hnRNPs) and a large range of auxiliary RNA-binding proteins (Cramer et al., 2001). The next and often final stage is of mRNA polyadenylation. Polyadenylation comprises two processes, including cleaving mRNA during RNA pol II elongation and adding a poly(A) tail to the 3’ end. The function of polyadenylation is to protect RNAs from degradation and provide its transfer to the cytoplasm for subsequent translation (Stewart, 2010).

mRNA processing is thought to occur cotranscriptionally, in which transcription and processing are not consecutive, but simultaneous. Following the co-transcriptional processing including capping, splicing and polyadenylation, mRNA is exported from nucleus to cytoplasm where it eventually serves as templates for protein synthesis via translation (Stewart, 2010).

1.2 Splicing

Precursor messenger RNA (pre-RNA) splicing is one of the most dynamic processes in eukaryotic cell biology (McManus and Graveley, 2011). Splicing plays a major role in gene expression. Splicing consists of the removal of introns, whilst exons are ligated together in order to form mature protein-coding mRNA (Will and Luhrmann, 2011). Splicing involves a series of reactions that are catalysed by the spliceosome. The spliceosome is one of the most complicated machineries in the cell. There are two type of spliceosome. The major spliceosome is responsible for the removal of introns that harbour consensus splice site sequences. The minor spliceosome targets a less abundant class of non-canonical intron, which differ in splice sites from the consensus ones.

The spliceosome contains approximately 150-300 proteins and five small nuclear RNAs within the U1, U2, U4/U6, U5 and ribonucleoproteins (snRNPs)—all of which combine to create a large complex machine containing RNA and protein (Wahl et al., 2009). The snRNPs designated U1, U2, U4/U6 and U5 are the major spliceosome components required for spliceosomal formation. The 5’ and 3’ splice sites (ss) are required for the assembly of the spliceosome. The snRNAs catalyse the splicing reaction, and direct recruitment to the 5’ and 3’ splice sites (ss) to help define intron/exon boundaries (McManus and Graveley, 2011) (Wahl et al., 2009). In humans, CAG/GURAGU (in which

R=purines and GU is invariant) is the consensus sequence for the 5'ss, whereas YAG/G (in which Y=pyrimidine and AG is invariant) is the consensus sequence at the 3'ss. There are also two other components to the 3' ss, including an upstream polypyrimidine tract (PPT), which has a variable length, and a branch point sequence containing an adenosine which is usually located 18-40 nucleotides upstream from the PPT (Figure 1.2 A) (Will and Luhrmann, 2011).

The 5' splice site and 3' splice site core elements are bound by specific components of the spliceosome. U1 snRNA hybridises with the 5'ss whereas U2 snRNA and the two subunits from U2AF recognise the 3'ss, PPT and branch point (BP) respectively. Splice site selection can be influenced by variations in the splice site sequence; splice sites which closely resemble the consensus sequence are more efficiently recognised by spliceosomal proteins and snRNAs than splice site which have considerable deviations from the consensus sequence figure 1.2B (Will and Luhrmann, 2011).

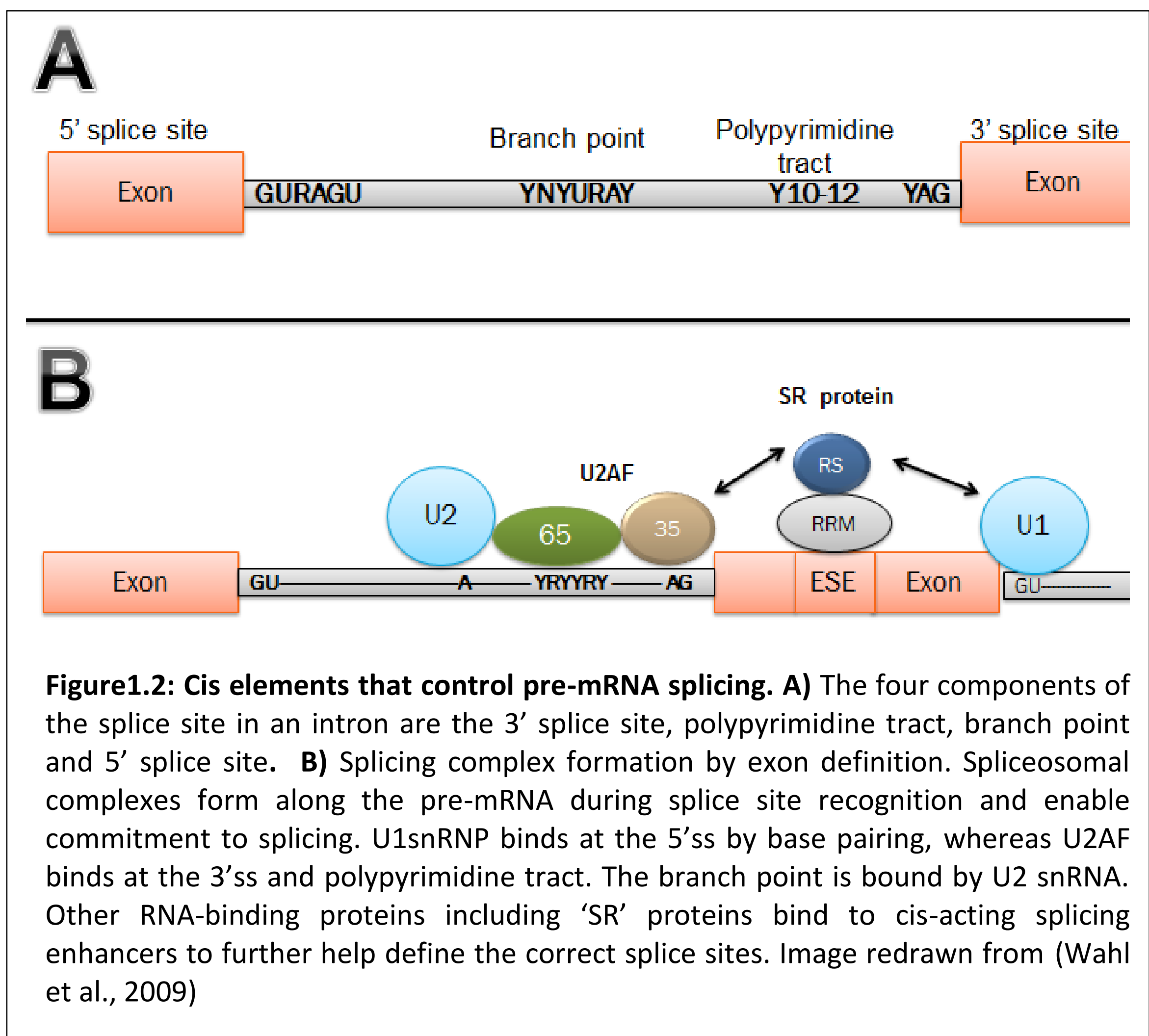


Figure1.2: Cis elements that control pre-mRNA splicing. A) The four components of the splice site in an intron are the 3' splice site, polypyrimidine tract, branch point and 5' splice site. **B)** Splicing complex formation by exon definition. Spliceosomal complexes form along the pre-mRNA during splice site recognition and enable commitment to splicing. U1snRNP binds at the 5'ss by base pairing, whereas U2AF binds at the 3'ss and polypyrimidine tract. The branch point is bound by U2 snRNA. Other RNA-binding proteins including 'SR' proteins bind to cis-acting splicing enhancers to further help define the correct splice sites. Image redrawn from (Wahl et al., 2009)

1.2.1 Splicing mechanism

Three major processes constitue the splicing mechanism: spliceosome assembly on the pre-mRNA, the splicing reaction *per se* and ultimate intron release (Moore and Sharp, 1993). The splicing reaction itself occurs in two steps, both of which are trans-esterification reactions (Will and Luhrmann, 2011). The first trans-esterification reaction occurs due to a nucleophilic attack by the 2'OH group of the conserved adenosine within the branch site, with the conserved guanine of the 5' splice site at the exon-intron junction (Moore and Sharp, 1993). As a result, the exon 1-intron junction is cleaved, and the lariat intermediate which is still linked to exon 2 is formed (Will and Luhrmann, 2011). The second trans-esterification occurs when the 3'OH group of the released upstream exon attacks the last nucleotide of the intron at the 3ss. Consequently, the two exon sequences are ligated together, and the intron sequence is released as a lariat structure figure 1.3 (Moore and Sharp, 1993).

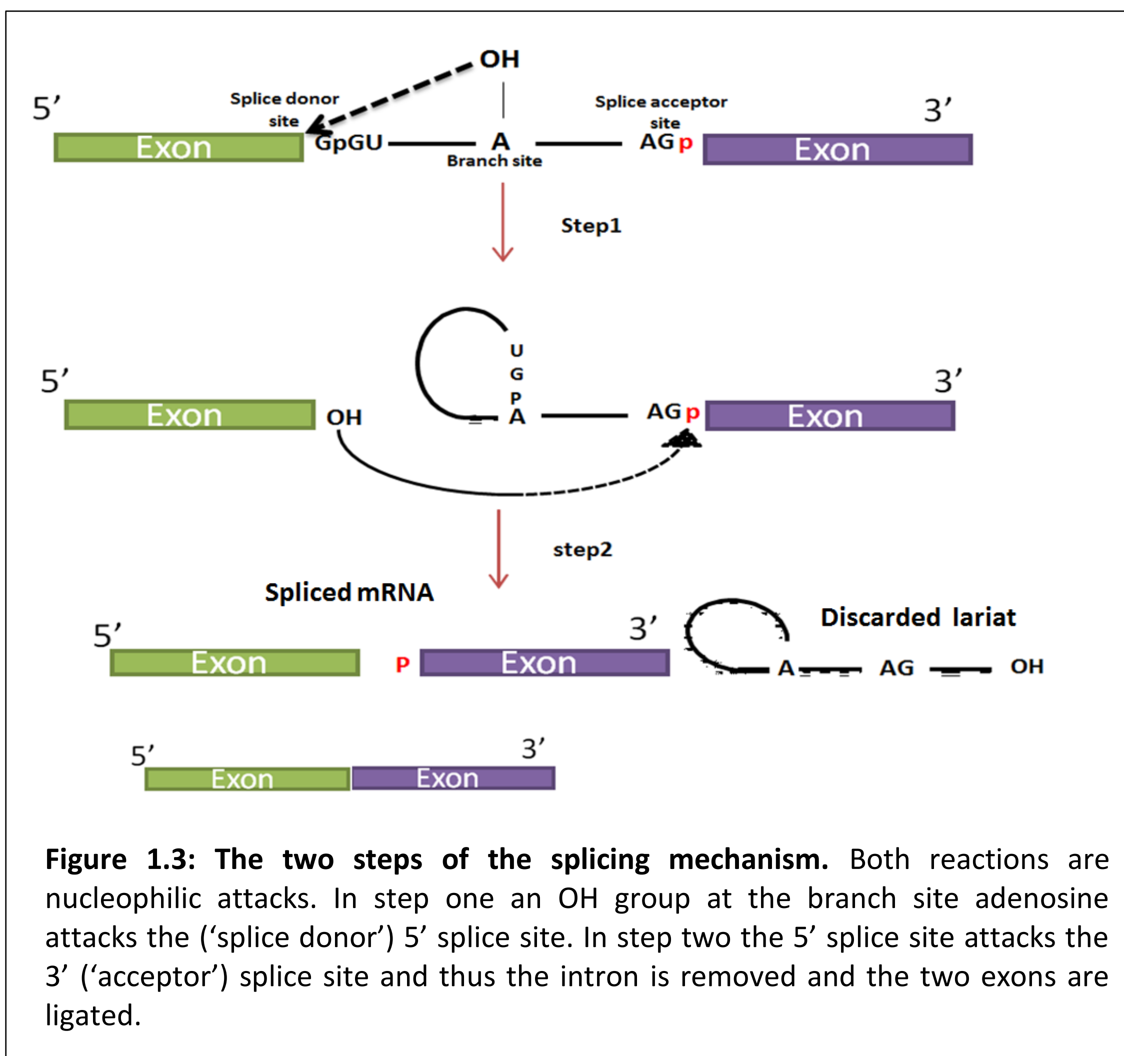


Figure 1.3: The two steps of the splicing mechanism. Both reactions are nucleophilic attacks. In step one an OH group at the branch site adenosine attacks the ('splice donor') 5' splice site. In step two the 5' splice site attacks the 3' ('acceptor') splice site and thus the intron is removed and the two exons are ligated.

The spliceosome consist of five distinct intermediate protein complex process called H, E, A, B and C that assemble sequentially as in figure 1.4 (Reed, 2000, Wahl et al., 2009). Splicing initiates with un-spliced pre mRNA forming H complex, consisting of pre-mRNA bound by the non-specific association of a group of RNA-binding proteins called hnRNPs (Black, 2003). After that, E complex forms in which recognition of the 5' splice consensus splice site is driven by U1 snRNP in an ATP-independent manner. Other RNA- binding proteins (including members of the SR protein family) interact with U1 snRNP (Soller, 2006). In addition to the U1-5'splice site interaction, splicing factor 1/branch point bound protein (SF1/BBP) binds to the branch point and interacts with U2 the auxiliary factor (U2AF) heterodimer. U2 auxiliary factor (U2AF) heterodimer consists of two subunits: a 65kDa subunit (U2AF65) which recognizes the polypyrimidine tract (PPT) and a 35kDa subunit (U2AF35) which recognizes the 3' splice site consensus sequence (Valadkhan, 2007).

E complex is followed by A complex in which U2 snRNP interacts with the branch point. Subsequently, the branch point adenosine is bulged out of the U2snRNP-premRNA duplex and is ready to attack the 5ss. The following step involves U4/U6/U5 tri-snRNA binding to the spliceosome in order to form the B complex, which promotes the first transesterification catalytic step of splicing. Within complex C, in which the second catalytic step also takes place when the upstream 5' splice site ligates to the 3' splice site in order to splice the exons and release the intron lariat intermediate (Wahl et al., 2009).

Overall the U2, U5, U6 snRNPs and other splicing factors are important in splicing reactions as well as other supportive proteins including ATPases, helicases and DExD/H-box proteins which drive spliceosome assembly. When splicing is completed, the lariat intron is degraded, while the snRNPs are recycled and used in the subsequent splicing reactions (Reed, 2000, Soller, 2006, Wahl et al., 2009).

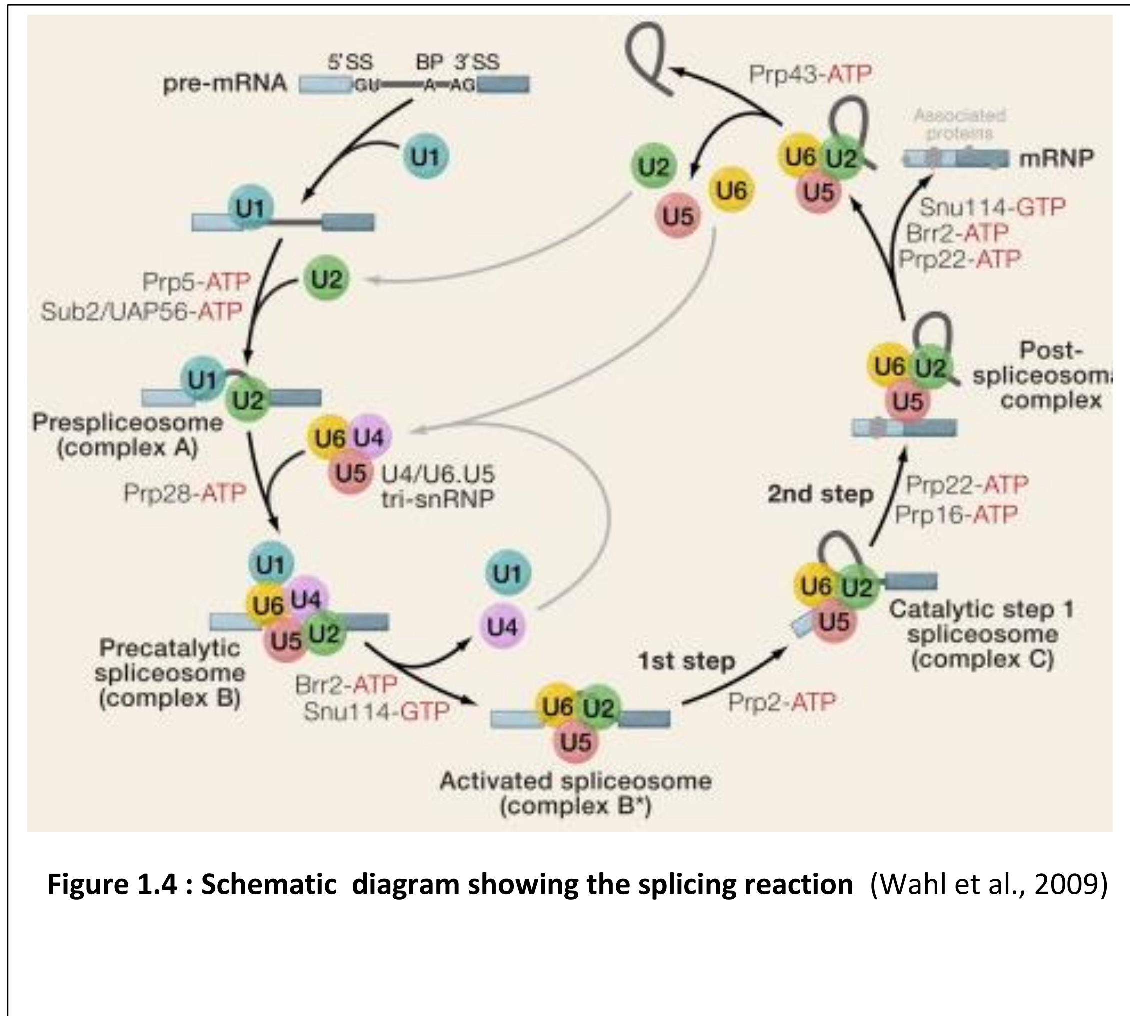


Figure 1.4 : Schematic diagram showing the splicing reaction (Wahl et al., 2009)

1.2.3 The minor spliceosome

Although the majority of eukaryotic introns are removed by the major spliceosome, there are a small number of (non-canonical) introns that are removed by a minor spliceosome. The minor spliceosome consist of four specific snRNPs called U11/U12/U4atac, U6 ATAC and U5 which the latter is the only snRNP shared by both types of spliceosomes (minor and major) (Patel and Steitz, 2003). In general terms, the function of the minor splicisome is similar to that the major spliceosome. The major spliceosomal splices U2-dependent introns which are recognised by U2 snRNP. These introns usually start with a GT nucleotide and finish with AG bases. On other hand, the minor splicisome recognises U12-dependent introns due to intron recognised by U12 type. U12 introns typically start with AT and finish with AC nucleotides (Turunen et al., 2013, Tarn and Steitz, 1996, Yu and Steitz, 1997)

1.2.4 Alternative splicing

The average gene contains approximately 10 introns which need removing from an average length of 120kb, to reduce the size to an average of 2kb mRNA after splicing. This large number of exons and surplus sequence attests to the remarkable fidelity of splicing and also leaves a large margin for variation. Alternative splicing is a process by which RNA diversity is generated by joining different combinations of splice sites from a single gene. Alternative splicing is now known to have a significant function in gene expression. It was estimated that alternative splicing affects 60% of human genes (Srebrow and Kornblihtt, 2006). Since then, with developments in bioinformatics analysis and advanced sequencing techniques, that estimated percentage figure is now increased, such that approximately 95% of multi-exon genes are now thought to undergo alternative splicing (Pan et al., 2008). However the proportion of these alternative events that are truly functional is not yet established.

Changes to splicing patterns play a significant role in gene expression and the function of various proteins. In addition splicing changes can be regulated to play a significant role in facilitating responses to changes in external stimuli, for example in CD44, a transmembrane glycoprotein (Pan et al., 2008). It has been reported that growth factor signalling has an effect on production of CD44 isoforms, specifically inclusion or exclusion of variable exon V6, and that this creates a positive feedback as the V6 containing isoform is more active as a co-receptor for growth factor (Cheng et al., 2006).

Exon skipping is one of the most prevalent patterns to occur during splicing (Figure 1.5B) (Reddy, 2007). Like all forms of alternative splicing, exon skipping can lead to a change in the reading frame and a truncated protein, or the change can be an in frame removal/addition of a number of amino acid residues (Srebrow & Kornblihtt, 2006). Another kind of alternative splicing involves mutually exclusive exons; this splicing pattern occurs when one exon is retained and one exon is skipped, leading to the generation of two different isoforms (Figure 1.5 C)(A. S. Reddy, 2007). The retention of an intronic sequence is a relatively rare form of AS in vertebrates but it too can occur due to the silencing of splice sites within the intron (Srebrow & Kornblihtt, 2006). Yet

another pattern of alternative splicing involves alternative 5' or 3' splice sites. This leads to deletion of part of the 3' end of the upstream exon or part the 5' end of the downstream exon (Pan et al., 2008) (Figure 1.5 E,F).

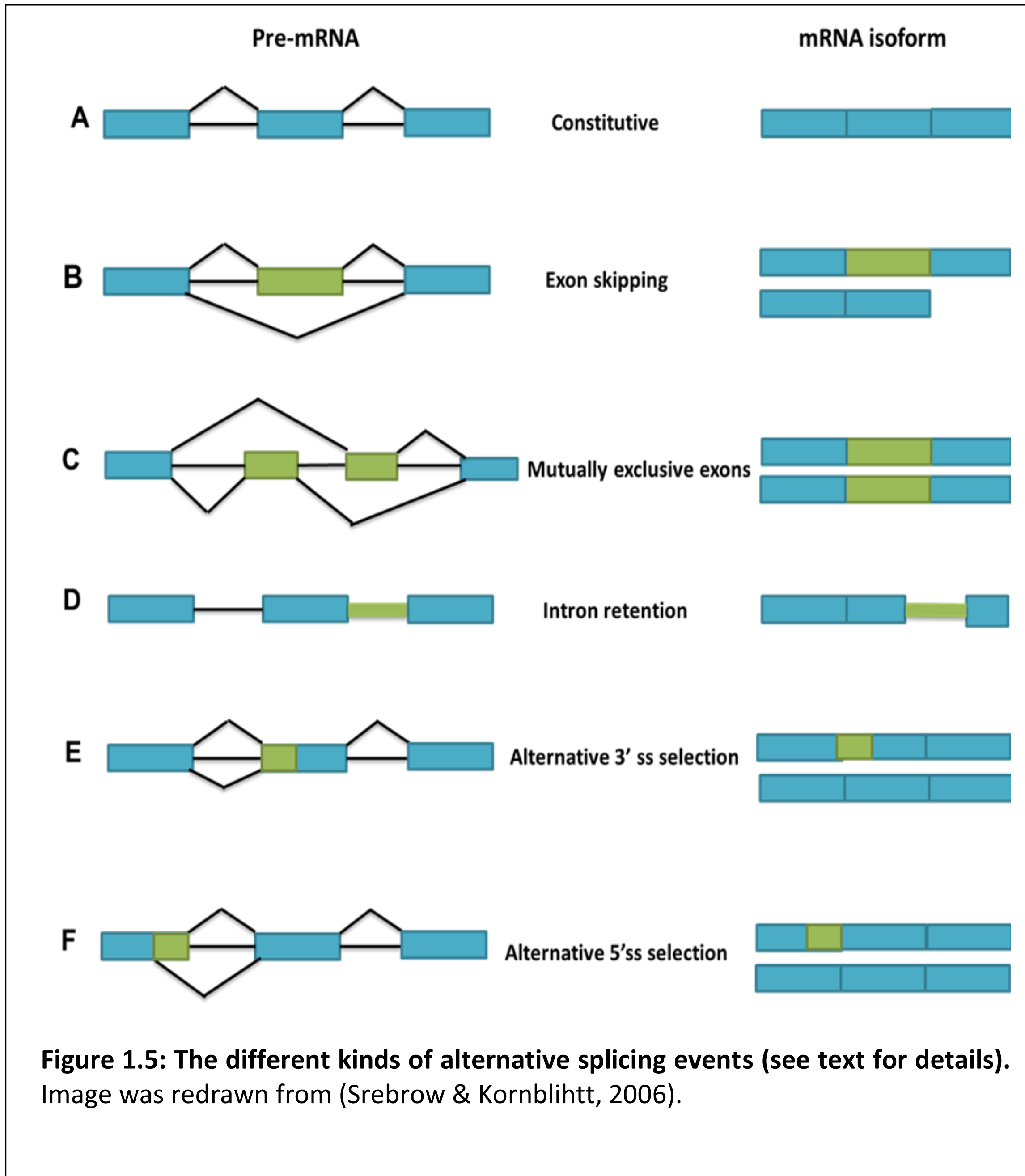


Figure 1.5: The different kinds of alternative splicing events (see text for details). Image was redrawn from (Srebrow & Kornblihtt, 2006).

1.3 RNA-protein interaction:

RNA-binding by RNA-binding proteins (RBPs) regulate the metabolism of RNA in eukaryotes (Anko and Neugebauer, 2012). RBPs contribute to all nuclear mRNA processing steps including capping, splicing, and polyadenylation as well as mediating the transport of mRNAs from the nucleus to the cytoplasm and regulation of their translation. RBPs form large ribonucleoprotein complexes (RNPs) on the RNA in order to regulate different steps of RNA metabolism. RNPs are classified according to whether they bind in the nucleus or in the cytoplasm and are thus called heterogeneous nuclear RNPs (hnRNPs) ,which also form on pre-mRNA, or messenger RNPs (mRNPs) respectively (Janga and Mittal, 2011).There are a variety of motifs that bind to RNA including RNA recognition motifs (RRMs), K homology (KH) domains and zinc fingers which help to facilitate and control the localization and processing of RNA. Figure 1.6 summarises the mechanisms controlled by RBPs in a cell (Janga & Mittal, 2011).

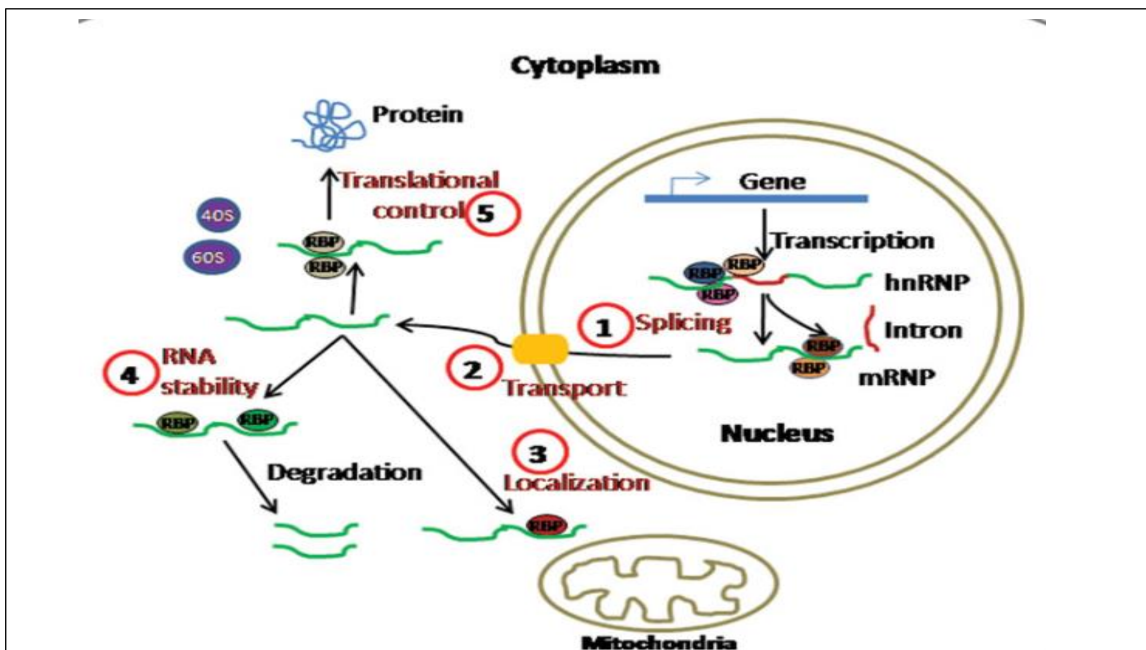


Figure 1.6: RNA-binding protein processing in various post-transcriptional processes at different locations in eukaryotic cells. RBPs are involved in 5 major processes that are represented by red circles. 1) RBPs play a major role in splicing of mRNA in the nucleus. 2) After that mRNA export from nucleus to the cytoplasm by various RBPs. In cytoplasm, 3) RBPs contribute to mRNA localization in compartments such as the mitochondria. 4) RBPs are also responsible for RNA stability and lead to degradation. 5) RBPs are also involved in translational control. Image adapted from (Janga and Mittal, 2011).

1.4 The regulation of splicing

There are several factors that affect splice site choice, including cis-acting elements and trans-acting factors (Smith and Valcarcel, 2000). Splicing patterns are affected by the expression of RNA-binding proteins that can be expressed at different stages of development and in different tissues. Other factors which can directly affect the splicing outcome include the rate of transcriptional elongation, epigenetic factors and the formation of secondary structures within RNA. Secondary structure, which forms within the pre-mRNA, has a significant impact on splicing activation. It can have a similar effect on splicing as the way that RNA-binding proteins prevent sequence recognition (Warf and Berglund, 2010). Secondary structures of RNA can also exert effects on splicing by changing the distance between auxiliary elements and splice sites. Epigenetic factors, including histone modifications, can influence splicing decisions. Histones interact with the snRNP as a complex during co-transcriptional spliceosome assembly and this facilitates the correct assembly of the pre-spliceosome (Luco et al., 2011).

Although the regulation of alternative splicing is complicated, the improvement and development of technology including high throughput sequencing and advances in bioinformatics analysis have facilitated the identification of splicing regulatory motifs (splicing code), which includes all binding sites for RNA-binding splicing repressors and activators within a pre-mRNA, as well as the splice sites themselves.

1.5Cis elements: splicing enhancers and silencers

1.5.1 Strength of splice site and position

Strength of splice sites can be quantified as the degree of complementarity between splice sites and U1 and U2 snRNAs. In general, constitutive exons which are always included into the mature mRNA transcript have strong splice sites whereas alternative exons are flanked by weak splice sites. The development computer programs have facilitated the predication of the strength of splicing sites (Smith and Valcarcel, 2000).

1.5.2 Cis- elements activators and repressors

Although the strength of splice sites has significant effect on splice site choice, cis-acting sequences near regions of alternative splicing are also extremely important in splice site choice. These sequences can form binding sites for trans-acting factors (Smith and Valcarcel, 2000). Cis-acting elements include splicing enhancers, which can be either exonic splicing enhancers (ESE) or intronic splicing enhancer (ISE); and silencers, which can be either exonic splicing silencers (ESS) or intronic splicing silencers (ISS). Generally, enhancers support weak alternative splice sites and assist spliceosome and trans-acting factors. ESS and ISS recruit splicing repressor proteins. Both silencer and enhancer sequences play a significant role in recognising both constitutive and alternative exons figure 1.7 (Matlin, Clark, & Smith, 2005)(Martinez-Contreras et al., 2007).

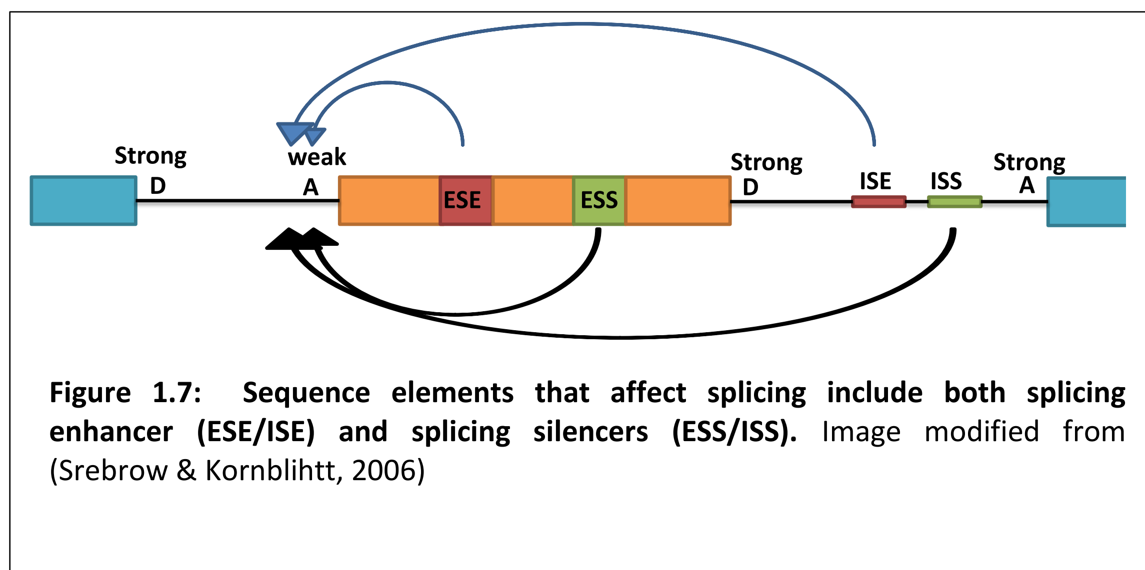


Figure 1.7: Sequence elements that affect splicing include both splicing enhancer (ESE/ISE) and splicing silencers (ESS/ISS). Image modified from (Srebrow & Kornblihtt, 2006)

1.5.3 Trans-acting RBPs

Trans-acting RNA-binding protein factors can significantly affect splicing depending on the concentration and localisation of the RNA-binding protein. One group of trans-acting factors acting at cis elements are the SR protein family that play a major role in binding splicing enhancers (acting at ESEs) in a concentration- and phosphorylation-

dependant manner (Smith & Valcarcel, 2000). Long non-coding RNA (IncRNA) have also been reported to influence splicing by modulating the phosphorylation of SR-proteins (Tripathi et al., 2010). The two most recognised families of trans-acting factors that have an effect on recognition of splicing silencers or enhancers are the SR proteins and heterogeneous ribonucleoproteins (hnRNPs) (Tripathi et al., 2010). These proteins are expressed at varying levels in different tissues. In contrary to SR proteins, hnRNPs are mostly seen as splicing regulators that bind RNA to repress splicing (although there are many exceptions). HnRNP proteins can contain RRM domains, RGG (arginine-glycine-glycine) domain and KH (hnRNP K homology) domains as well as others. The presences of hnRNPs on the RNA affect the ability of the core splicing machinery to engage with splice sites (Martinez-Contreras et al., 2007).

SR proteins have RRM regions that bind to RNA and auxiliary regions rich in RS (arginine-serine) dipeptides that are thought to be mostly involved in protein-protein interactions that assist assembly of functional splicing complexes (Tripathi et al., 2010). When SR proteins bind to sequences within exons, they act to recruit components of the spliceosome and thus to activate splicing. There is no clear distinction, however, as RNA-binding proteins of either class (hnRNP or SR) can act as activators or repressors which either inhibit assembly or activity of spliceosomal components at the splice site. Furthermore, an individual SR protein or hnRNP can act as a repressor or as an activator depending on its binding location in an alternative exon (Martinez-Contreras et al., 2007).

1.6 Other factors that affect splicing

1.6.1 Intron/exon structure

There are other factors which might affect splice site recognition including the structure of introns and exons (Sterner et al., 1996). Effects of Introns and exons on splice site recognition can be based on their size. When intron size is small, the splice sites are recognised across the intron. This is called intron definition. When the intron is long, the splice sites are recognised across the exon, this is called exon definition. There are a variety of intron size including eukaryotes and human cells. The average

exon size in the human is roughly 170 base pairs, whereas the average intron size is 11,000 base pairs (Sakharkar et al., 2004) (Berget, 1995, Sterner et al., 1996).

1.6.2 RNA secondary structure

RNA structure has significantly impact on splicing. Usually RNA is single stranded; however it can fold back in order to adopt secondary and tertiary structures that can involve up to hundreds of nucleotides (Buckanovich and Darnell, 1997). RNA structure can affect the outcome of pre-mRNA splicing by inhibiting or activating spliceosomal assembly. Secondary structures forming within the pre-mRNAs, can have an effect on splicing similar to RNA-binding proteins so on to prevent sequence recognition (Warf & Berglund, 2010). RNA secondary structures also exert an effect on splicing by changing the distance between auxiliary elements and the splice sites such as survival motor neuron 2 (SMN2)(Buckanovich and Darnell, 1997).

1.6.3 Other factors effect on splicing

Transcription is another factor that affects splicing (Wang and Cooper, 2007). Splicing is influenced by the speed of transcription, whether the elongation is rapid or slow. If the elongation of transcription is rapid, alternative exons will not be recognised before downstream competing exons are transcribed, and so will be skipped. If transcription is slow, alternative exons may be recognised by spliceosomes and will be included. Epigenetic factors, including histone modifications, can influence splicing decisions through affecting transcription. Histones interact with snRNP complexes during co-transcriptional spliceosome assembly and this facilitates the correct assembly of the pre-spliceosome (Luco, Allo, Schor, Kornblihtt, & Misteli, 2011).

Although many RNA-binding proteins are ubiquitously expressed, a variety of tissue-specific RNA-binding splicing regulators have been identified including neuro-oncological ventral antigen NOVA1, polypyrimidine tract-binding protein PTBP1, FOX protein (RBFOX) and muscleblind (MBNL) protein (Singh et al., 2004). To add further complexity, most of these important splicing factors fall into multi-gene families made up of several paralogues.

1.6.3.1 PTB family proteins

Polypyrimidine tract-binding protein (PTBP1) and its homologues form one of the most significant RNA-binding protein families that have been studied in mammals (Valcarcel and Gebauer, 1997a). The prototypical member is known as PTBP1 or HnRNPI, and generally functions as a splicing repressor. PTB proteins also function in a large number of diverse cellular processes including polyadenylation, mRNA stability, mRNA localization and translation (Sawicka et al., 2008b). PTBP binds to splicing silencers, at pyrimidine rich motifs such as UCUU or CUCUCU in the RNA, to mediate splicing repression of many alternatively spliced pre-mRNAs (Noiret et al., 2012). PTB sites are highly overrepresented upstream of brain-specific exons, implying that down-regulation of PTB in brain depresses these exons and causes their brain-specific inclusion (Castle et al., 2008, Cheung et al., 2009). In some cases, PTB repression is mediated by a single silencer; however, in the majority of cases, multiple PTB binding sites are present. In vertebrates, PTBP1 is widely expressed and belongs to a family of RNA-binding proteins including two paralogs, PTBP2 and PTBP3, which are expressed in a more tissue-restricted manner (Noiret et al., 2012). PTBP2 (previously known as nPTB or brPTB) is expressed mainly in neurons, but also in testis and at lower levels in skeletal muscle. On the other hand, PTBP3 (also known as ROD1) is mainly expressed in embryonic and adult hematopoietic organs (Sawicka et al., 2008). The three PTBP paralogues have a >70% amino acid sequence identity and a common arrangement of four RRM-type domains (Noiret et al., 2012). Furthermore, they have a functional link in the sense that they act as repressors of at least some exons (Noiret et al., 2012).

1. 6.3.2 RBFOX family proteins:

Another significant RNA-binding proteins that play a major role in alternative splicing regulation are called RBFOX (Gallagher et al., 2011). Like the PTBP proteins, The RBFOX family has three members: RBFOX1, RBFOX2 and RBFOX3, that have similar RNA-binding domains and are conserved between flies and men (Damianov and Black, 2010). The RBFOX proteins are among the most sequence-specific RNA-binding proteins known, binding to the hexanucleotide UGCAUG. RBFOX proteins display the

classic asymmetric splicing activity of many splicing factors; alternative exons are repressed by RBFOX-binding directly upstream of the exon and RBFOX proteins enhance exon recognition when binding downstream of the exon (Venables et al., 2009). Thus RBFOX splicing regulation can be exquisitely controlled by placement of its binding sites either upstream or downstream of alternative exons. RBFOX proteins regulate alternative splicing in a tissue-specific manner including in neurons and muscle where they are highly expressed. RBFOX2 has also been defined to regulate splicing in embryonic stem cells and embryos (Venables et al., 2013). RBFOX has been reported to play a significant role in disease including cancer and autism by affecting alternative splicing. Comparing RBFOX1 expression in normal brain tissues and autism samples, it was concluded that RBFOX1/2 are likely implicated in Autism (OMIM database *605317 and The Autism Genome Project). RBFOX also has been reported to have a significant role in cancers including breast and ovarian cancer where a large program of alternative splicing was found concomitant with down regulation of RBFOX2 (Venables et al., 2009)

1.6.3.3 The 'Muscleblind' MBNL family:

Muscleblind (MBNL) is another of the best characterised and most important RNA-binding proteins (RBPs) that plays a major role in regulating alternative splicing (Ho et al., 2004). Like the other RBPs discussed above, the MBNL family has three members: MBNL1, MBNL2 and MBNL3, which have nearly identical RNA-binding domains (Ho et al., 2004). Like RBFOX, MBNL acts as an activator or as a repressor based on its binding position (downstream or upstream of the exon respectively). MBNL regulates AS in specific tissues including skeletal muscle, cardiac muscle and the nervous system (Han et al., 2013). It has also recently been shown that MBNL down regulation is essential for embryonic stem cell derivation and subsequent MBNL upregulation is needed for re-differentiation (Han et al., 2013; Venables et al., 2013). MBNL plays a significant role in diseases including muscular dystrophy. (Fardaei et al., 2002) appear that there was a connection between MBNL and Myotonic dystrophies (DM1) and (DM2). DM1 occurs primarily due to an expansion of CTG repeats in DMPK 3'UTR whereas DM2 results from an intronic CCTG expansion in another gene called ZFN9. Muscleblind has

significant impact in both DM1 and DM2 due to its sequestration at CUG expanded repeats which localise to nuclear foci in both diseases (Fardaei et al., 2002; H. Jiang, Mankodi, Swanson, Moxley, & Thornton, 2004). The resulting reduction in the normal function of MBNL1 is thought to underlie DM1 pathology.

1.6.3.4 Epithelial Splicing Regulatory Proteins (ESRPs)

Epithelial Splicing Regulatory Proteins (ESRPs) are RNA-binding proteins that have significant roles in regulating alternative spicing events associated with epithelial cells (Ishii et al., 2014). The ESRP family includes ESRP1 and ESRP2 which are highly conserved paralogs containing three RNA Recognition Motif (RRM) domains (Warzecha et al., 2010a). ESRP1 and ESRP2 are involved in epithelial-mesenchymal transition (EMT) by maintenance of epithelial cell-specific isoforms. ESRP1 and ESRP2 down regulation are essential for EMT progression in a mammary epithelial cell line. It has been reported that ESRPs bind to UG-rich motifs. ESRPs regulate alternative splicing expression in epithelial cell and cancer cells (Ishii et al., 2014).

1.7 Alternative splicing in ITGA6 Integrins

Previous results by Julian Venables data ((Venables et al., 2013b). showed that both MBNL and RBFOX control ASE in 15 genes including ITGA6 during stem cell differentiation. My project focused on ITGA6 protein to study in depth how alternative splicing is regulated RNA-binding proteins. In this section I will review the latest finding an integrin including the alternative splicing of ITGA6.

1.8 Integrins

Integrins are a large family of heterodimeric cell surface glycoprotein receptors that mediate the attachment between a cell and its surroundings including the extracellular matrix (ECM) (Barczyk et al., 2010). Integrin dimers are formed by the association of an alpha and a beta subunit. There are 18 known α subunit genes and 8 known β subunit genes (de Melker and Sonnenberg, 1999, van der Flier and Sonnenberg, 2001). Integrin heterodimers are formed by non-covalent association of α and β subunits that each straddles the cell membrane through a single-pass, type I, transmembrane protein domain (de Melker & Sonnenberg, 1999). Each integrin subunit consists of three major domains: the large extracellular domain, the single membrane-spanning transmembrane domain and the intracellular cytoplasmic tail domain (van der Flier & Sonnenberg, 2001). The integrin extracellular domain functions by binding proteins in the extracellular environment, while the intracellular cytoplasmic tail domain interacts with the intracellular environment. The integrin family plays a significant role in controlling biological and cellular functions such as cell adhesion, migration, proliferation, cell differentiation and apoptosis by assembling the actin cytoskeleton inside the cell and modulating the signal transduction pathway arriving from outside the cell (van der Flier & Sonnenberg, 2001).

There are various ways of categorizing integrins based on either their ligand-binding properties or their subunit compositions. The extracellular domains of the α and β integrins play major roles in the specificity of integrin binding to the extracellular matrix (ECM) components (i.e. collagens, fibronectins and laminins) (Figure 1.8) (Barczyk et al., 2010). Table 1.1 shows the ligand specificity of different integrins. One subfamily of the integrins, which includes a1b1, a2b1, a10b1 and a11b1, are the major collagen receptors, whereas the integrin a3b1, a6b1, a6b4 and a7b1 subunits represent the primary laminin receptors (Johnson et al., 2009). The major fibronectin receptors are formed by integrins, a5b1, a8b1, aIIbb3, and the avb. This group binds the arginine-glycine-aspartic acid (RGD) cell adhesion sequence, which commonly appears in extracellular matrix (ECM) 15 components. Nevertheless, there are certain integrins that bind to the same extracellular ligands with different affinities, e.g., integrin α1β1, which binds to collagen and laminins (Johnson et al., 2009).

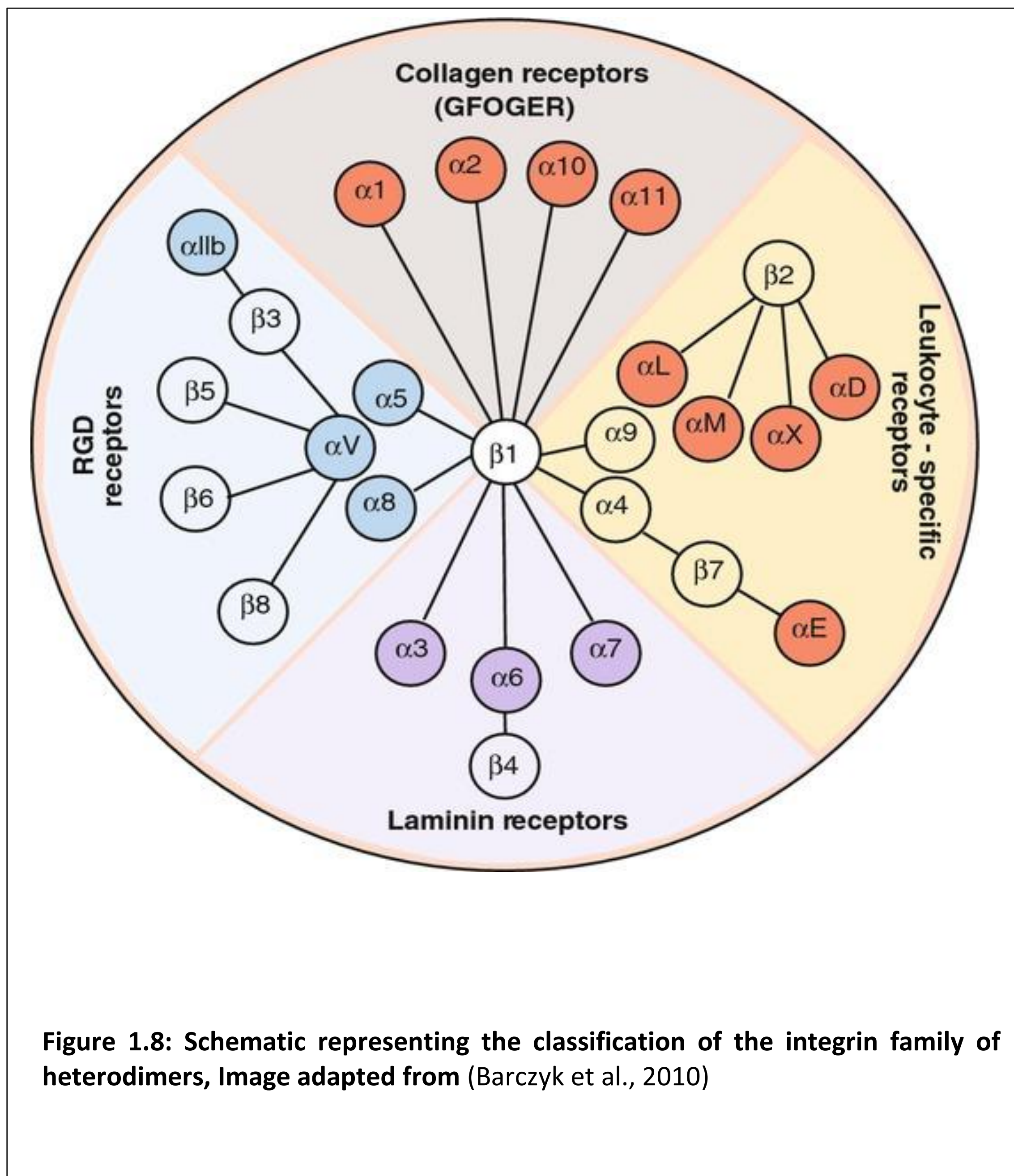


Figure 1.8: Schematic representing the classification of the integrin family of heterodimers, Image adapted from (Barczyk et al., 2010)

Family	β subunit	α subunit	αβ complex	Receptor	I domain	Ligand	RGD recognition	Distribution
VLA proteins	β_1	α_1	$\alpha_1\beta_1$	VLA-1	+	–	?	Ubiquitous
		α_2	$\alpha_2\beta_1$	VLA-2	+	L, C	?	Ubiquitous
		α_3	$\alpha_3\beta_1$	VLA-3	−	C, Fn, entacin	?	Ubiquitous
		α_4	$\alpha_4\beta_1$	VLA-4	−	Fn, VCAM-1	−	Ubiquitous
		α_5	$\alpha_5\beta_1$	VLA-5	−	Fn	+	Ubiquitous
		α_6	$\alpha_6\beta_1$	VLA-6	−	L	RGD	Ubiquitous
		α_7	$\alpha_7\beta_1$					
		α_8	$\alpha_8\beta_1$			Fn, Vn, Tn		
		α_9	$\alpha_9\beta_1$			Tn		Epithelial cells
						Fn, Vn		Muscle cells
		α_v	$\alpha_v\beta_1$				+	
Leukocytes proteins	β_2	α_L	$\alpha_L\beta_2$	LFA-1	+	ICAM-1, -2, -3		Leukocytes
		α_M	$\alpha_M\beta_2$	MAC-1	+	C3b, Fb, ICAM-1	−	Neutrophil, monocyte, LGL
		α_X	$\alpha_X\beta_2$	P150, 95	+	C3b, LPS	−	M, monocyte
		α_D	$\alpha_D\beta_2$		+	ICAM-3	−	M
	β_7	α_4	$\alpha_4\beta_7$					
		α_H	$\alpha_H\beta_7$	LPAM-1		Fn, VCAM-1		Lymphocytes
		α_E	$\alpha_E\beta_7$		+	E-cadherin, adhesion		Lymphocytes
Cytoadhesin	β_3	α_v	$\alpha_v\beta_3$	Vitronectin	−	Vn, Fb, Fn, OPN	+	Ubiquitous
						VWF, Tn, Thr		
		α_{II}b	$\alpha_{II}b\beta_3$	GpIIb/IIIa		Fb, Fn, VWF, Vn	+	Platelets
Other combination	β_5	α_v	$\alpha_v\beta_5$			Vn, Fn	+	Ubiquitous
	β_6	α_v	$\alpha_v\beta_6$			Fn, Tn	?	Lung-epithelial cells
	β_8	α_v	$\alpha_v\beta_8$			Vn	+	
	β_4	α_6	$\alpha_6\beta_4$			L	+	Epithelial cells

Table 1.1: Classification of integrins. From (Reddy and Mangale, 2003**).**

1.8.1 Integrin structure

1.8.1.1Extracellular domains

Integrin extracellular domains consist of more than 700 amino acid residues for the α subunit and 1000 amino acid residues in their β subunits (Reddy and Mangale, 2003). The subunits are organised in order to form a globular ligand-binding N-terminal head that ‘stands’ on two long and extended C-terminal ‘legs’ (Reddy and Mangale, 2003). The C-terminal legs link to the cytoplasmic and transmembrane domains of each subunit figure 1.9A (Barczyk et al., 2010). The extracellular domain of the α subunit consists of seven repeated homologous segments (of approximately 60 amino acids) that fold into a seven-bladed β propeller ‘head’ domain and two ‘calf’ domains (van der Flier & Sonnenberg, 2001). In addition, half of the α subunits have an extra independently folding domain termed the I ’inserted’ or ‘A’ domain, of approximately 200 amino acids (van der Flier & Sonnenberg, 2001). The I-domain contributes to ligand binding, and similar domains are also found in trimeric G protein α subunits and in small G proteins (Reddy and Mangale, 2003).

The extracellular parts of the β subunits also contain an I-like domain and these too play a significant role in integrin ligand binding. I-domains, which are found in both α and β subunits, have a conserved “metal ion-dependent adhesion site” (MIDAS). The MIDAS contributes to protein ligand binding with divalent metal cations (Mg2+). Integrin ligand binding alters the coordination of the metal ion and also shifts the I-domains from a resting to an open and active conformation; increasing ligand affinity and integrin activation occurs as a result (Reddy and Mangale, 2003) figure 1.9B. Taking together, the divalent cations binding sites appear in the C-terminus in four or three repeats and these appear to mediate integrin ligand binding (van der Flier and Sonnenberg, 2001).

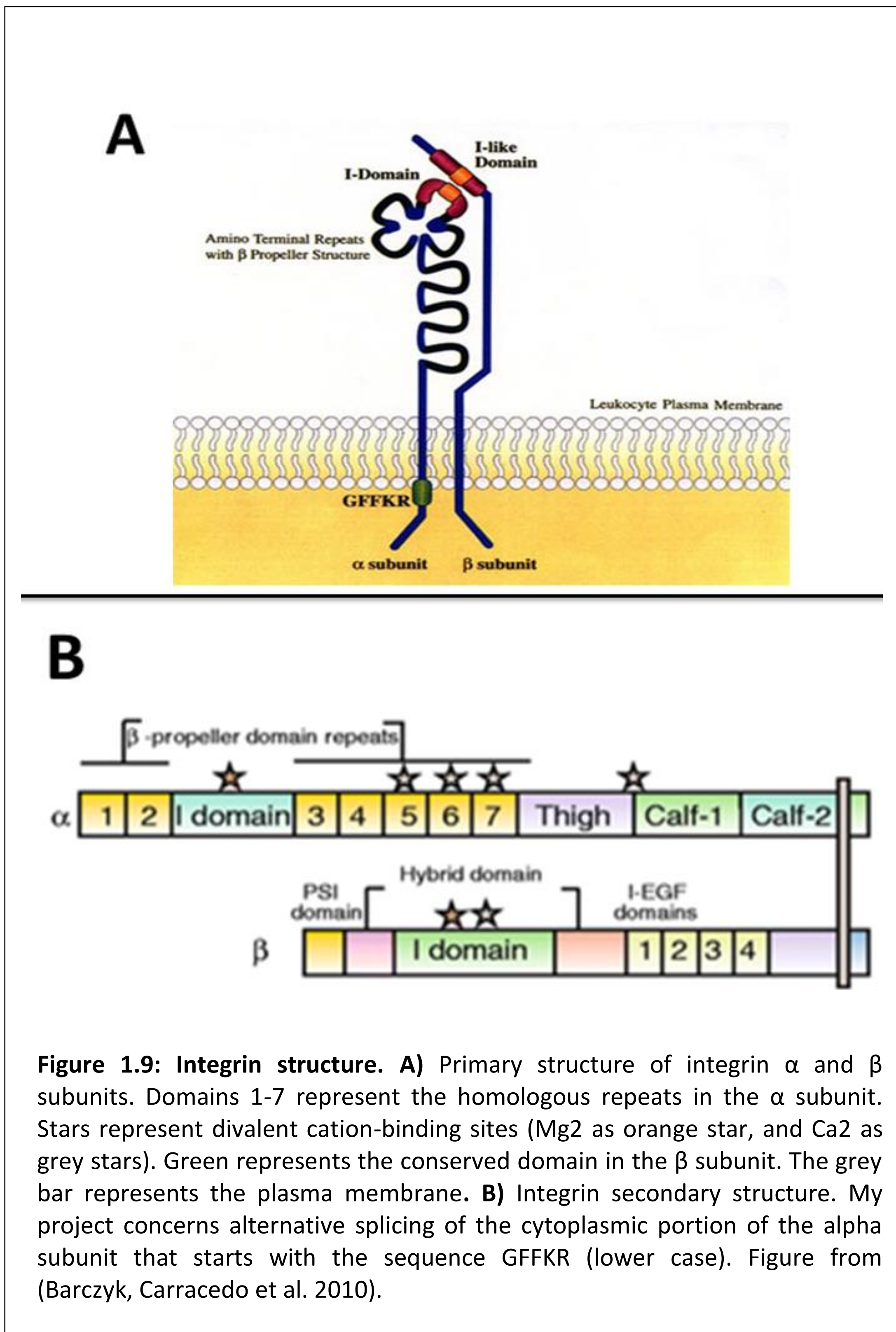


Figure 1.9: Integrin structure. A) Primary structure of integrin α and β subunits. Domains 1-7 represent the homologous repeats in the α subunit. Stars represent divalent cation-binding sites (Mg2 as orange star, and Ca2 as grey stars). Green represents the conserved domain in the β subunit. The grey bar represents the plasma membrane**. B)** Integrin secondary structure. My project concerns alternative splicing of the cytoplasmic portion of the alpha subunit that starts with the sequence GFFKR (lower case). Figure from (Barczyk, Carracedo et al. 2010).

1.8.1.2Cytoplasmic domains

The integrin cytoplasmic domain is generally much shorter than the extracellular domain, being about 20-50 residues in the α subunits and variable in the β subunits between 15 to 65 amino acid residues with the exception of β4 subunits which consists of approximately 1000 amino acids (Calderwood et al., 2003).The crystal structure of the cytoplasmic domain has not been resolved at high-resolution, but the cytoplasmic domain structure has been determined by nuclear magnetic resonance (NMR). Integrin β cytoplasmic tails are highly homologous, while integrin α cytoplasmic tails are strikingly divergent (Berman et al., 2003). GFFKR and HDR(R/K)E are conserved sequences next to the transmembrane region that form a salt bridge between the α subunit at arginine (R) and a β subunit at aspartic acid (D) respectively to help assemble and stabilise the heterodimer (Sastry and Horwitz, 1993). Within β subunit tails, there are two well-defined motifs, including a membrane proximal NPxY and a membrane distal of the NxxY motif (de Melker & Sonnenberg, 1999). These motifs bind proteins that contain a phosphotyrosine–binding domain and they act as binding sites for multiple integrin binding proteins, such as talin and the kindlins figure 1.10 (de Melker & Sonnenberg, 1999).

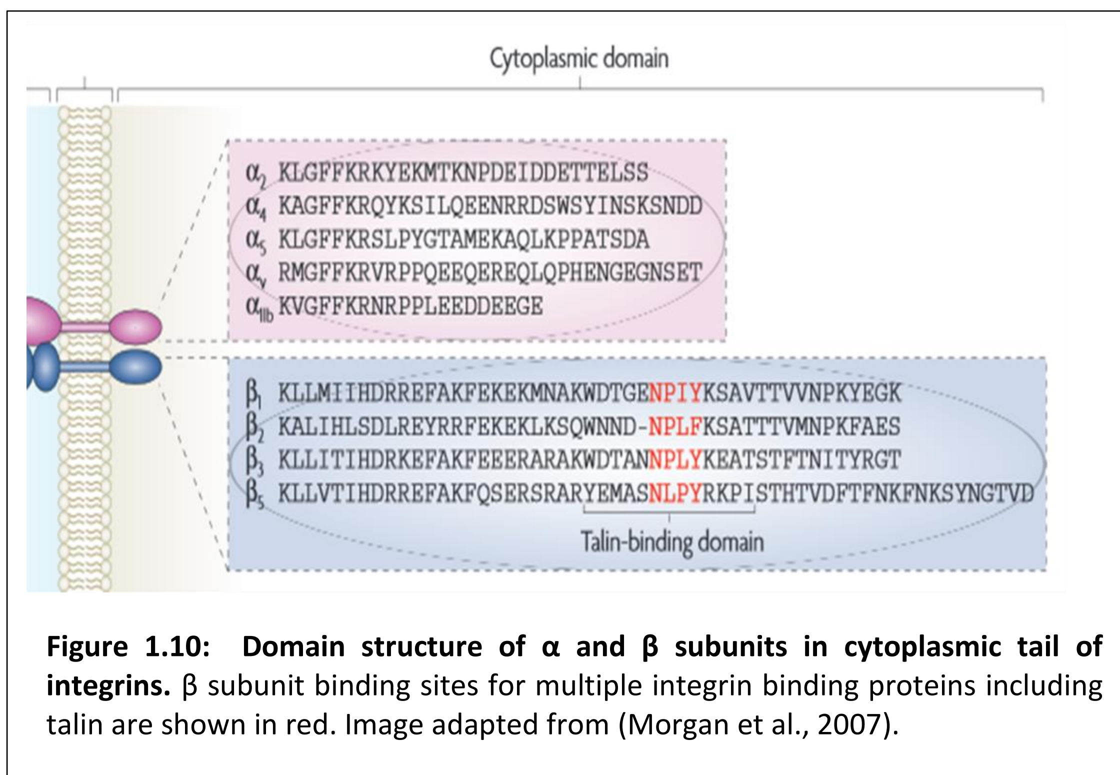


Figure 1.10: Domain structure of α and β subunits in cytoplasmic tail of integrins. β subunit binding sites for multiple integrin binding proteins including talin are shown in red. Image adapted from (Morgan et al., 2007).

1.8.3 Integrin cytoplasmic domains and cell signalling

It has long been known that integrins function as receptors which form a link between the ECM and cytoplasm of a cell. Integrins are involved in various functions, including signal transduction (Sastry & Horwitz, 1993). Integrins have advantages over other cell receptors in that they are able to shift between high and low-affinity ligand binding states. Integrin cytoplasmic domains play major roles in transducing two types of signalling (Calderwood et al., 2003; Sastry & Horwitz, 1993). The first type, which is called outside–in signalling, transmits signalling from ECM to the cell. This occurs due to integrin binding to extracellular ligands which lead to changes in the conformation of the integrin. Many ligands are multivalent, this causes to Integrins to cluster (Askari et al., 2009). Taken together, these events combine and lead to intracellular signals. This has a significant impact on modulation of cellular responses, including gene expression, proliferation and cytoskeletal organization (Askari et al., 2009). The second type of signalling, which is called inside-out signalling, plays a major role in cell physiology. Inside-out signalling affects the specificity and affinity of integrins for their extracellular ligands (Sastry & Horwitz, 1993). This occurs due to conformational changes that are induced in the integrin heterodimer. Integrin cytoplasmic domain tails bind with a huge number of cytoskeletal and signalling proteins including talin and kindlin which play a major role in separating the cytoplasmic tails which leads to integrin activation(Luo, Carman, & Springer, 2007). Figure 1.11 summarises the mechanism of integrin signalling.

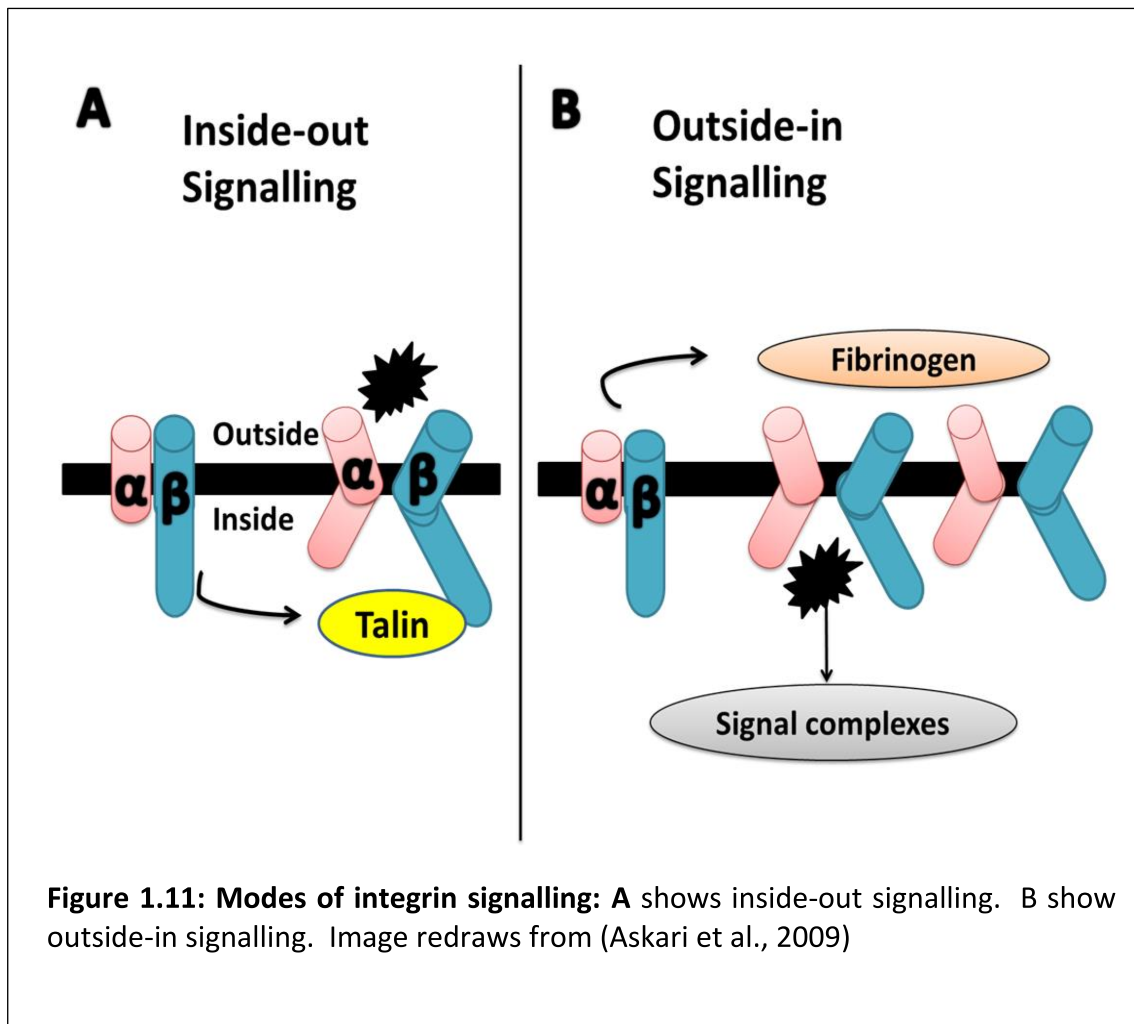


Figure 1.11: Modes of integrin signalling: A shows inside-out signalling. B show outside-in signalling. Image redraws from (Askari et al., 2009)

1.8.4 α subunits in the cytoplasmic domain:

Integrins are composed of α and β subunits in a finite but large number of different combinations which are involved in different kinds of cell–cell and cell-matrix interactions (Shaw, Turner et al. 1995). The α subunit in cytoplasmic domains are conserved in different species with only weak overall homology between the different genes. Integrin α3 and α6 cytoplasmic domains are good examples which have been studied. It has been reported that α3 and α6 cytoplasmic domains have highest homology among the alpha domains (Sastry & Horwitz, 1993). The GFFKR motif just adjacent to the membrane is the most strikingly conserved part of the α cytoplasmic domains. The α7 cytoplasmic domain is the longest α cytoplasmic domain with 77 amino acid residues. The α7 cytoplasmic has several short regions of homology with other protein such as tyrosine phosphatase. Taken together, α3, α6 and α7 have shared motifs in their cytoplasmic domains (Sastry & Horwitz, 1993) .

1.9 Integrin α6 (ITGA6)

The subject of this project, the integrin α6 subunit gene is located on chromosome 2 at 2q31.1 with a total length of 78870 bp (Pulkkinen et al., 1997). Integrin α6 does not dimerise with all beta variants. Integrin α6 mostly combines with two specific β subunits α6β1 and α6β4 and both are receptors for the laminin family of extracellular matrix proteins (Hogervorst et al., 1993). It has been found that α6β1 is expressed in platelets, epithelial cells and several of other cell types. α6β1 is also involved in adhesion (Hogervorst et al., 1993). α6β4 integrin is present in various epithelial tissues, endothelia and peripheral nerves. α6β4 is also present in the hemidesmosomes of epidermal cells which strongly suggested a function of α6β4 in adhesion of cells to the extracellular matrix (Hogervorst et al., 1991). ITGA6 is implicated in cancer cells and serves as a signalling receptors that triggers signalling cascades that enhance survival, invasion and metastasis (reviewed in Mercurio et al., 2001).

The two forms of α6 integrin studied heavily are the α6A and α6B mRNA splice variants (Figure 1.12). These two isoforms have distinct cytoplasmic domains. Both α6A and α6B consist of a large extracellular domain (991 amino acids) and a transmembrane domain (23 amino acids) but they differ in their cytoplasmic domains (36 amino acids for alpha and 54 for beta) (Hogervorst et al., 1993). The CFFKR sequence is identical in both α subunits encoded by either of the α6A or α6B specific exons which then diverge downstream figure 1.12 (Hogervorst et al., 1991). The A and B forms are differentially expressed throughout the body, however a mouse deleted for the alternative exon expressed the B (exon skipped) form everywhere surprisingly was fertile and normal with just slight quantitative differences in ex vivo assays (Gimond et al., 1998). A role for ITGA6 in cancer was suggested as they serve as signalling receptors that triggers signalling cascades that enhance survival, invasion and metastasis (reviewed in Mercurio et al., 2001). ITGA6 was found to be necessary for tumorigenicity of a stem cell like population within the MCF-7 cell line (Cariati et al., 2008) and it also regulated glioblastoma stem cells (Lathia et al., 2010). It was also found that α6 integrin high expression is a biomarker in breast cancer cells. α6 subunit signalling regulates the processes in tumourigenes including proliferation and metastasis, thus suggesting the two splice variants of α6 integrin have specific functions. Taking together these studies raise the importance of definition how alternative splicing of ITGA6 is regulated.

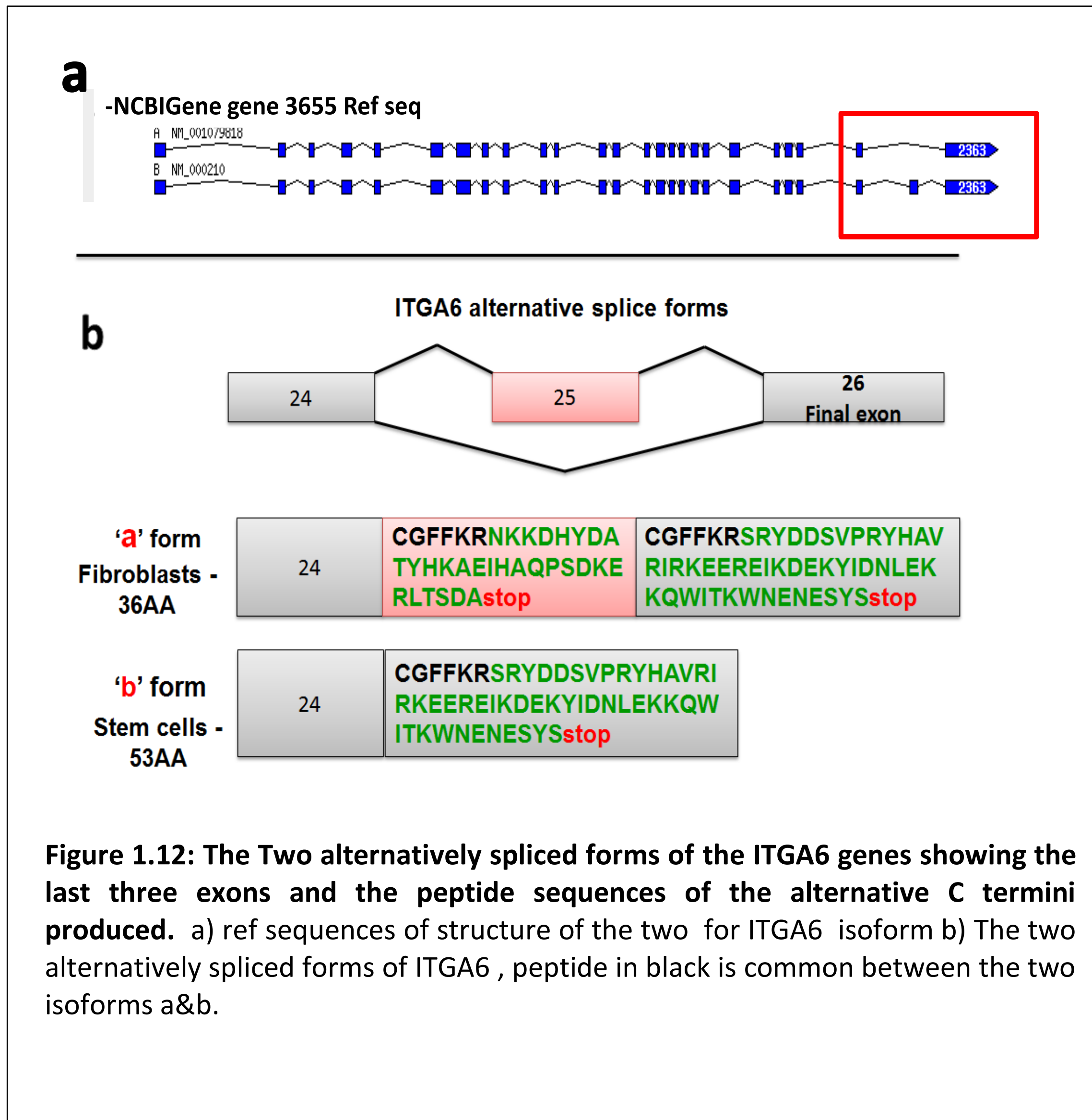


Figure 1.12: The Two alternatively spliced forms of the ITGA6 genes showing the last three exons and the peptide sequences of the alternative C termini produced. a) ref sequences of structure of the two for ITGA6 isoform b) The two alternatively spliced forms of ITGA6 , peptide in black is common between the two isoforms a&b.

1.10 Research aims and Hypothesis

The preliminary data prior to starting this thesis suggested the following hypotheses:

- Since the original papers showed a splicing difference between fibroblasts and stem cells. We hypothesised that these splicing differences might be regulated during development in whole animals, and also modelled in other cell types.
- The original work indicated that the ITGA6 splicing pattern was controlled by two nuclear RNA-binding proteins RBPs (MBNL & RBFOX). We hypothesised there might be other RBPs that control this splicing choice, and these RBPs might work by direct binding to ITGA6 mRNA.
- Previous studies showed that RBPs binding downstream of a regulated exon usually activated (which include alternative exon) ,whereas RBP binding upstream inhibited splicing(which exclude alternative exon) . We hypothesised that PTBP binding sites identified in sites downstream of the ITGA6 regulated exon might be inhibitory.

1.10.1 Aims:

- To monitor splicing patterns for ITGA6 in development.
- To monitor splicing patterns for ITGA6 in different cell lines.
- To determine which RBPs control ITGA6 alternative splicing
- To identify the PTBP binding site which inhibits the ITGA6 alternative exon, and if this is downstream of the regulated exon.

Chapter 2: Investigation of orthologues of stem cell regulated exons in zebrafish

2.1 Introduction

Alternative splicing plays an important role in the gene function and in complex organisms, such as mammals. Alternative splicing is implicated in germ cell development including stem cell and fibroblast differentiation. Recently, data by Venables et al (2013) determine alternative splice events that were important for stem cell differentiation. It was shown that a large portion of alternative splicing changes during fibroblast differentiation from stem cells (Venables et al., 2013b). This research employed high-throughput RT-PCR with a RefSeq database, which facilitated the identification of alternative splicing events that changed in the full spectrum of high and low gene expression (Venables et al., 2013b). The technique started with a list of 81 genes including *ITGA6* that had alternative splicing events in human tissue. Those alternative splicing events were controlled by the *MBNL* and *RBFOX* regulation factors. *ITGA6* with nine genes, including *PLOD2, CLSTN1, ATP2A1, PALM, KIF13A, FMNL3, PPIP5K1, MARK2* and *FNIP1,* which showed alternative splicing events between stem cells and fibroblasts, were conserved in evolution (Figure 2.1). Most of these 10 genes were subject to gene duplications prior to vertebrate radiation, and the alternative exons appeared ‘shortly’ beforehand (Venables et al., 2013b). This chapter build on data generated in (Venables et al., 2013) and takes an evolutionary approach to further investigate these genes.

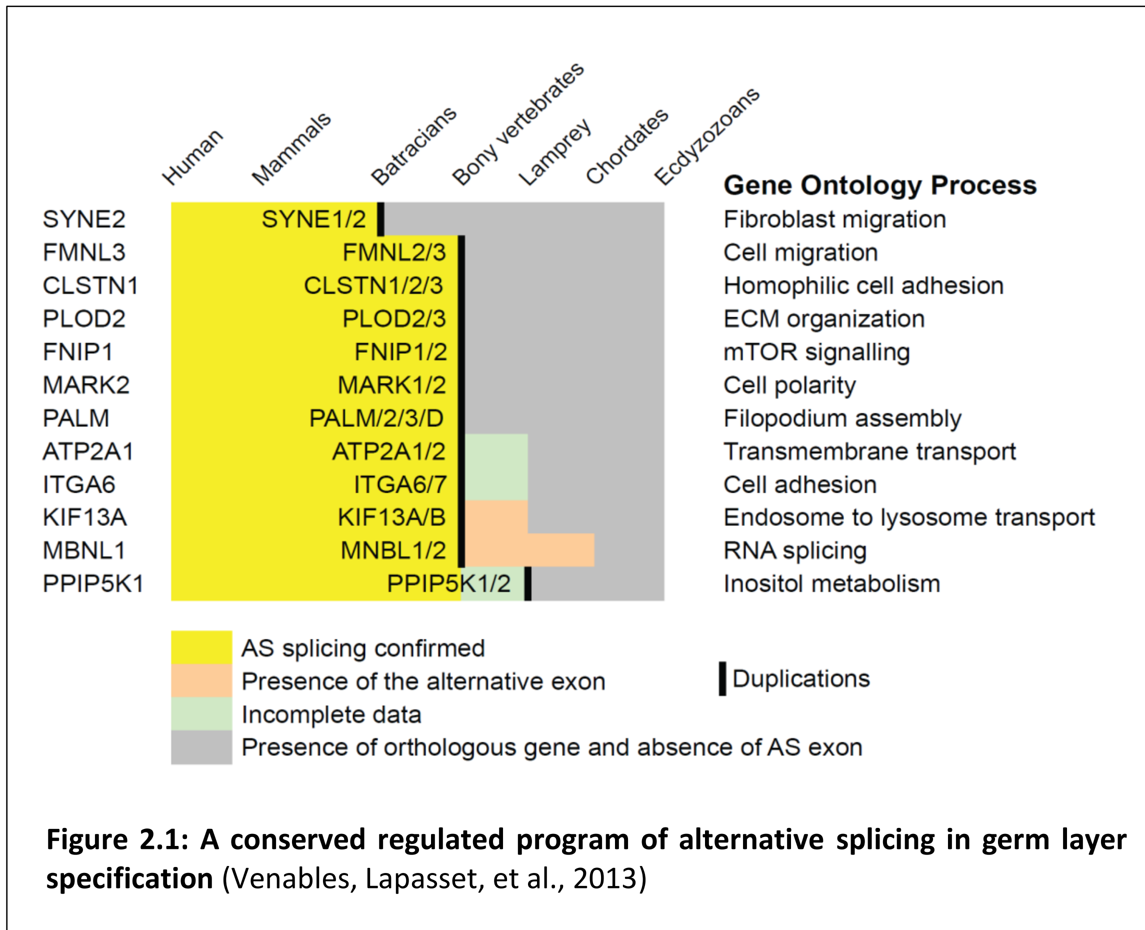


Figure 2.1: A conserved regulated program of alternative splicing in germ layer specification (Venables, Lapasset, et al., 2013)

To conduct this study, it was important to find a good model. The zebrafish has been reported as an excellent vertebrate model system for investigating alternative splice events that are similarly regulated in both fish and man. The rationale is that conserved regulated alternative splicing events are likely to be extremely important for the differentiation and development of the body plan of all vertebrates including man; thus, these events will likely have medical significance.

High-throughput RT-PCR is the gold standard technique for identifying alternative splicing changes. Based on(Venables et al., 2013) data, primers to detect orthologous splicing events in zebrafish were designed using sequences identified by Philippe Fort, from the CRBM, Montpellier, for the presence or absence of both of two predicted isoforms for the 10 genes. The conserved splice events (*KIF13* and *ITGA6*) were studied during the first two days of zebrafish development. *ITGA6* was one of the 10 ASEs that

shifted from one isoform to another during stem cell differentiation and zebrafish embryogenesis. Thus, these isoforms are likely to be of importance for differentiation, development and fundamental cellular processes.

2.2 Aims:

- To Investigate alternative splicing events of 10 genes orthologues of the stem cell regulated exons in zebrafish in order to find a good model gene to investigate alternative splicing mechanisms.
- To investigate an alternative splicing event of *ITGA6* paralogues and orthologues in zebrafish development

2.3 Materials and Methods

2.3.1 Zebrafish dissection method:

Dissection was performed on zebrafish embryos (24 and 48 hours post-fertilisation). Firstly embryos were put in PBS in order to maintain osmotic balance. After that the embryonic shield was cut by using sharpened forceps. Next, needles (26 gauge) were used to dissect the zebrafish. Finally heads and tails were collected by pipet and placed in a new Eppendorf tube containing Trizol (Figure 2.2)

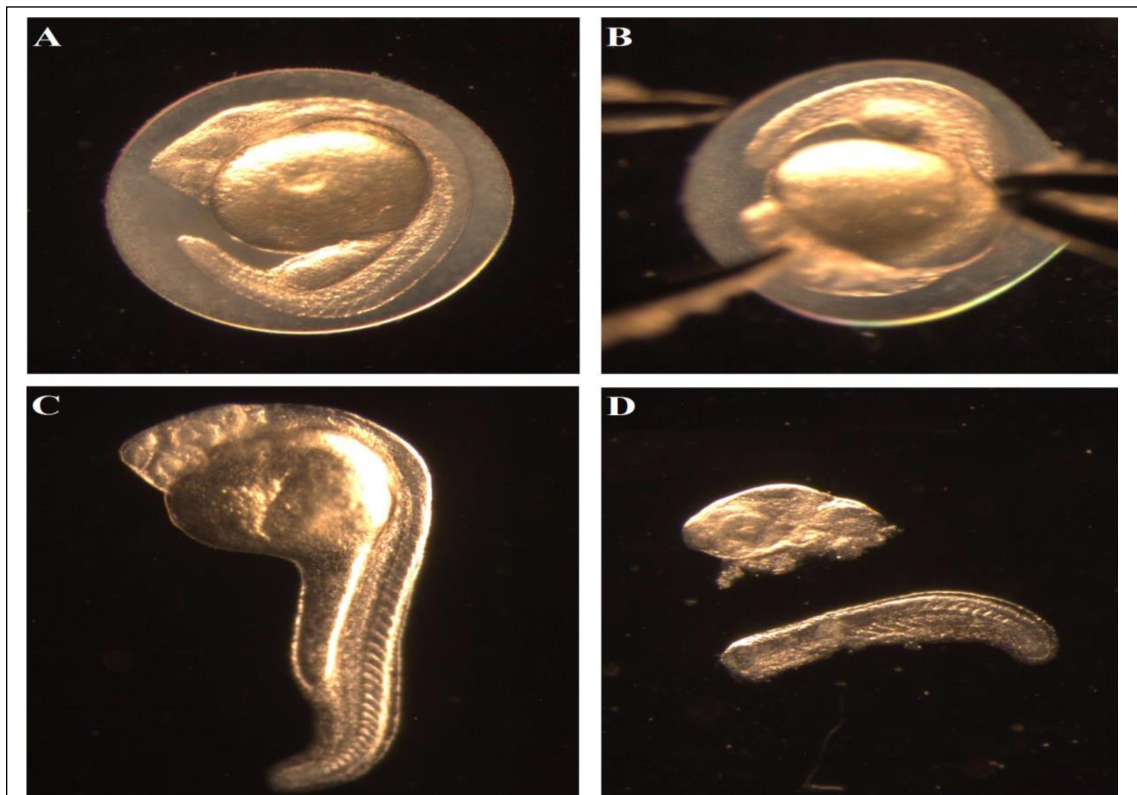


Figure 2.2: The steps of head and tail dissection from zebrafish embryos after 24 hours. A Zebrafish embryo is covered by shield. B. The embryonic shield is pulled away. C Zebrafish embryo at 24 hours. D. head and tail after dissection. Photos taken from a previous dissection by Alicia Madgwick**.**

2.3.2 RNA extraction:

200 µl of Trizol (Life technology) was added to tissue or cells in an Eppendorf and vortexed, then incubated at room temperature for 5 minutes and 40µl chloroform was added. Samples were vortexed again and incubated at room temperature for 5 minutes. Samples were then centrifuged at 13,000rpm for 15 mins at 4°C. The supernatant was carefully transferred (approximately 100 µl, avoiding the interface) to a new Eppendorf tube and 100µl isopropanol was added. Next, samples were briefly vortexed and incubated at room temperature for 5 min. Samples were vortexed again and centrifuged at 13,000rpm for 15 min at 4°C. After that, the aqueous phase was removed. The pellet was washed by adding 100µl of 70% ethanol and spun for 5 min at 4°C. The liquid was removed and the pellet was allowed to air dry. Finally, 10 or 20 µl of Diethylpyrocabonate(DEPC)-treated dH20 was added to re-suspend the pellet and the concentration was quantified on a 'Nanodrop' spectrophotometer.

2.3.4 Geneious program

'Geneious' is a bioinformatics program with multiple functions. It was used to align different paralogues and homologues of ITGA6. It was also used to interpret DNA sequencing results. Geneious links to NCBI and BLAST database, which it uses for downloading a nucleotide or protein sequence into the database used here.

2.3.5 Primer design

For the subsequent PCR, primers were designed to amplify both alternative splice forms simultaneously so the ratio of the two could be calculated. The primers were thus designed in adjacent exons to the alternative exon so that they would only amplify mRNA, because pre-mRNA and contaminating genomic DNA both contain introns which are sufficiently long that their products are at a disadvantage and therefore will not compete significantly. Primers were designed to target the same alternative splicing events in the orthologous genes in zebra fish by using the Primer3 program. Homologous exon sequences, on which to design the primers, were provided

by Dr. Philippe Fort (Montpellier). Primers were synthesised by Integrated DNA Technology (IDT) (Table 2.1). The 10 genes in zebrafish were homologues of the 10 alternative splice variants shown to shift between stem cell and fibroblasts (Venables et al., 2013b) (Table 2.2)

Table 2.1: primers for orthologues of the 10 genes

Primer	Sequence
MARK2 (F)	ACCAGCACAAATCGAAGCAG
MARK2 (B)	AGGCAACAGGGACACGCT
PPIP5K1 (F)	CCGAATCTTCAGGACTACGC
PPIP5K1 (B)	GGGCATTATGCAGTGTTTCC
FMNL3 (F)	GCGGGAATTTCTGAATGATG
FMNL3 (B)	CACTAGGCGGGAGTTCTTCA
FNIP1 (F)	GCAGCAGTATTTGTGGGAGTC
FNIP1 (B)	TCCAGGCATGTCCATTGG
KIF13A (F)	TGCCACTTATGGTTGAAGCCA
KIF13A (B)	TGCATCTGACCACCTCTCCCTT
PALM (F)	ACAAGCGAGTCTCCAACACG
PALM (B)	GTCCGCTTTGTGGATGAGTT
ATP2A1 (F)	AGTTCGTTGCTCGGAACTACC
ATP2A1 (B)	GCCTGAAGATGTGTCACTATCG
ITGA6 (F)	ATCATCCTAGTGGCTATTCTCGC
ITGA6 (B)	ACTGTCATCGTACCTAGAGCGT
PLOD2 (F)	TTTGTTCGTGATAAACTGGATCC
PLOD2 (B)	GCAGTGGATAATAGCCTTCCAA
CLSTN1 (F)	CACAGAGAACGACAACACCG
CLSTN1 (B)	CGAATGACTCCCTCACCAGT
CTTN (F)	ACAGACAAGACAAATGTGCCC
CTTN (B)	TATCCATCCGATCCTTCTGC

Table 2.2: ITGA6 paralogues and orthologue zebrafish primers

Primer	Sequence
zITGA3A (F)	TATGCTGGGCGTGATTGTTA
zITGA3A (B)	GCCTAAGCGAGACAGTTTGC
zITGA3B (F)	AGTTCCCCTCTGGATCATC
zITGA3B (B)	CCCAGAGTTTCTTGCTGAGG
zITGA6A (F)	TGGCTGTGTTAGCTGGAATC
zITGA6A (B)	ACGTGGTCATCCACTGCTTC
zITGA6B (F)	TTGCTGGGCTTACTGGTCTT
zITGA6B (B)	CATTTCGTCTTTGCCTGACA
zITGA7 (F)	CGCTGCTCGTGTGTTTACTG
zITGA7 (B)	ATGAGTGTGTCGTCCAGCAG

2.3.6 PCR

Standard ‘endpoint’ PCR was performed with Invitrogen ‘Platinum Taq’ hot start PCR enzyme following the manufacturer’s instructions. Table 2.3 shows the 5X stock buffer made for all PCR master mixes table 2.4. The final PCR mix used to amplify alternative splicing products of cDNA is shown in table 2.5.

Table 2.3 : 5xbuffer stock (1ml)

Components	volume
Platinum Taq 10X buffer	500 µl
25mM dNTPs	40 µl
50mM $MgCl_2$	150 µl
water	310 µl

Table 2.4: PCR component for 50µlx reaction

Component	Volume
5Xbuffer	10µl
H2O	34.75µl
Primers (50µM mixed forward and reverse primers)	1µl
cDNA (12.5ng/ul)	4µl
Platinum Taq polymerase	0.25µl

Table 2.5: PCR procedure

PCR steps	Temperature °C	Time	Cycle
Initial enzyme activation and DNA denaturing step	95	2 min	1
Denaturing	94	30 sec	35
Annealing	55	30 sec	35
Elongation	72	30 sec	35
Final extension	72	2min	

2.3.7 Gel electrophoresis

PCR products (12 µl) were run on 1.5% agarose for product size estimated. Gel preparation was done following the protocol in table 2.6 , 2.7&2.8.

Table 2.6: Electrophoresis gel protocol (1.5%)

Component	Volume
Water	126ml
10X TBE	14ml
Agarose	2.1g
Ethidium bromide 0.5 µg/ml	11µl

Table 2.7: 10X TBE (1litre) components

Component	Volume
Tris base	108g
Boric acid	55g
0.5M EDTA (at pH8)	40ml
Water	960ml

Table 2.8: 0.5M EDTA (at pH8)

Component	Volume
EDTA	74.49g
Water	400ml
NaOH was added until the solution reached pH8	

2.4 Results:

2.4.1 Alternative splicing in 10 genes in zebrafish head and tail.

Before beginning the study, 10 genes, including *ITGA6,PLOD2, CLSTN1, ATP2A1, PALM, KIF13A, FMNL3, PPIP5K1, MARK2* and *FNIP1,* which showed an alternative splicing event in stem cell differentiation, were chosen for analysis in zebrafish. Due to the importance of alternative splicing in stem cells and the relevance of stem cell cultures to model differentiation and development, I began my PhD project with the aim of further investigation of these genes in order to identify a good paradigm gene which regulated by different proteins and has impact function in cells in order to study alternative splicing mechanism. To first profile the extent of alternative splicing events, a panel of 10 genes with alternative splicing events across zebrafish tissue development including a 24 hours post-fertilization were studied.

To investigate potential alternative splicing events of orthologues in these 10 genes zebrafish embryogenesis, I collected zebrafish heads and tails at 24 hours post-fertilization (the full experiment details are detailed in the methods and materials section). RT-PCR with specific primers designed across each alternative splicing events was employed to evaluate the ratio of inclusion and exclusion exons. *PPIP5K1, CLSTN1, PALM, FNIP* and *PLOD2* express only the short form (exclusion exon) in both head and tails of zebrafish. *ATP2A1* expresses two isoforms (inclusion and exclusion) in only tail, whereas *FMNL3* express two forms (inclusion and exclusion) in only head of zebrafish. *MARK2* was not expressed at all in head and tail of zebrafish. The interestingly, *KIF13A* and *ITGA6,* out of the 10 exons tested, showed specific alternative splicing between the heads and tails. *KIF13A* and *ITGA6* showed both mRNA isoforms (exon inclusion and exclusion) in tails; however, only the short form (exclusion exon) was visible in heads (Figure 2.3).

Because *KIF13A* and *ITGA6* were the only genes that showed specific alternative splicing between heads and tails at 24 hours post-fertilization, it was decided to look at these ASEs in different stages of zebrafish development to find exactly when splicing patterns shifted. Samples were taken from the zebrafish at different stages in the first two days post-fertilization (2, 4, 5, 11, 15, 28 and 48 hours). Here, Interestingly, *ITGA6* mRNA isoforms gradually shifted towards increasing inclusion (long form), from two

hours to 15 hours, and towards full reversion to exon skipping at 48 hours. *KIF13* gradually shifted towards increasing inclusion (long form) from 15 hours to 48 hours (Figure 2.4).

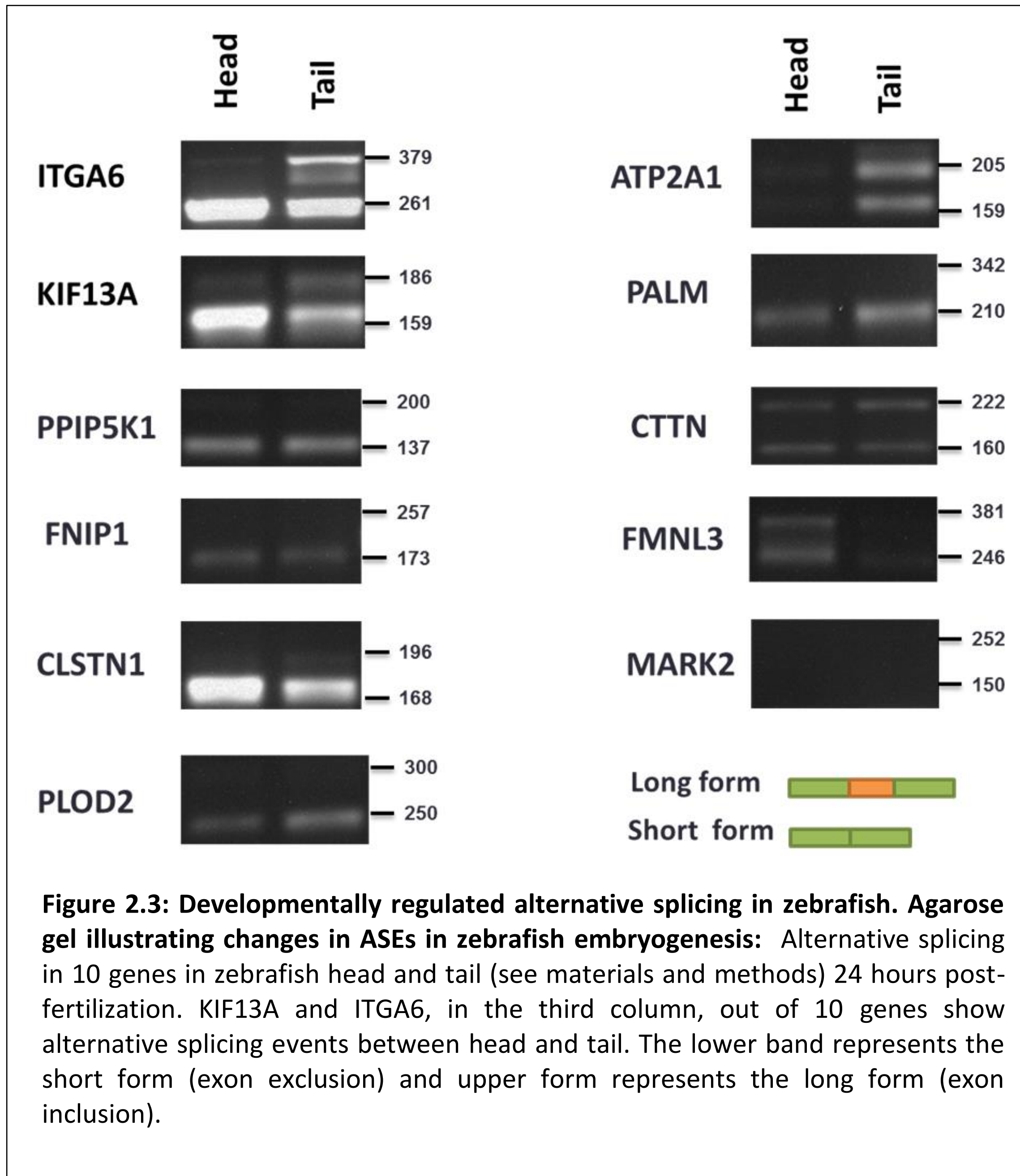


Figure 2.3: Developmentally regulated alternative splicing in zebrafish. Agarose gel illustrating changes in ASEs in zebrafish embryogenesis: Alternative splicing in 10 genes in zebrafish head and tail (see materials and methods) 24 hours post-fertilization. KIF13A and ITGA6, in the third column, out of 10 genes show alternative splicing events between head and tail. The lower band represents the short form (exon exclusion) and upper form represents the long form (exon inclusion).

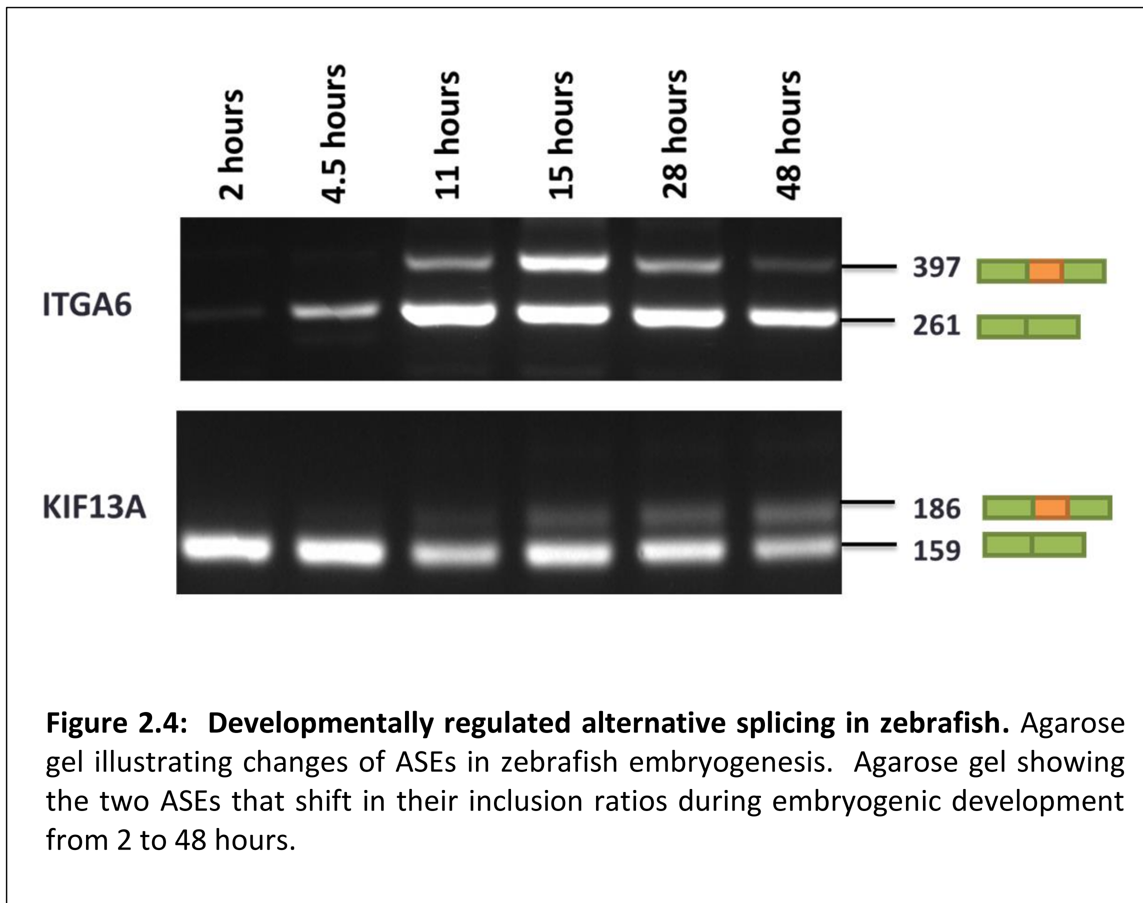


Figure 2.4: Developmentally regulated alternative splicing in zebrafish. Agarose gel illustrating changes of ASEs in zebrafish embryogenesis. Agarose gel showing the two ASEs that shift in their inclusion ratios during embryogenic development from 2 to 48 hours.

2.4.2 Investigation of *ITGA6* alternative splicing in vertebrates

Since *ITGA6* splicing shifts most strongly in zebrafish embryogenesis out of the previously identified conserved stem cell splicing events (Venables et al., 2013b) we decided to concentrate our efforts on investigation of this gene. Therefore Dr Philippe Fort (CRBM Montpellier, France) launched a deeper investigation into the evolution of this alternative splice in the *ITGA6* C-terminus and its paralogues in different species. By mining EST and genomic sequences *ITGA3, ITGA6* and *ITGA7* each show two different forms due to alternative splicing in all vertebrate species. Furthermore, in zebrafish, there are two paralogous *ITGA6* genes and two paralogues of the *ITGA3* gene, so there is homologous alternative splicing of 5 genes in zebrafish (Figure 2.5). By alignment of the sequence of all C-termini in different species, Dr. Fort showed that the two alternative C-termini produced by alternative splicing (a and b) cluster together for *ITGA3*, *ITGA6* and *ITGA7.* The similarity in amino acid of the 'a' forms of *ITGA6* and *ITGA3* are extremely strong as are the similarity between the 'b' forms of these genes across the three different genes and across different vertebrate species. The a and b forms of *ITGA7* also cluster with the respective a and b forms of the other genes although with slightly less similarity (Figure 2.6). Overall these amazing results provide the unprecedented conclusion that alternative splice forms in C-termini of *ITGA3,* 6 and 7 are more fundamentally different (and likely more functionally important) than any distinction between the 3 genes (or 5 genes in zebrafish) themselves. This observation led us to investigate the function and regulation of this conserved alternative splice choice in *ITGA6.*

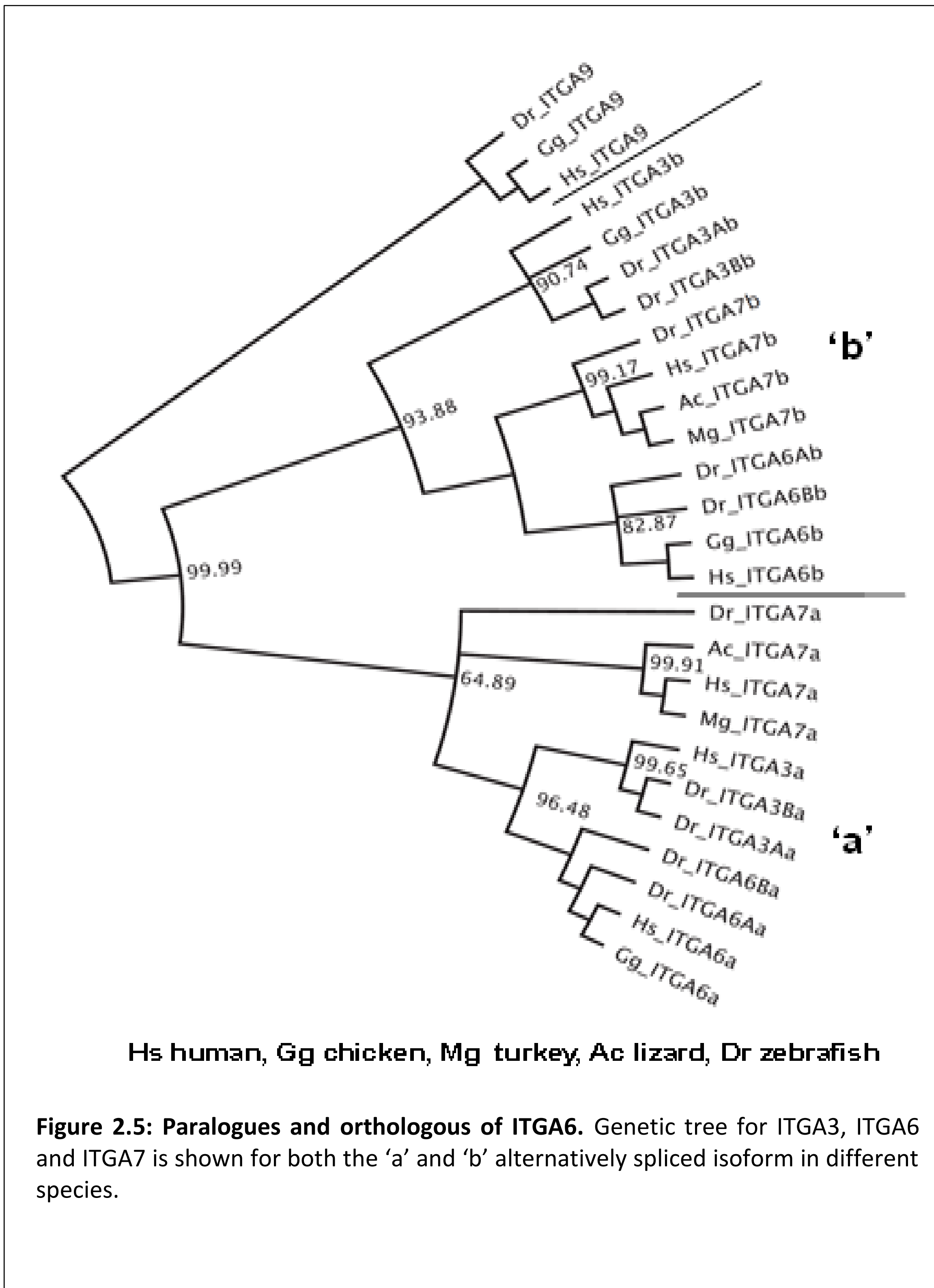


Figure 2.5: Paralogues and orthologous of ITGA6. Genetic tree for ITGA3, ITGA6 and ITGA7 is shown for both the 'a' and 'b' alternatively spliced isoform in different species.

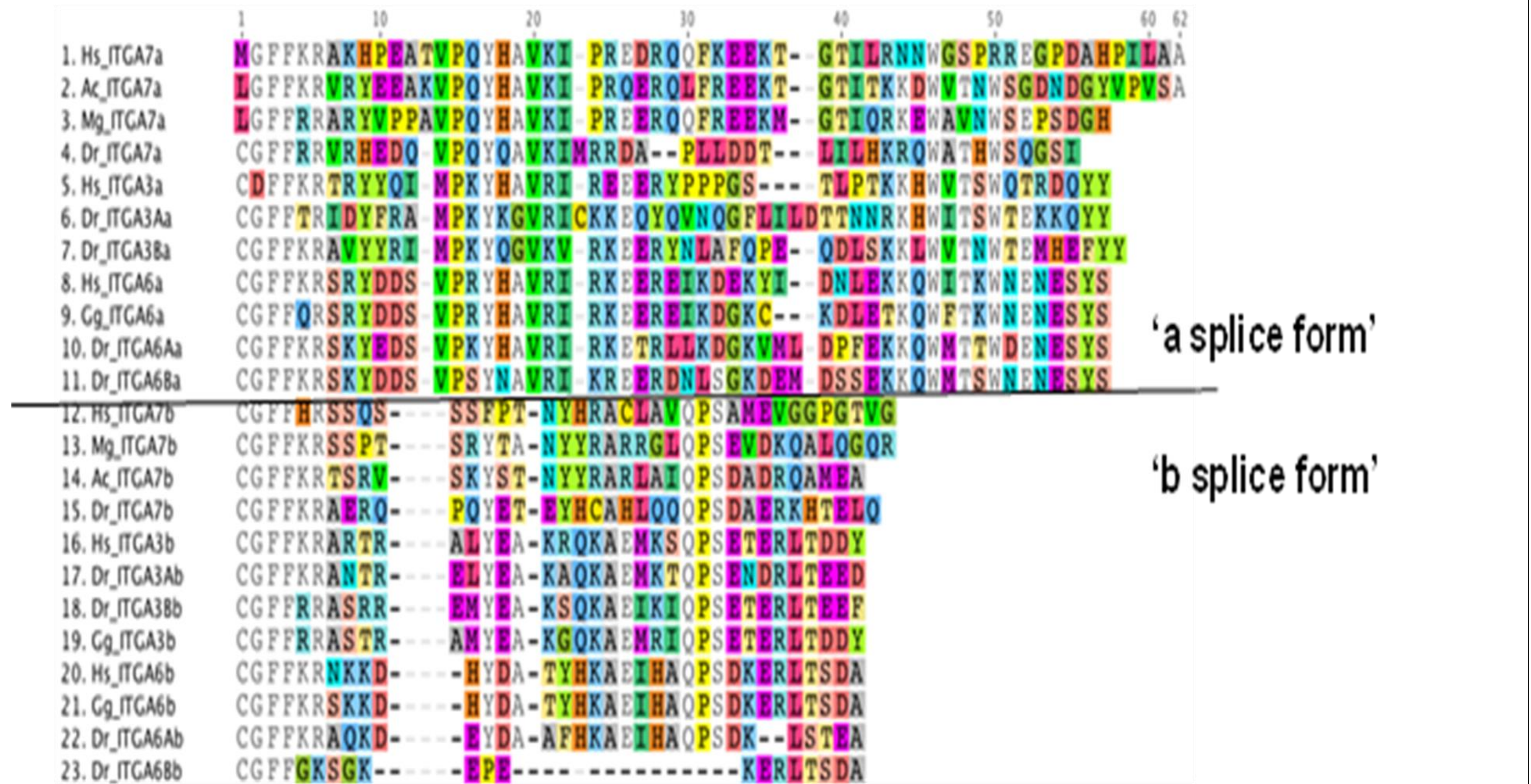


Figure 2.6: Alignment of the sequence Paralogues and orthologous of ITGA6. C-termini of various vertebrate species are shown. Hs human, Gg chicken, Mg turkey, Ac lizard, Dr zebrafish

2.4.3 Investigation of paralogues and orthologues of *ITGA6* in Zebrafish Head and Tail after Post- Fertilization.

The previous bioinformatics and RT-PCR for *ITGA6* revealed that there are 5 homologues and paralogues of *ITGA6* with regulated splicing in their downstream regions. To investigate alternative splice events, the regulated splicing of these *ITGA6* homologues and paralogues were tested in head and tail of 24 hours post-fertilization zebrafish by RT-PCR. 2 genes (*ITGA6A* and *ITGA6B*) showed alternative splicing patterns between head and tail. The other homologues, *ITGA3A* and *ITGA3B* and *ITGA7*, appear to be expressed at very low levels such that they were barely detected by RT-PCR. *ITGA6A* showed long form (exon inclusion) in tail, whereas the short form (exon exclusion) appears in both head and tails. *ITG6B* also expresses the long form in tail while the short form is present in both head and tail (Figure 2.7).

Although *ITGA6A* and *ITGA6B* showed alternative splicing event differences between head and tail, it was important to define them in zebrafish development, as each stage might show a dynamic pattern over this time frame. The original PCR I had done on zebrafish development was just for the *ITGA6A* gene. Therefore I re-performed the analysis for both *ITGA6A* and *ITGA6B* genes individually. The cDNA from different stages in embryogenesis was investigated by using RT-PCR and agarose gel electrophoresis. For *ITGA6A*, the smaller band (292bp), which represents the short form (exon exclusion), is visible in all development stages from 4.5 up to 48 hours, whereas the upper long form (exon inclusion) appears gradually from 11 hours up to 15 hours, before starting to disappear again from 15 to 48 hours (Figure 2.8 A). On the other hand the *ITGA6B* gene is also regulated, but in a completely different way. For *ITGA6B*, the major lower band (137bp), representing the short form (exon exclusion), is present in all stages, while the long form (186bp) appeared only from 24 hours post fertilization (Figure 2.8B)

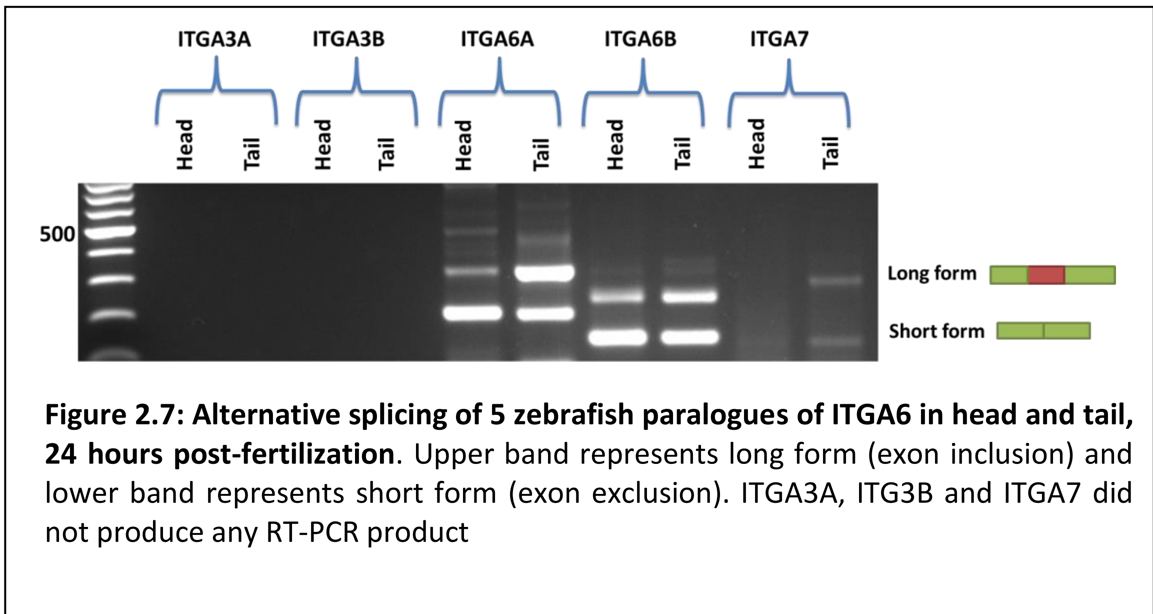


Figure 2.7: Alternative splicing of 5 zebrafish paralogues of ITGA6 in head and tail, 24 hours post-fertilization. Upper band represents long form (exon inclusion) and lower band represents short form (exon exclusion). ITGA3A, ITG3B and ITGA7 did not produce any RT-PCR product

A
2 hours
4.5 hours
11 hours
15 hours
28 hours
48 hours
ITGA6A
292
195
ITGA6B
186
137
Long form
short form

B
Alternative splicing
2
1
ITGA6B
ITGA6A
Hours
2 4.5 11 15 28 48

Figure 2.8 : Alternative splicing for ITGA6A and ITGA6B in development stages from 2 hours to 48 hours post-fertilization in zebrafish. A) Agarose gel illustrating alternative splicing for ITGA6A and ITGA6B in zebrafish development. The long form (292bp) clearly appears at 15 hours whereas the short form (195bp) is visible at all stages from 11 to 48 hours in ITGA6A. In ITGA6B, the long form (186bp) is visible just at 28 and 48 hours, whereas the short form (137bp) clearly shows in all stages**. B)** The chart shows different of alternative splicing event between ITGA6A and ITGA6B during zebrafish development stages.

2.5 Discussion

2.5.1 Validation of a novel alternative splicing event from Zebrafish development:

Interest in alternative splicing is gaining momentum and now widely considered as an important mechanism for controlling gene expression. During cell differentiation, it is known that gene expression can be changed at different levels including alternative splicing (Han et al., 2013, Venables et al., 2013b). The purpose of this study was to investigate alternative splicing during cell differentiation more deeply than has been done before. As conservation of regulated alternative splicing events has been reported across evolution (especially across vertebrate radiation) (Merkin et al., 2012), there is clearly a potential to investigate vertebrate alternative splicing by using the tractable genetics of the zebrafish model organism.

The Zebrafish model has several advantages, which allow for an understanding of the functions of alternative splicing in development. One advantage of using the Zebrafish embryonic model is that development occurs rapidly until 72 hours post-fertilisation and all major organs appear after 36 hours and hatching occurs from 12 -36 hours. Between 2 hours post fertilization (hpf) and 48hpf, zebrafish embryos develop rapidly through many stages (Figure 2.9). In the first 2 hours post fertilisation, cells multiply to reach the 64-cell stage. The epiblast forms from around 5hpf. Between 10-24hpf the segmentation period occurs and organogenesis commences and somite development takes place after 24hpf until 48hpf (Kimmel et al., 1995).

We hypothesised that alternative splicing may play a significant role in embryonic development in zebra fish. Therefore I investigated the orthologues of 10 genes in zebrafish development, for which alternative splicing events have been documented in stem cell differentiation (Venables, et al., 2013). *ITGA6* and *KIF13* had specific alternative splicing changes between head and tail after 24 hours of development. Exon exclusion was seen just in in both head and tail; however exon inclusion was just seen in tail which might give a clue that both isoforms of mRNA have an important role in embryonic devolvement of zebrafish tail (which is mostly composed of muscle). Since *ITGA6* and *KIF13* had specific alternative splicing differences between head and

tail, I went on to observe their splicing patterns during a time course of zebra fish embryonic development. Interestingly, the *ITGA6* long form peaked at 15hps, which may suggest that the longer form plays a significant role in forming the epiblast, and more generally has an impact on cellular function during embryonic development of vertebrates.

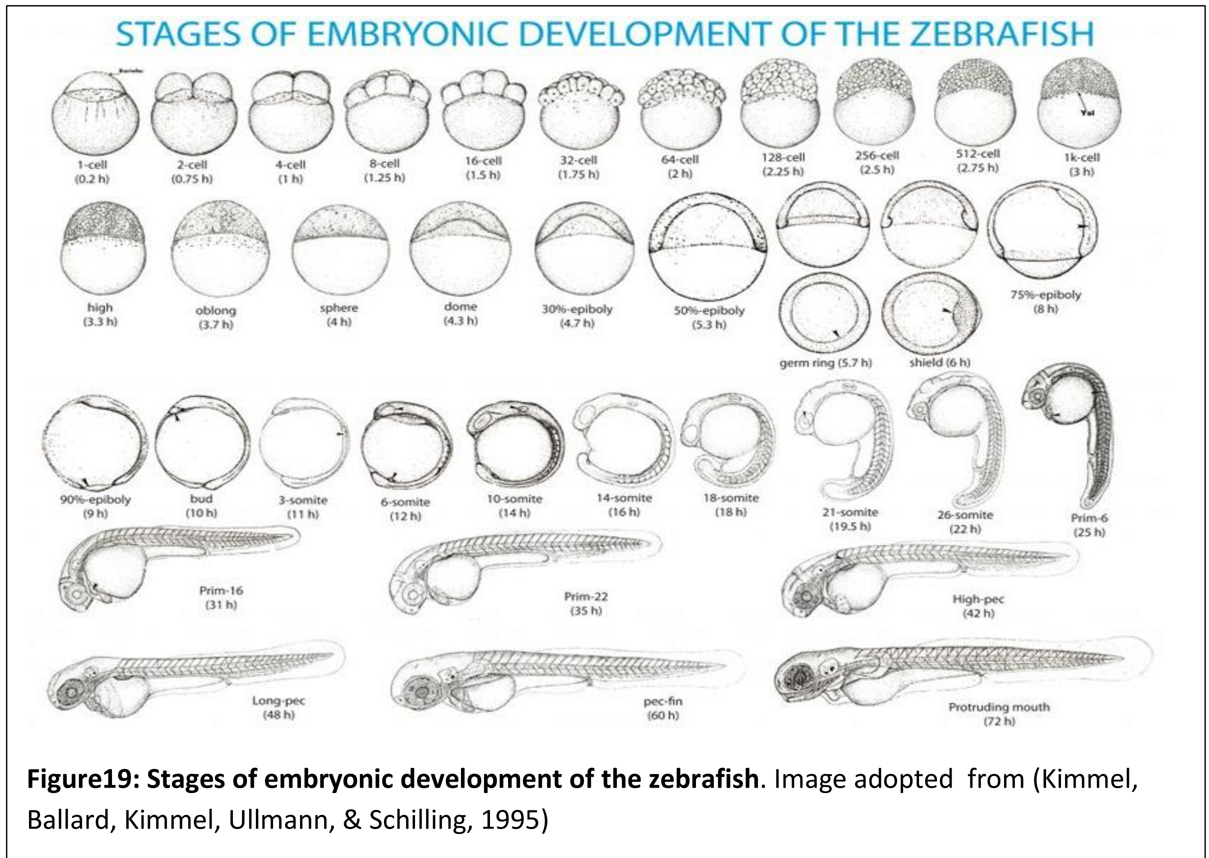


Figure19: Stages of embryonic development of the zebrafish. Image adopted from (Kimmel, Ballard, Kimmel, Ullmann, & Schilling, 1995)

2.5.2 ITGA6 alternative splicing pattern in zebrafish

ITGA6 belongs to the integrin family, which are dimeric cell-surface proteins composed of an alpha chain and a beta chain (Sastry and Horwitz, 1993). It has been reported that there are three paralogues in mammals for *ITGA6*, including *ITGA3* and *ITGA7*, which have a close evolutionary relationship (Sastry & Horwitz, 1993). Interestingly, we found that there are five close paralogues of *ITGA6* in zebrafish, compared with the three close genes in *ITGA3*, *ITGA6* and *ITGA7* in mammals. This is caused by the ancient duplication of the *ITGA6* and *ITGA3* genes. Amazingly, two C-terminal isoforms of all these proteins are conserved even in zebrafish. The two C-terminal isoforms including , the 'a' and 'b' forms are more closely related between paralogues and orthologues than the orthologues and paralogues themselves. In other words, the The two C-terminal isoforms (a and b) has more conserved between paralogues and orthologues , whereas the two C-terminal isoforms (a and b) has less conserved between the paralogues genes or orthologues.

Alternative splicing of *ITGA6* has been reported during stem cell differentiation (Venables et al., 2013). Alternative splicing *ITGA3* and *ITGA7* have been reported in mouse embryogenesis. In this study, the five paralogues of *ITGA6* in the head and tail 24 hours post-fertilization and in the embryogenesis developmental stages of zebrafish were observed. *ITGA6A* and *ITGA6B* had similar specific alternative splicing, with the long form appearing only in tail.

Do both *ITGA6* A and B have the same alternative splicing events and functions? Some studies have reported that *ITGA6* A and B have the same function in mouse embryogenesis (Gimond et al., 1998). However, they showed different timings of these splicing events during embryogenesis, which suggested that splicing control of both *ITGA6* genes also might play a major role during zebrafish embryonic development. There is some evidence showing that *ITGA6* isoforms A and B are involved in different functions of cancer stem cells (Gimond et al., 1998).

ITGA3 and *ITGA6* show alternative splicing patterns in mouse embryos during development. Interestingly, I found that *ITGA3* and *ITGA7* have low expression in zebrafish embryogenesis. It is tempting to speculate that the major functional alpha integrins in embryogenesis in zebrafish are provided by *ITGA6A*, *ITGA6B* and their four respective alternatively spliced isoforms (*ITGA6Aa*, *ITGA6Ab*, *ITGA6Ba* and *ITGA6Bb*). However, in other organisms, the important functions might be provided by a different combination of the *ITGA3*, *ITGA6* and *ITGA7* genes, albeit with essential functions for the a and b isoforms.

2.6 Chapter Summary

In this chapter, the different pattern splicing for *ITGA6, PLOD2, CLSTN1, ATP2A1, PALM , KIF13A, FMNL3, PPIP5K1, MARK2 and FNIP1* have been shown these an alternative splicing event in stem cell differentiation in zebrafish. ITGA6 and KIF13A were the only genes that had different patterns of splicing between head and in tail after 48 hours post-fertilization in zebrafish. It was found that ITGA6 has a different pattern of splicing during the development stage of the zebrafish.

We identified thatITGA6 has five paralogs in zebrafish. These have interestingly conserved two C-terminal isoforms. Only ITGA6 a and b of the five paralogues has a different expression pattern between head and tail after 24 hours of development. They also showed different timings of these splicing events during embryogenesis.

Chapter 3: Identification of PTBP as a novel regulator of the ITGA6 splicing switch

3.1 Introduction

ITGA6 plays an important role in many aspects of a cell's biology, including cell signalling and interactions with the cell matrix. It has been reported that *ITGA6* was one of 15 ASEs that shift from one isoform to another during stem cell differentiation (Venables et al., 2013b). Both isoforms (*ITGA6A* and *ITGA6B*) are involved in regulating tumorigenesis, including proliferation and metastasis (Goel et al., 2014). It has been suggested that the two splicing variants of α6 integrin have a specific function in tumorigenesis (Goel et al., 2014). These studies raise the importance of knowing how alterative splicing of *ITGA6* is regulated. In this chapter, the aim was to investigate how *ITGA6* splicing is regulated.

The alternative exon 25 of *ITGA6* has been reported to be controlled by two separate splicing factors, Muscleblind-like (*MBNL1*) and Forkhead transcription factors (*RBFOX2*), which also involved in cellular differentiation (Venables et al., 2013b). Both MBNL1 and RBFOX2 proteins can act as repressors or activators of alternative exons based on binding site location. MBNL1 and RBFOX2 proteins bind downstream of the exon to enhance inclusion and upstream to inhibit inclusion. Venables et al. (2013) showed that MBNL1 and RBFOX2 regulated the enhanced splicing inclusion of the *ITGA6* exon during stem cell differentiation.

(Goel et al., 2014) showed that there is a third regulation factor, epithelial splicing regulator protein (*ESRP1*), which also regulates *ITGA6* alternative exon. ESRP1 is part of a positive feedback loop downstream of vascular endothelial growth factor *VEGF*. It has been reported that ESRP binds downstream of the ITGA6 alternative exon to act as an activator in breast cancer cell lines (Warzecha et al., 2010b). Thus there are three splicing factors which enhance the ITGA6 alternative exon: MBNL1, RBFOX2 and ESRP1.

Although it has been reported that the *ITGA6* alternative exon is regulated, or activated, by MBNL, RBFOX and ESRPs, other important splicing factors are poly pyrimidine tract-binding proteins PTBP1 and PTBP2, which are expressed in all cells except mature neurons. PTBP1 and 2 are part of a family of three splicing factors that have largely redundant functions and which auto-regulate each other's expression. The homologues of PTBP form one of the most significant RBP families that have been studied in mammals (Valcarcel and Gebauer, 1997b). The prototypical member PTBP, or HnRNPI, generally functions as a splicing repressors. PTBP also functions in a number of diverse cellular processes, including polyadenylation, mRNA stability, mRNA localisation and translation (Sawicka et al., 2008a). PTBP binds to splicing silencers at pyrimidine-rich motifs, such as UCUU or CUCUCU in the RNA, to mediate splicing repression in a long list of alternatively spliced pre-mRNAs (Noiret et al., 2012b). In this chapter, I aimed to investigate whether PTBP regulates alternative splicing of *ITGA6*.

3.2 Aims

- To investigate splicing pattern of *ITGA6* in different cancer cell lines.
- To confirm that RBFOX, MBNL1, ESRPs activate *ITGA6* alternative exon in cancer cell lines.
- To test if PTBP acts as a potential novel regulator for the *ITGA6* alternative exon, and investigate whether PTBP acts as an activator or inhibitor.

3.3 Materials and Methods

3.3.1 Quantitative real time PCR (qPCR)

Quantitative real time PCR was used to determine gene expression using a SYBR green PCR master mix kit (applied Biosystems) and an applied for Biosystems fast real time PCR machine. In this chapter, qPCR was used to determine gene expression of the splicing regulation factors inclusion PTBPs and MBNL1 in cancer cell lines. RNA samples were collected from HeLa and MCF7 cells 72 hours after of siRNA transfection. After RNA extraction, Maxima Reverse Transcriptase First strand cDNA synthesis kit (Thermo Scientific) including enzyme mix and 5x buffer mix was used in order to make cDNA. RT-PCR was performed following the manufacturer’s instructions (Table 3.1 &3.2). cDNA was diluted in 1/20 using RNase-free water (Ambion) and used for PCR. After that the 9 ul of PCR master mix was added per well to a 96 well qPCR plate containing 1ul diluted cDNA per well. qPCRs were performed using a minimum of 3 replicate samples per sample. A control was used for each master mix which had no template. The plate was loaded into the instrument and standard cycling conditions were performed (Table 3.3 &3.4). After that, gene expression was calculated using software (Applied Biosystems) which calculated ct value and CSDS using the average of three reference genes, *GAPDH*, *Tubulin* and *Actin*. All primers used for quantitative real-time PCR (qPCR) are provided in Table 3.5.

Table 3.1: components of reverse transcription

components	Volume
RNA sample (variable volume)	3µg
5x mix	4µl
Enzyme mix	2µl
water	To complement RNA to 14 µl

Table 3.2: Temperature sequence of RT reaction

steps	Temperature (°C)	Time(min)
annealing	25	10
elongation	50	30
Termination	85	5
cold	4	∞

Table 3.3: components of qPCR

Components	Volume
2X SYBR Green PCR Master Mix	5µl
Forward primer (10µM)	1µl
Reverse primer (10µM)	1µl
RNase-free water	2µl
Total	9µl

Table 3.4 : steps of qPCR reaction

Steps	Temperature (°C)	Time	Cycles
Heat activation	50	2min	
Initial denaturation	95	10 min	
Denaturation	95	15 second	
Annealing/elongation	60	1 min	Cycling to step 3 45

Table 3.5: Primers designed to targeted regulatory proteins for qPCR

Primers	Sequences
PTBP1variable exon3 F	GCCTGACCAAGGACTACGG
PTBP1 variable exon 3 Rev	CCCCATTGCTGGAAAACA
PTBP2variable exon3 F	GTGGCTCGGTTCTTGTGA
PTBP2 variable exon 3 Rev	TGCCTGAGAGTAGTTCGTCA
RBFOX2 variable exon3 F	TCATCTATCCGTTTGGTTTAT
RBFOX2 variable exon 3 Rev	TGGCGTCAGGAGTTGTTGTC
ESRP1 variable exon3 F	AGCACTACAGAGGCACAAACA
ESRP1 variable exon 3 Rev	TGGAGAGAAACTGGGCTACC
ESRP2 variable exon3 F	GGGAGTTCGCCACAGATATTC
ESRP2 variable exon 3 Rev	AGCCATAAATGCTCTGTCCG
MBNL1 variable exon F	GCTGTTAGTGTCACACCAATTCG
MBNL1 variable exon Rev	AGGCGATTACTCGTCCATTTTC
ACTB For (reference gene 1)	CATCGAGCACGGCATCGTCA
ACTB Rev (reference gene 1)	TAGCACAGCCTGGATAGCAAC
GAPDH For (reference gene 2)	ATCATCCCTGCCTCTACTGG
GAPDH Rev (reference gene 2)	GTCAGGTCCACCACTGACAC
α6A f	CCACATATCACAAGGCTGAG
α6A Rev	CACTGTCATCGTACCTAGAG
ITGA6 constitutive f	CTGGGATCTTGATGCTTGCT
ITGA6 constitutive Rev	GCAGTTTGGGTACTGTGAAGC

3.3.2 Cell culture

3.3.2.1 Cell lines

In this chapter, three human cancer cell lines including: MCF7 (catalogue number: ATCC-HTB-22), MDA-MB-231 (catalogue number: ATCC-HTB-26) and HeLa (catalogue number: ATCC-CCL-2) were used. All this cell lines were purchased from the American type Culture collection (ATCC) and LGC stander, Europe

3.3.2.2 MCF-7

The human cell line MCF7 was originally established from pleural effusion of a 69 years old caucasian female patient with breast adenocarcinoma (Soule et al., 1973). MCF7 is a tumorigenic breast cancer cell line but less invasive than MDA-MB231 (below), early staged and ER and PR positive.

3.3.2.3 MDA-MB231

The human cell line MDA-MB231 was originally established from pleural effusion of a 51 year old female patient with breast cancer adenocarcinoma (Cailleau et al., 1978). It is tumorigenic breast epithelia with invasive characteristics in early stage. ER and PR are negative in this cell line.

3.3.2.4 Hela

HeLa is a tumorigenic cervix epithelia cell line originally derived from a 31 years old black female patient with cervical cancer. This cell line is often used in splicing assays due to high efficiency observed for transfection.

3.3.2.5 Cell culture condition

Cell culture was performed in a class 11 laminar flow microbiological safety cabinet. All cell line were grown in (25cm2 and 75cm2) flask in a 37°C incubator containing 5% CO_2. DMEM (Dulbecco's Modified Eagle's Media) without phenol red (purchased from PAA) with 10% FBS and 1% penicillin streptomycin added (purchased from Sigma-Aldrich) was used as a culture medium for all cell lines including MDA-MB231, MCF-7 and HeLa.

3.3.2.1 .6 Cell line maintenance

Every 3 to 5 days on average, cells were passaged at roughly 70-80% confluency in flasks. To do passaging, growth on media (DMEM) was removed and the cells were washed with sterile 1x phosphate buffered saline (PBS) (Sigma-Aldrich). 2mM trypsin-EDTA (Sigma-Aldrich) was added to the cell and incubated for 5 min at 37°C. Growth media was added in order to stop the effect of trypsin. This was followed by centrifugation of the cells at 200xg for 5 min in order to collect detached cells. The supernatant was removed and the pellet was suspended in complete growth media and placed in new flask at a ratio 1:5 (diluted cells to new culture media).

3.3.2.1.7 Cryopreservation of cells

To generate a continuous stock of cells, cells were routinely frozen at early passage number. After cells reached roughly a confluency of 70-80% in flasks, the cells were passaged and aliquots placed in cryoprotective media, which consists of 95% FBS with 5% dimethl sulphoxide (DMSO) (Sigma- Aldrich). This was frozen and stored at -80°C. When required, frozen cells were thawed. Thawing was performed rapidly at 37°C in a water bath, and media removed by centrifugation at 200xg for 5 min. Cells were resuspended in complete growth media and plated in new tissue culture flasks.

3.3.2.1.8 Capillary gel electrophoresis (Qiaxcell)

Capillary gel electrophoresis (Qiaxcell) is advanced fully automates sensitive system replaces traditional labor-intensive gel analysis of DNA and RNA using ready gel cartridges. It was performed using 5 µl of RT-PCR samples diluted with 5 µl Qiaxcell dilution buffer (Qiagen). This was followed by loading samples in the multi-capillary electrophoresis system. Qiaxcell biocalculator software was used to determine the size and concentration of each PCR product.

3.3.3 Calculation of percentage splicing inclusion (PSI %)

In this chapter, the percentage of splicing inclusion was calculated using following formula figure 3.1

$$\frac{\text{Concentration of 'exon included' PCR band (ng/ul)}}{\text{(Concentration of 'exon included' + 'exon excluded' PCR bands (ng/ul))}} \text{ X100 = PSI (\%)}$$

3.3.4 Gel electrophoresis

In this chapter, gel electrophoresis was used following the protocol described in chapter 2.

3.3.5 Western immunoblotting

Western immunoblotting was used to determine protein expression levels. PTBP1 and MBNL1 proteins were detected by western blotting using rabbit monoclonal PTBP1 antibody (Abcam, ab63697) and rabbit polyclonal MBNL1 antibody (Abcam, ab45889). This was followed by secondary α-rabbit HRP (all antibodies 1:1000 dilution). Efficiency of loading and transfer were measured using actin immunoblotting and GAPDH. In brief, after harvesting cells, cell pellets were lysed in 2X SDS loading buffer. This was followed by boiling at 100°C for 5 min. Proteins were separated by 10 % SDS-PAGE, followed by transfer to Hybond-P membranes (GE). Membranes were blocked for one

hour in block solution (Tris Buffer Saline (TBST) 5% with non-fat dry milk). The membrane was washed three times for 5 minutes each in TBST and then probed with primary antibody overnight at 4°C. Following this, membranes were washed three times for 5 minutes with TBST, and further probed with secondary antibody for one hour. Finally membranes were washed three times using TBST for 5 minutes. This was followed ECL detection using ECL western blotting detection reagent (Amersham). Excess ECL was then removed and filters were exposed on photographic film (Kodak). Films were developed in a Compact X4 developer (Xograph Imaging Systems).

3.3.6 siRNA Transfection:

Knockdown for the endogenous PTBP1, PTBP2, MBNL1, RBFOX2, ESRPS1 and ESRPS2 was achieved by transfecting MCF-7 and HeLa cell lines with silencer selected pre-designed siRNA (Integrated DNA Technologies (IDT)). SiRNA *PTBP1* (NM_031990)(hs.Ri.PTBP1.13.1) *PTBP2* (hs.Ri.PTBP2.13.2) (NM_021190), *MBNL1* (hs.Ri.MBNL1.13.2)(NM_021038), *RBFOX2* (hs.Ri.RBFOX2.13.2) (NM_001082578), *ESRP1* (hs.Ri.ESRP1.13.3) (NM_001122827) and *ESRP2* (hs.Ri.ESRP2.13.2)(NM_024939) were transfected using Lipofectamine RNAiMAX™ reagent (Thermofisher). Control (untreated) cells were treated with transfection reagent only. Transfection was performed following the manufacturers instructions (Table 3.6).

Table 3.6: Transfection protocol

Transfection protocol Reaction mix	Components	volume	Incubation time
Mix 1	Opti-MEM® Medium	150µL	5minutes
	Lipofectamine RNAiMAX reagent	9 µL	
Mix2	Opti-MEM® Medium	150 µL	5minutes
	siRNA (10 µM)	3 µL(30pmol	
Mix 3	Mix 1	160 µL	20 minutes
	Mix 2	155 µL	
Short Centrifugation Add 250 µL to each well in 6 well plate and incubate at 37c overnight			

3.4 Results

3.4.1 Splicing pattern of ITGA6 in different cell lines (HeLa, MCF7 and MDA-MB231), stem cells and fibroblasts

The alternative splicing of *ITGA6* has been reported to have a different patterns between stem cells and fibroblasts (Venables et al., 2013). I sought to confirm this result, and to also test the splicing pattern in cancer cell lines. I performed RT-PCR using primers that were designed to match the constitutive exons that flanked the alternative exon of *ITGA6*. This assay used RNA from the HeLa, MCF7 and MDA-MB231 cancer cell lines, as well as pre-existing cDNA from stem cells and fibroblasts as described by Venables et al. (2013). The details of the cancer cell lines, stem cells and fibroblasts are explained in the methodology chapter. The splicing pattern of the RNA expressed in each cell line was analysed using capillary gel electrophoresis. Two bands, one for exon inclusion and one for exon exclusion, could be detected by gel electrophoresis. The concentration between the isoforms was calculated using a multi-capillary Qiaxcell gel electrophoresis system (Figure 3.1).

The *ITGA6* inclusion percentage was calculated and plotted on a barchart (Figure3. 1a). The pattern of splicing *ITGA6* inclusion (long form) is predominant in fibroblast and breast cancer cell lines, including MCF7 and MDA-MB231. However, the exclusion pattern of the *ITGA6* (short form) is the predominantly form in stem cells and the HeLa cancer cells line. Taken together, the breast cancer cell lines proved to be a good model to study how the alternative exon of *ITGA6* can be activated as they roughly express the same levels of both isoforms of *ITGA6*, whereas the HeLa cancer cell line proved to be a good model to study how the *ITGA6* alternative exon can be inhibited as they have less exon inclusion.

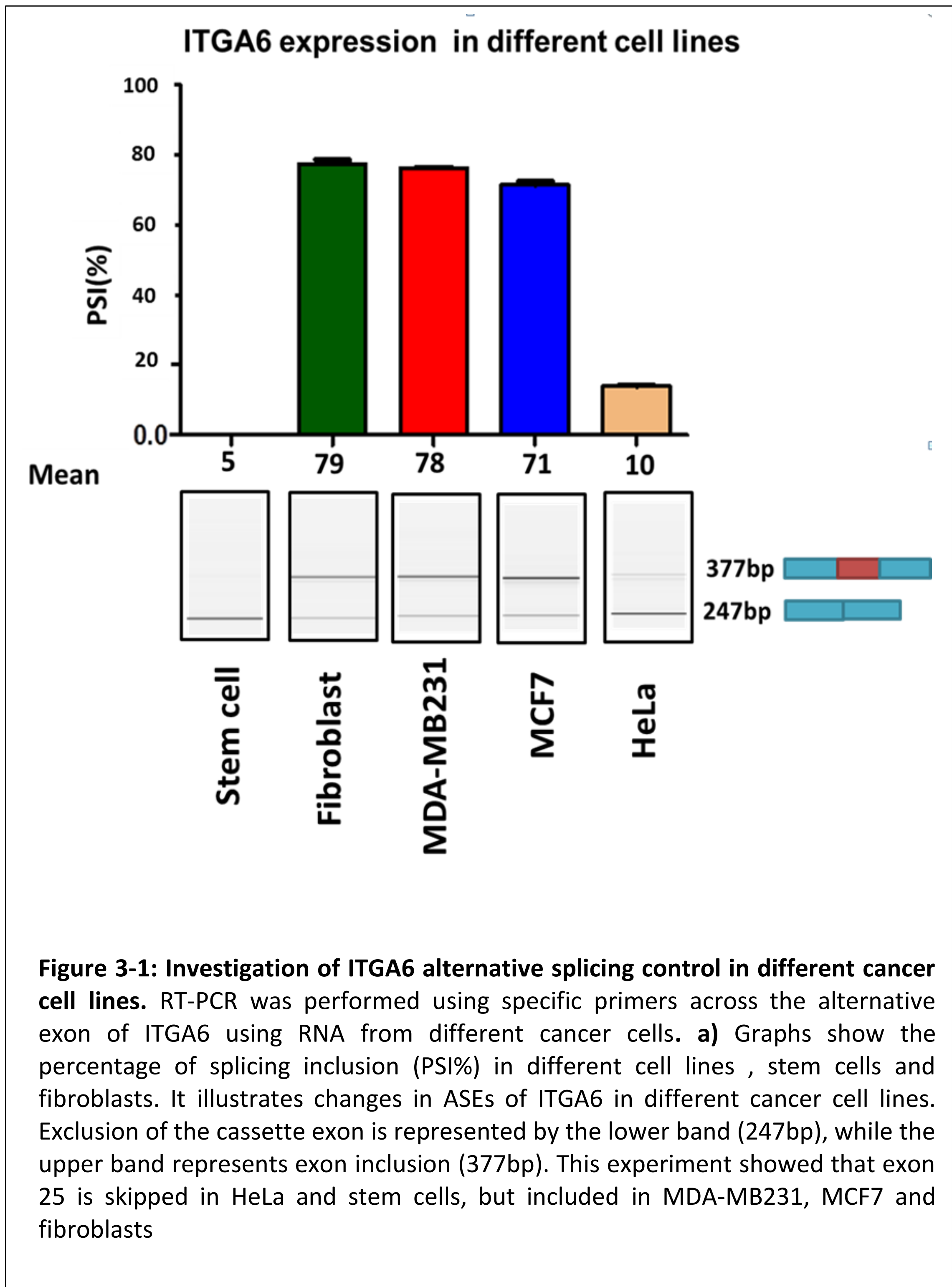


Figure 3-1: Investigation of ITGA6 alternative splicing control in different cancer cell lines. RT-PCR was performed using specific primers across the alternative exon of ITGA6 using RNA from different cancer cells**. a)** Graphs show the percentage of splicing inclusion (PSI%) in different cell lines , stem cells and fibroblasts. It illustrates changes in ASEs of ITGA6 in different cancer cell lines. Exclusion of the cassette exon is represented by the lower band (247bp), while the upper band represents exon inclusion (377bp). This experiment showed that exon 25 is skipped in HeLa and stem cells, but included in MDA-MB231, MCF7 and fibroblasts

3.4.3 SiRNA knockdown of *MBNL1, RBFOX2*, and *PTBPs* in HeLa, MCF7 and MDA-MB231 cell lines

The *ITGA6* alternative exon is regulated by MBNL and RBFOX1 during stem cell differentiation (Venables et al., 2013). Since *ITGA6* alternative splicing showed different patterns in different cancer cell lines, The knockdowns of the endogenous regulation factors, including MBNL1, RBFOX2, PTBPs and ESRPs, were tested to see if these would affect the splicing patterns of *ITGA6.*

To begin this study, siRNA targeted against these splicing factors were transfected into HeLa, MCF7 and MDA-MB231 cell lines. After 72 h, the cells were harvested, and a RT-PCR was performed using specific primers located on the flanking constitutive exon followed by gel electrophoresis in order to analyse any alternative splicing changes (Figure 3.2). As expected, the *ITGA6* splicing shifted towards the short form (exclusion of the alternative exon) upon knockdown of either MBNL1 or RBFOX2 in MCF7 and MDA-MB231 cells compared with the control (with no siRNA) (Figure 3.2a,b). However, the *ITGA6* shifted from the long form (inclusion of the alternative exon) to the short form (exclusion of the exon) upon knockdown of RBFOX2 in HeLa cell line compared with the control (Figure 3.2c).. On other hand, the ITGA6 splicing shifted towards the long form (inclusion of the alternative exon) upon double knockdown of PTBP1 and PTBP2 in MDA-MB231, MCF7 and HeLa cells compared with the control.

Overall, these results confirmed that the *ITGA6* alternative spliced exon is enhanced by *MBNL1* and *RBFOX2*. The results also showed for the first time that PTBP inhibits this exon. The ITGA6 alternative exon is thus under the control of three major splicing factors.

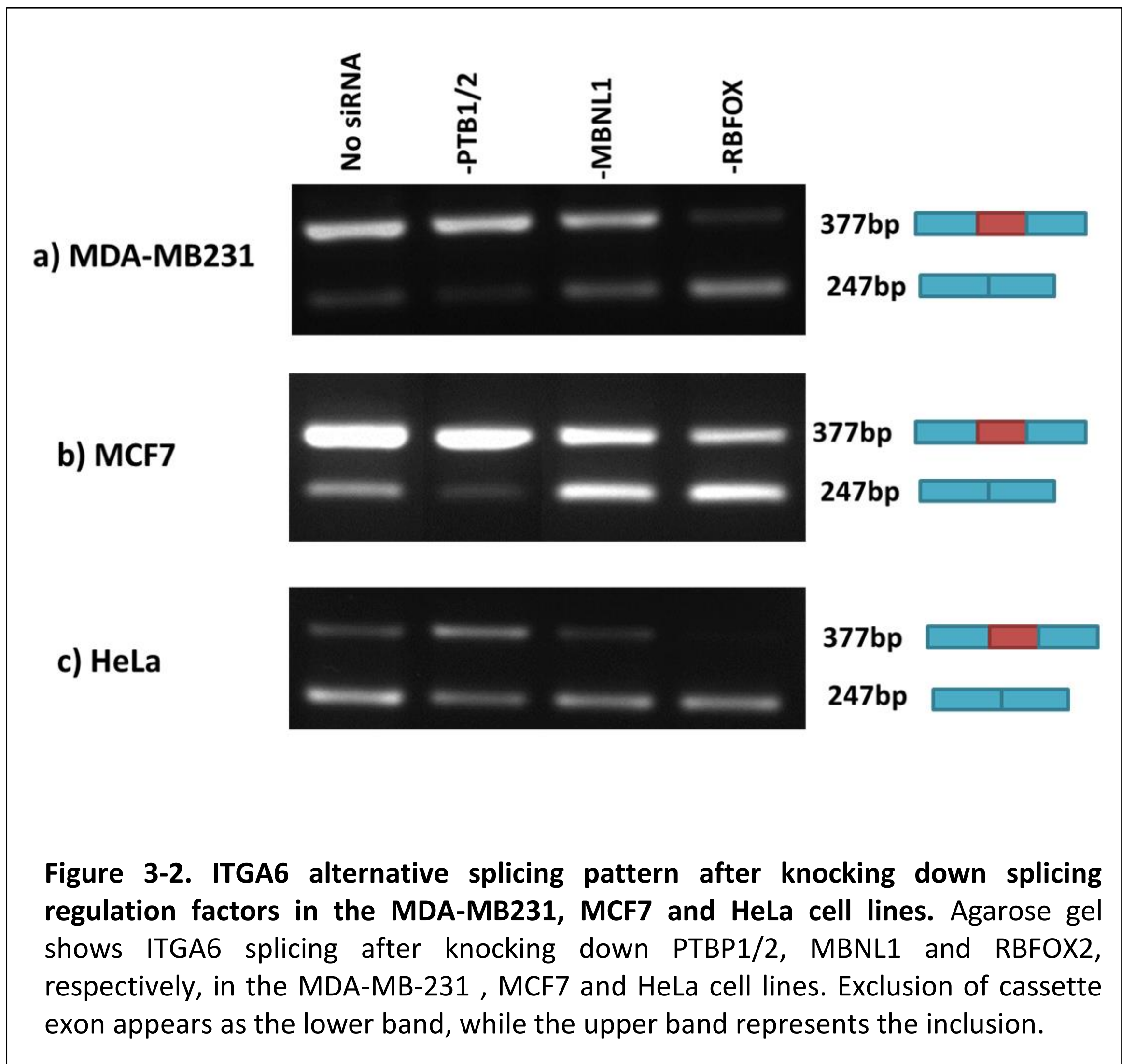


Figure 3-2. ITGA6 alternative splicing pattern after knocking down splicing regulation factors in the MDA-MB231, MCF7 and HeLa cell lines. Agarose gel shows ITGA6 splicing after knocking down PTBP1/2, MBNL1 and RBFOX2, respectively, in the MDA-MB-231 , MCF7 and HeLa cell lines. Exclusion of cassette exon appears as the lower band, while the upper band represents the inclusion.

3.4.3.1 Confirmation that splicing regulatory factors affect (activate) the alternative exon of ITGA6 in MCF7 cell lines

The *ITGA6* alternative exon has different splicing patterns in different cancer cell lines. Splicing regulatory factors, including MBNL1, RBFOX2 and PTBPs have been shown to regulate the exon. The *ITGA6* alternative exon has also been reported to be enhanced by ESRP1, which is part of the positive feedback loop downstream of VEGF. I decided to confirm this result using the MCF7 cell line due to the clearly of splicing events in pervious results.

To begin this test, the siRNAs specific for MBNL1, RBFOX, PTBP and ESRP were transfected (in triplicate) into the MCF7 cell line. After 72 h, the cells were harvested and RNA purified. Reverse transcription of the RNA was then performed. The splicing pattern of the endogenous RNA was analysed using capillary gel electrophoresis. *ITGA6* showed two bands on the gel: a long form (inclusion of the exon) and a short form (exclusion of the exon). Ratios of the two bands were determined using a multi-capillary Qiaxcell gel electrophoresis system in order to calculate the expression of both isoforms of *ITGA6* splicing. The percentages of the alternative splicing pattern were calculated and plotted on a graph (Figure 3.3). The method of calculation is described in method section 3.3.3.

As expected, the *ITGA6* alternative exon had shifted from the long form (inclusion) to the short form (exclusion) upon knockdown of MBNL1 and RBFOX2 compared with the control (untreated or negative siRNA) cell. Double knockdown of ESRP1 and ESRPS2 did not show any change in the splicing pattern of ITGA6 compared with the control cells. This result confirms that the *ITGA6* alternative exon is enhanced by MBNL1 and RBFOX2 but suggests ESRP proteins are not involved.

In contrast, the *ITGA6* alternative exon shifted from the short form (exclusion) to the long form (inclusion), with a double knockdown of PTBP1 and PTPB2 compared with the negative siRNA treated cell. This result confirms that *ITGA6* alternative splicing is repressed by PTBPs.

Overall, these results confirmed that the *ITGA6* alternative splicing exon is enhanced by MBNL and RBFOX, and inhibited by PTBPs.

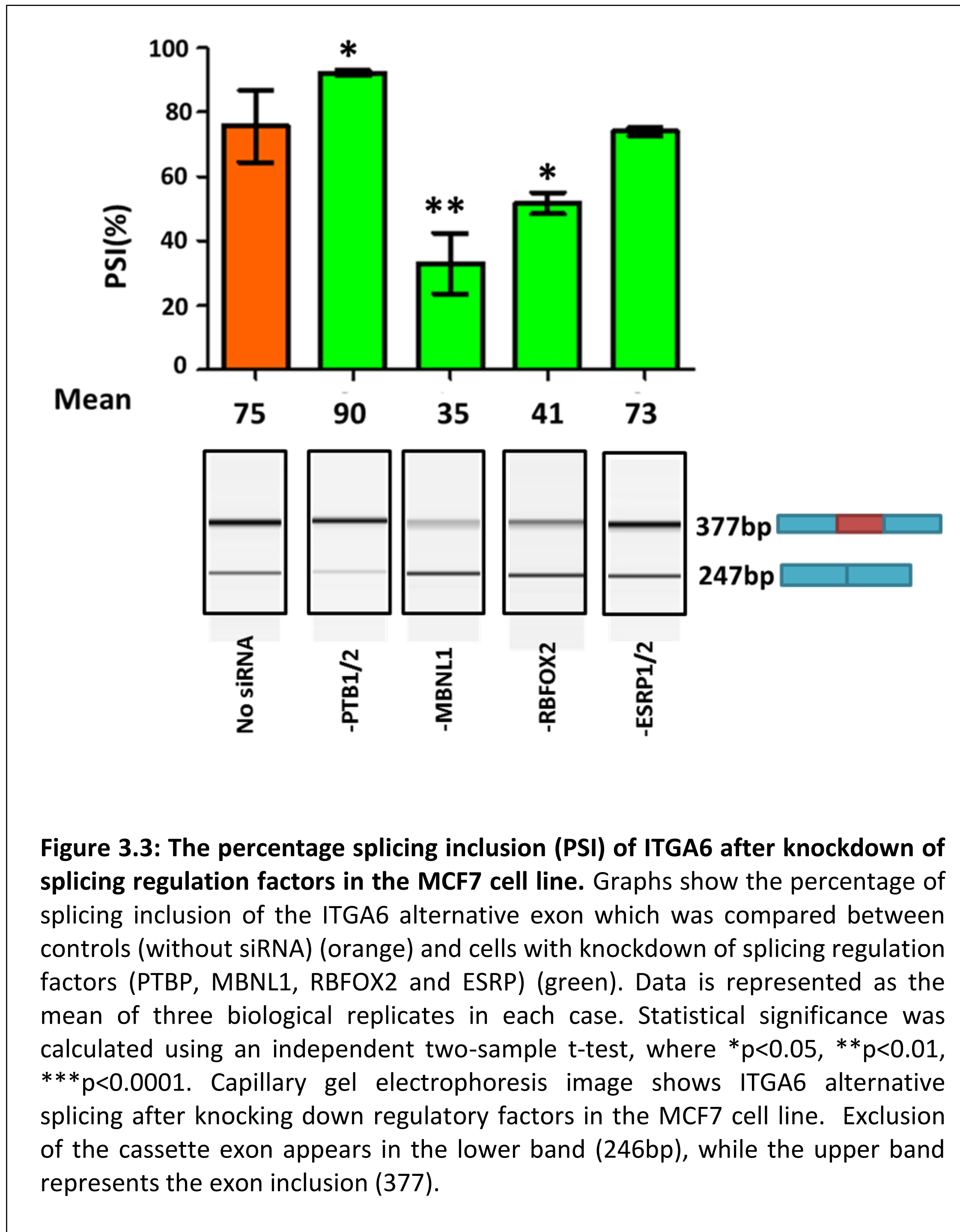


Figure 3.3: The percentage splicing inclusion (PSI) of ITGA6 after knockdown of splicing regulation factors in the MCF7 cell line. Graphs show the percentage of splicing inclusion of the ITGA6 alternative exon which was compared between controls (without siRNA) (orange) and cells with knockdown of splicing regulation factors (PTBP, MBNL1, RBFOX2 and ESRP) (green). Data is represented as the mean of three biological replicates in each case. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. Capillary gel electrophoresis image shows ITGA6 alternative splicing after knocking down regulatory factors in the MCF7 cell line. Exclusion of the cassette exon appears in the lower band (246bp), while the upper band represents the exon inclusion (377).

3.4.3.1.1 Monitoring MBNL knockdown in MCF7 cell lines

As previously described result (3.4.3.1), *ITGA6* splicing is activated by MBNL1 and RBFOX2 in the MCF7 cell lines. Because the knockdown of MBNL1 in MDA-MB231, MCF7 and HeLa cancer cell lines showed a clear shift from the long form to the short form compared with knockdown of RBFOX2*,* studied was focused on MBNL1 as a strong repressor factor. While the knockdown of MBNL1 was shown to enhance splicing the *ITGA6* alternative exon via quantitative RT-PCR, it was important to confirm the efficiency of the knockdown of this protein.

The siRNA specific to MBNL1 was transfected into the MCF7 cell line, and levels of this protein analysed via Western immunoblotting using a rabbit polyclonal antibody against MBNL and a rabbit monoclonal antibody against GAPDH as a control. 72 h after of transfection, efficient depletion of MBNL1 was observed in the transfected cells and compared with the negative control siRNA transfected cells (Figure 3.4A).

Although the MBNL knockdown efficiency was confirmed by Western immunoblotting, this was also confirmed by qPCR as well. MBNL1 siRNAs were transfected into MCF7 cells. After 72 hours, cells were harvested and cDNA was made from each cell. Using specific primers for MBNL1 (see Section 3.3.1), qPCR was performed to determine the *MBNL* expression. After the fold change was calculated (refer to method Section 3.3.3), the efficiency of MBNL depletion had significantly changed compared with the untreated cells (Figure 3.4B).

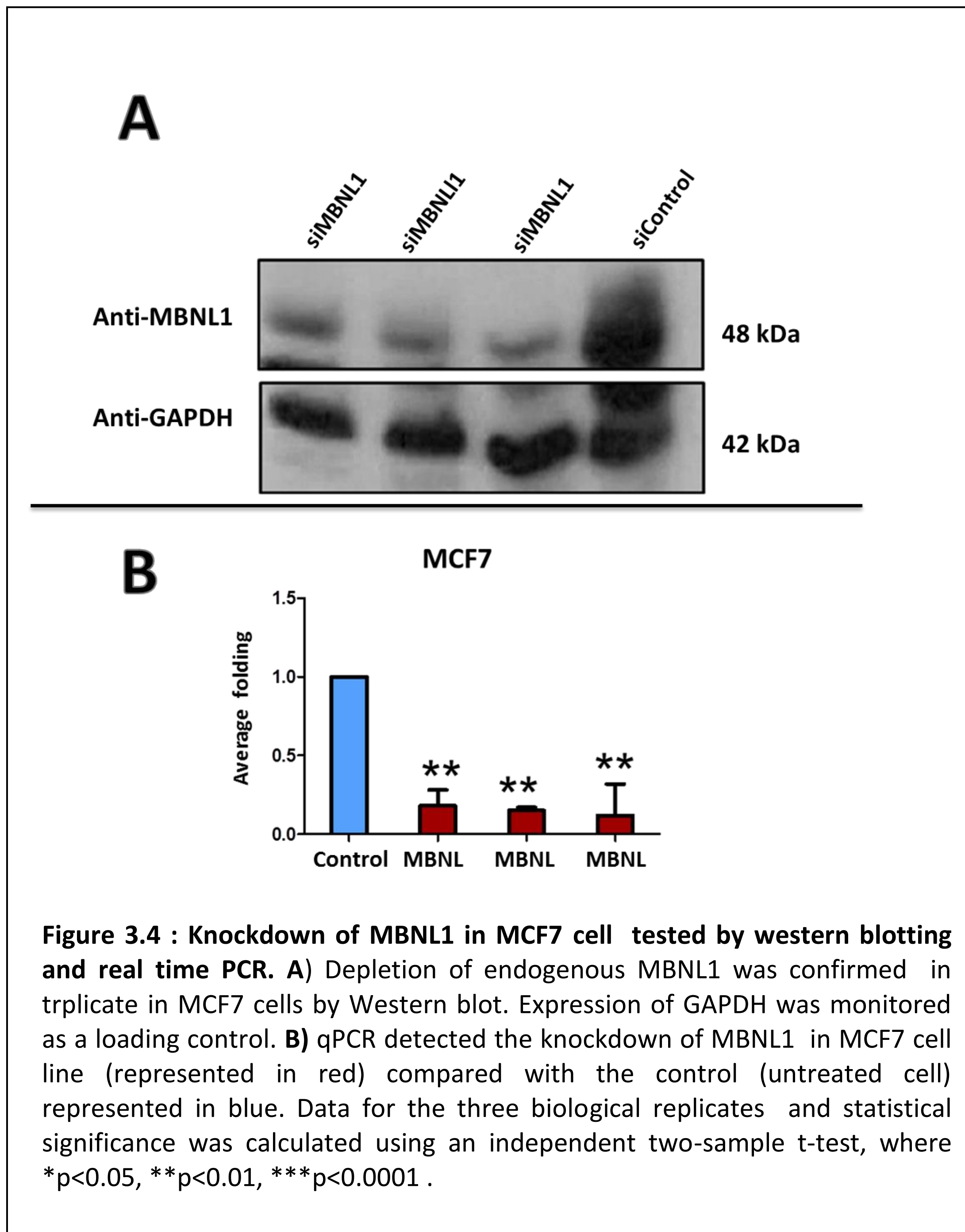


Figure 3.4 : Knockdown of MBNL1 in MCF7 cell tested by western blotting and real time PCR. A) Depletion of endogenous MBNL1 was confirmed in trplicate in MCF7 cells by Western blot. Expression of GAPDH was monitored as a loading control. **B)** qPCR detected the knockdown of MBNL1 in MCF7 cell line (represented in red) compared with the control (untreated cell) represented in blue. Data for the three biological replicates and statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001 .

3.4.3.2 Investigating the regulatory factors that inhibit the alternative exon of ITGA6 in the HeLa cell line

Previous experiments in different cell lines identified that the splicing regulatory factors, MBNL, RBFOX and ESRPs, activated the alternative exon of ITGA6, and showed that PTBP acted as an inhibitor for the exon (refer to Section 3.4.3). The PTBP1 and PTBP2 double knockdown showed significant changes in the splicing pattern of *ITGA6* which inhibit alternative exon (more inclusion exon) in the MCF7 cell line (Figure 3. 3). However, it was important to test what happened in other cell lines that might have a different pattern of splicing control.

I investigated the splicing pattern of *ITGA6* using the HeLa cancer cell line, which had more exclusion for the alternative exon of ITGA6. siRNAs specific for MBNL1, RBFOX, PTBP (PTBP1&PTBP2), ESRP (ESRP1&ESRP2) were transfected in triplicate to the HeLa cell line. After 72 h, the cells were harvested and RNA was reverse transcribed. The splicing pattern of the endogenous RNA was analysed using capillary gel electrophoresis and the concentration for each form (long form and short form) was calculated using a multi-capillary QIAxcel gel electrophoresis system. The percentages of the alternative splicing pattern were calculated and plotted on a barchart.

As expected, the *ITGA6* alternative exon shifted from the long form (inclusion of the exon) to the short form (exclusion of the exon) upon the double knockdown of PTBP1 and PTBP2 compared with untreated cells. In other words, the double knockdown of PTBP1 and PTBP2 resulted in a significant decrease in percent splicing in (PSI) of the *ITGA6* exon from 68% to 41%. This result confirmed that PTBP (PTBP1 and PTBP2) act as repressors of the *ITGA6* alternative exon (Figure3.5).

The knockdown of MBNL1 resulted in a significant decrease in PSI from 68% to 82% in the *ITGA6* exon as it shifted from the long form to the short form. However, the knockdown of RBFOX2 and ESRPs (ESRPs1 and ESRPS2 together) did not show an effect on the *ITGA6* alternative exon compared with untreated cells. This result confirmed that the *ITGA6* alternative exon was activated mainly by MBNL1.

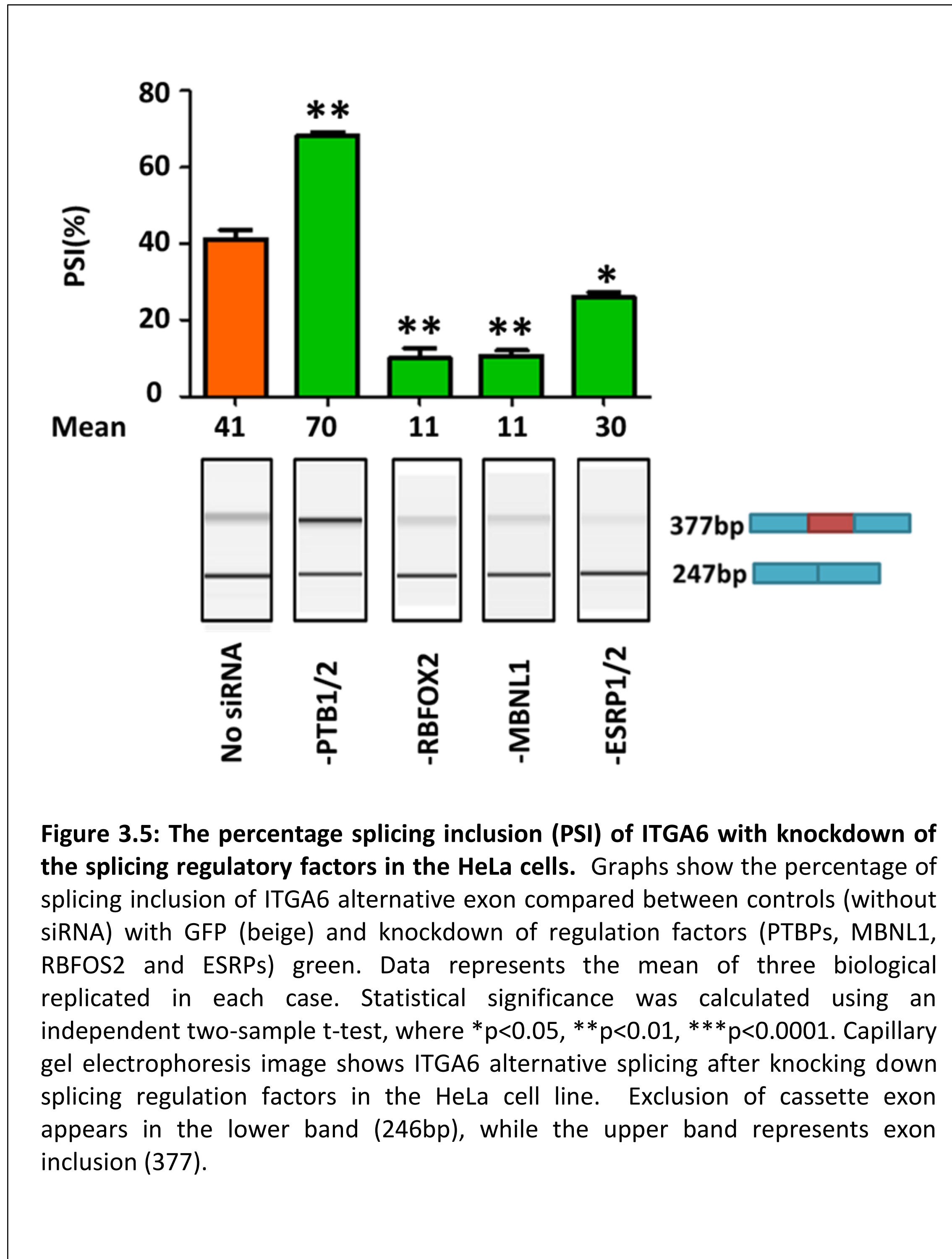


Figure 3.5: The percentage splicing inclusion (PSI) of ITGA6 with knockdown of the splicing regulatory factors in the HeLa cells. Graphs show the percentage of splicing inclusion of ITGA6 alternative exon compared between controls (without siRNA) with GFP (beige) and knockdown of regulation factors (PTBPs, MBNL1, RBFOS2 and ESRPs) green. Data represents the mean of three biological replicated in each case. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. Capillary gel electrophoresis image shows ITGA6 alternative splicing after knocking down splicing regulation factors in the HeLa cell line. Exclusion of cassette exon appears in the lower band (246bp), while the upper band represents exon inclusion (377).

3.4.3.2.1 Monitoring PTBP knockdown in HeLa cell lines

It is important to confirm the previous experiment (section 3.4.3.2) that indicated that the depletion of PTBP1 and PTBP2 repressed the ITGA6 alternative exon in cancer cell lines. siRNAs for PTBP1 were transfected into the HeLa cell line and PTBP1 analyzed using Western immunoblotting with a rabbit polyclonal antibody against PTBP1 and a rabbit monoclonal antibody against actin as a control. The expression of PTBP1 protein was detected on the Western immunoblot, and corresponded to the expected size of 57 kDa for the full length of PTBP1 in the control (untreated cells). However, this band was significantly reduced after knockdown of PTBP1*,* which indicated that this band was specific for PTBP1 compared with the control (Figure 3.6a).

Although the PTBPs knockdown efficiency was confirmed by Western immunoblotting, I wanted to confirm this result by doing qPCR. The PTBP1 and PTBP2 double siRNAs were transfected to the HeLa cell line. After 72 h, the cells were harvested and cDNA was made from each cell. Using specific primers for PTBP1, qPCR was performed to determine PTBP1 expression. The levels of PTBP1 significantly changed compared with the untreated cells after depletion (Figure 3.6b).

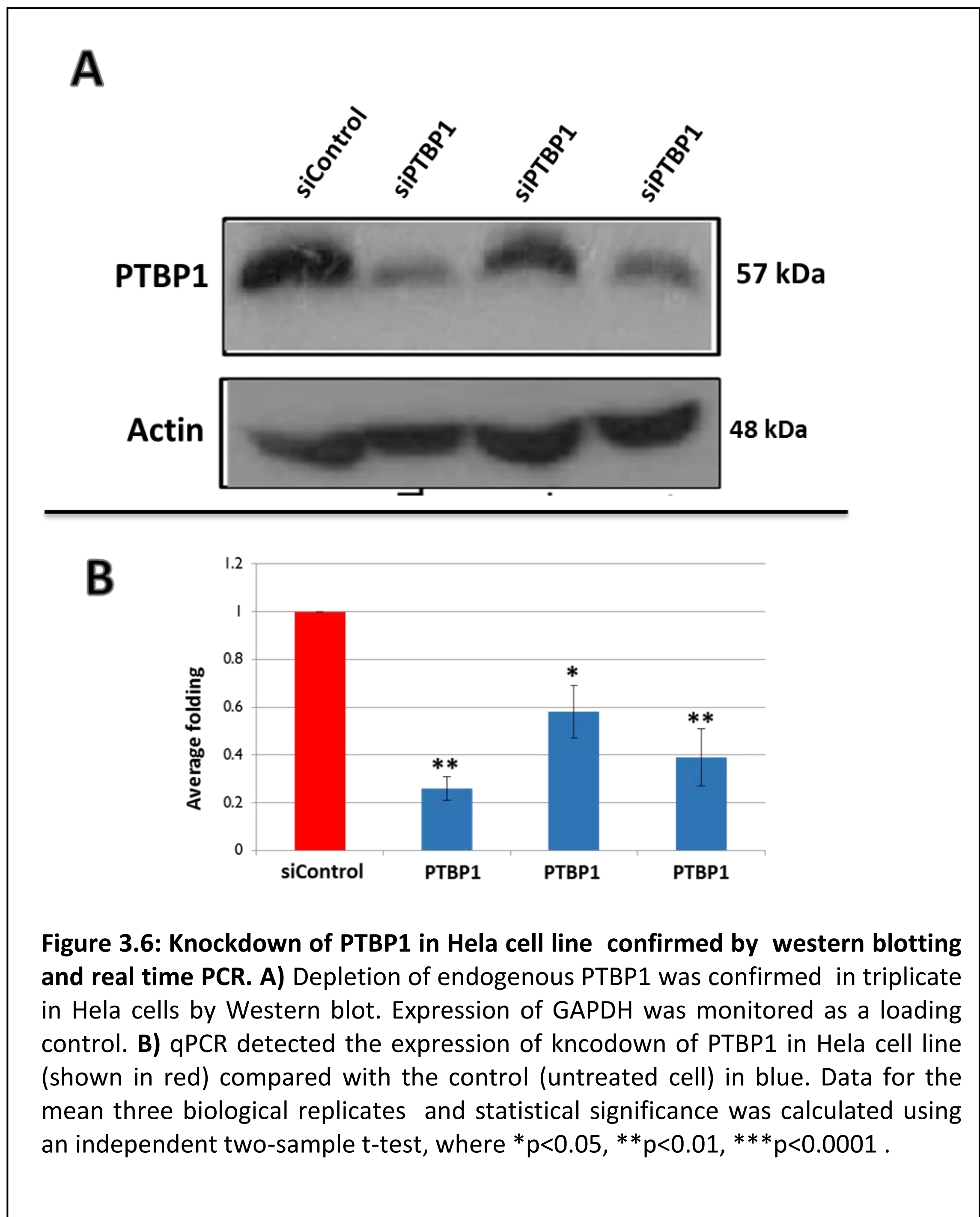


Figure 3.6: Knockdown of PTBP1 in Hela cell line confirmed by western blotting and real time PCR. A) Depletion of endogenous PTBP1 was confirmed in triplicate in Hela cells by Western blot. Expression of GAPDH was monitored as a loading control. **B)** qPCR detected the expression of kncodown of PTBP1 in Hela cell line (shown in red) compared with the control (untreated cell) in blue. Data for the mean three biological replicates and statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001 .

3.4.2 Measurement of the relative levels of factors that regulate *ITGA6* alternative splicing

I have demonstrated the *ITGA6* alternative splicing pattern in cancer cell lines, stem cells and fibroblasts and also shown that an alternative exon of *ITGA6* was activated by RBFOX2, MBNL1 and ESRP and inhibited by PTBPs in the cancer cell line. Since these regulatory factors have significant implications for the alternative splicing pattern, it was important to investigate the level of protein regulatory factors, including PTBP, RBFOX2, MBNL1 and ESRP, in the different cell lines.

I analysed the expression of splicing regulation factors (the PTBP, RBFOX2, MBNL1 and ESRP isoforms) in different cell lines, including MCF7, MDA-MB231 and HeLa. I also used pre-existing cDNA for the stem cells and fibroblast, as demonstrated by Venables et al. (2013), and performed real-time quantitative PCR (qPCR) using the specific primers for the regulation factors (see primers, table 3.5, in the Methods section).
To determine the expression of regulation factors that control *ITGA6* alternative exon in different cell lines, I compared expression levels in cell types that have variant *ITGA6* isoform expression. MCF7 breast cancer cell lines were previously shown to have roughly the same level of expression for both the *ITGA6* isoforms (inclusion and exclusion for alternative exon) (see Section 3.4.1). I compared MCF7 cell line with MDA-MB231 and HeLa cell lines as well as fibroblast and stem cells for the expression of the splicing factors PTB, RBFOX2, MBNL1 and ESRP.

After calculating the fold change, the expression of PTBP1 was found to be highest in the HeLa cell lines, whereas the expression of PTBP2 was extremely high in the HeLa cell lines compared with the other cell lines. The expression of MBNL1 had roughly the same range of expression level in the different cancer cell lines, but was expressed at an extremely low levels in stem cells. RBFOX2 was expressed at a high level in MCF7 compared with the other cancer cell lines, stem cells and fibroblasts. ESRP1 exhibited roughly the same level of expression in the different cell lines; however, it had an extremely low expression in fibroblast. Finally, ESRP2 was expressed at a significantly high level in the Hela cell lines.

Overall, PTB (PTBP1 and PTBP2) were found to be expressed significantly higher in HeLa cell lines, which suggests that a differential expression of PTB might explain the splicing pattern of *ITGA6* in HeLa cells. On other hand, PTB, RBFOX, MBNL and ESRP were expressed at roughly the same levels in MCF7 and MDA-MB231. This suggests that the roughly equivalent expression of two isoforms of *ITGA6* might explain the splicing pattern of *ITGA6* in the MCF7 and MDA-MB231 cell lines. PTB, MBNL, RBFOX and ESRP1 were expressed at low levels for fibroblast and stem cells, while ESRP1 was expressed at a high level in the stem cells (Figure 3.7).

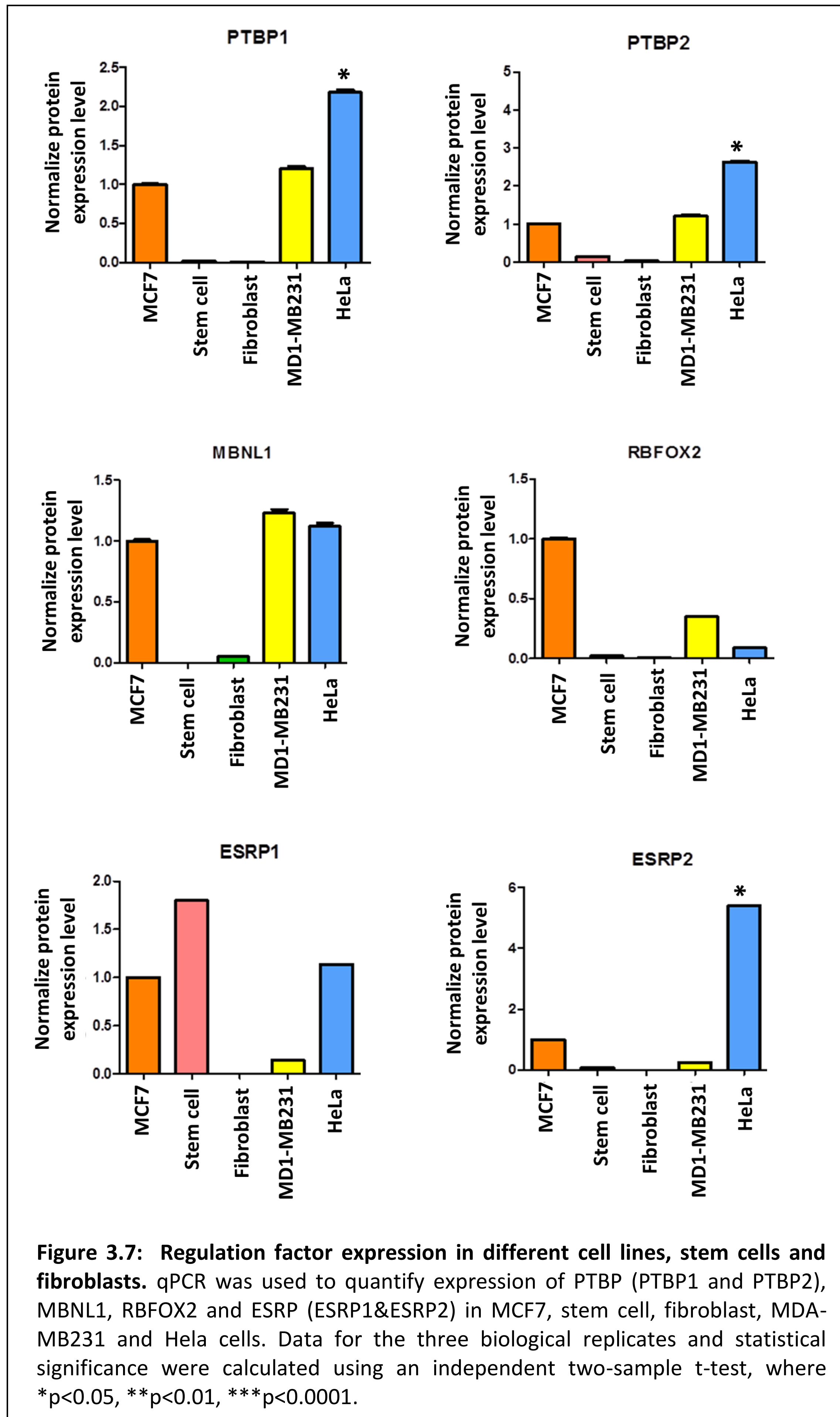


Figure 3.7: Regulation factor expression in different cell lines, stem cells and fibroblasts. qPCR was used to quantify expression of PTBP (PTBP1 and PTBP2), MBNL1, RBFOX2 and ESRP (ESRP1&ESRP2) in MCF7, stem cell, fibroblast, MDA-MB231 and Hela cells. Data for the three biological replicates and statistical significance were calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001.

3.5 Discussion

3.5.1Investigation of ITGA6 alternative splicing in different cancer cell lines

In this chapter, I describe an in-depth investigation of the mechanism and regulation of endogenous integrin α6 (ITGA6) alternative splicing. Cancer cell lines were selected for this study because they are easy to grow and induce to express most genes and transcripts. The MDA-MB231 and MCF7 breast cancer cell lines were selected since ITGA6 may be expressed in breast carcinoma cells under stress conditions through the upregulation of vascular endothelial growth factor (VEGF) expression, either at the level of transcription or translation. MDA-MB231 and MCF7 were both originally isolated from metastatic breast cancer cell lines; however, MCF7 is less aggressive than MDA-MB231 (Chung et al., 2002; Chung et al., 2004a). We also used the HeLa cervical cancer cell line since other researchers have used it successfully for studying alternative splicing mechanisms.

We started by investigating different cancer cell lines in order to determine the differences in their *ITGA6* alternative splicing patterns. During differentiation of stem cells into fibroblasts, the *ITGA6* splicing mechanism switches from inclusion to exclusion of the alternative exon (Venables et al., 2013a). In the previous chapter, we identified different *ITGA6* splicing patterns between the head and tail of a zebrafish embryos as well as changes in these patterns during development. In order to determine whether different ITGA6 splicing patterns exist in different cell lines, reverse transcription polymerase chain reaction (RT-PCR) was performed with a primers flanking alternative exon 25 of *ITGA6*. We found that both *ITGA6* splicing patterns (inclusion and exclusion) occur for the alternative exon 25, with roughly the same percentages of each pattern in both breast cancer cell lines. This suggests that breast cancer cell lines are good models to study ITGA6 alternative splicing since, like the zebrafish tail, they express two forms of *ITGA6*. Both isoforms of mRNA may play an important role in breast cancer, which is supported by findings that *ITGA6a* and *ITGA6b* have different functions in cancer stem cells (Goel et al., 2014). The HeLa cell line, on other hand, exhibited more exclusion than inclusion of the alternative exon, indicating that the shorter form of *ITGA6* may be more important in cervical cancer.

The HeLa cell line, therefore, might be a good model to study the repression of *ITGA6* alternative splicing.

To summarise, we confirmed that *ITGA6* exhibits both alternative splicing patterns in the breast cancer cell lines MCF7 and MDA-MB231, and showed for the first time that a particular splicing pattern is established in the HeLa cell line. These results confirm that cancer cell lines are good models to investigate alternative splicing mechanisms controlling *ITGA6*.

3.5.2 How is *ITGA6* alternative splicing regulated?

ITGA6 splicing is regulated by MBNL1 and RBFOX during differentiation of stem cells into fibroblasts (Venables et al., 2013). Epithelial splicing regulatory proteins (ESRPs) regulate *ITGA6* alternative exon splicing in breast cancer cell lines (Chang et al., 2007). Therefore, we screened the effects of these known splicing regulatory proteins by transfection of siRNA with knockdown of the MBNL1, RBFOX2 and ESRP into cancer cell lines. *ITGA6* alternative exon 25 was then strongly excluded in the MCF7 and MDA-MB231 breast cancer cell lines, confirming the previous finding that the *ITGA6* alternative splicing pattern is positively regulated by MBNL1 and RBFOX2 during cell differentiation (Venables et al., 2013). Chang et al. (2007) also confirmed that the inclusion of the *ITGA6* alternative exon is activated by ESRP in breast cancer cell lines.

MBNL1 is involved in myotonic dystrophy and modulates splicing during muscle and heart development. RBFOX is a splicing factor that is implicated in the epithelial-to-mesenchymal transition. ESRP1 and ESRP2 are known to regulate alternative splicing during epithelial and mesenchymal differentiation. The splicing activity of RNA-binding proteins (RBPs), such as MBNL1, RBFOX and ESRP, are flexible according to their binding site (Dredge et al., 2005; Zhang et al., 2008a; Goers et al., 2010; Llorian et al., 2010). MBNL1, RBFOX and ESRP enhance exon inclusion when they bind downstream of the exon but inhibit it when they bind upstream (Figure 3.8).

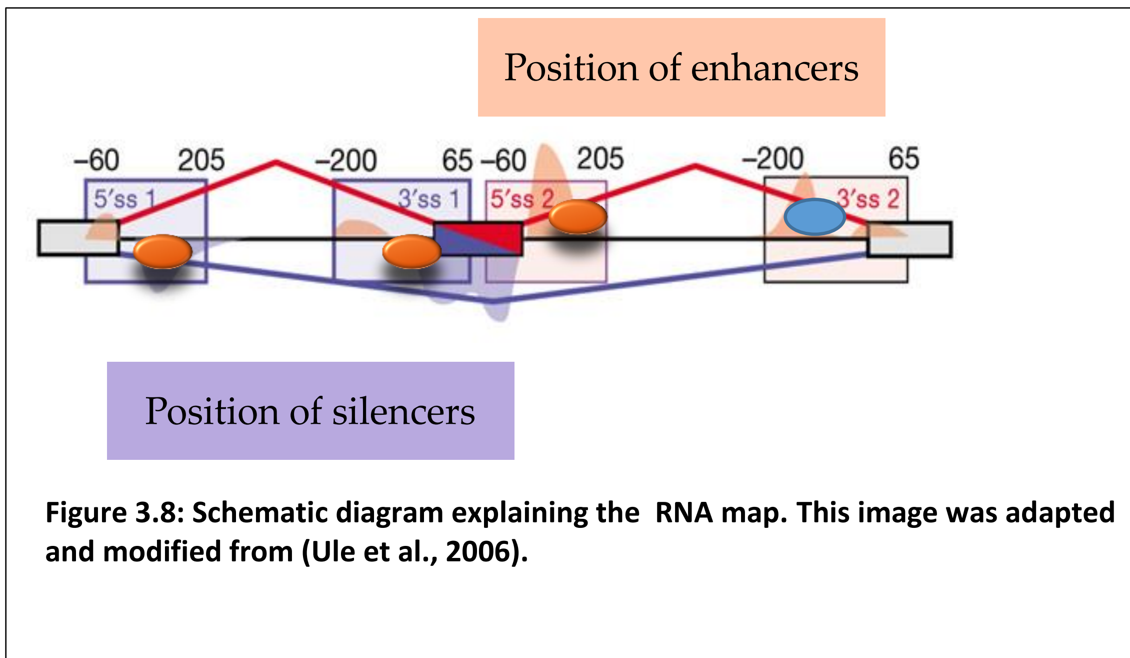


Figure 3.8: Schematic diagram explaining the RNA map. This image was adapted and modified from (Ule et al., 2006).

3.5.3 Identification of PTBP as a novel regulator

To fully understand the *ITGA6* alternative splicing mechanism, it was important to find other regulatory factors which perform opposite functions to ESRP, MBNL1 and RBFOX2. Polypyrimidine tract-binding protein (PTBP) is widely known as a repressor. Therefore, we performed transfection of siRNA with knockdown of PTBP in breast cancer cell lines. Surprisingly, knockdown of PTBP strongly reduced the exclusion of exon 25 from *ITGA6* in the breast cancer cell lines, which indicates that PTBP inhibits the alternative splicing of *ITGA6* exon 25. Knockdown of PTBP caused an even greater degree of exon inclusion in the HeLa cell line. PTBP1 is widely expressed and belongs to a family of RBPs that includes two other paralogs, PTBP2 and PTBP3, which are expressed in a more tissue-restricted manner (Noiret et al., 2012). PTBP binds to pyrimidine-rich upstream elements in the RNA, which usually act as repressors for the alternative splicing of pre-mRNA. However, there are also indications that PTBP can be involved in splicing modulation through the binding of downstream pyrimidine-rich elements.

We have observed the ITGA6 alternative splicing patterns after knockdown of the regulation factors MBNL1, RBFOX, ESRP and PTBP in triplicate experiments in both the HeLa and MCF7 cell lines. The results showed clearly that *ITGA6* alternative splicing is activated by MBNL1 and inhibited by PTBP in both cell lines; the protein expression was tested and confirmed using western blotting. These findings will contribute to the understanding the mechanisms of *ITGA6* alternative splicing, following the discovery of combinatorial control of alternative splicing which describes how proteins can interact with each other in order to regulate alternative exon splicing (Smith & Valcarcel, 2000).

The PTBP and MBN1 regulatory factors can also affect gene function via splicing. *ITGA6a* is involved in the proliferation of colon cancer cells (Groulx et al., 2014), and *ITGA6b* is involved in cancer stem cell function by means of the VEGF loop (Goel et al., 2014). In the VEGF loop, *ITGA6b* expression is repressed by the ESRP1 factor and associated with the VEGF signalling pathway, which represses *ITGA6a* expression and sustains *ITGA6b* expression in breast cancer (Figure 3.9). Our results thus suggest there might be other factors, including PTBP and MBNL1, involved in the splicing of different *ITGA6* variants.

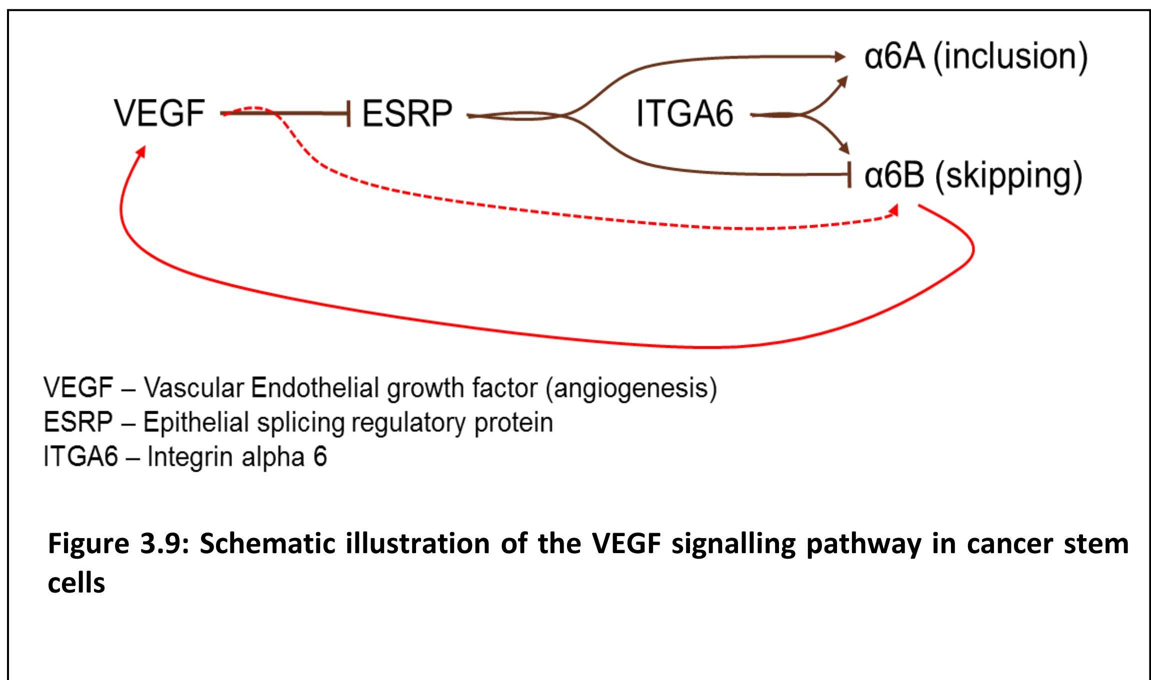


Figure 3.9: Schematic illustration of the VEGF signalling pathway in cancer stem cells

The last purpose of this chapter is to describe our investigation of the expression of different regulation factors which might affect *ITGA6* splicing. We performed quantitative polymerase chain reaction (qPCR) to quantitate expression of

PTBP1, PTBP2, MBNL1, RBFOX2, ESRP1 and ESRP2 in the different cell lines. ESRP2, PTBP1 and PTBP2 were significantly expressed in the HeLa line. PTBP1 was not expressed in fibroblasts nor in stem cells. MBNL1 was expressed in roughly the same amounts in the different cell lines, while RBFOX was expressed in greater amounts in MCF7. In summary, PTBPs had the highest expression in the HeLa cell line, whereas MBNL and RBFOX were expressed at roughly the same levels in HeLa and MCF7. These experiments suggested that the expression level of PTBP may be the factor that differentiates these cell lines in respect to *ITGA6* splicing patterns.

Our overall findings show that *ITGA6* exhibits different alternative splicing patterns in cancer cell lines, just as in zebrafish fibroblasts and stem cells. The *ITGA6* alternative exon is activated by the three regulatory proteins MBNL1, RBFOX and ESRP and repressed by PTBP regulatory factors. PTBP may also be involved in cell differentiation, as it was highly expressed in all cancer cell lines.

Chapter 4: Establishment of a minigene system to study regulation of the *ITGA6* alternative exon

4.1 Introduction

In the previous chapter, it was demonstrated that *ITGA6* alternative splicing produces two isoforms, with exon 25 included or excluded in different cancer cell lines. It was also shown that splicing factors, including MBNL1, RBFOX2, ESRPs and PTBPs, regulated the *ITGA6* endogenous alternative exon. In this chapter, using a minigene system, I carried out an in-depth investigation of how the *ITGA6* alternative splicing mechanism occurs. I hypothesised that establishing a minigene model could help me to investigate how *ITGA6* alternative exons can be regulated and to identify the regulation factors' binding sites.

The minigene system is the gold standard method for exploring RNA splicing mechanisms (Cooper, 2005). This approach includes the generation of plasmid constructs containing parts of genes or entire genes. These constructs are used to investigate the process of splicing and the related control mechanisms. Minigene constructs were prepared by digesting both the insert (template amplified target region) and the vector plasmid with a restriction enzyme (usually selected based on making compatible sticky ends); these were then ligated together to obtain the constructs. These constructs have the ability to be introduced into different cell lines to study the splicing mechanisms. The advantages of minigenes include that they can help to test for cell-, tissue- or species-specific splicing effects (Cooper, 2005). The minigene method can also test the function of sequences of variable size, from 10 nucleotides up to several kilobases. Although the minigene method is used to test the splicing effect of nucleotide changes on exon inclusion/exclusion, it also can be used to test the effect of splicing site alterations, ESE/ESS sites and even in-treatment trials (Cooper, 2005).

Previously (Venables et al., 2013b), it was found that RBFOX1 and MBNL1 regulate the splicing of genes, including *ITGA6*, in embryonic stem cells and embryos. The RBFOX proteins are among the most sequence-specific RNA-binding proteins; and bind to the hexanucleotide UGCAUG. The MBNL protein also exhibits high affinity in binding to YGCY sites, which are likely to be found approximately every 256 bases. Both RBFOX and MBNL1 have been observed to act as repressors for exons when binding downstream, whereas they act as activators when binding upstream of the exons. ESRPs have also been reported to regulate alternative splicing of *ITGA6* in breast cancer cell lines. ESRPs have been reported to activate *ITGA6* alternative exons through binding to UGG-rich motifs downstream of the *ITGA6* alternative exon (Goel et al., 2014).

In this Chapter, based on the data from the previous Chapter, I aimed to establish an *ITGA6* minigene, consider the candidate binding sites to investigate the relationship between the splicing regulator proteins and assess how they regulate *ITGA6* in cancer cell lines. Our novel findings in chapter 2 showed that the *ITGA6* alternative exon was inhibited by PTBPs. PTBP proteins, which generally function as repressors, have an affinity for binding at pyrimidine-rich motifs, such as UCUU or CUCUCU, to mediate splicing repression in a long list of alternatively spliced pre-mRNAs (Noiret et al., 2012a). I aimed to identify the PTBP binding site that regulates *ITGA6* alternative splicing. PTBPs binding upstream of the alternative exon generally repress splicing. In the final part of this chapter, I investigate the possibility that PTBPs binding downstream of the ITGA6 exon might control splicing patterns of *ITGA6* by analysing a mutant PTBP binding site in the minigene system. The *ITGA6* minigene was investigated using cancer cell lines, including Hela, MDA-MB231 and MCF7.

4.2 Aims

The aims of this chapter were to:

- To establish minigene construct for *ITGA6* that would mimic the endogenous gene.
- To investigate how alternative splicing regulators control this *ITGA6* minigene
- To identify the binding site of PTBP downstream of the *ITGA6* alternative exon and investigate the mechanism of splicing regulation.

4.3 Methodology

4.3.1 ITGA6 Minigene essay:

4.3.1.1 Primer design

The *ITGA6* alternative exon and roughly 332 nucleotides of flanking upstream intron and 900 nucleotides of downstream intron sequence were selected for cloning into the pXJ41 vector. Primers were designed using Primer3 program in the following link: http://bioinfo.ut.ee/primer3-0.4.0/primer3/ . *Mfe1* restriction sites were included at the 5' end of each primer. The same procedure was also followed using different different primers to clone different sizes of *ITGA6*. Primers are listed in Table 4.1.

Table 4.1: Primers designed to clone different size of ITAG6 regulatory

Primer	Sequence
ITGA61.3Kb F	TAGCTAGCACAATTGTGCAGGAAGAAACTACCCAAA
ITGA61.3Kb B	TAGCTAGCACAATTGCCCACGTAAATGGTGAAAGG
ITGA6ShortB	TAGCTAGCACAATTGAAAAGCCCATCATGCTGATAA
ITGA64Kb F	TAGCTAGCACAATTGAAGTTGGAGGCAGGACCAG
ITGA64Kb B	TAGCTAGCACAATTGGGGATCAACGGAAGAGAACA
ITGA6minigeneScreener	GCCACTGCATTGTTTCATTG

4.2.1.2PCR amplification

PCR was performed with a PCR Phusion Kit from Thermo Scientific, following the manufacturer's instructions for 50µl reactions (Table 4.2). Amplification was carried out in SensoQuest thermo cyclers by the program shown in Table 4.3.

Table 4.2: Components of cloning PCR

Reagent	Volume µl
5xCG buffer	10.0
dNPs	1.0
10µM Forward Primer	2.5
10µM Reverse Primer	2.5
DMSO	1.0
Template(Human genomic DNA)	1.0
Phusion DNA Polymerase	0.5
H2O	31.5

Table 4.3: Steps of PCR reaction for cloning

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

4.3.1.3 Gel electrophoresis

PCR products were run on 1.1% agarose gels. The correct product size was cut out and purified from the gel using QIAquick Gel Extraction Kit according to manufacturer’s instruction (Qiagen).

4.3.1.4 Restriction digests

Restriction digests were performed using *EcoRI* (New England Biolabs) restriction enzyme in order to digest the PXJ41 vector. Samples were digested using *Mfe1* restriction enzyme (New England Biolabs) following the manufacturer’s instructions (Table 4.4 and 4.5). Digestions were incubated for 3 hours at 37 °C. After that, digested vector and constructed were ligated following the manufacturer’s instructions table and incubated at room temperature overnight (Table 4.6).

Table 4.4: Constructs Restriction Digest

DNA	26 µl
NEBuffer	5µl
BSA (X10)	5µl
MfeI restriction enzyme	2µl
Total	50µl

Table 4.5: Vector Restriction Digest

DNA	26 µl
EcoRI buffer	5µl
BSA (X10)	5µl
EcoRI restriction enzyme	2µl
Total	50µl

Table 4.6: Component for ligation

Insert (digested and clean PCR)	37.9µl
Vector (PXJ41)	2µl
T4 DNA Ligase	2µl
10XT4 Buffer	5µl
H2O	3.1µl
TOTAL	50µl

4.3.1.5 Molecular cloning

Competent DH5α *E.coli* and ligation mixes (Inserting DNA and cut Vector pXJ41) were mixed, and transformation was achieved using a heat shock protocol. This protocol was done by sequential different incubations on ice, 42°C and ice for 10, 1 and 10 min respectively. After that, 1 ml of Luria-Bertani (LB) was added and cells were incubated at 37°C for 1 hour. Next bacteria were spread on to a plate of LB agar with 50µg/ml ampicillin and incubated at 37°C overnight. Next day, single colonies were picked

individually and screened using primers binding to the vector (Table 4.7) and one cloning primer from insert in order to confirm that bacteria contained a correct plasmid (Table 4.8 for PCR conditions). Plasmid were picked into 5 ml LB broth plus 50µg/ml ampicillin and incubated at 37°C overnight in a shaker. The next day, tubes were centrifuged at 5,000rpm for 2 minutes in order to get a pellet of cells. Finally a QIAprep Mini-prep Kit was used following manufacturer's instructions in order to extract supercoiled plasmid DNA from cells. Plasmid DNA was sequenced using vector primers (pXJ41F and pXJ41F, Table 4.7) by Source Bioscience, Oxford.

Table 4.7: pXJ41 vector primers for colony screening and for sequencing

Primer	Sequence
PXJ41F	GCTCCGGATCGATCCTGAGAACT
PXJ41B	GCTGCAATAAACAAGTTCTGCT

Table 4.8: Steps for cloning screening PCR

Step	Temperature(°C)	Time	Cycle Number
Heat activation	95	2 mint	1
Denaturation	95	1 mint	35
Annealing	58	1 mint	35
Extension	72	1 mint	35
End Stage Extension	72	5 min	1
Cooling	15	∞	1

4.3.1.6 Site- directed mutagenesis of the ITGA6 minigenes

Site-directed mutagenesis was conducted in order to mutate the PTBP downstream binding site using the wildtype *ITGA6* minigene. As a template, primers designed in order to delete the PTBP binding site are shown (Table 4.12 and 4.13). Site-directed mutagenesis was conducted in two PCR reactions (PCR1 and PCR2) in order to create two fragments which overlap desired complementary mutations on opposite strands. A third reaction PCR3 was then performed using the fragments from PCR1 and PCR2 as template in order to create a full-length insert with the mutation. Finally, the fragment containing the mutant on both strands was cloned into empty pXJ41 vector. A figure 4.1 showing principle of site-directed mutagenesis

A 50µl reaction was performed for each PCR (PCR1, PCR2 and PCR3) using Phusion Kit from Thermo Scientific. The PCR master mix and conditions used are shown in tables 4.9 , 4.10 and 4.11. Amplifications were carried out in SensoQuest thermo cyclers by the program shown in (Table 4.3). A QIAquick PCR Purification Kit was used to purify all PCR products following the standard manufacturer’s protocol.

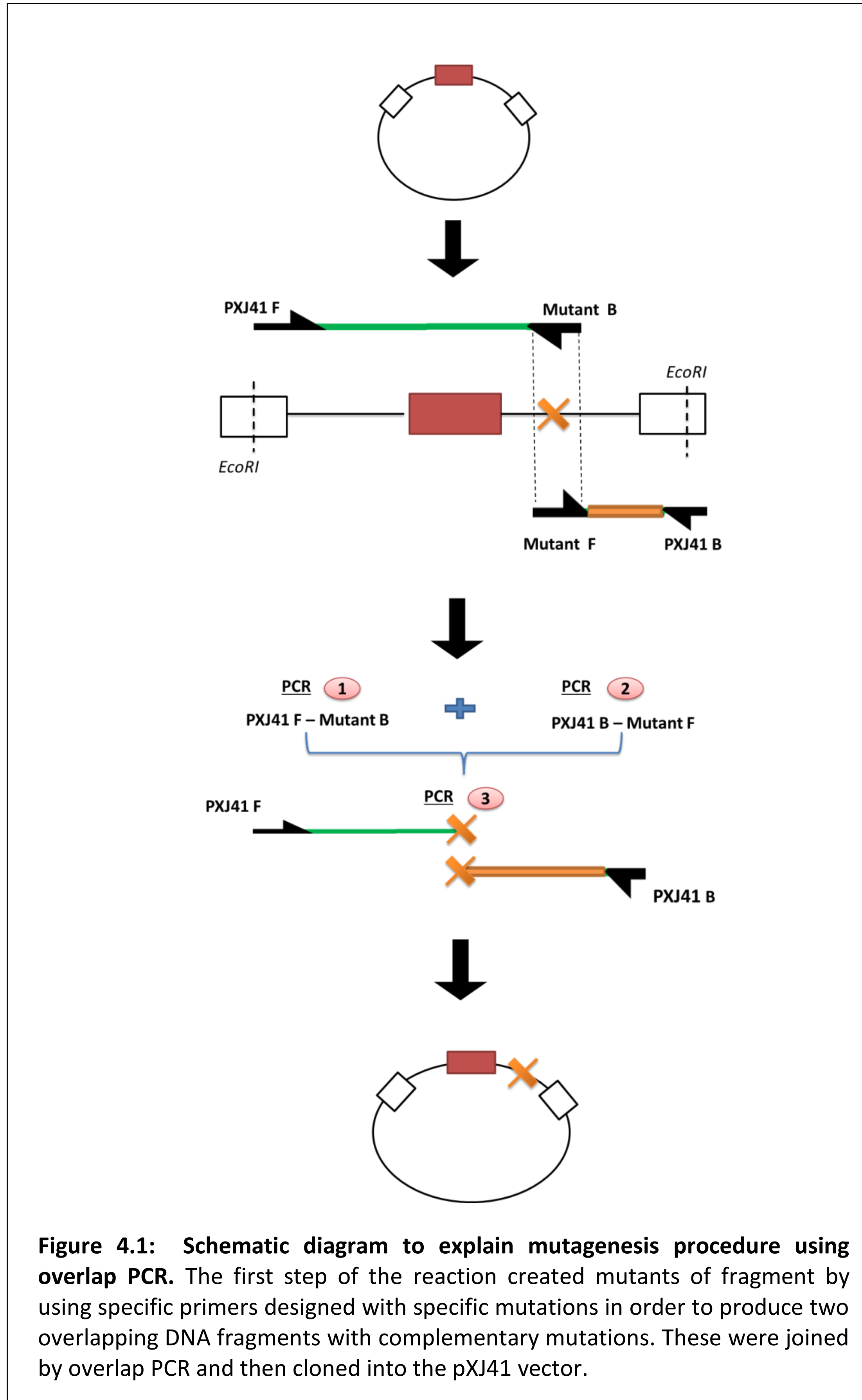


Figure 4.1: Schematic diagram to explain mutagenesis procedure using overlap PCR. The first step of the reaction created mutants of fragment by using specific primers designed with specific mutations in order to produce two overlapping DNA fragments with complementary mutations. These were joined by overlap PCR and then cloned into the pXJ41 vector.

Table 4.9: PCR1

Reagent	Volume µl
5xCG buffer	10.0
dNPs	1.0
Pxj41F	2.5
Mutant Reverse	2.5
DMSO	1.0
Template (wildtype minigene, 20ng/µl)	1.0
Phusion DNA Polymerase	0.5
H2O	31.5

Table 4.10: PCR2

Reagent	Volume µl
5xCG buffer	10.0
dNPs	1.0
pXJ 41B	2.5
Mutant Forward primer	2.5
DMSO	1.0
Template (wildtype minigene, 20ng/µl)	1.0
Phusion DNA Polymerase	0.5
H2O	31.5

Table 4.11:PCR3

Reagent	Volume µl
5xCG buffer	10.0
dNPs	1.0
pXJ 41F	2.5
pXJ 41B	2.5
DMSO	1.0
Template 1 – (PCR 1)	1.0
Template 2 – (PCR 2)	1.0
Phusion DNA Polymerase	0.5
H2O	30.5

Table 4.12: Primers was used for overlapping PCR (mutagenesis) Del PTBPA

Primer	Sequence
PXJ41 F	GCTCCGGATCGATCCTGAGAACT
DelPTBPA B	CAA GCC TTCCCC ATC ACTTTA TAG CA CCT CAC CAA GAG AAACAC AAC ATTTCC
DelPTBPA F	TGC TAT AAA GTG ATG GGG AAG GCT TG
PXJ41 B	GCTGCAATAAACAAGTTCTGCT

Table 4.13: primers was used for overlapping PCR (mutagenesis) Del PTBPB

Primer	Sequence
PXJ41 F	GCTCCGGATCGATCCTGAGAACT
DelPTBPB B	TCC CTG TTC AGG GTATTT TTT GTA CTA GTAA CTTTGG TAA GGG ATC ATC TTC TGA AGT
DelPTBPB F	CTAGTACAA AAA ATA CCCTGA ACAGGG A
PXJ41 B	GCTGCAATAAACAAGTTCTGCT

4.3.2 siRNA transfection

SiRNA transfections in this chapter were used to knockdown PTBP1/2 and MBNL1 proteins in MCF7 and Hela cells. The protocol of siRNA transfection was described previously in chapter 2, methods section.

4.3.3 Western blot

In this chapter Western immunoblotting was used to confirm PTBP expression after transfection with the flag epitope. These used mouse monoclonal antibodies against flag (F3165) (1:1000 dilution) as followed by Secondary anti-mouse antibody (antibodies at 1:1000 dilution). This was performed following the protocol which was described in chapter 2, method section.

4.3.4 Transfection

A variety of plasmids including mutant plasmids and wild type were transfected into the MCF7 breast cancer cell line. Transfection was performed using Lipofectamine® 2000 (Thermofisher). The transfection was performed following the manufacturer's instruction. Table 4.14

Table 4.14: Transfection protocol for ITGA6 minigene: first we mixed (Mix1) and (Mix2) together and incubated for 10-20 min. after that, we spin it for short time and then transfected to cell.

Transfection protocol Reaction mix	Components	volume	Incubation time
Mix 1	Opti-MEM® Medium	80µL	5minutes
	Lipofectamine® 2000reagent	4 µL	
Mix2	Opti-MEM® Medium	80 µL	5minutes
	Plasmid 1 µg	3 µL	
Mix 3	Mix 1	85µL	20 minutes
	Mix 2	83 µL	
Short Centrifugation Add 250 µL to each well in 6 well plate and incubate at 37°C overnight			

After 24 hours incubation, we checked for transfection efficiency for cell that accepted GFP-linked constructs using a fluorescent microscope. Next, cells were washed and aliquots pelleted in order to do RNA extractions.

4.4 Results

4.3.1 ITGA6 alternative exon minigenes respond to splicing control similarly to the endogenous ITGA6 alternative exon

Minigene constructs are tools used for the identification of the regulatory factors that control splicing and regulate alternative splicing. In the previous chapter, the endogenous *ITGA6* alternative exon (exon 25) was shown to be regulated by factors like MBNL, RFOX, PTB1 and PTB2. I began the work in this chapter with the aim of establishing a minigene construct for the *ITGA6* alternative exon that could be used to investigate regulatory factors.

Based on preliminary data from (Venables et al., 2013b) the splicing inclusion of the *ITGA6* exon was enhanced by MBNL1 and RBFOX2; moreover, as (Goel et al., 2014) reported, the GU-rich binding site for ESRP protein enhances splicing of the *ITGA6* alternative exon. Given these results, I cloned the alternative exon with the surrounding 300 bp upstream and 700 bp flanking downstream intron from genomic DNA. The exon and conserved flanking intron sequence was inserted between two exons in the pXJ41 plasmid vector to create the minigene construct labelled ITGA61.3Kb. This genomic region contains several candidates splicing factor binding sites (MBNL1, RBFOX2, ESRPs) that were analysed by Dr Venables (Newcastle University; Figure 4.2).

Subsequently, I transfected the *ITGA6* minigene into several cancer cell lines, including MDA-MB231, MCF7, and HeLa with plasmid expressing splicing factors in or empty vectors as control. RT-PCR was performed using the appropriate primers, which were designed to distinguish the minigene transcript from the endogenous one. Each transfection was performed in triplicate to observe any changes, and the mRNA splicing patterns were analysed using gel electrophoresis. This produced 2 products. The upper band included the *ITGA6* exon, whereas the lower band represented the PCR product without *ITGA6* exon 25 (Figure 4.3). This experiment showed that this exon is indeed recognised and spliced within the context of the minigene in different cell lines with good efficiency compared to the endogenous gene.

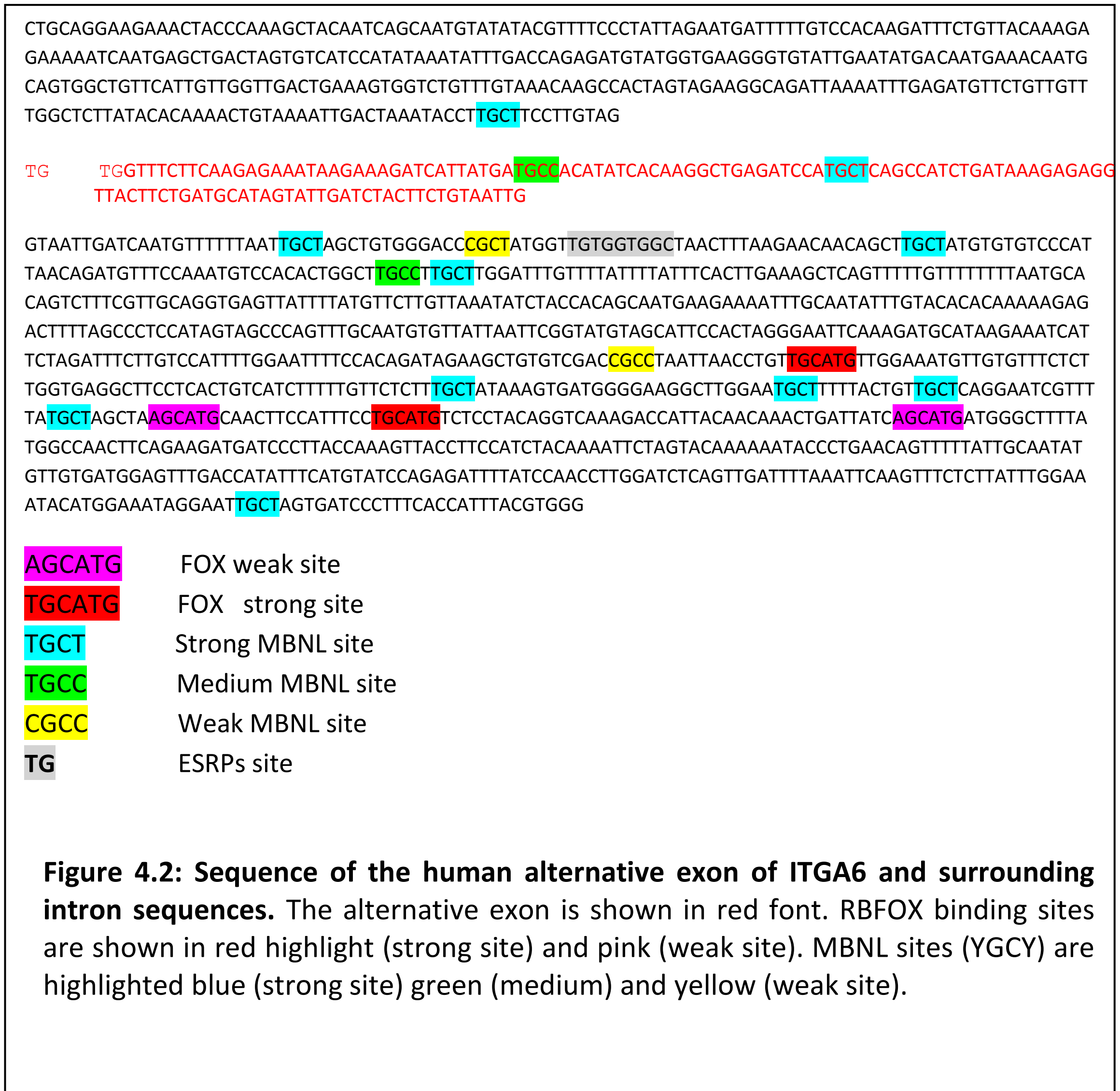


Figure 4.2: Sequence of the human alternative exon of ITGA6 and surrounding intron sequences. The alternative exon is shown in red font. RBFOX binding sites are shown in red highlight (strong site) and pink (weak site). MBNL sites (YGCY) are highlighted blue (strong site) green (medium) and yellow (weak site).

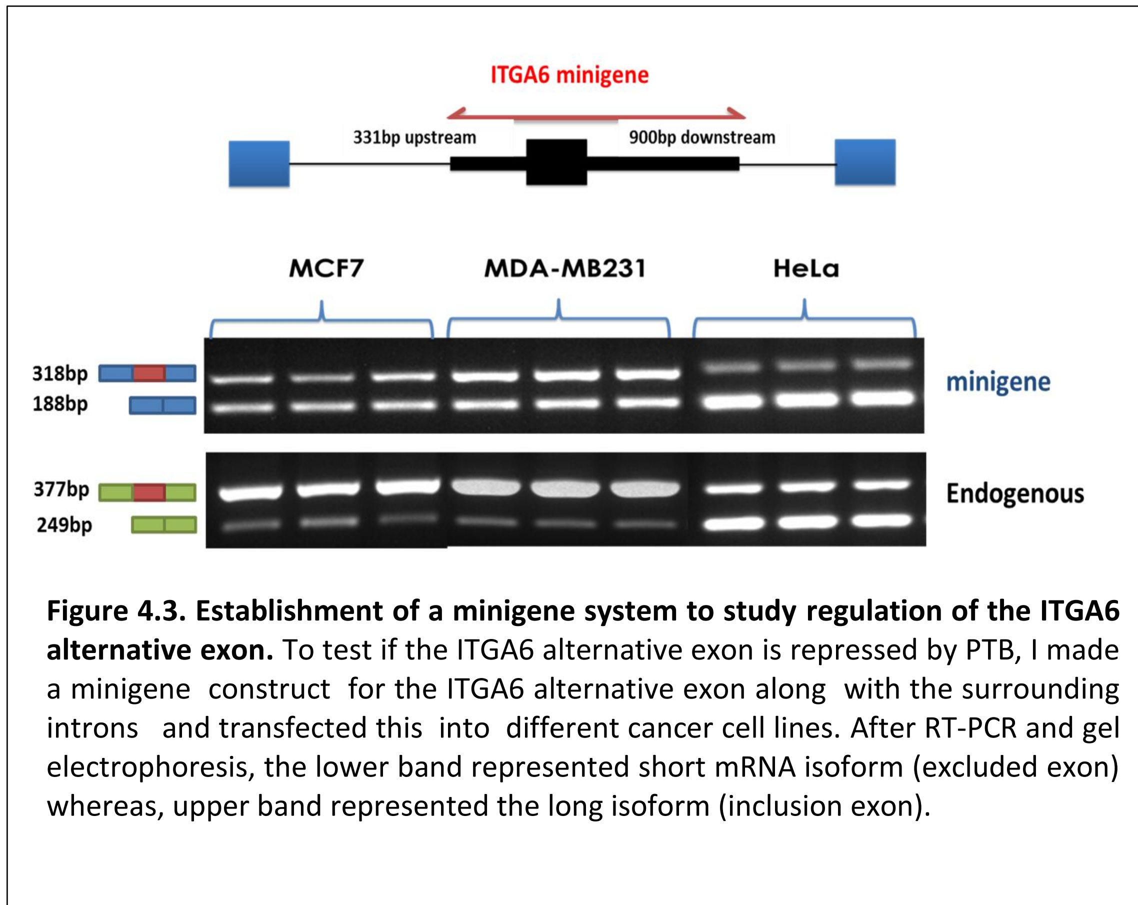


Figure 4.3. Establishment of a minigene system to study regulation of the ITGA6 alternative exon. To test if the ITGA6 alternative exon is repressed by PTB, I made a minigene construct for the ITGA6 alternative exon along with the surrounding introns and transfected this into different cancer cell lines. After RT-PCR and gel electrophoresis, the lower band represented short mRNA isoform (excluded exon) whereas, upper band represented the long isoform (inclusion exon).

4.4.1.1 A ITGA61.3Kb is an appropriate minigene to study the regulatory factors that control ITGA6 alternative exon 25

The pervious result indicate that the *ITGA6* minigene mimics the endogenous *ITGA6* alternative exon. Given these results, I sought to further identify a good minigene that would define the response regions that are bound by the splicing factors. Subsequently, with help from Dr Venables, a strategy was established to clone different sizes of the *ITGA6* alternative exon with variable lengths of surrounding introns to obtain the ITGA6 minigene that would best facilitate investigation of the splicing mechanism.

Based on bioinformatics data from Dr Venables, I decided to clone the *ITGA6* alternative exon with different sizes of surrounding introns, whether upstream or downstream of the alternative exon. I cloned a larger minigene containing a greater amount of the intronic sequence to determine whether this contained further regulatory sequences; in addition, I cloned short minigenes in the hope that they would be as useful to allow a more precise definition of the control regions. For the longer clone, I amplified the exon with 1500 bp upstream and 2300 bp downstream intron from the genomic DNA, and cloned this into a pXJ41 exon trap plasmid, to create the minigene called ITGA64Kb (Figure 4.4). I also cloned ITGA6short, an ITGA6 alternative exon with deletions removing the expected binding sites for RBFOX and MBNL, with 300 bp upstream and 225 bp downstream flanking intron (Figure 4.4).

Following this, the two minigenes ITGA64Kb and ITGA6short were transfected in cancer cell lines, including Hela, MDA-MB231 and MCF7. RT-PCR was performed using appropriate primers designed on the β globin exons of the pXJ41 plasmid vector to discriminate the minigene transcript from the endogenous context. Each transfection was performed in triplicate to observe any change in splicing pattern, and the mRNA splicing patterns were analysed using gel electrophoresis. The upper band included ITGA6 exon), whereas the lower band represented the PCR product without ITGA6

exon (Figure 4.5). The ITGA64kb showed variable exon inclusion between the cell lines, whereas ITGA6short did not show exon inclusion in any cell lines. Overall the original

ITGA61.3kb minigene showed the best exon inclusion when compared with the two other *ITGA6* minigenes, and showed similar patterns of splicing compared to the endogenous *ITGA6.*

A

B

ITGA61.3Kb

331bp upstream

900bp downstream

ITGA6short

ITGA64Kb

Figure 4.4: Cloning a series of constructs of different lengths containing the ITGA6 alternative exon and surrounding intron sequences. a) A sequence of ITGA6 alternative exons (red font) and surrounding intron sequences (black font). The yellow line represents ITGA6short region cloning. The ITGA61.3Kb cloning region is shown in blue, while the ITGA64kb cloning region is shown in red**. b)** A diagram showing the different cloning regions of the ITGA6 alternative exon.

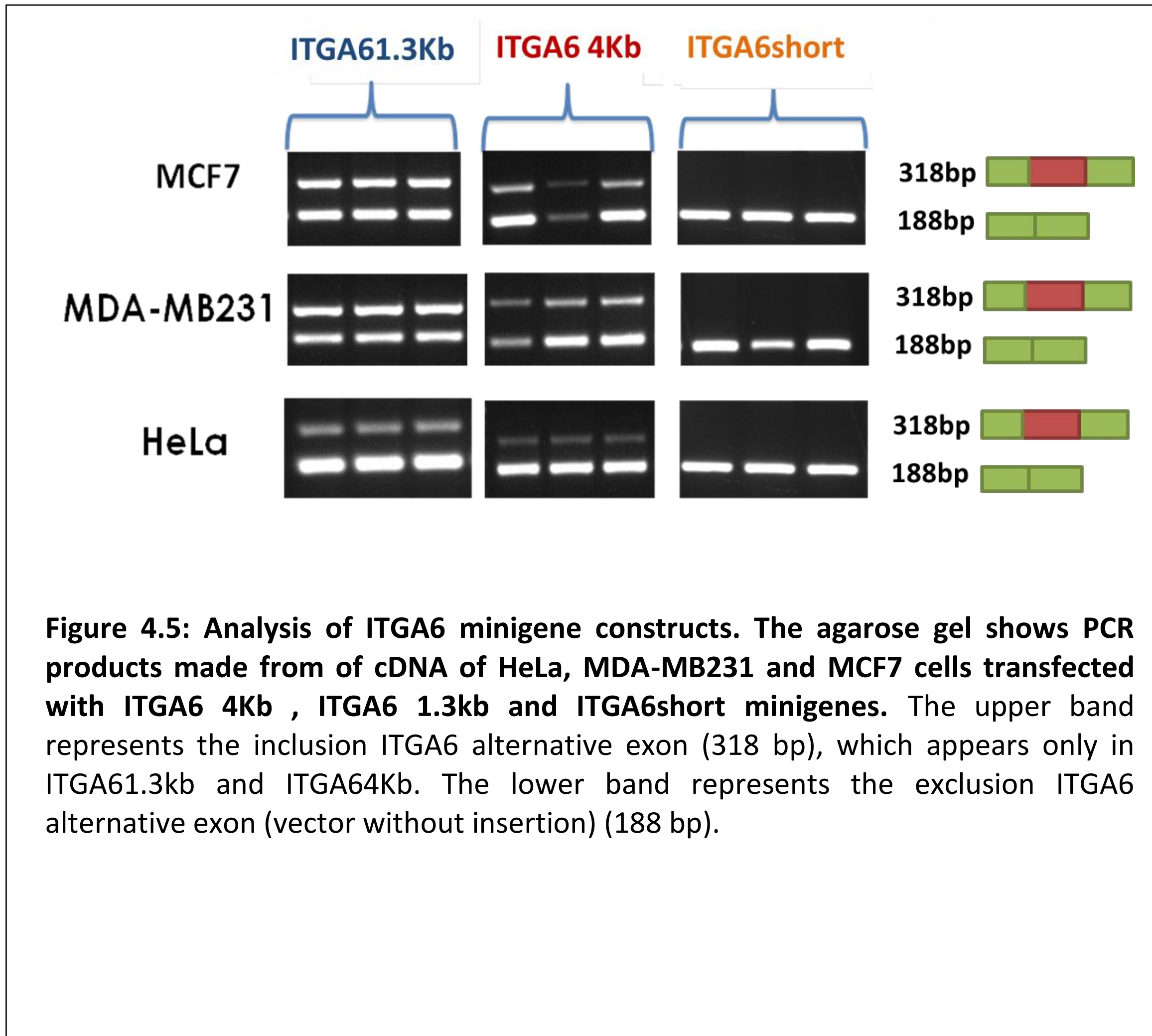


Figure 4.5: Analysis of ITGA6 minigene constructs. The agarose gel shows PCR products made from of cDNA of HeLa, MDA-MB231 and MCF7 cells transfected with ITGA6 4Kb , ITGA6 1.3kb and ITGA6short minigenes. The upper band represents the inclusion ITGA6 alternative exon (318 bp), which appears only in ITGA61.3kb and ITGA64Kb. The lower band represents the exclusion ITGA6 alternative exon (vector without insertion) (188 bp).

4.4.2 The *ITGA6* alternative exon encoded with a minigene response to changes in regulation factors similarly to the endogenous *ITGA6* alternative exon .

The ITGA61.3kb minigene construct showed strong inclusion for the alternative *ITGA6* exon in the MCF7 and MDA-MB-231 cancer cell lines compared to HeLa cells. To determine the factors that regulate the ITGA6 alternative exon in endogenous, I used the ITGA61.3kb minigene construct. This also had the advantage of being smaller than the ITGA64kb minigene insert, so there was less sequence to consider binding sites within.

Based on our findings in chapter 3, where it was shown that the *ITGA6* alternative exon is activated by RBFOX2, MBNL1, ESRPs and repressed by PTBPs, I sought to determine whether these potential splicing factors regulate the ITGA61.3Kb minigene in the cell lines. The MCF7 cell line, in which *ITGA6* can be expressed, was shown to be a good model for studying endogenous alternative splicing of *ITGA6* since it expresses both of the two mRNA isoforms. The MCF7 cell lines also showed successful expression for the *ITGA6* minigene. Subsequently, the *ITGA6* minigene was transfected into MCF7 cells the same time as these same cells were co-transfected with PTBP, MBNL1, RBFOX2 and ESRP siRNAs to knock down these endogenous proteins. RT-PCR was performed using the appropriate primers, which were designed to distinguish the minigene transcript from the endogenous. In the upper band, the minigene-derived transcripts contain the alternative exon, whereas the lower band represented the PCR product without exon insertion. Pattern of splicing from the minigene ITGA61.3Kb alternative exon shifted clearly from inclusion to more exon exclusion with the knockdown of MBNL1, while the knockdowns of ESRPs and RBFOX2 did not show an effect. Similar results were obtained for the endogenous mRNA. In addition, siRNA of PTBPs did not show a clear effect for either the minigene or endogenous gene. Overall, the knockdown of MBNL1 alone showed a clear shift in alternative exons compared to the non-siRNA control (Figure 4.6).

Based on novel findings presented in the previous chapter, it was predicted that PTBPs act as repressors for the endogenous *ITGA6* alternative exon. The PTBP knockdown removed all inhibition of *ITGA6* exon inclusion in MCF7; however, the splicing shift was small due to the high baseline level of exon inclusion in the cell line. To observe a greater shift, I turned to the HeLa cell line, which exhibits a low level of exon inclusion. Since PTBs were shown to clearly inhibit the endogenous *ITGA6* alternative exon in the HeLa cell line, I sought to determine whether I could also observe the same inhibition using the *ITGA6* minigene. Thus, I co-transfected the *ITGA6* minigene into HeLa cells with double knockdown of PTB1 and PTB2. RT-PCR was performed using primers designed across the alternative exons for endogenous *ITGA6,* and to detect mRNAs from the minigene respectively. As expected, *ITGA6* splicing shifted towards the long form with double knockdown of PTB1 and PTB2. Moreover, as anticipated, PTB1 and PTB2 shifted splicing to the short form in the *ITGA6* minigene (Figure 4.7).

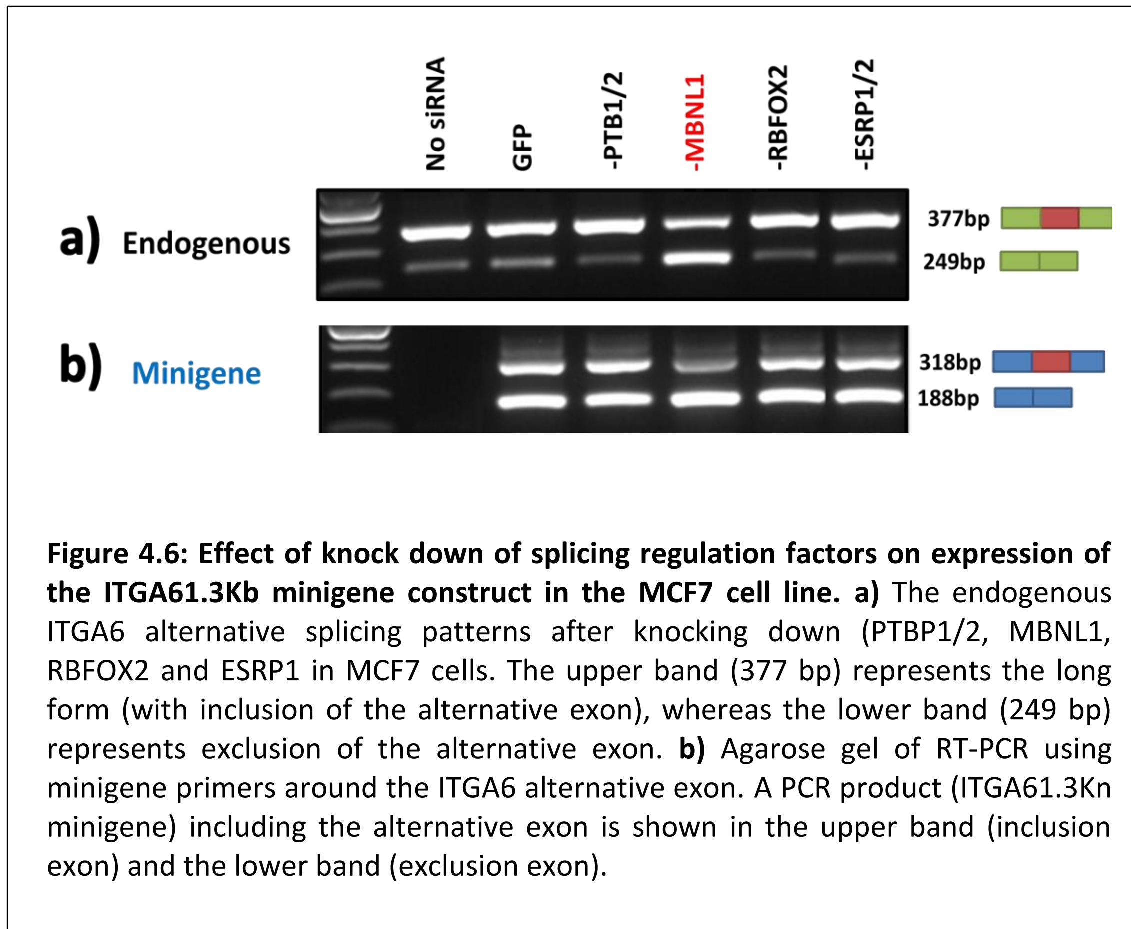


Figure 4.6: Effect of knock down of splicing regulation factors on expression of the ITGA61.3Kb minigene construct in the MCF7 cell line. a) The endogenous ITGA6 alternative splicing patterns after knocking down (PTBP1/2, MBNL1, RBFOX2 and ESRP1 in MCF7 cells. The upper band (377 bp) represents the long form (with inclusion of the alternative exon), whereas the lower band (249 bp) represents exclusion of the alternative exon. **b)** Agarose gel of RT-PCR using minigene primers around the ITGA6 alternative exon. A PCR product (ITGA61.3Kn minigene) including the alternative exon is shown in the upper band (inclusion exon) and the lower band (exclusion exon).

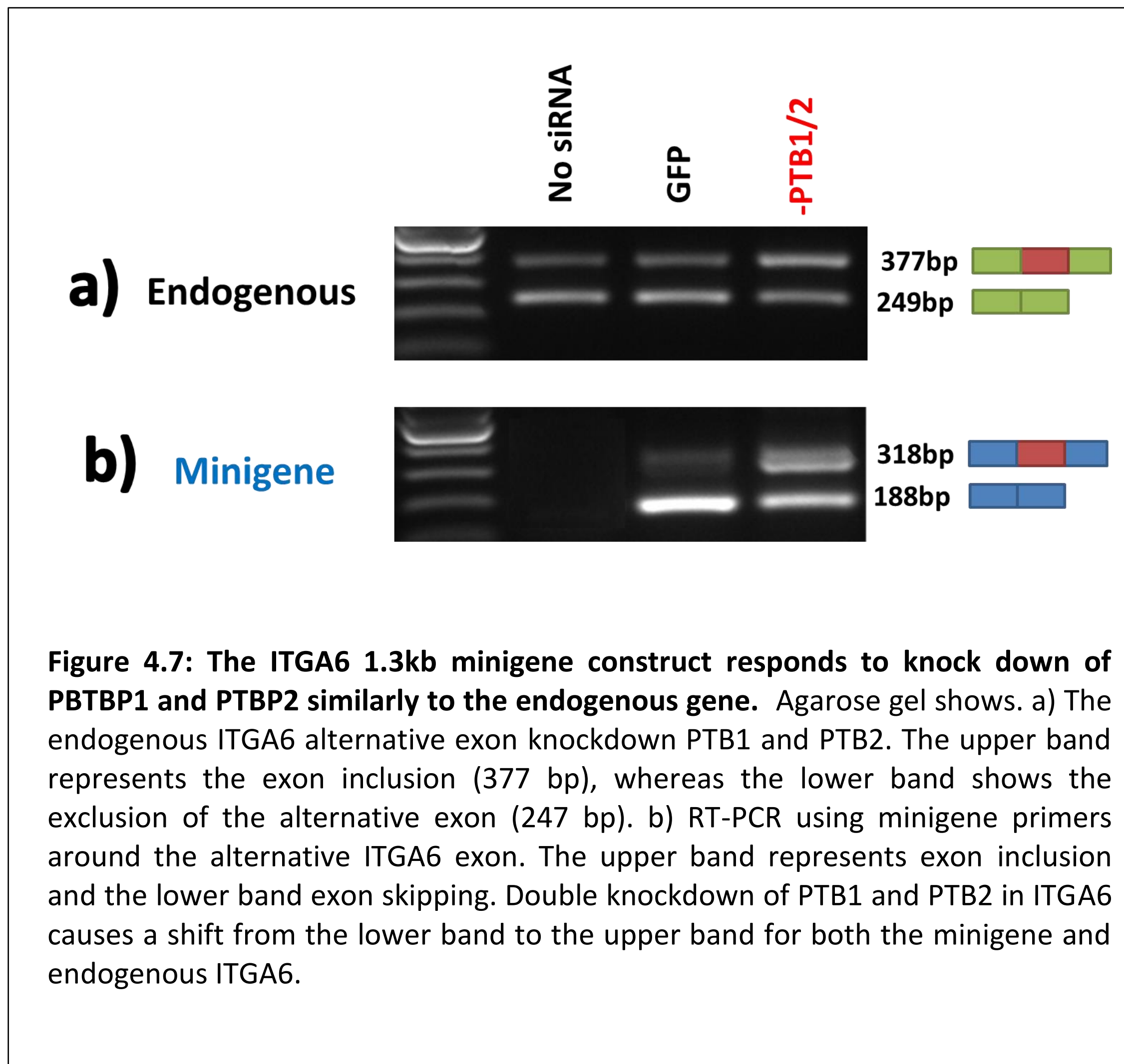


Figure 4.7: The ITGA6 1.3kb minigene construct responds to knock down of PBTBP1 and PTBP2 similarly to the endogenous gene. Agarose gel shows. a) The endogenous ITGA6 alternative exon knockdown PTB1 and PTB2. The upper band represents the exon inclusion (377 bp), whereas the lower band shows the exclusion of the alternative exon (247 bp). b) RT-PCR using minigene primers around the alternative ITGA6 exon. The upper band represents exon inclusion and the lower band exon skipping. Double knockdown of PTB1 and PTB2 in ITGA6 causes a shift from the lower band to the upper band for both the minigene and endogenous ITGA6.

4.4.2.1 Confirmation of PTB as a novel regulator of the ITGA6 alternative exon

The *ITGA6* 1.3kb minigene construct showed a response to two regulatory factors, namely MBNL1 and PTBP1, in the MCF7 and HeLa cell lines, respectively. PTBP acted as a repressor of the *ITGA6* alternative exon, whereas MBNL1 acted as an activator for both the minigene and endogenous mRNA. Therefore, these proteins contributed to the inclusion or exclusion of *ITGA6* exon 25.

The prediction that the knockdowns for these proteins cause inclusion or skipping for exon 25 of the *ITGA6* minigene was previously considered in this work (4.4.2). In this section, complementary experiments on the overexpression of these regulatory factors, including PTBP and MBNL1, was carried in the MCF7 cell line out to confirm the functional role of these proteins. I performed co-transfection of the ITGA6 minigene (ITGA61.3Kb) and expression vectors encoding PTBPs (PTBP1, PTBP2) and MBNL1 in triplicate in a breast cancer cell line. RT-PCR was performed by using specific primers across flanking exons for the minigene. The upper band includes the *ITGA6* exon, whereas the lower band represented the PCR product without *ITGA6* exon. Multicapillary QIAxcel electrophoresis was used to calculate the concentration ratio between the two forms of *ITGA6* splicing. The percentages of alternative splicing patterns were calculated and plotted on a graph.

The overexpression of PTBP1 co-transfected with the ITGA6 minigene (ITGA61.3Kb) showed a significant decrease in per cent splicing in (PSI) of the *ITGA6* minigene exon 25 from 14 to 4%, compared with the ITGA61.3Kb minigene alone (Figure 4.8). The overexpression of PTBP2 co-transfected with the ITGA6 minigene did not show a change. As expected, overexpression of MBNL1 with the ITGA61.3Kb minigene showed a significant increase in PSI for the ITGA6 minigene alternative exon, changing from 14 to 30% inclusion, compared with the minigene alone. Overall, these results confirm that the *ITGA6* minigene is activated by MBNL1 and inhibited by PTBP1.

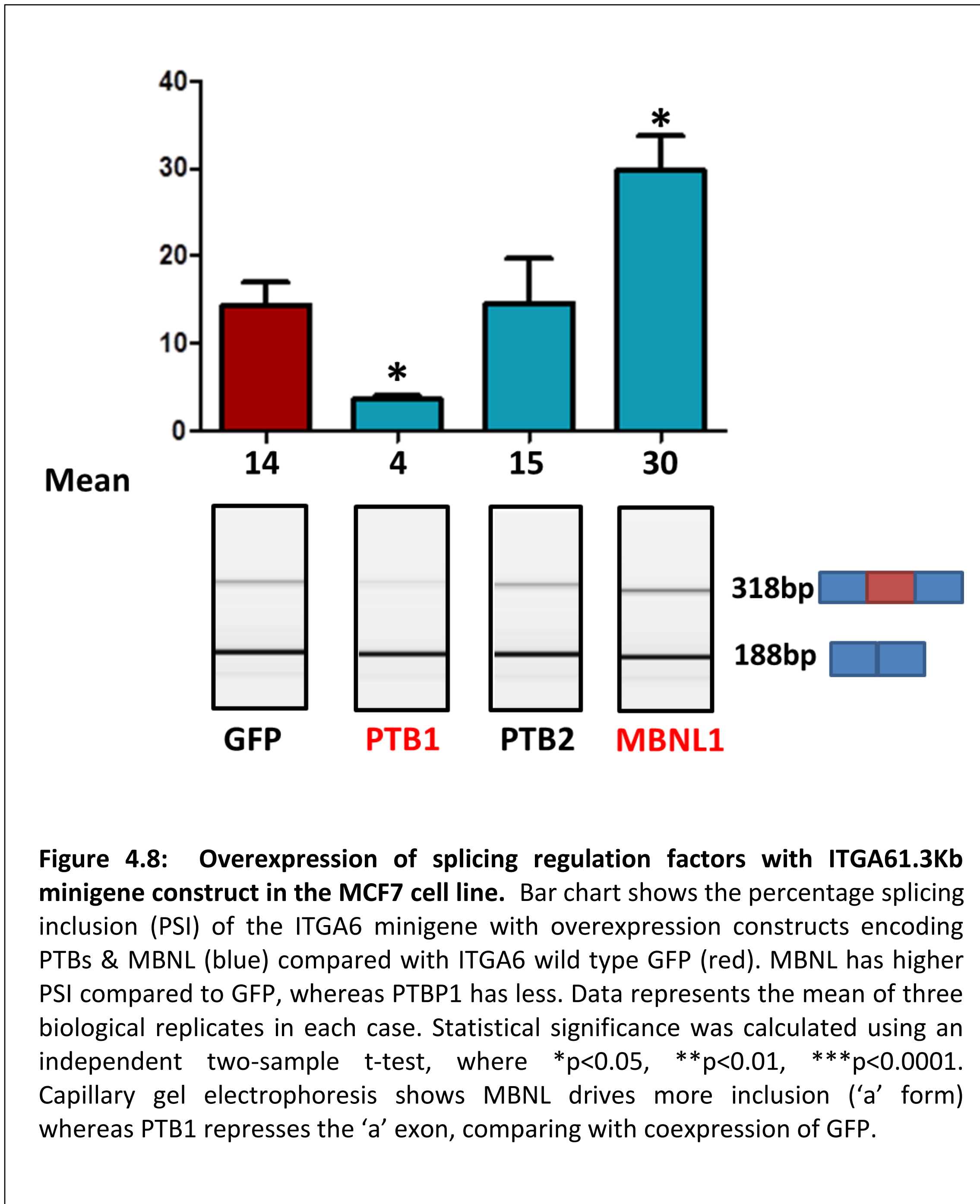


Figure 4.8: Overexpression of splicing regulation factors with ITGA61.3Kb minigene construct in the MCF7 cell line. Bar chart shows the percentage splicing inclusion (PSI) of the ITGA6 minigene with overexpression constructs encoding PTBs & MBNL (blue) compared with ITGA6 wild type GFP (red). MBNL has higher PSI compared to GFP, whereas PTBP1 has less. Data represents the mean of three biological replicates in each case. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. Capillary gel electrophoresis shows MBNL drives more inclusion ('a' form) whereas PTB1 represses the 'a' exon, comparing with coexpression of GFP.

4.4.3 Effects of PTBP overexpression/KD in different cell lines on splicing pattern from the minigene

In previous experiments, I investigated the splicing regulators for the *ITGA6* alternative exon in cancer cell lines. MBNL1, RBFOX2, ESRPs and PTBPs regulated the *ITGA6* alternative exon. Following on from that work, MBNL1 and PTBPs were shown to regulate the splicing of both the minigene-encoded and endogenous *ITGA6*. Since PTBPs represent novel, not previously reported regulators for the *ITGA6* alternative exon, I sought to investigate alternative splicing regulation in more detail and identify the binding sites of the PTBP regulatory factor.

To test the hypothesis that an RNA binding site for PTBP proteins caused skipping of the exon, it was important to find a good cell line for more investigation. Experiments on silencing and overexpression of PTBP proteins were conducted on cancer cell lines to identify the binding sites for PTBPs. In the MCF7, MDA-MB237 and HeLa cell lines, I performed co-transfection of the ITGA6 minigene (ITGA61.3Kb) and expression vectors encoding PTBPs (PTBP1, PTBP2); at the same time, the ITGA6 minigene was co-transfected with PTBP siRNAs to knockdown protein. Following this, RT-PCR was performed using specific primers across the flanking exons. Different patterns between two RT-PCR bands were observed using multicapillary QIAxcel gel. The percentages of the alternative splicing patterns were calculated and plotted on a graph (The method of calculation is described in the method results 3.3.3).

By observing the knockdown and overexpression of PTB proteins, it was found that the *ITGA6* minigene showed a significant increase in included exon after knockdown of PTBPs and a significant decrease in the alternative exon with overexpressed PTBPs compared with the ITGA6 minigene alone in the MCF7 cell line (Figure 4.9a). In the HeLa cell line, inclusion the alternative exon of the *ITGA6* minigene was significantly decreased with overexpression of PTPBs, compared with the *ITGA6* minigene alone. However, knockdown of PTBPs with the ITGA6 minigene did not show significant

difference for the alternative exon compared with the ITGA6 minigene transfected alone (Figure 4.9b). As expected, in MDA-MB-231 cells the alternative exon of the *ITGA6* minigene was significantly decreased with the overexpression of PTBPs, whereas it showed no change with knockdown PTBPs compare with the ITGA6 minigene alone(Figure 4.9c). Hence, overall, MCF7 was the only cell line that showed a significant alternative exon of ITGA6 with overexpression and silencing of the PTBP regulatory factor

.

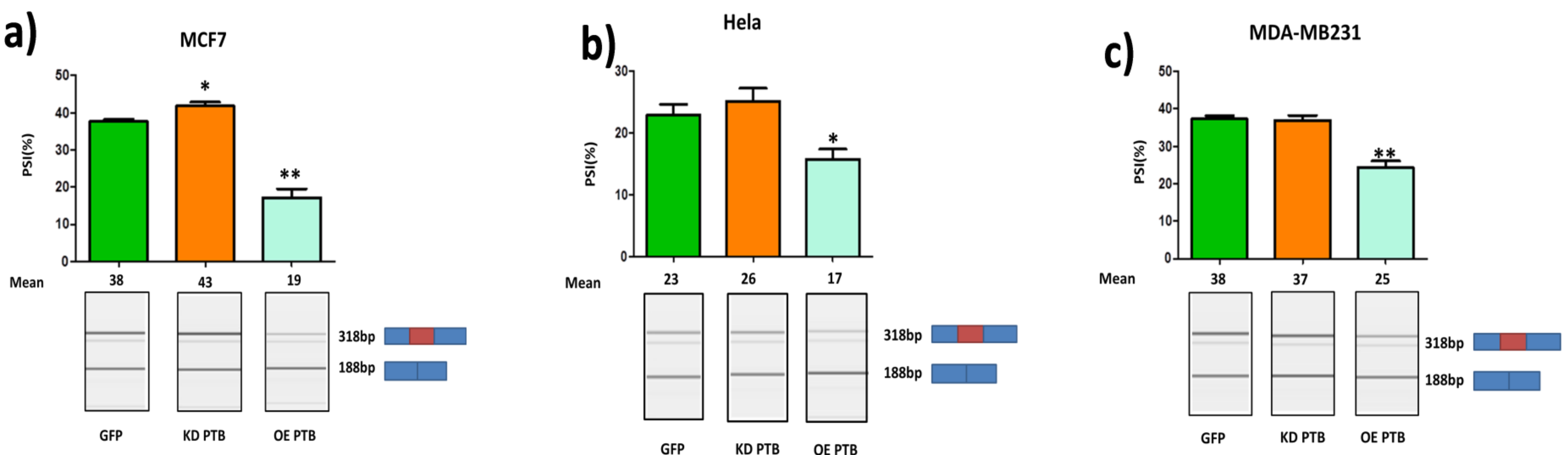


Figure 4.9: Effects of PTBP over-expression and knockdown in different cell lines on minigene. Bar chart shows the percentage splicing inclusion (PSI) from the ITGA6 minigene after knockdown (orange) and overexpression(bright blue) of PTBP regulator factors comparing with wild type GFP (green) in A) MDA-MB-231, B) MCF7 & C) HeLa cell lines. Data represent the mean of three biological replicates in each case. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001.

4.4.4 The downstream binding site for PTB represses splicing of the 'a' exon through RNA protein interactions

Since the alternative exon of the *ITGA6* minigene successfully exhibited a significant response to knockdown and overexpression of PTBPs in the MCF7 breast cancer cell line, I decided to investigate whether PTBP proteins could bind directly to ITGA6. To investigate the binding site, it was important first to identify candidate binding sites using bioinformatics. It has been reported in several studies that PTBPs binding upstream or within the alternative exon act as a splicing repressor (Amir-Ahmady et al., 2005). However, an analysis of PTBP binding sites at the *ITGA6* alternative exon using a CLIP experiment, which was performed by Dr C.W. Smith (Cambridge University), identified two candidate binding sites for PTBPs downstream of the *ITGA6* alternative exon (Figure 4.10 a).

To investigate whether these putative binding sites for PTB were required for splicing repression of the *ITGA6* alternative exon, PTBP candidate binding sites (CTTCCTCACTGTCATCTTTTTGTTCTCTT and TCCATCTACAAAATT) were mutated within the wildtype *ITGA6* minigene. To do this, I used overlapping PCR with specific primers to delete the first PTBP binding site (Del PTBPA). For deletion of the second PTBP binding site (Del PTBPB), I used gblock to synthesise the sequence after the PTBP binding site (TCCATCTACAAAATT) and directly overlap it with the upstream product. Overlapping PCR and gblock cloning are explained in the methodology section.

After generating minigenes with deletions for the PTBP binding sites, each minigene was transfected into the MCF7 breast cancer cell line. After 48 hours, the cells were harvested and the RNA was reverse transcribed. Following this, RT-PCR was performed using specific primers across the flanking exon. Splicing patterns were observed by gel electrophoreses, and the concentration of inclusion of the *ITGA6* minigene alternative exon was calculated using a multicapillary QIAxcel gel electrophoresis device. The concentration of inclusion was calculated and presented as a graph (Figure 4.10b).

The results of the mutant minigene transfection are shown in Figure 10b. Unexpectedly, mutants removing the downstream binding site for PTBPs, whether Del PTBPA or Del PTBPB, showed less inclusion for the alternative exon from minigene compared with wildtype ITGA61.3kb. In other words, there was a small reduction in splicing activation by PTBPs in mutants A and B. This means that although PTB overall expression represses the alternative exon of ITGA6, these actual binding sites activate alternative exon as well .

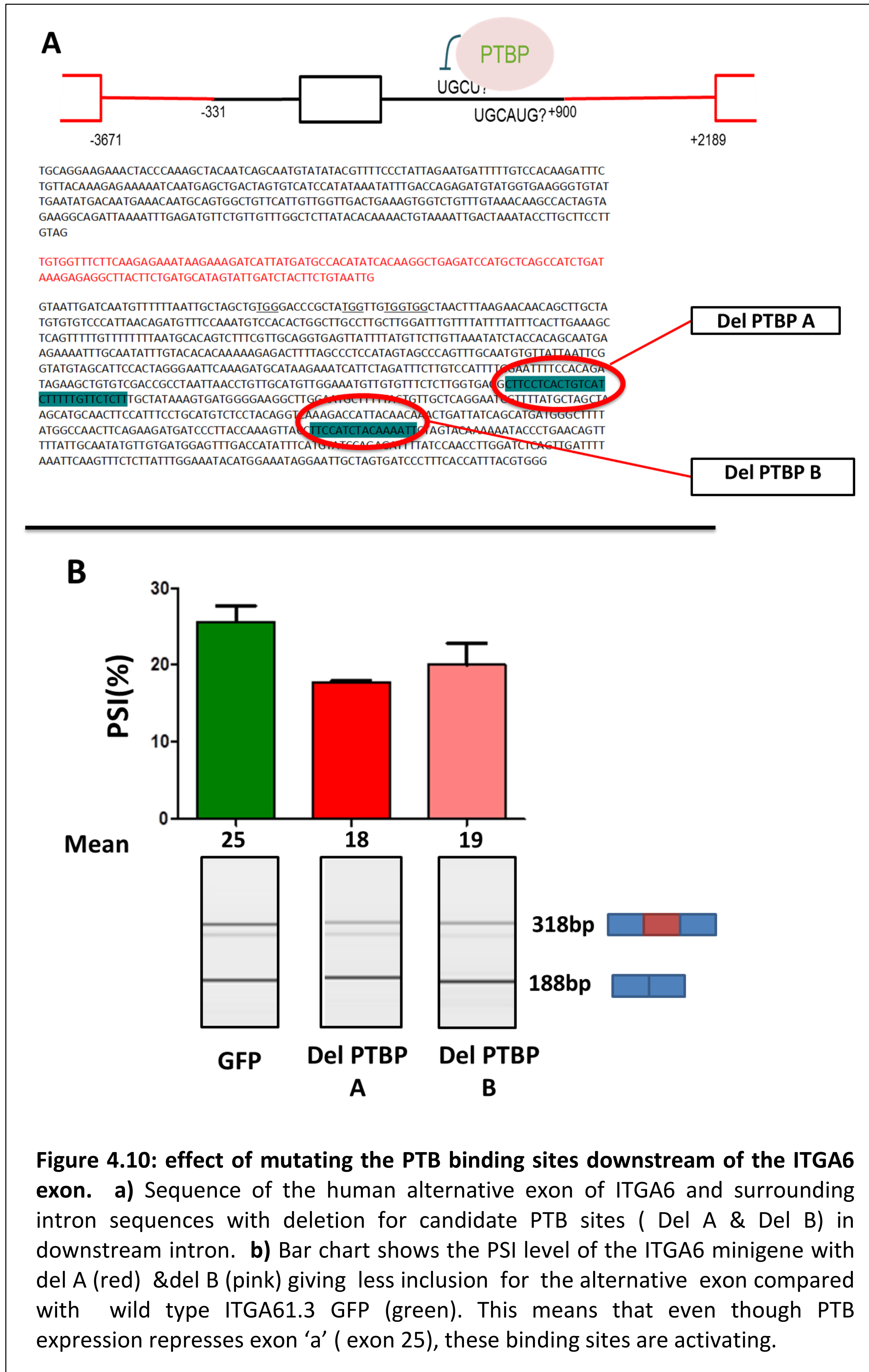


Figure 4.10: effect of mutating the PTB binding sites downstream of the ITGA6 exon. **a)** Sequence of the human alternative exon of ITGA6 and surrounding intron sequences with deletion for candidate PTB sites (Del A & Del B) in downstream intron. **b)** Bar chart shows the PSI level of the ITGA6 minigene with del A (red) &del B (pink) giving less inclusion for the alternative exon compared with wild type ITGA61.3 GFP (green). This means that even though PTB expression represses exon 'a' (exon 25), these binding sites are activating.

4.4.4.1 PTBP may function in ITGA6 splicing by sequestering other splicing regulators

PTBPs usually controls alternative splicing based on its binding site location. While PTBPs act as a repressor when binding upstream or within the exon, they function to activate splicing from downstream binding sites (Amir Ahmed et al., 2005)(Boutz et al., 2007, Llorian et al., 2010, Xue et al., 2009) . Since deletion of PTB binding sites downstream of the ITGA61.3kb alternative exon showed the opposite response (less inclusion for alternative exon), we predict that these sites normally activate the ITGA6 exon, but that protein interactions of PTB may act to repress the ITGA6 alternative exon by impacting other splicing regulators.

PTBPs have four RNA recognitions motifs (RRMs), which recognise pyrimidine-rich RNAs and allow splicing regulation. It has been reported that all PTBP RRMs can specifically recognise a CUCUCU ligand based on the NMR structure (Mickleburgh et al., 2014). RRM1 and RRM3 recognise the motif YCU, whereas RRM2 recognises core CU. RRM4 is the least specific motif, with recognition of 5-YC-3 (Mickleburgh et al., 2014). PTBs can also bind to other splicing regulators and perhaps inhibit their function (Figure 4.11).

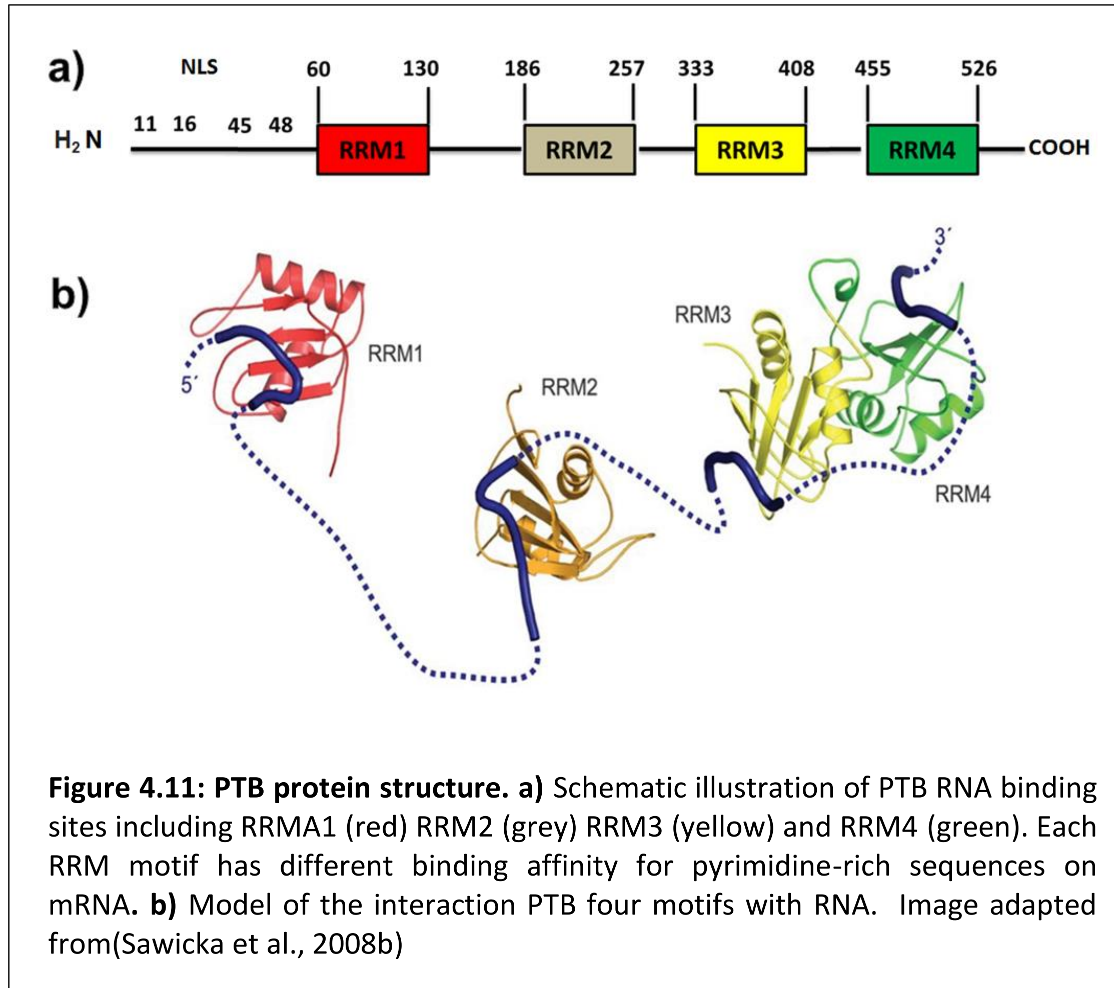


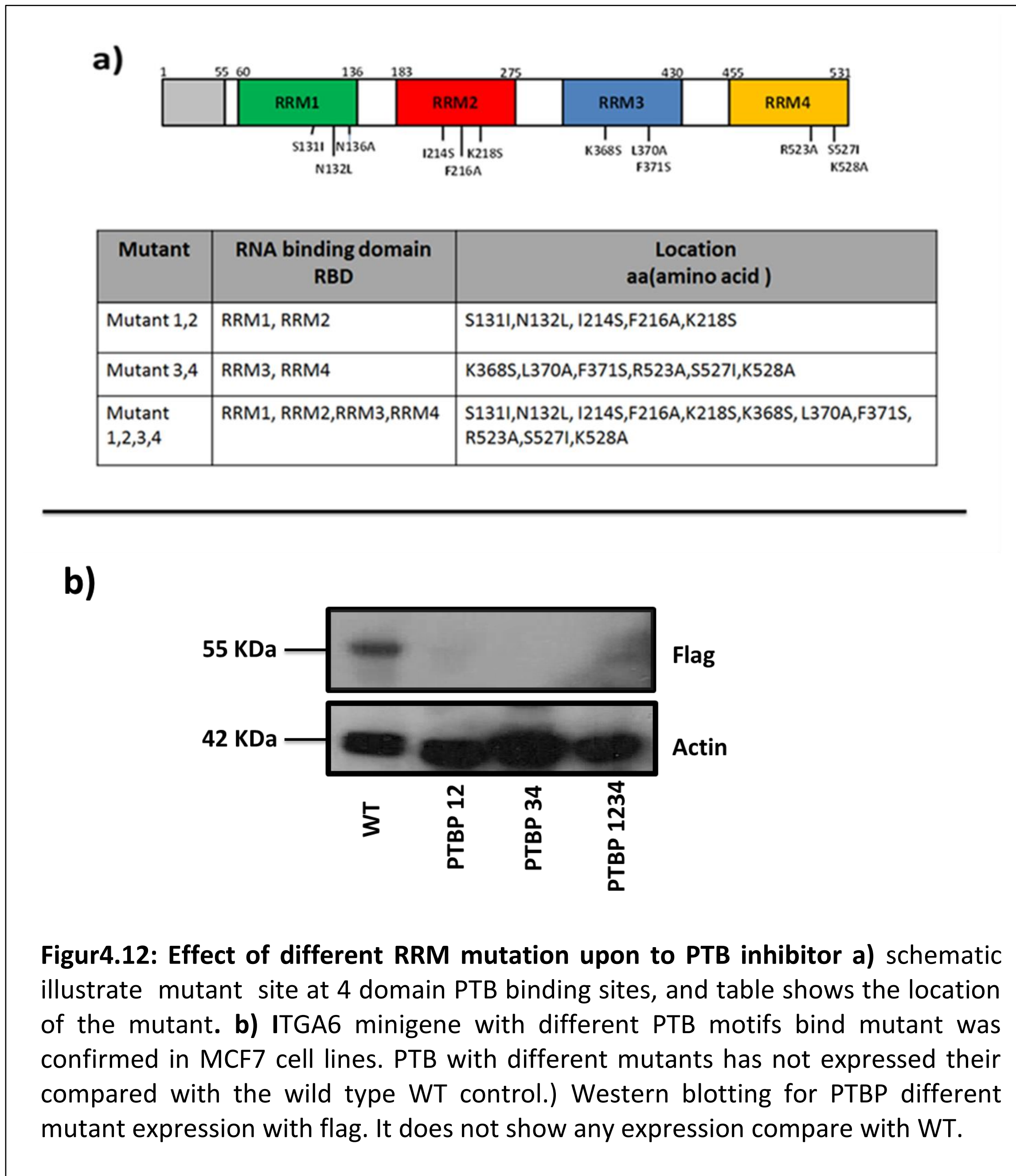
Figure 4.11: PTB protein structure. a) Schematic illustration of PTB RNA binding sites including RRMA1 (red) RRM2 (grey) RRM3 (yellow) and RRM4 (green). Each RRM motif has different binding affinity for pyrimidine-rich sequences on mRNA**. b)** Model of the interaction PTB four motifs with RNA. Image adapted from(Sawicka et al., 2008b)

To find out what part of PTB protein is responsible for responsible for repressing the *ITGA6* exon, PTBPs with several mutant motifs, including RRM1, RRM2, RRM3 and RRM4, were generously provided by Dr C.W.J. Smith. All constructs were cloned within the MluI/AvrII sites of the FLAG-NLS-MS2-ABM vector, which originates from the pCINeo and contains a FLAG tag in the 5' end of the gene and a NLS signal at the 3' end.

The ITGA61.3kb minigene alternative exon, which was previously shown to be inhibited by PTBPs, was co transfected with expression vectors encoding PTBPs in the MCF7 breast cancer cell line (figure 4.12a). These expression vectors encoded different RRM mutants in the following order: WT; mutant 1, 2, mutant3, 4; mutant 1, 2, 3, 4; RRM12L (N-term half); and RRM34 (C-term half). The transfection approach is explained in the methods section.

It was Important to confirm that at the protein level, PTBS RRM1,2, RRM3,4 and RRM1234 were generated with Flag. To accomplish this, western immunoblotting was performed using mouse monoclonal antibody against Flag and rabbit monoclonal against Actin as the control. On western blotting, the expression of the flag protein was only detected and corresponded to the expected size of 55 kDa in the WT. However, the RRM mutants (RRM1,2 RRM3,4 and RRM1234) did not show any bands (Figure 4.12b).

.



Mutant	RNA binding domain RBD	Location aa(amino acid)
Mutant 1,2	RRM1, RRM2	S131I,N132L, I214S,F216A,K218S
Mutant 3,4	RRM3, RRM4	K368S,L370A,F371S,R523A,S527I,K528A
Mutant 1,2,3,4	RRM1, RRM2,RRM3,RRM4	S131I,N132L, I214S,F216A,K218S,K368S, L370A,F371S, R523A,S527I,K528A

Figur4.12: Effect of different RRM mutation upon to PTB inhibitor a) schematic illustrate mutant site at 4 domain PTB binding sites, and table shows the location of the mutant**. b) I**TGA6 minigene with different PTB motifs bind mutant was confirmed in MCF7 cell lines. PTB with different mutants has not expressed their compared with the wild type WT control.) Western blotting for PTBP different mutant expression with flag. It does not show any expression compare with WT.

4.5 Discussion

In this chapter, I explained the design, assembly and validation of the *ITGA6* minigene to study the mechanism of *ITGA6* alternative splicing and its use to identify a binding site for the splicing regulator PTBP. The use of a minigene assay provided the flexibility to identify information about the protein-RNA interaction via mutating and rearranging the gene sequence. Many studies have used minigene assays for splicing, such as (Steffensen et al., 2014) for the BRCA1 gene, (Ulzi et al., 2014) for the CLCN1 gene and (Grellscheid et al., 2011) for Tra2β target *HIPK3*.

This chapter's first aim was to establish an *ITGA6* minigene that mimics the endogenous *ITGA6* alternative exon. To test this aim, the minigene was created using a genomic DNA fragment including the *ITGA6* alternative exon and surrounding intron, which included cloning RNA-binding proteins' candidate sites. This construct transfected into standard cancer cell lines like MCF7, MDA-MB237 and HeLa cells. In vitro expression and alternative splicing of minigene mRNA transcripts were confirmed and found to have a similar splicing pattern to the endogenous gene. This provides strong evidence that the *ITGA6* minigene represents a good model for a splicing study, due to showing roughly the same percentage of ITGA6 alternative exon inclusion as the endogenous gene after transfection into three different cancer cell lines.

Therefore, to get a good minigene to investigate the *ITGA6* alternative exon, I made two additional *ITGA6* minigene constructs with variable intronic lengths (ITGA64K.B and ITGA6 short). These two minigenes were transfected to standard cancer cell lines as well as to the original *ITGA6* minigene (called ITGA61.3KB). ITGA61.3K.B was still the best model in the three cancer cell lines due to the consistency and high efficiency of minigene expression in multiple independent transfections of the exon inclusion expression compared with the ITGA64Kb and the ITGA6 short construct. ITGA64k.b expression was not consistent in all cell lines, and this might be due to the intron length, which is 1413 nucleotides upstream and 1081 nucleotides downstream. The Houston Departments of Pathology and Molecular and Cellular Biology (2005) reported that the primary elements for regulating alternative splicing are typically within 200–

300 nucleotides upstream and/or downstream of the regulated exon. Furthermore, ITGA6 short did not show any exon inclusion. This could be due to the short length of downstream alternative exon, which is 198 nucleotides.

This chapter's second aim was to investigate factors that regulate the *ITGA6* alternative exon. MBNL1, RBFOX2 and ESRPs activated the endogenous *ITGA6* alternative exon, whereas PTBP was a repressor for *ITGA6* alternative exon. Therefore, I performed a co-transfection siRNA knockdown for PTBP, MBNL1, RBFOX2 and ESRP regulation factors with the *ITGA6* minigene in MCF7 and HeLa cell lines. MBNL1 and PTBP regulated the *ITGA6* minigene alternative exon just like the endogenous one. MBNL1 activated *ITGA6* minigene alternative exon, whereas PTBP repressed the *ITGA6* minigene alternative exon. This result was confirmed by transfected *ITGA6* minigene with overexpression constructs for PTBs and MBNL in MCF7. This demonstrated that MBNL drives more inclusion exon 25 ('a' form), whereas PTBP2 represses the exon25 when compared with the co-expression of GFP.

These results showed that ITGA61.3kb minigene is a potential paradigm for the concept of combinatorial control of alternative splicing (Smith & Valcarcel, 2000). The MBNL protein has a high affinity in binding to YGCY sites (TGCT Strong MBNL site, TGCC Medium MBNL site, CGCC Medium MBNL site). The alternative exon of *ITGA6* minigene was activated by MBNL1, consistent with MBNL1 being bound downstream of the alternative exon. The study supports MBNL1 acting as an activator when it is bound downstream of the alternative exon. (Figure 4.13 suggestions the possible binding site for MBNL1 in ITGA6)

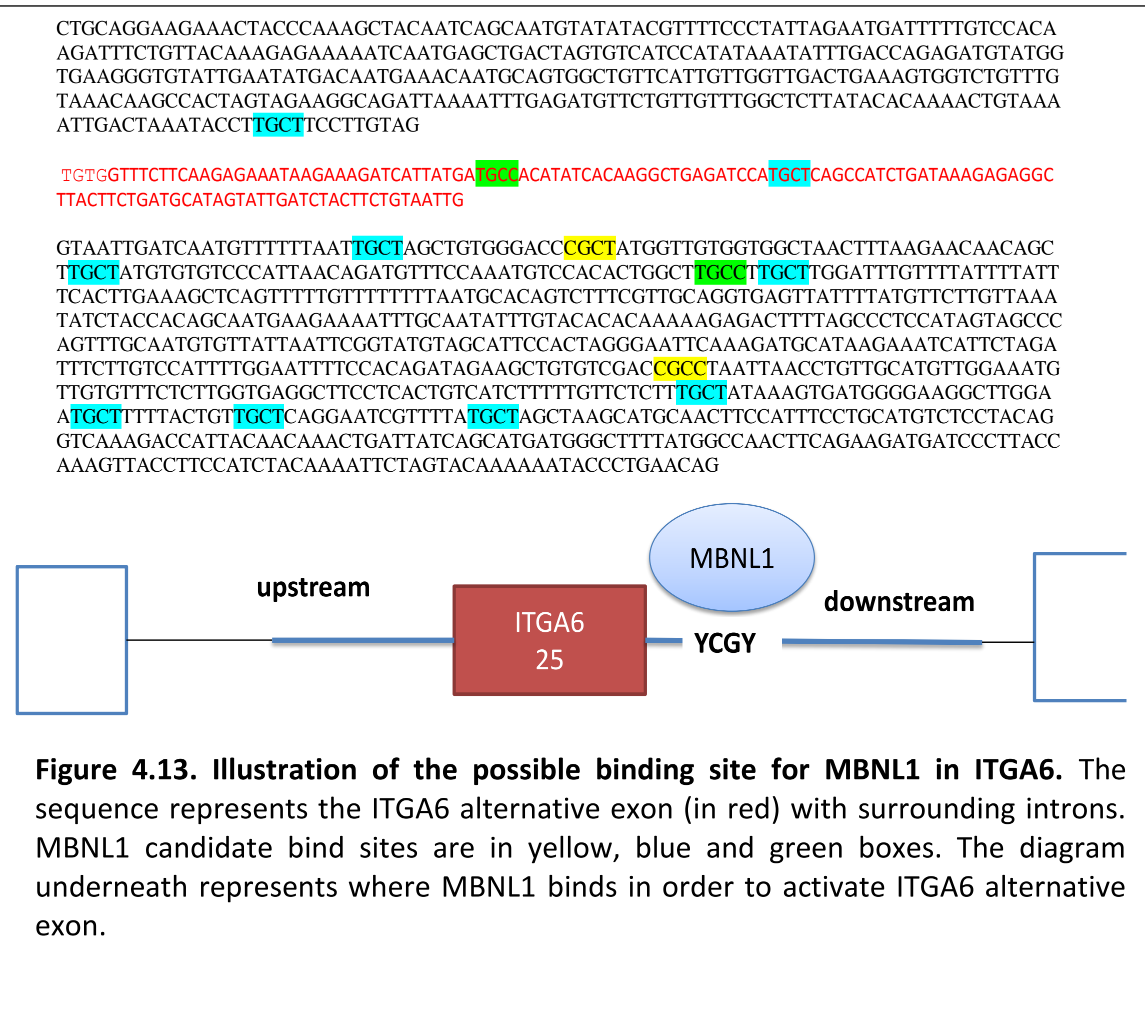


Figure 4.13. Illustration of the possible binding site for MBNL1 in ITGA6. The sequence represents the ITGA6 alternative exon (in red) with surrounding introns. MBNL1 candidate bind sites are in yellow, blue and green boxes. The diagram underneath represents where MBNL1 binds in order to activate ITGA6 alternative exon.

However, PTBPs were also confirmed as novel regulators (repressors) of the *ITGA6* alternative exon after their overexpression. Since PTBP is a novel regulator for *ITGA6*, I used the end of this chapter to investigate PTBP binding and how it can regulate the *ITGA6* alternative exon. PTB overexpression and knockdown was combined with *ITGA6* minigene co-transfection into three standard cancer cell lines, including MCF7, MDA-MB231 and HeLa cell lines. Of these, the MCF7 breast cancer cell line was the best cell line for studying PTBP regulation for the *ITGA6* minigene due to the high-quality response to both PTBP overexpression and depletion. In other words, MCF7 cell lines were the only cell lines that responded to both PTBP overexpression and depletion.

This might be due to the high expression of PTBP in MCF7. Thus, MCF7 was shown to be a good cell line for studying splicing in minigenes.

PTBP is well characterised as a splicing suppressor for alternative exons, particularly in muscles and neuron tissues in which PTB is high. The same studies also reported that PTB acts as a repressor (Smith & Valcarcel, 2000) for exon 3 of the ITM gene. Other studies (Smith & Valcarcel, 2000) showed that PTB represses alternative exons in different genes. PTB usually binds at pyrimidine-rich motifs, such as UCUU or CUCUCU, in the RNA at upstream (near to the 3’ splice site) alternative exon in brain and neuro tissues. To test and identify the PTB-binding sites, we used CLIP-seq results that identified the direct RNA targets for PTB-binding sites in ITGA6 minigene sequences from our collaboration with Dr Smith’s lab at Cambridge University. They had shown that PTB candidate-binding sites were in downstream alternative exon from ITGA6, with the intronic regions near the 5’ splice site (Figure4.15). Also downstream of the alternative exon are PTB consensus binding sites. However, according to current rules, these would be predicted to activate the 25 alternative exon, not repress it. To test whether these downstream sites could repress 25 exon, we mutated them by deleting downstream PTBP candidate sites (Del A & Del B) from an *ITGA6* minigene and transfected it into MCF7 cells. The *ITGA6* alternative exon had less inclusion than with the wild-type minigene.

From our siRNA and over-expression data for PTBP proteins, we expected to see more inclusion for the alternative exon after binding site mutation; however, with mutated PTB-binding sites, we saw less inclusion for the alternative exon. This means that even though PTB expression represses the alternative exon overall, these downstream binding sites are activating.

Several different mechanisms have been proposed for how PTB represses splicing, including PTB interaction with other regulatory proteins. We hypothesise that PTB might interact with other proteins to repress the *ITGA6* alternative exon, and specifically to titrate away splicing activator proteins needed for this activation. This would be due therefore to protein-protein interactions, not protein-RNA interactions.

PTB binds to RNA on four RNA-recognition motifs, which recognise pyrimidine-rich RNAs. To determine what part of PTB proteins was responsible for repressing ITGA6 exon, we transfected PTB with a mutant in different RRM (RRM1.RRM2, RRM3 and RRM4, which were generated and provided by Dr C. W. Smith) in MCF7 cancer cell lines. However the confirmation test for those mutations in protein levels did not show expression.

Taken as a whole, these results indicate that PTB might interact with other regulatory proteins to repress the *ITGA6* alterative exon. Other studies have shown that PTBP can repress alternative exon splicing via interaction with other regulatory factors, including Nova-1 and Nova-2 (Polydorides et al., 2000), Raver1 (Gromak et al., 2003) and MRG15 (Luco et al., 2010). PTB-binding sites are located downstream of the ITGA6 alternative exon within a cluster of binding sites for other proteins, including MBNL1, RBFOX and ESRPs. As a result, PTB might also follow the mechanism that suggests MBNL1 and PTB cooperate to repress alternative exon during splicing. (Gooding et al., 2013) conducted a study that showed MBNL1 interacts with PTB to repress the splicing of TPM1 exon 3. Combining our findings with those of other studies, such as (Gooding et al., 2013), we predict that PTB might interact with other regulatory proteins, likely MBNL1 or RBFOX, to repress the ITGA6 alternative exon as a complex (Figure 4.14).

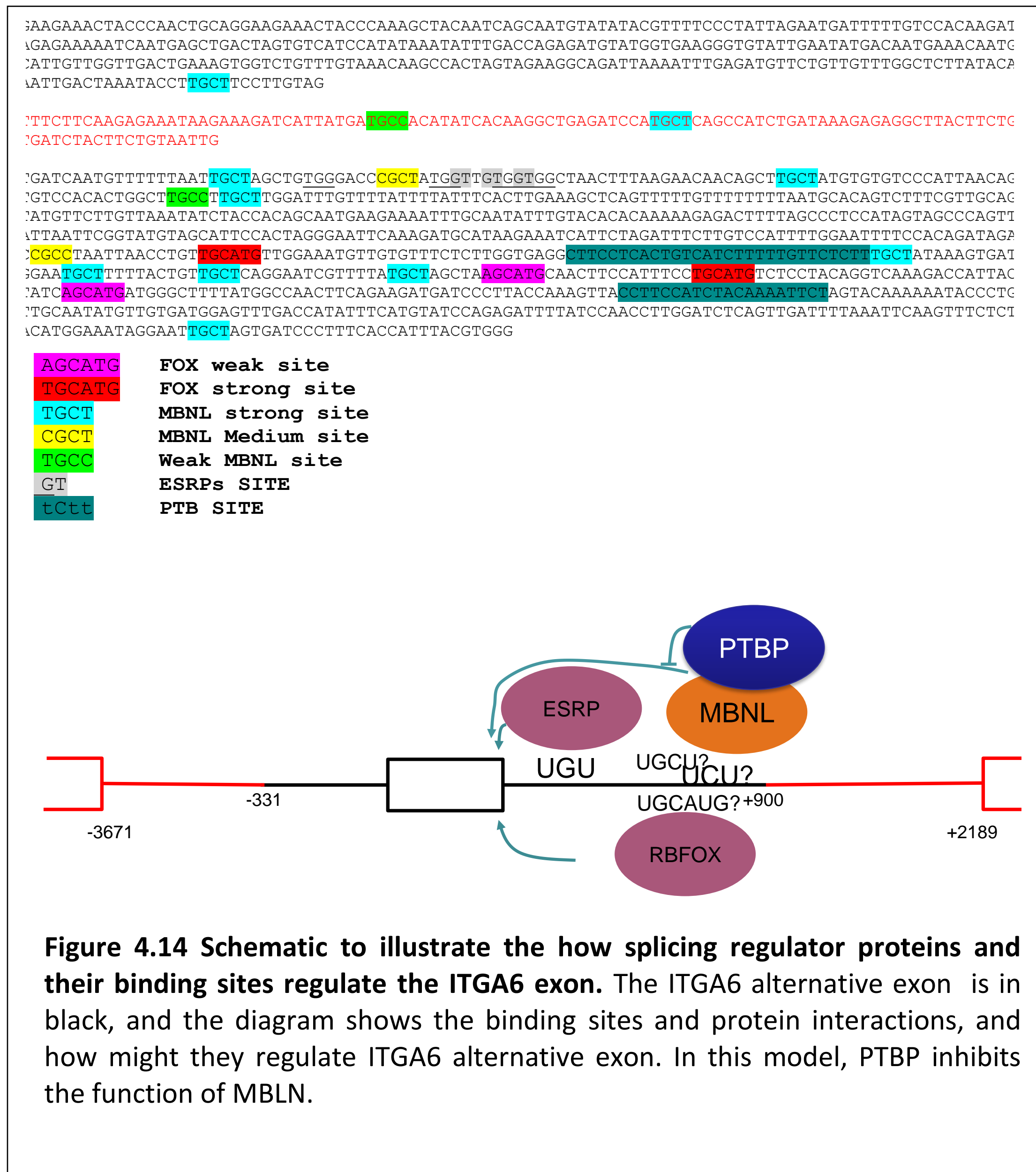


Figure 4.14 Schematic to illustrate the how splicing regulator proteins and their binding sites regulate the ITGA6 exon. The ITGA6 alternative exon is in black, and the diagram shows the binding sites and protein interactions, and how might they regulate ITGA6 alternative exon. In this model, PTBP inhibits the function of MBLN.

To summarise, this chapter established the gold model of an ITGA6 minigene, which allowed us to investigate how the ITGA6 alternative exon can be regulated. MBNL1 and PTB were the factors that regulated *ITGA6* alternative exon both endogenously and in the minigene. Furthermore, PTB, which is a novel regulatory factor for the *ITGA6* alternative exon, was identified to bind downstream of the alternative exon and yet to repress the alternative exon. By deleting the candidate-binding sites, PTB was shown to activate the alternative exon through these downstream sites. The exact mechanism through which PTB can repress ITGA6 alternative exon was not identified; however, several hypotheses exist for how PTB can regulate alternative splicing, including through binding to specific individual binding sites to control exon inclusion, or through the indirect repression via interaction with other regulatory proteins, of which our data would support the latter model. We predict that PTB interacts with other regulatory proteins, whether MBNL1 or RBFOX2, to indirectly regulate ITGA6 alternative exon, as well as through direct binding to the downstream binding sites.

Chapter 5 General discussion and conclusion

Alternative splicing is now widely considered an important mechanism for controlling gene expression in the eukaryotic cells. During cell differentiation, gene expression can be changed in different ways, including alternative splicing. Recently, abnormal alternative splicing has been reported to play a major role in causing some serious conditions, including cancer, infertility, muscular and neurological disease. Alternative splicing occurs via regulatory proteins, each of which has a specific mechanism. In this thesis, I have focused on studying alternative splicing mechanisms in depth.

To enable investigation of fundamentally important alternative splicing mechanisms, it is important to find a gene which has an important function in development and has been conserved across many species during evolution. I started my project by using data from (Venables et al., 2013b), who discovered alternative splicing events conserved across the evolutionary tree for 10 genes, including *PLOD2, CLSTN1, ATP2A1, PALM, KIF13A, FMNL3, PPIP5K1, MARK2, FNIP1* and *ITGA6*, which are all important in stem cell biology. As these alternative splicing events have been conserved, they are likely to be extremely important in the cell differentiation and body plan development of all vertebrates, including humans; thus, these events probably also have medical significance. I investigated the above genes in early zebrafish development to find alternative splicing events with regulation mechanisms that have been conserved between fish and humans. I found that zebrafish integrin α6 (*ITGA6*) and kinesin family member 13A (*KIF13A*) had alternative splicing mechanisms that differed between the head and tail after fertilization of the egg. However, in my research, *ITGA6* was the only gene which changed its splicing pattern during the course of embryonic development. This result suggested that *ITGA6* was an appropriate gene on which to focus the majority of my study of alternative splicing mechanisms.

ITGA6 belongs to the integrin family, which is a group of dimeric cell surface proteins that are each composed of an alpha and a beta chain. Due to ancient gene duplication, *ITGA6* has three mammalian paralogues, including *ITGA3* and *ITGA7*. I found that *ITGA6* has five paralogues in zebrafish. The presences of two alternative C-terminals isoforms of ITGA6 are conserved among all vertebrate, even in zebrafish. Five paralogues of *ITGA6* were seen in the zebrafish head and tail during development after fertilization; only *ITGA6A* and *ITGA6B* showed similar alternative splicing patterns, suggesting that they may have similar functions.

ITGA6 splicing patterns in human cancer cell lines were determined to see if they were similar to the different splicing profiles observed in zebrafish and stem cell development. These studies indicated important differences in splicing profiles between HeLa, MCF7 and MDA-MB-231 cells. These might be important. Some studies have reported that *ITGA6A* and *ITGA6B* have the same function in mouse embryogenesis. However, there is evidence that the ITGA6 A and B isoforms have different functions in cancer stem cells. In cancer, ITGA6 may be necessary for the tumorigenicity of a stem cell–like population within the MCF-7 cell line (Cariati et al., 2008), and ITGA6 may also regulate glioblastoma stem cell differentiation (Lathia et al. 2010). High ITGA6 expression is a biomarker in breast cancer cells.

Alternative splicing mechanisms of *ITGA6* within cancer cell lines was also studied to understand how they are regulated. Alternative splicing of *ITGA6* creates two different cytoplasmic C-termini in fibroblasts and stem cells (Cooper et al., 1991, Tamura et al., 1991). It has already been shown that *ITGA6* exon 25 is under the control of three separate splicing factors. In two parallel studies of alternative splicing during stem cell differentiation, Ben Blencowe's laboratory showed that the knockdown of the splicing regulatory factor MBNL1 increases the efficiency of stem cell differentiation (Han et al., 2013), and Venables et al. showed that the MBNL1 and RBFOX2 splicing regulatory factors enhance the inclusion of the *ITGA6* exon 25 in fibroblasts (Venables et al., 2013) . A recent study showed that the epithelial splicing regulatory protein (ESRP) enhances the inclusion of this exon as well (Goel et al., 2014).

We aimed to examine these regulation factors and try to identify others using cancer cell lines. We confirmed that the MBNL1, RBFOX2 and ESRP regulatory factors activate the inclusion of the alternative exon 25 of *ITGA6* in breast and cervical cancer cell lines. Normally, the splicing activities of these factors are controlled according to their binding sites (Dredge et al., 2005; Zhang et al., 2008a; Goers et al., 2010; Llorian et al., 2010). MBNL1, RBFOX and ESRP enhance exon inclusion when they bind downstream of the exon but inhibit it when they bind upstream. ITGA6a is involved in the proliferation of colon cancer cells, and ITGA6b is involved in cancer stem cell function by means of a feedback loop involving VEGF. If ITGA6 is repressed by the ESRP factor, it might also be repressed by MBNL1, RBFOX or PTBP, which could affect VEGF signalling.

After confirming the activating effects of the above splicing factors in cancer cell lines, we hoped to identify a novel regulation factor that represses the *ITGA6* alternative exon 25. We identified PTB as a novel regulator of this splicing event. Polypyrimidine tract-binding protein (PTBP) is known to be a repressive splicing factor. Experiments involving the siRNA knockdown of PTBP in breast cancer cell lines strongly reduced the skipping of exon 25 from *ITGA6*, indicating that PTBP is inhibitory. This is in direct contrast to *ITGA6* alternative exon 25 being activated by MBNL1, RBFOX and ESRP.

It was important to understand how these RNA splicing regulators regulate alternative splicing of ITGA6. Consequently, we aimed to identify the PTBP binding site and examine how it can regulate the inclusion of the *ITGA6* alternative exon 25. PTBP usually binds at pyrimidine-rich motifs in the RNA, such as UCUU or CUCUCU, and binds upstream of the alternative exon (near the 3’ splice site) to act as a repressor. The hypothesis was that if PTBP binds downstream, it may still function as a repressor of the *ITGA6* alternative exon 25, perhaps by interacting with other regulatory factors. To test this hypothesis, I established a minigene for the *ITGA6* alternative exon 25 with introns surrounding it upstream and downstream. This 1.3-kb *ITGA6* minigene behaved as if it were endogenous in terms of splicing control, showing inhibition by PTBP and activation by MBNL1. Two candidate binding sites for PTBP were identified, both downstream. However, mutagenesis of these

PTBP candidate binding sites reduced the inclusion of the *ITGA6* alternative exon 25. This means that PTBP is actually activating this exon through direct RNA binding, even though it has an overall repressive effect. Studies have shown that PTBP can repress alternative exon splicing via interaction with other regulatory factors, including RAVER1 (Gromak et al., 2003), MRG15 (Luco et al., 2010) and NOVA1 and NOVA2 (Polydorides et al., 2000). PTB might thus be indirectly repressing splicing of the ITGA6 exon through protein-protein interactions, possibly with MBLN or RBFOX, to inhibit the splicing activity of these proteins.

Given the finding that the PTBP binding sites are located downstream of the *ITGA6* alternative exon 25 within a cluster of binding sites for other proteins, including MBNL1, RBFOX, ESRP and PTB suggested in two hypotheses. One hypothesis is that PTB binding downstream of the alternative exon can repress alternative exon splicing via direct binding to RNA. This was proven not to be the case. In fact, PTB binding downstream of the regulated exon seemed to weakly enhance this exon, similarly to other examples of PTB splicing regulation. The alternative hypothesis was that PTB might repress ITGA6 splicing through indirect interactions. In this case and we suggest PTB interacts with regulatory proteins such as MBNL1 that normally activate the inclusion of the alternative exon 25, and this interaction prevents MBLN from functioning properly (Figure 5.1). Supporting this hypothesis, MBNL1 protein interactions with PTBP have been reported (Gooding et al., 2013).

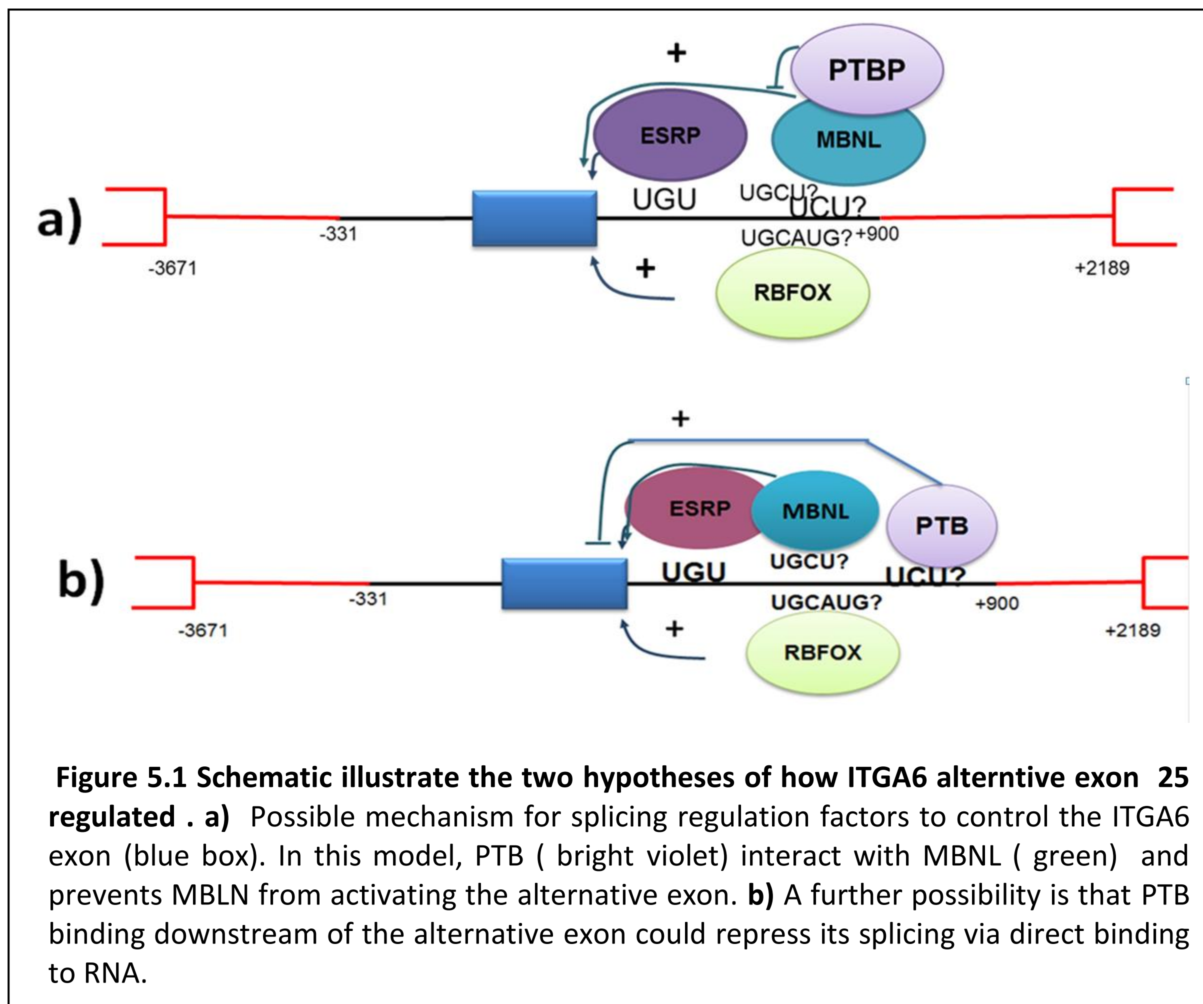


Figure 5.1 Schematic illustrate the two hypotheses of how ITGA6 alterntive exon 25 regulated . a) Possible mechanism for splicing regulation factors to control the ITGA6 exon (blue box). In this model, PTB (bright violet) interact with MBNL (green) and prevents MBLN from activating the alternative exon. **b)** A further possibility is that PTB binding downstream of the alternative exon could repress its splicing via direct binding to RNA.

To conclude, the research has demonstrated the principle of combinatorial regulation of alternative splicing in *ITGA6*. We found that the *ITGA6* alternative splicing mechanism is conserved across evolutionary history. I confirmed that the *ITGA6* alternative exon 25 is activated by RBFOX, MBNL1 and ESRP in cancer cell lines as well as stem cells and fibroblasts. PTBP was identified as a novel regulatory factor for *ITGA6* and also found the sequence in *ITGA6* through which the repression of alternative splicing by PTBP occurs. I also developed a hypothesis for how *ITGA6* may be regulated by PTBP and MBNL1 in cancer cell lines. The findings regarding PTBP regulation of alternative splicing by interacting with other proteins will prove an exciting basis for further research.

Bibliography

AMIR-AHMADY, B., BOUTZ, P. L., MARKOVTSOV, V., PHILLIPS, M. L. & BLACK, D. L. 2005. Exon repression by polypyrimidine tract binding protein. *RNA,* 11**,** 699-716.

ANKO, M. L. & NEUGEBAUER, K. M. 2012. RNA-protein interactions in vivo: global gets specific. *Trends Biochem Sci,* 37**,** 255-62.

ASKARI, J. A., BUCKLEY, P. A., MOULD, A. P. & HUMPHRIES, M. J. 2009. Linking integrin conformation to function. *J Cell Sci,* 122**,** 165-70.

BARCZYK, M., CARRACEDO, S. & GULLBERG, D. 2010. Integrins. *Cell Tissue Res,* 339**,** 269-80.

BERGET, S. M. 1995. Exon recognition in vertebrate splicing. *J Biol Chem,* 270**,** 2411-4.

BERMAN, A. E., KOZLOVA, N. I. & MOROZEVICH, G. E. 2003. Integrins: structure and signaling. *Biochemistry (Mosc),* 68**,** 1284-99.

BLACK, D. L. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem,* 72**,** 291-336.

BOUTZ, P. L., CHAWLA, G., STOILOV, P. & BLACK, D. L. 2007. MicroRNAs regulate the expression of the alternative splicing factor nPTB during muscle development. *Genes Dev,* 21**,** 71-84.

BUCKANOVICH, R. J. & DARNELL, R. B. 1997. The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. *Mol Cell Biol,* 17**,** 3194-201.

CAILLEAU, R., OLIVE, M. & CRUCIGER, Q. V. J. 1978. Long-Term Human Breast Carcinoma Cell Lines of Metastatic Origin - Preliminary Characterization. *In Vitro-Journal of the Tissue Culture Association,* 14**,** 911-915.

CALDERWOOD, D. A., FUJIOKA, Y., DE PEREDA, J. M., GARCIA-ALVAREZ, B., NAKAMOTO, T., MARGOLIS, B., MCGLADE, C. J., LIDDINGTON, R. C. & GINSBERG, M. H. 2003. Integrin beta cytoplasmic domain interactions with phosphotyrosine-binding domains: a structural prototype for diversity in integrin signaling. *Proc Natl Acad Sci U S A,* 100**,** 2272-7.

CARIATI, M., NADERI, A., BROWN, J. P., SMALLEY, M. J., PINDER, S. E., CALDAS, C. & PURUSHOTHAM, A. D. 2008. Alpha-6 integrin is necessary for the tumourigenicity of a stem cell-like subpopulation within the MCF7 breast cancer cell line. *Int J Cancer,* 122**,** 298-304.

CASTLE, J. C., ZHANG, C., SHAH, J. K., KULKARNI, A. V., KALSOTRA, A., COOPER, T. A. & JOHNSON, J. M. 2008. Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nat Genet,* 40**,** 1416-25.

CHENG, C., YAFFE, M. B. & SHARP, P. A. 2006. A positive feedback loop couples Ras activation and CD44 alternative splicing. *Genes Dev,* 20**,** 1715-20.

CHEUNG, H. C., HAI, T., ZHU, W., BAGGERLY, K. A., TSAVACHIDIS, S., KRAHE, R. & COTE, G. J. 2009. Splicing factors PTBP1 and PTBP2 promote proliferation and migration of glioma cell lines. *Brain,* 132**,** 2277-88.

COOPER, H. M., TAMURA, R. N. & QUARANTA, V. 1991. The major laminin receptor of mouse embryonic stem cells is a novel isoform of the alpha 6 beta 1 integrin. *J Cell Biol,* 115**,** 843-50.

COOPER, T. A. 2005. Use of minigene systems to dissect alternative splicing elements. *Methods,* 37**,** 331-40.

CRAMER, P., SREBROW, A., KADENER, S., WERBAJH, S., DE LA MATA, M., MELEN, G., NOGUÉS, G. & KORNBLIHTT, A. R. 2001. Coordination between transcription and pre-mRNA processing. *FEBS Letters,* 498**,** 179-182.

DAMIANOV, A. & BLACK, D. L. 2010. Autoregulation of Fox protein expression to produce dominant negative splicing factors. *RNA,* 16**,** 405-16.

DARNELL, J. E., JR. 2013. Reflections on the history of pre-mRNA processing and highlights of current knowledge: a unified picture. *RNA,* 19**,** 443-60.

DE MELKER, A. A. & SONNENBERG, A. 1999. Integrins: alternative splicing as a mechanism to regulate ligand binding and integrin signaling events. *Bioessays,* 21**,** 499-509.

DJEBALI, S., LAGARDE, J., KAPRANOV, P., LACROIX, V., BOREL, C., MUDGE, J. M., HOWALD, C., FOISSAC, S., UCLA, C., CHRAST, J., RIBECA, P., MARTIN, D., MURRAY, R. R., YANG, X., GHAMSARI, L., LIN, C., BELL, I., DUMAIS, E., DRENKOW, J., TRESS, M. L., GELPI, J. L., OROZCO, M., VALENCIA, A., VAN BERKUM, N. L., LAJOIE, B. R., VIDAL, M., STAMATOYANNOPOULOS, J., BATUT, P., DOBIN, A., HARROW, J., HUBBARD, T., DEKKER, J., FRANKISH, A., SALEHI-ASHTIANI, K., REYMOND, A., ANTONARAKIS, S. E., GUIGO, R. & GINGERAS, T. R. 2012. Evidence for transcript networks composed of chimeric RNAs in human cells. *PLoS One,* 7**,** e28213.

FARDAEI, M., ROGERS, M. T., THORPE, H. M., LARKIN, K., HAMSHERE, M. G., HARPER, P. S. & BROOK, J. D. 2002. Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum Mol Genet,* 11**,** 805-14.

GERSTEIN, M. B., BRUCE, C., ROZOWSKY, J. S., ZHENG, D., DU, J., KORBEL, J. O., EMANUELSSON, O., ZHANG, Z. D., WEISSMAN, S. & SNYDER, M. 2007. What is a gene, post-ENCODE? History and updated definition. *Genome Res,* 17**,** 669-81.

GIMOND, C., BAUDOIN, C., VAN DER NEUT, R., KRAMER, D., CALAFAT, J. & SONNENBERG, A. 1998. Cre-loxP-mediated inactivation of the alpha6A integrin splice variant in vivo: evidence for a specific functional role of alpha6A in lymphocyte migration but not in heart development. *J Cell Biol,* 143**,** 253-66.

GOEL, H. L., GRITSKO, T., PURSELL, B., CHANG, C., SHULTZ, L. D., GREINER, D. L., NORUM, J. H., TOFTGARD, R., SHAW, L. M. & MERCURIO, A. M. 2014. Regulated splicing of the alpha6 integrin cytoplasmic domain determines the fate of breast cancer stem cells. *Cell Rep,* 7**,** 747-61.

GOODING, C., EDGE, C., LORENZ, M., COELHO, M. B., WINTERS, M., KAMINSKI, C. F., CHERNY, D., EPERON, I. C. & SMITH, C. W. 2013. MBNL1 and PTB cooperate to repress splicing of Tpm1 exon 3. *Nucleic Acids Res,* 41**,** 4765-82.

GRELLSCHEID, S., DALGLIESH, C., STORBECK, M., BEST, A., LIU, Y., JAKUBIK, M., MENDE, Y., EHRMANN, I., CURK, T., ROSSBACH, K., BOURGEOIS, C. F., STEVENIN, J., GRELLSCHEID, D., JACKSON, M. S., WIRTH, B. & ELLIOTT, D. J. 2011. Identification of evolutionarily conserved exons as regulated targets for the splicing activator tra2beta in development. *PLoS Genet,* 7**,** e1002390.

GROMAK, N., RIDEAU, A., SOUTHBY, J., SCADDEN, A. D., GOODING, C., HUTTELMAIER, S., SINGER, R. H. & SMITH, C. W. 2003. The PTB interacting protein raver1 regulates alpha-tropomyosin alternative splicing. *EMBO J,* 22**,** 6356-64.

GROULX, J. F., GIROUX, V., BEAUSEJOUR, M., BOUDJADI, S., BASORA, N., CARRIER, J. C. & BEAULIEU, J. F. 2014. Integrin alpha6A splice variant regulates proliferation and the Wnt/beta-catenin pathway in human colorectal cancer cells. *Carcinogenesis,* 35**,** 1217-27.

HAN, H., IRIMIA, M., ROSS, P. J., SUNG, H. K., ALIPANAHI, B., DAVID, L., GOLIPOUR, A., GABUT, M., MICHAEL, I. P., NACHMAN, E. N., WANG, E., TRCKA, D., THOMPSON, T., O'HANLON, D., SLOBODENIUC, V., BARBOSA-MORAIS, N. L., BURGE, C. B., MOFFAT, J., FREY, B. J., NAGY, A., ELLIS, J., WRANA, J. L. & BLENCOWE, B. J. 2013. MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. *Nature,* 498**,** 241-5.

HO, T. H., CHARLET, B. N., POULOS, M. G., SINGH, G., SWANSON, M. S. & COOPER, T. A. 2004. Muscleblind proteins regulate alternative splicing. *EMBO J,* 23**,** 3103-12.

HOGERVORST, F., ADMIRAAL, L. G., NIESSEN, C., KUIKMAN, I., JANSSEN, H., DAAMS, H. & SONNENBERG, A. 1993. Biochemical characterization and tissue distribution of the A and B variants of the integrin alpha 6 subunit. *J Cell Biol,* 121**,** 179-91.

INTERNATIONAL HUMAN GENOME SEQUENCING, C. 2004. Finishing the euchromatic sequence of the human genome. *Nature,* 431**,** 931-45.

ISHII, H., SAITOH, M., SAKAMOTO, K., KONDO, T., KATOH, R., TANAKA, S., MOTIZUKI, M., MASUYAMA, K. & MIYAZAWA, K. 2014. Epithelial splicing regulatory proteins 1 (ESRP1) and 2 (ESRP2) suppress cancer cell motility via different mechanisms. *J Biol Chem,* 289**,** 27386-99.

JANGA, S. C. & MITTAL, N. 2011. Construction, structure and dynamics of post-transcriptional regulatory network directed by RNA-binding proteins. *Adv Exp Med Biol,* 722**,** 103-17.

JOHNSON, M. S., LU, N., DENESSIOUK, K., HEINO, J. & GULLBERG, D. 2009. Integrins during evolution: evolutionary trees and model organisms. *Biochim Biophys Acta,* 1788**,** 779-89.

KALSOTRA, A. & COOPER, T. A. 2011. Functional consequences of developmentally regulated alternative splicing. *Nat Rev Genet,* 12**,** 715-29.

KIMMEL, C. B., BALLARD, W. W., KIMMEL, S. R., ULLMANN, B. & SCHILLING, T. F. 1995. Stages of embryonic development of the zebrafish. *Dev Dyn,* 203**,** 253-310.

LANDER, E. S. 2011. Initial impact of the sequencing of the human genome. *Nature,* 470**,** 187-97.

LATHIA, J. D., GALLAGHER, J., HEDDLESTON, J. M., WANG, J., EYLER, C. E., MACSWORDS, J., WU, Q., VASANJI, A., MCLENDON, R. E., HJELMELAND, A. B. & RICH, J. N. 2010. Integrin alpha 6 regulates glioblastoma stem cells. *Cell Stem Cell,* 6**,** 421-32.

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. *Nat Struct Mol Biol,* 17**,** 1114-23.

LUCO, R. F., ALLO, M., SCHOR, I. E., KORNBLIHTT, A. R. & MISTELI, T. 2011. Epigenetics in alternative pre-mRNA splicing. *Cell,* 144**,** 16-26.

MANDAL, S. S., CHU, C., WADA, T., HANDA, H., SHATKIN, A. J. & REINBERG, D. 2004. Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *Proc Natl Acad Sci U S A,* 101**,** 7572-7.

MARTINEZ-CONTRERAS, R., CLOUTIER, P., SHKRETA, L., FISETTE, J. F., REVIL, T. & CHABOT, B. 2007. hnRNP proteins and splicing control. *Adv Exp Med Biol,* 623**,** 123-47.

MCMANUS, C. J. & GRAVELEY, B. R. 2011. RNA structure and the mechanisms of alternative splicing. *Curr Opin Genet Dev,* 21**,** 373-9.

MERKIN, J., RUSSELL, C., CHEN, P. & BURGE, C. B. 2012. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science,* 338**,** 1593-9.

MICKLEBURGH, I., KAFASLA, P., CHERNY, D., LLORIAN, M., CURRY, S., JACKSON, R. J. & SMITH, C. W. 2014. The organization of RNA contacts by PTB for regulation of FAS splicing. *Nucleic Acids Res,* 42**,** 8605-20.

MOORE, M. J. & SHARP, P. A. 1993. Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature,* 365**,** 364-8.

MORGAN, M. R., HUMPHRIES, M. J. & BASS, M. D. 2007. Synergistic control of cell adhesion by integrins and syndecans. *Nat Rev Mol Cell Biol,* 8**,** 957-69.

NILSEN, T. W. & GRAVELEY, B. R. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature,* 463**,** 457-63.

NOIRET, M., AUDIC, Y. & HARDY, S. 2012a. Expression analysis of the polypyrimidine tract binding protein (PTBP1) and its paralogs PTBP2 and PTBP3 during Xenopus tropicalis embryogenesis. *Int J Dev Biol,* 56**,** 747-53.

NOIRET, M., AUDIC, Y. & HARDY, S. 2012b. Expression analysis of the polypyrimidine tract binding protein (PTBP1) and its paralogs PTBP2 and PTBP3 during Xenopus tropicalis embryogenesis. *International Journal of Developmental Biology,* 56**,** 747-753.

PAN, Q., SHAI, O., LEE, L. J., FREY, B. J. & BLENCOWE, B. J. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet,* 40**,** 1413-5.

PATEL, A. A. & STEITZ, J. A. 2003. Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol,* 4**,** 960-70.

POLYDORIDES, A. D., OKANO, H. J., YANG, Y. Y., STEFANI, G. & DARNELL, R. B. 2000. A brain-enriched polypyrimidine tract-binding protein antagonizes the ability of Nova to regulate neuron-specific alternative splicing. *Proc Natl Acad Sci U S A,* 97**,** 6350-5.

PULKKINEN, L., KIMONIS, V. E., XU, Y., SPANOU, E. N., MCLEAN, W. H. & UITTO, J. 1997. Homozygous alpha6 integrin mutation in junctional epidermolysis bullosa with congenital duodenal atresia. *Hum Mol Genet,* 6**,** 669-74.

REDDY, A. S. 2007. Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu Rev Plant Biol,* 58**,** 267-94.

REDDY, K. V. & MANGALE, S. S. 2003. Integrin receptors: the dynamic modulators of endometrial function. *Tissue Cell,* 35**,** 260-73.

REED, R. 2000. Mechanisms of fidelity in pre-mRNA splicing. *Curr Opin Cell Biol,* 12**,** 340-5.

SAKHARKAR, M. K., CHOW, V. T. & KANGUEANE, P. 2004. Distributions of exons and introns in the human genome. *In Silico Biol,* 4**,** 387-93.

SASTRY, S. K. & HORWITZ, A. F. 1993. Integrin cytoplasmic domains: mediators of cytoskeletal linkages and extra- and intracellular initiated transmembrane signaling. *Curr Opin Cell Biol,* 5**,** 819-31.

SAWICKA, K., BUSHELL, M., SPRIGGS, K. A. & WILLIS, A. E. 2008a. Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein. *Biochemical Society Transactions,* 36**,** 641-647.

SAWICKA, K., BUSHELL, M., SPRIGGS, K. A. & WILLIS, A. E. 2008b. Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein. *Biochem Soc Trans,* 36**,** 641-7.

SINGH, N. N., ANDROPHY, E. J. & SINGH, R. N. 2004. The regulation and regulatory activities of alternative splicing of the SMN gene. *Crit Rev Eukaryot Gene Expr,* 14**,** 271-85.

SIOMI, H. & DREYFUSS, G. 1997. RNA-binding proteins as regulators of gene expression. *Curr Opin Genet Dev,* 7**,** 345-53.

SMITH, C. W. & VALCARCEL, J. 2000. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci,* 25**,** 381-8.

SOLLER, M. 2006. Pre-messenger RNA processing and its regulation: a genomic perspective. *Cell Mol Life Sci,* 63**,** 796-819.

SOULE, H. D., VAZQUEZ, J., LONG, A., ALBERT, S. & BRENNAN, M. 1973. Human Cell Line from a Pleural Effusion Derived from a Breast Carcinoma. *Journal of the National Cancer Institute,* 51**,** 1409-1416.

SREBROW, A. & KORNBLIHTT, A. R. 2006. The connection between splicing and cancer. *J Cell Sci,* 119**,** 2635-41.

STEFFENSEN, A. Y., DANDANELL, M., JONSON, L., EJLERTSEN, B., GERDES, A. M., NIELSEN, F. C. & HANSEN, T. 2014. Functional characterization of BRCA1 gene variants by mini-gene splicing assay. *Eur J Hum Genet,* 22**,** 1362-8.

STERNER, D. A., CARLO, T. & BERGET, S. M. 1996. Architectural limits on split genes. *Proc Natl Acad Sci U S A,* 93**,** 15081-5.

STEWART, M. 2010. Nuclear export of mRNA. *Trends Biochem Sci,* 35**,** 609-17.

TAMURA, R. N., COOPER, H. M., COLLO, G. & QUARANTA, V. 1991. Cell type-specific integrin variants with alternative alpha chain cytoplasmic domains. *Proc Natl Acad Sci U S A,* 88**,** 10183-7.

TARN, W. Y. & STEITZ, J. A. 1996. Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science,* 273**,** 1824-32.

TRIPATHI, V., ELLIS, J. D., SHEN, Z., SONG, D. Y., PAN, Q., WATT, A. T., FREIER, S. M., BENNETT, C. F., SHARMA, A., BUBULYA, P. A., BLENCOWE, B. J., PRASANTH, S. G. & PRASANTH, K. V. 2010. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell,* 39**,** 925-38.

TURUNEN, J. J., NIEMELA, E. H., VERMA, B. & FRILANDER, M. J. 2013. The significant other: splicing by the minor spliceosome. *Wiley Interdiscip Rev RNA,* 4**,** 61-76.

ULZI, G., SANSONE, V. A., MAGRI, F., CORTI, S., BRESOLIN, N., COMI, G. P. & LUCCHIARI, S. 2014. In vitro analysis of splice site mutations in the CLCN1 gene using the minigene assay. *Mol Biol Rep,* 41**,** 2865-74.

VALADKHAN, S. 2007. The spliceosome: caught in a web of shifting interactions. *Curr Opin Struct Biol,* 17**,** 310-5.

VALCARCEL, J. & GEBAUER, F. 1997a. Post-transcriptional regulation: the dawn of PTB. *Curr Biol,* 7**,** R705-8.

VALCARCEL, J. & GEBAUER, F. 1997b. Post-transcriptional regulation: The dawn of PTB. *Current Biology,* 7**,** R705-R708.

VAN DER FLIER, A. & SONNENBERG, A. 2001. Function and interactions of integrins. *Cell Tissue Res,* 305**,** 285-98.

VENABLES, J. P., BROSSEAU, J. P., GADEA, G., KLINCK, R., PRINOS, P., BEAULIEU, J. F., LAPOINTE, E., DURAND, M., THIBAULT, P., TREMBLAY, K., ROUSSET, F., TAZI, J., ABOU ELELA, S. & CHABOT, B. 2013a. RBFOX2 is an important regulator of mesenchymal tissue-specific splicing in both normal and cancer tissues. *Mol Cell Biol,* 33**,** 396-405.

VENABLES, J. P., KLINCK, R., KOH, C., GERVAIS-BIRD, J., BRAMARD, A., INKEL, L., DURAND, M., COUTURE, S., FROEHLICH, U., LAPOINTE, E., LUCIER, J. F., THIBAULT, P., RANCOURT, C., TREMBLAY, K., PRINOS, P., CHABOT, B. & ELELA, S. A. 2009. Cancer-associated regulation of alternative splicing. *Nat Struct Mol Biol,* 16**,** 670-6.

VENABLES, J. P., LAPASSET, L., GADEA, G., FORT, P., KLINCK, R., IRIMIA, M., VIGNAL, E., THIBAULT, P., PRINOS, P., CHABOT, B., ABOU ELELA, S., ROUX, P., LEMAITRE, J. M. & TAZI, J. 2013b. MBNL1 and RBFOX2 cooperate to establish a splicing programme involved in pluripotent stem cell differentiation. *Nat Commun,* 4**,** 2480.

WAHL, M. C., WILL, C. L. & LUHRMANN, R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell,* 136**,** 701-18.

WANG, G. S. & COOPER, T. A. 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet,* 8**,** 749-61.

WARF, M. B. & BERGLUND, J. A. 2010. Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem Sci,* 35**,** 169-78.

WARZECHA, C. C., JIANG, P., AMIRIKIAN, K., DITTMAR, K. A., LU, H., SHEN, S., GUO, W., XING, Y. & CARSTENS, R. P. 2010a. An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. *EMBO J,* 29**,** 3286-300.

WARZECHA, C. C., JIANG, P., AMIRIKIAN, K., DITTMAR, K. A., LU, H. Z., SHEN, S. H., GUO, W., XING, Y. & CARSTENS, R. P. 2010b. An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. *Embo Journal,* 29**,** 3286-3300.

WILHELM, M., SCHLEGL, J., HAHNE, H., GHOLAMI, A. M., LIEBERENZ, M., SAVITSKI, M. M., ZIEGLER, E., BUTZMANN, L., GESSULAT, S., MARX, H., MATHIESON, T., LEMEER, S., SCHNATBAUM, K., REIMER, U., WENSCHUH, H., MOLLENHAUER, M., SLOTTA-HUSPENINA, J., BOESE, J. H., BANTSCHEFF, M., GERSTMAIR, A., FAERBER, F. & KUSTER, B. 2014. Mass-spectrometry-based draft of the human proteome. *Nature,* 509**,** 582-7.

WILL, C. L. & LUHRMANN, R. 2011. Spliceosome structure and function. *Cold Spring Harb Perspect Biol,* 3.

XUE, Y., ZHOU, Y., WU, T., ZHU, T., JI, X., KWON, Y. S., ZHANG, C., YEO, G., BLACK, D. L., SUN, H., FU, X. D. & ZHANG, Y. 2009. Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol Cell,* 36**,** 996-1006.

YU, Y. T. & STEITZ, J. A. 1997. Site-specific crosslinking of mammalian U11 and u6atac to the 5' splice site of an AT-AC intron. *Proc Natl Acad Sci U S A,* 94**,** 6030-5.