An Investigation Into The Function and Regulation Of ERG Exon 7b

Samantha Jumbe

A thesis submitted in partial fulfilment of the requirements of the University of the West of England, Bristol for the degree of Doctor of Philosophy

Department of Applied Science,

Faculty of Health and Applied Sciences, University of the West of England

June 2019

ABSTRACT

The ETS family transcription factor ERG is a key oncoprotein in bone, blood, vascular and most notably prostate cancer where it is activated in at least 50% of cases. Of the several splice isoforms of ERG with variable biological activity those that include the cassette exon 7b are associated with aggressiveness and progression of disease in prostate cancer. Inclusion of exon 7b adds 24 amino-acids in frame to the central 'alternative domain' which also contains binding sites for other transcriptional regulators.

Alignment of the amino acids of exon 7b showed it is evolutionary conserved in echinoderms emphasizing its functional importance. Splice switching oligonucleotides (SSO) targeting the splice sites for exon 7b were designed to induce exon 7b skipping. Successful SSO-induced skipping of exon 7b in the osteosarcoma MG63 cell line resulted in decreased cell migration, invasion and proliferation and increased apoptosis *in vitro.* ERG was shown to bind to the promoter of tissue non-specific alkaline phosphatase, a marker of cell differentiation, and SSO-induced skipping of exon 7b attenuated its expression. Moreover several splicing regulatory elements and proteins were identified using bioinformatics prediction methods and an RNA pull down of the 3´ splice site of exon 7b identified several potential splicing regulatory proteins for this exon.

This study provides evidence that the inclusion of exon 7b, which encodes the full length transactivation domain enhances the oncogenic activity of ERG *in vitro*. Further refinement of *in vivo* experimentation is required but if successful SSOs that target oncogenes could potentially be developed as therapeutic agents. The study also highlights the requirement to understand splicing

regulation of disease associated splicing events in ETS transcription factors and has provided pilot data for further study of *ERG* exon 7b splicing regulation.

For those I love

ACKNOWLEDGEMENTS

My sincere gratitude and appreciation go to my supervisors Dr Michael Ladomery, Dr Ian Wilson and Dr Bahareh Vahabi. Mike your tireless commitment and encouragement have been spurred me along and allowed me to develop my confidence and capabilities as a scientist. Thank you for your always open door, never ending time and patience and mostly for being there for me in the most difficult time. Ian for help with techniques and statistics and helping me remember that science is a creative discipline. For a listening ear and fresh perspectives - Bahareh thank you.

I would like to acknowledge and thank my collaborators. Dr Sebastian Oltean (University of Exeter) and Dr Sean Porazinski (Ladomery Lab) for the xenograft work, Dr Lee Spraggon for advice on design and testing of SSOs and the RNAomics Platform team at the University of Sherbrooke, Canada for the high-throughput PCR analysis.

I would like to thank the Malawi Government Scholarship Fund for funding this research and the University of the West of England for providing the research facilities to conduct the research in. Thank you to Kayla Friedman and Malcolm Morgan of the Centre for Sustainable Development, University of Cambridge, UK for producing the Microsoft Word thesis template used to produce this document.

Thanks go to the superhero technical team of the CRIB labs with special thanks to David Corry and Dr Jeff Davey. All the fantastic members of Ladomery Lab and CRIB- thank you for the laughs, ideas and inspiration.

So many friends who are like family – thank you for your prayers and so much more. My parents, sisters, brothers, nieces and nephew who are my most precious possession and have provided unsurpassable sustenance and wisdom. My niece Nina for always checking up on how my work on finding a cure for cancer was going (it's still a work in progress). I could not be without all of your remarkable love and belief in me. A special mention for my late uncle Luke, and other men like him who fought prostate cancer. Your braveness was my inspiration to embark on this journey.

"To Him who is able to do immeasurably more than all we ask or imagine, according to his power that is at work within us"

Ephesians 3:20 (NIV)

ABBREVIATIONS AND ACRONYMS

1,25D calcitriol the active metabolite of vitamin D3

AMKL acute megakaryoblastic leukaemia

AML acute myeloid leukaemia

AP1 activator protein 1

AS alternative splicing

ASO antisense oligonucleotide

ASPCR alternative splicing using endpoint PCR coupled with microcapillary electrophoresis

BCL-X B-cell lymphoma-extra-large

BPS branch point site

BRAF B-Raf Proto-Oncogene, Serine/Threonine Kinase

BRCA1 breast cancer 1

CDK cyclin dependent kinase

cDNA complementary DNA

CELF CUGBP ELAV-like family member

ChIP chromatin immunoprecipitation

CLK cdc2-like kinase

c-Myc MYC Proto-Oncogene, BHLH Transcription Factor

CRISPR-Cas 9 Clustered Regularly Interspaced Short Palindromic Repeats CRISPR associated protein 9

CXCL12 C-X-C motif chemokine ligand 12

CXCR4 C-X-C motif chemokine receptor 4

DAPI 4′,6-diamidino-2-phenylindole

DEF docking site for ERK

DMD Duchenne muscular dystrophy

DNA deoxyribonucleic acid

DSCAM Downs syndrome cell adhesion molecule

ELAVL1 ELAV Like RNA Binding Protein 1

EMT epithelial-mesenchymal transition

ENCODE The Encyclopedia of DNA Elements

ERG ETS-related gene or v-ets avian erythroblastosis virus E26 oncogene homolog

ERK extracellular signal-regulated kinase

ESE exonic splicing enhancer

ESS exonic splicing silencer

ETS E-26 transformation specific

FAM98B Family With Sequence Similarity 98 Member B

FHBP (3S)1-fluoro-3-hydroxy-4-(oleoyloxy)butyl-1-phosphonate, a synthetic analogue of LPA

FITC Fluorescein isothiocyanate

G4 G-quadruplex

GTEx Genotype-Tissue Expression

HDR homology-directed repair

hnRNP heterogeneous nuclear ribonucleoprotein

ISE intronic splicing enhancer

ISS intronic splicing silencer

kDa kilodalton

LPA lysophosphatidic acid

MAPK mitogen-activated protein kinase

mRNA messenger RNA

MSI2 musashi RNA binding protein 2

ng nanogram

NHEJ non-homologous end joining

nM nanomolar

NRG Neuroglian

Nucleic acid bases A adenosine, **T** thymine, **G** guanine, **C** cytosine and **U** uracil

PCa prostate cancer

PCBP poly (rC) binding protein

p-NP p-nitrophenol

p-NPP p-nitrophenylphosphate

PRAD prostate adenocarcinoma

Pre-mRNA pre-messenger RNA

PS phosphorothioate

PSI percent spliced in

PSM peptide spectrum match

PTBP1 polypyrimidine tract-binding protein 1

qPCR quantitative polymerase chain reaction

RBP RNA binding protein

RNAPII RNA polymerase II

RNA ribonucleic acid

RNA-seq RNA sequencing

RNPS1 RNA binding protein with serine rich domain 1

RPKM reads per kilobase transcript per million

RT-PCR reverse transcriptase polymerase chain reaction

SDS-PAGE sodium dodecyl sulfate polyacrylamide gel electrophoresis

SELEX systematic evolution of ligands by exponential enrichment

SF splice factor

SF3A1 splice factor 3 subunit 1

SF3B4 splice factor 3b subunit 4

SFPQ Splicing factor proline and glutamine rich

siRNA small interfering RNA

Sm smith antigen

SMA Spinal muscular atrophy

smRNA small nuclear RNA

snRNP small nuclear ribonucleoprotein

SNW1 SNW domain containing 1

SR serine-arginine rich proteins

SR-A1 steroid receptor RNA activator 1

SRE splicing regulatory element

SRPK serine-arginine protein kinase

SRSF serine-arginine splice factor

SSO splice switching oligonucleotide

SXL Sex-lethal

T-ALL T-cell acute lymphoblastic leukaemia

TARBP1 transactivation response RNA-binding protein 1

TCGA The Cancer Genome Atlas

TCR T-cell receptor

TNSALP tissue non-specific alkaline phosphatase

UTR untranslated region

VDR vitamin D receptor

Wnt Wingless-related integration site

µM micromolar

CONTENTS

1 Introduction.....21

1.1 Alternative splicing.....22

1.1.1 Prevalence of alternative splicing..... 24

1.1.2 Specificity and complexity of splicing..... 25

1.1.3 Splicing regulatory elements and proteins involved in splicing regulation..... 27

1.1.4 The splicing reaction..... 32

1.1.5 Alternative splicing and cancer..... 38

1.2 E26-transformation specific (ETS) family of transcription factors 41

1.2.1 Transcription factors and gene expression..... 41

1.2.2 Structure and binding specificity of ETS transcription factors..... 43

1.2.3 Expression and biological function of ETS transcription factors... 48

1.3 The ERG transcription factor.....50

1.3.1 Structure and isoforms..... 50

1.3.2 ERG and Cancer..... 53

1.3.3 Function of ERG..... 67

1.4 Aims of this Study.....72

2 Materials and Methods.....73

2.1 Cell Lines.....74

2.1.1 Cell counting and viability assessment 74

2.2 Cell treatments 75

2.3 Designing and using splice switching oligonucleotides.....75

2.4 RNA extraction 76

2.5 cDNA synthesis 77

2.6 Standard PCR and Gel Electrophoresis 78

2.7 Isolating and Sequencing PCR products 79

2.8 Quantitative PCR.....79

2.9 RNAi high-throughput RT-PCR screen.....79

2.10 Protein isolation and preparation 80

2.11 SDS-PAGE and Immunoblotting.....81

2.12 RNA Pull-Down.....81

2.13 siRNA knockdown of splice factors.....82

2.14 Resazurin cell viability assay 83

2.15 Immunofluorescence analysis of Ki-67 84

2.16 Caspase 3/7 assay 85

2.17 Transwell invasion and migration assay 85

2.18 Alkaline Phosphatase assay.....86

2.19 Chromatin immunoprecipitation assay.....87

2.20 Xenograft mouse model and in vivo tumour growth.....88

2.21 Bioinformatic analysis 89

2.21.1 Databases used for expression analysis 89

2.21.2 Gene sequence annotation 89

2.21.3 Sequence alignments 90

2.21.4 Prediction of splicing regulatory elements 90

2.21.5 Splice factor binding prediction 90

2.22 Statistical Analysis 90

3 Analysis of *ERG* exon 7b sequence and expression 92

3.1 Introduction 93

3.1.1 Modification of alternative splicing using splice switching oligonucleotides (SSOs) 98

3.2 *ERG* exon 7b is evolutionarily conserved in *ERG* orthologues 104

3.3 ERG exon 7b expression in prostate cancer 115

3.4 ERG exon 7b is alternatively spliced in several malignant cell lines 117

3.5 ERG exon 7b can be skipped with splice switching oligonucleotides (SSOs) targeted to the 5´ and 3´ splice sites 121

3.6 Summary of findings 125

4 Function of ERG exon 7b splice isoforms 128

4.1 Introduction 129

4.2 SSO-induced ERG exon 7b skipping alters cell growth in MG63 cells 131

4.3 SSO-induced ERG exon 7b skipping alters cell proliferation in MG63 cells 133

4.4 SSO-induced ERG exon 7b skipping induces apoptosis in MG63 cells 139

4.5 SSO-induced ERG exon 7b skipping alters cell migration in MG63 cells 143

4.6 SSO-induced ERG exon 7b skipping alters cell invasion in MG63 cell 145

4.7 ERG exon 7b skipping reduces tumour growth in an MG63 xenograft mouse model.....147

4.8 SSO-induced ERG exon 7b exon skipping alters *TNSALP* activity in MG63 cells in a time-dependent manner.....149

4.9 SSO-induced ERG exon 7b skipping alters the expression of *TNSALP* and other bone maturation genes in MG63 cells.....154

4.10 Summary of findings.....158

5 Regulation of *ERG* exon 7b splicing.....163

5.1 Introduction.....164

5.2 Analysis of ERG exon 7b acceptor and donor splice sites 165

5.3 Prediction of *cis*-acting splicing features that regulate exon 7b splicing 167

5.4 Prediction of potential *trans*-acting splice factors that regulate exon 7b splicing 170

5.5 The knockdown of several RBPs significantly changes *ERG* exon 7b splicing in MCF7 cells.....174

5.6 Several proteins interact with the 3´ splice site of ERG exon 7b exon *in vitro*176

5.7 Knockdown of PTBP1, SRSF5 and hnRNPF did not significantly alter ERG exon 7b splicing.....180

5.8 Summary of findings.....183

6 Discussion and Future Work 186

6.1 Evolutionary conservation of ERG and its exon 7b.....188

6.2 Role of ERG exon 7b in oncogenic processes 190

6.3 Therapeutic potential of SSOs.....195

6.4 Role of exon 7b in osteoblast maturation 197

6.5 Regulation of exon 7b inclusion.....200

6.6 Main conclusions.....203

6.7 Future Work.....205

6.7.1 Further study of the role of exon 7 in conjunction with exon 7b . 205

6.7.2 Further investigation of the regulation of ERG exon 7b regulation 208

6.7.3 The role of ERG in osteogenesis of osteoblasts 209

7 References 211

8 Appendices 261

Sequences of primers used in this study 262

Sequencing results for PCR products in this this study 263

Normalisation and quantification procedure for qPCR 264

PSI and ΔPSI values for high throughput RNAi experiment 268

Proteins identified in mass spectrometry of RNA pull down 270

Summary of splicing regulatory proteins identified in this study 273

Publications arising from this thesis 276

LIST OF TABLES

Table 1 Gene names and locations of the ETS transcription family (adapted from Yates et al., 2017) 45

Table 2 RBPs from mass spectrometry of RNA pull down of exon 7b acceptor site 179

LIST OF FIGURES

Figure 1.1 The splicing reaction.....**Error! Bookmark not defined.**

Figure 1.2 The regulation and patterns of alternative splicing (adapted from Park et al., 2018).....**Error! Bookmark not defined.**

Figure 1.3 Antisense oligonucleotide chemistry and mode of action ..... **Error! Bookmark not defined.**

Figure 1.4 The ETS transcription factor family (adapted from Wei et al., 2010 and Hollenhorst et al., 2011)..... 44

Figure 1.5 *ERG*: gene structure and isoforms..... 52

Figure 3.1 *ERG* expression in human tissues according to the Genotype-Tissue Expression project (GTEx)..... 95

Figure 3.2 *ERG* exon 7b expression in human tissues (GTEx)..... 97

Figure 3.3 Amino acid alignment of human ERG protein isoforms 108

Figure 3.4 Amino acid alignment of ERG in representative vertebrate species 110

Figure 3.5 Amino acid alignment of human ERG and echinoderm ERG isoforms 112

Figure 3.6 Amino acid alignment of ERG in primitive species..... 113

Figure 3.7 ERG exon 7b PSI values in prostate adenocarcinoma patients . 116

Figure 3.8 ERG exon 7b expression in haematological cell lines 118

Figure 3.9 ERG RNA expression in human cell lines..... 120

Figure 3.10 Schematic of ERG exon 7b splicing pattern and SSO sequences used in this study 123

Figure 3.11 Titration of ERG exon 7b SSOs in MOLT4 and MG63 cells 124

Figure 4.1 Growth curve of MG63 cells treated with ERG exon 7b SSO 132

Figure 4.2 Proliferation index of PRAD patients from TCGA database 135

Figure 4.3 Proliferation of MG63 cells following SSO treatment 136

Figure 4.4 Ki-67 expression in MG63 cells following 72 hours of SSO treatment 138

Figure 4.5 Caspase 3/7 expression following 48 hours of SSO treatment in MG63 cells 141

Figure 4.6 Caspase 3/7 expression following 72 hours of SSO treatment in MG63 cells 142

Figure 4.7 ERG exon 7b skipping causes reduction in MG63 cell migration 144

Figure 4.8 ERG exon 7b skipping causes reduction in MG63 cell invasion . 146

Figure 4.9 Growth of MG63 xenograft mouse model tumours treated with ERG exon 7b SSO 148

Figure 4.10 TNSALP expression is attenuated after 48 hours of SSO induced ERG exon 7b skipping in MG63 cells 151

Figure 4.11 TNSALP expression is attenuated after 72 hours of SSO induced ERG exon 7b skipping in MG63 cells 152

Figure 4.12 TNSALP expression is attenuated after 96 hours of SSO induced ERG exon 7b skipping in MG63 cells 153

Figure 4.13 Expression of TNSALP, OPN and RUNX2 genes after 48h of 1,25D and FHBP co-treatment in the presence of SSOs. 156

Figure 4.14 ERG binds to the promoter of *TNSALP* 157

Figure 5.1 *ERG* exon 7b splice sites compared across metazoan species.. 166

Figure 5.2 ESE prediction analysis for *ERG* exon 7b 169

Figure 5.3 RBPs predicted to bind to ERG exon 7b and its flanking introns 172

Figure 5.4 Predicted SF and RBP regulators of *ERG* exon 7b alternative splicing 173

Figure 5.5 RNAi knockdown of RBPs change ERG exon 7b inclusion in MCF-7 cells 175

Figure 5.6 RNA pull down of ERG exon 7b acceptor site and surrounding sequence 178

Figure 5.7 RNAi knockdown of SRSF5, hnRNPF and PTBP1 does not change ERG exon 7b alternative splicing..... 182

LIST OF APPENDICES

Sequences of primers used in this study 262

Sequencing results for PCR products in this this study 263

Normalisation and quantification procedure for qPCR 264

PSI and ΔPSI values for high throughput RNAi experiment..... 268

Proteins identified in mass spectrometry of RNA pull down..... 270

Summary of splicing regulatory proteins identified in this study..... 273

Publications arising from this thesis..... 276

1 Introduction

1.1 Alternative splicing

The information of almost all living organisms is stored as DNA (deoxyribonucleic acid), a very stable molecule. However, the expression of the genetic information held in DNA requires the synthesis (transcription) of a more unstable and dynamic molecule called RNA (ribonucleic acid). Similar to DNA in structure, RNA has a different base (uracil, U in place of thymine, T) and sugar backbone where the hydroxyl group is in the 2´ position making RNA more chemically reactive (Holley *et al.*, 1965). DNA is transcribed into RNA and this can then be translated into proteins where necessary. This fascinating mechanism of molecular biology is the means by which our genetic information is expressed. We now know that one gene can produce several proteins and this is because RNA is not a straightforward copy of the DNA but rather it undergoes processing allowing for the production of different RNA molecules from the one gene.

The Nobel Prize winning discovery was made independently by Richard Roberts and Phillip Sharp in 1977. Adenovirus, responsible for causing many a common cold, was used as a model as it could be grown rapidly with ease and had a relatively short, simple genome that could be easily isolated. By comparing the Adenovirus *hexon* gene to its RNA molecule the experiment revealed that genes were interrupted by segments of DNA that were not present in the messenger RNA (mRNA) molecule (Berget, Moore and Sharp, 1977; Chow *et al.*, 1977) something that was contrary to the thinking at the time. We now call these interrupting sequences introns and the parts of the gene that generally end up in mRNA are called exons. Exons are joined, or ‘spliced’

together very precisely. However, further research revealed that alternative RNA transcripts can be made from a single processed primary transcript (Berget *et al.*, 1978). This process of defining which sequences constitute the final RNA transcript is alternative splicing (AS).

A mature mRNA molecule is made up of a protein coding sequence (open reading frame) flanked by a 5´ untranslated region (UTR) and 3´ UTR. The start of the coding region is demarcated by AUG or sometimes CUG (Lobanov *et al.*, 2010). AS not only affects the coding potential of RNA, it also affects the 5´ UTR, 3´ UTR, adjusting the rates of translation, stability and localisation of mRNA molecules. During transcription, other pre-mRNA processing events occur which are connected to splicing. The formation of the 5´ cap promotes the recruitment of the splicing machinery, and the process of polyadenylation is also closely linked to the splicing of the final intron (Schwer and Shuman, 1996). Each exon-exon junction has exon junction complexes placed on them during splicing which support RNA export and enhance mRNA translation (Luo *et al.*, 2001). Many years after the discovery of AS not only are we beginning to comprehend how essential it is to gene expression but also how understanding this process can aid us when determining the causes of disease as well as establishing novel contexts for potential therapies.

1.1.1 Prevalence of alternative splicing

Upon the sequencing of the human genome it became apparent that the same number of genes present does not necessarily correlate with the evolutionary hierarchy of an organism. For example, *Arabidopsis thaliana* plants have an estimated 25,000 protein coding genes comparable to the estimated 20-24,000 genes in humans. Human genes, however, are characterised by much greater alternative splicing. Of the more than 20,000 genes in the human genome over 92% undergo AS (E. T. Wang *et al.*, 2008). The highly flexible nature of AS; the selective combination of exons means that tens, hundreds or even thousands of isoforms can be produced from one gene. In comparison with fruit fly, mice or worms, humans were found to have the highest average number of mRNA isoforms per gene and the other organisms genomes did not undergo AS to nearly the same extent as humans (Lee and Donald C. Rio, 2015). Interestingly only 1.2% of the genomic sequence contributes to exonic sequence (Venter *et al.*, 2001) further showing how alternative splicing is an important mechanism for creating proteomic complexity from a relatively limited protein coding genome. It also partially alludes that AS, as a means of the regulating gene expression, contributes to the specificity, depth and extent of the proteome of an organism.

1.1.2 Specificity and complexity of splicing

The expression of isoforms is dependent on the conditions in and around the cell. This means some isoforms require external cues to be expressed or are only expressed during a certain developmental stage or in a specific tissue type. A number of Drosophila genes have been revealed to be tissue specific. The determination of sex requires the presence of Sex-lethal (SXL) and other slice factors (Penalva and Sánchez, 2003). ELAV-like protein 1 (ELAVL1) is a neuron-specific RBP that was identified as a regulator of *Neuroglian* (*NRG*) splicing, being required for the production of the neural specific isoform (Koushika, Lisbin and White, 1996).

Not only are tissue specific RBPs important but the activation of signalling cascades can regulate AS. A signalling cascade is generally activated by a ligand that may bind to an extracellular or transmembrane receptor. The receptor then activates intracellular proteins activating the cascade that may ultimately result in the activation of a nuclear target that alters splicing patterns. An example is the splicing of the *CD44* gene upon activation of the T-cell receptor (TCR). The TCR activates the ERK (extracellular signal-regulated kinase)/MAP (mitogen-activated protein) kinase signalling pathway leading to ERK phosphorylating SRC-associated substrate in mitosis of 68kDa (Sam68) which has RNA binding activity. Phosphorylated Sam68 can then bind to the *CD44* pre-mRNA and promote inclusion of the v5 exon (Cheng and Sharp, 2006).

Similarly, the presence and function of SR proteins can be regulated by serine-arginine protein kinase (SRPK) and cdc2-like kinase (CLK) families which can

be activated as part of more extensive signalling pathways (Jakubauskiene *et al.*, 2015). The phosphorylation of newly synthesised SR proteins by SRPKs is required for them to be transported to the nuclear speckles, a nuclear sub-compartment where they are stored. Subsequently hyperphosphorylation by CLKs releases SR proteins from the speckles and facilitates their recruitment to the transcription sites and pre-mRNA molecules (Naro and Sette, 2013).

1.1.3 Splicing regulatory elements and proteins involved in splicing regulation

There are several modes of AS; exon skipping, mutually exclusive exons splicing, alternative usage of the 3´ and 5´ splice site and intron retention. The most common is exon skipping (Keren et al., 2010, Figure 1.2A). The mode of AS that occurs is influenced by the splicing regulatory elements (SREs) that exist within both intron and exon sequences. These and additional regulatory sequences can be referred to as *cis*-acting regulatory elements. They function to recruit splice factor (SF) proteins that may enhance or silence splicing. Hence the sequences are termed as follows: exonic splicing enhancers (ESEs), intronic splicing enhancers (ISEs), exonic splicing silencers (ESSs) and intronic splicing silencers (ISSs) (Lee and Donald C. Rio, 2015). As with splice site strength, bioinformatic tools have been developed to predict potential SREs and these tools can be informative for experimental validation (Fairbrother *et al.*, 2004; Desmet *et al.*, 2009).

In addition, *trans*-acting splicing factors are involved in the regulation of AS. RNA-binding proteins (RBPs) that are not part of the spliceosome can contribute to regulating splice site selection. These proteins can be classified into three main types; serine-arginine-rich (SR) proteins, heterogeneous nuclear ribonucleoproteins (hnRNPs) or tissue-specific RBPs. Classically SR proteins are splicing activators binding to ESEs and ISEs, whereas hnRNPs bind ISSs and ESSs to repress splicing (Black, 2003; Wang and Burge, 2008).

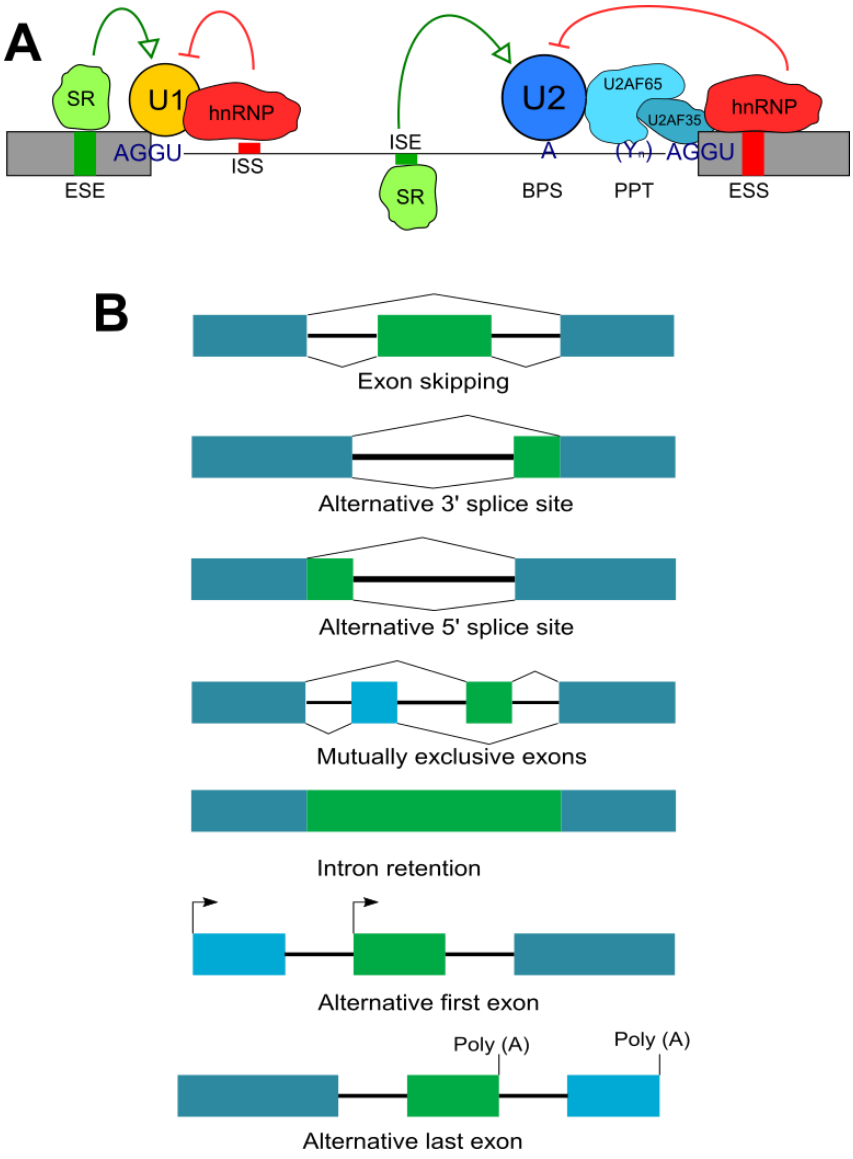


Figure 1.1 The regulation and patterns of alternative splicing (adapted from Park et al., 2018)

A) Alternative splicing is regulated by an extensive protein-RNA interaction network involving cis elements within the pre-mRNA and trans-acting factors that bind to these cis elements and complex. The most essential splicing signals within the pre- mRNA are the 50 splice site (50SS), 30 splice site (30SS), branch site (A), and polypyrimidine tract (Y(n)). The 50 and 30 splice sites have highly conserved GU and AG dinucleotides as the first and last two nucleotides of the intron, respectively. The U1 snRNP complex recognizes the 50 splice site, and the U2 snRNP complex recognizes the branch site. The U2AF proteins recognise the 30 splice site and polypyrimidine tract. Exonic splicing enhancers (ESEs), exonic splicing silencers (ESSs), intronic splicing enhancers (ISEs), and intronic splicing silencers (ISSs) are pre-mRNA cis regulatory motifs that recruit various RNA-binding proteins (e.g., SR and hnRNP proteins) to regulate alternative splicing. **B)** Main patterns of alternative splicing. Dark-blue boxes represent constitutively spliced exons. Light-blue and green boxes represent alternatively spliced exons.

ESEs close to splice site can encourage the binding of the U1 and U2AF snRNPs to their respective splice sites. When SR proteins are bound to an ESE they can recruit other splice factors that interact with the arginine and serine residue rich RS domain on the SR protein (Graveley, 2000). This generally increases exon inclusion. SR-depletion of splicing extracts showed that two SR proteins, SRSF1 and SRSF2 are required for pre-mRNA splicing, splice site selection and splice switching (Fu and Maniatis, 1990; Ge and Manley, 1990; Krainer, Conway and Kozak, 1990).

However, ESEs may have a wider ranging function in pre-mRNA splicing as they are expressed in constitutive exons (ones that are consistently conserved after splicing) as well as alternative ones. ISEs are equally important for enhancing splicing. The terminal exon in the *Calcitonin* gene has a conserved ISE that constitutes part of the sequence of a 5´ cryptic splice site facilitating the recognition of a 3´ splice site required for exon inclusion (Lou *et al.*, 1995). Overall enhancer SREs are able to activate and define both alternative and constitutive exons as well as weak and strong splice site by appropriately recruiting SR proteins and other splice factors.

Silencer SREs such as ISSs and ESSs typically recruit hnRNP proteins. Multimerization of hnRNPs across an exon can repress its inclusion by blocking access for the spliceosome and other auxiliary splicing factors. Polypyrimidine tract-binding protein 1 (PTBP1), also known as hnRNPI, has been characterised as a repressor of splicing often by binding to the polypyrimidine tract or in co-operation with splicing co-factors (Bothwell *et al.*, 1991; Spellman *et al.*, 2005). Systematic evolution of ligands by exponential enrichment (SELEX) is a method

for producing RNA or DNA strands that bind a specific target ligand, and this method has been used to identify silencers. U1 binding to the 5´ss was shown to be altered by the presence of splicing silencers affecting splice site choice (Yu *et al.*, 2008).

Whereas *cis*-acting element tend to either repress or enhance splicing, studies have revealed *trans*-acting proteins are actually able to perform both functions depending on the binding motif sequence and position of the binding site. An example of this is hnRNPL which was initially thought to only repress CD45 exon 4 splicing (House and Lynch, 2006) but has been shown to also enhance exon 4 and 5 inclusion by binding to weak splice sites (Motta-Mena, Heyd and Lynch, 2010). In addition the action of RBPs may involve interactions with snRNPs like U1 during spliceosome assembly and modulate their binding to the ss. PTBP1 represses splicing of the N1 exon in the proto-oncogene c-SRC through direct binding to U1 when it is part of the snRNP complex (Sharma *et al.*, 2011).

1.1.3.1 Splicing, Transcription and RNA processing

The transcription, 5´ capping and 3´ polyadenylation of an RNA molecule are closely linked to its splicing. Transcription is initiated by association of the RNA polymerase II (RNAPII) enzyme with a promoter sequence, where RNAPII and splice factors can interact. It has been suggested that the choice of promoter can change the profile of splice factors that associate with RNAPII (de la Mata *et al.*, 2003). The kinetics of transcription can also regulate splicing as the timing of the production of competing exons, intron length, types of transcriptional activators and presence of RNAPII elongation mutants have all been shown to alter splicing patterns (de la Mata *et al.*, 2003; Hicks *et al.*, 2006). An example of this is the earlier mentioned *Calcitonin* gene ISE which is additionally required for exon 4 recognition and efficient polyadenylation (Lou *et al.*, 1995).

1.1.4 The splicing reaction

The RNA molecule produced after the transcription of protein coding genes is called a pre-messenger RNA (pre-mRNA) and it is made up of introns and exons which are defined by sequence markers and elements (Berget, 1995). Indeed the formation of the spliceosome, the core machinery that carries out the splicing reaction, can only happen with the recognition of these key regulatory elements (Reed, 1996). Although there are over 100 other proteins in the major spliceosome, the binding of small nuclear ribonucleoproteins (snRNPs) is fundamental to the reaction. A snRNP is an RNA-protein complex, and the major spliceosome snRNPs are U1, U2, U4, U5 and U6. The minor spliceosome contains others such as U11 and U12. These snRNPs bind to key RNA elements in the pre-mRNA sequence in a stepwise fashion (Black, 2003).

The boundaries between introns and exons are defined by 3´ and 5´ splice sites (ss) also referred to as the acceptor and donor site respectively. The splice site splice siteare degenerate consensus sequences; YAG/GURAGU for the donor site and NYAG/G for the acceptor site in human genes (Zhang, 1998; Sun and Chasin, 2000). The Y represents a pyrimidine (U or cytosine, C), R is for adenosine (A) or guanine (G) and the / defines the actual splice site. Initially U1 snRNP binds to the 5´ splice site and U5 and U6 follow in subsequent stages. The detection of the 3´ splice site also relies on the recognition of a polypyrimidine tract, a pyrimidine rich sequence in the upstream intron where the U2 auxiliary factor (U2AF) splicing factor binds. U2AF is made up of a 65kDa (U2AF65) and 35kDa (U2AF35) subunit with the former binding to the polypyrimidine tract and the later to the AG of the 3´ss (Zhang, 1998). U2

snRNP binds across the branch point site (BPS) containing a conserved A flanked by a degenerate sequence (YNYUR**A**Y) (Query, Moore and Sharp, 1994).

The conservation of the splice site splice site sequence as well as the length of the polypyrimidine tract determine the strength of the ss. Indeed the stronger the conservation of these sites the more complementary and stronger affinity of the binding of the spliceosome components (Hertel, 2008). Several predictive matrices and tools have been developed to assess splice site splice site strength based on the dependencies of adjacent and distant nucleotides that constitute the splice site splice site (Shapiro and Senapathy, 1987; Yeo and Burge, 2004; Mani *et al.*, 2011). There is a growing role of bioinformatics programs such as Human Splice Finder which have been developed to predict splice sites and how mutations in them may affect splicing (Desmet *et al.*, 2009).

The splicing reaction is a two-step transesterification reaction that converts pre-mRNA to mature mRNA by the removal of introns. The reaction occurs in the nucleus and requires the assembly of various complexes at each stage to remove the intron and ligate the two exons remaining. The 2-hydroxyl group of the BPS attacks the phosphate of the phosphodiester bond at the 5´ splice site and forms a 2´-5´ phosphodiester bond between the BPS and 5´ splice site in the first step of the transesterification reaction. Next the 3-hydroxyl group of the upstream exon, which has been created in the first step, attacks the 3´ splice site. This ligates the 5´ and 3´ exons and the free intron is released as a lariat.

In order to facilitate the splicing reaction, the spliceosome must assemble on the pre-mRNA molecule in a stepwise manner. In mammals this is more

commonly via the U2 dependent spliceosome made up of U1, U2, U5 and U4/U6 snRNPs and other non snRNP proteins. A snRNP is a made up of a small nuclear RNA (snRNA) in close association with so-called smith antigen (Sm) proteins and other particle-specific proteins. There are seven Sm proteins, SmB/B', SmD, SmE, SmF, SmG, SmN, of decreasing magnitude and they are named after Stephanie Smith, the lupus patient in whom they were first discovered (Séraphin, 1995).

The binding of U1 to the 5´ splice site and non-snRNP proteins such as U2AF (auxiliary factor) to the BPs and polypyrimidine tract results in the formation of the initial E complex. Following on the U2 snRNP binds to the BPS and this is the pre-spliceosome or A complex. Next a pre-assembled tri-snRNP of U4/U6.U5 is recruited to make a pre-catalytic B complex that is converted to an activated B complex due to rearrangements caused by changes in RNA-RNA and RNA-protein interactions. This also destabilises the U1 and U4 snRNPs which dissociate from the complex. The first transesterification reaction takes place when catalytic activation occurs upon the recruitment of pre-mRNA-splicing factor ATP-dependent RNA helicase-like protein, Prp2. The reaction leads to the formation of the C complex catalysing the second step of the transesterification reaction. Finally the complex is remodelled and dissociation of the snRNP and associated proteins, which can be recycled occurs (Brow, 2002; Staley & Woolford, 2009, Figure 1.1).

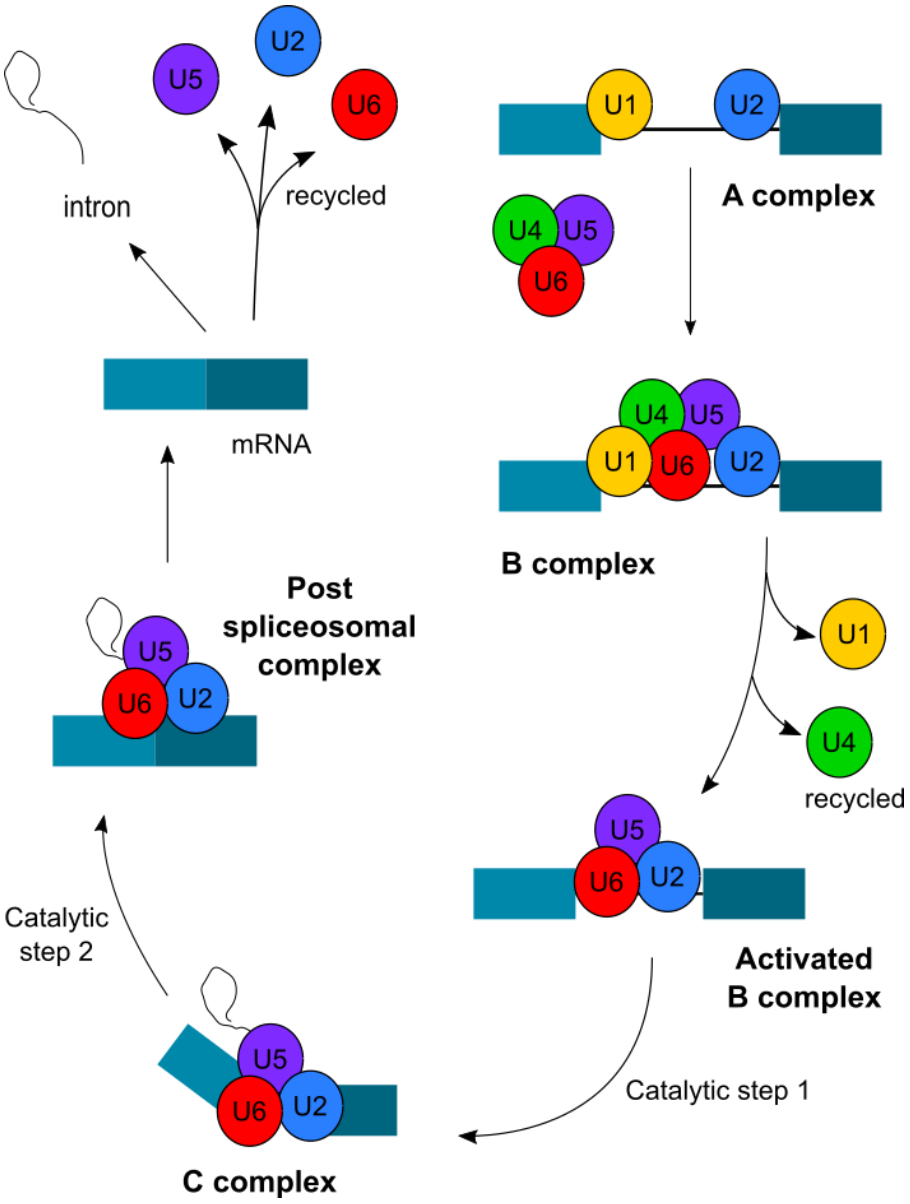


Figure 1.2 The splicing reaction

This schematic represents the splicing cycle. The splicing reaction results in the formations of mature mRNA and requires the removal of introns from pre-mRNA followed by ligation of exons. The sequential steps of the splicing reaction lead to the formation of several complexes composed of the pre-mRNA, the five snRNPs (U1, U2, U5, and U4/U6) and other spliceosome machinery not shown. A, B and C complexes are shown in this figure but there are more. U5 and U2/U6 constitute the catalytic core and catalyse the two nucleophilic attacks (catalytic steps). The reaction commences with the binding of the U1 snRNP to the 5′ splice site and U2 snRNP to the branch site and 3´ splice site, forming the A complex. The B complex forms by addition of the U4/U6.U5 tri-snRNP. Catalytic activation of B complex requires the departure of U1 and U4 and forms activated B complex. The two catalytic reaction happen with 5′ splice site cleavage leading to the formation of the C complex and subsequent exon ligation. The spliceosome disassembles, releasing the intron lariat and mature mRNA as well as recycling the snRNPs for subsequent reactions.

1.1.4.1 Secondary RNA structure, chromatin modification and splicing

Changes in RNA-RNA binding caused by secondary structure and chromatin changes affect AS mechanisms. Even though RNA is single stranded it is able to form secondary and tertiary structures as a result of interactions between many hundreds of nucleotides within the molecule. The effect of the structure can be both enhancement and repression of splicing as SREs or splice site may be concealed.

An example of this is the splicing regulation of the exon 6 cluster in the *Drosophila Downs syndrome cell adhesion molecule* (*DSCAM*) gene which has over 38,000 alternative mRNA transcripts. The RNA molecule has a binding site and selector sequence in the exon 6 region. Each of the 48 mutually exclusive exons of the exon 6 cluster has a unique selector sequence and this can base pair with the docking site forming a secondary structure of 1,000-14,000 nucleotides in size. This interaction is responsible for driving the mutually exclusive exon pairing in the gene (Graveley, 2005).

G-quadruplexes (G4) are G-rich four-stranded quadruplex structures that can form in RNA and these have been implicated in the splicing regulation of *B-cell lymphoma-extra-large* (*BCL-X*) gene. The gene has pro-and anti-apoptotic isoforms that can be generated through the use of alternative 5´ ss. A G4 has been located next to each of these alternative splice site and the administration of G4 targeting ligands switched splicing in BCL-X (Weldon *et al.*, 2018).

Chromatin remodelling can alter splicing either through phosphorylation of RNAPII or recruitment of splicing factors influenced by chromatin binding proteins. Furthermore, modification to the nucleosome can influence exon

definition either by histone methylation or increase in nucleosome occupancy. Indeed in some cases alternative exon are less methylated and this can lead to splice switching (Tilgner *et al.*, 2009; Luco *et al.*, 2010). The discovery that SRSF1 and SRSF3 are associated with interphase chromatin via histone-modification highlights the link between splicing regulation and chromatin modifications (Loomis *et al.*, 2009).

1.1.5 Alternative splicing and cancer

We know that many human diseases have a genetic origin or contribution, therefore the process of AS which is so crucial to gene expression is an obvious process that may be a target of some of these mutations. Not only splice site but also SREs and other *cis*-acting elements have been shown to have disease associated mutations in them. Changes in these sequences can prevent splicing, recruit a different profile of *trans*-acting splicing regulators or activate cryptic splice sites. Furthermore, mutation in the genes that encode for the splicing machinery, RBPs and even transcriptional proteins all have potential to change AS patterns and contribute to disease. However in the case of *trans*-acting splicing factors the effects of mutations may be masked by functional redundancy. It has been evidenced that splice factors from the same family may have similar functions at the same motif (Barberan-Soler *et al.*, 2011) or in the same tissue (Kuroyanagi *et al.*, 2007). Therefore loss of function in a splice factor can be rescued by another splicing factor in the same family and is why there is a low prevalence of disease as a result of such mutations.

Specific for cancer, several AS events have been identified and are being investigated to understand their role in pathogenesis and potential as a biomarker or therapeutic target. These changes often result in changes in expression of oncogenes and tumour suppressor genes. Oncogenes generally promote cell growth whereas tumour suppressor genes reduce growth rate or activate cell death. In cancer, mutations in these genes contribute to uncontrolled cell death either through activation or lack of control of the process. Proliferation, apoptosis, invasive ability, angiogenesis and evasion of

proliferation control and the immune system, common traits of cancerous cells, all have AS events associated with them that lead to aberrant gene expression.

Mutation of *cis*-elements occurs less commonly in contrast to *trans*-acting proteins which are often mutated in several cancer types. For example, in malignant melanoma patients more than half express a V600E mutation in the B-Raf proto-oncogene, serine/threonine kinase (BRAF) oncogene which is a component of the MAPK/ERK signalling cascade (Zhang and Manley, 2013).

Perhaps the best known angiogenesis gene is vascular endothelial growth factor (VEGF) which has both angiogenic and anti-angiogenic isoforms. These isoforms can be produced via differential inclusion of exons (for example exon 8) as well as alternative use of 3´ ss. Further to this the splicing of VEGF is reliant on control by splice factors such as SRSF1. Protein kinase C signalling induced by insulin like growth factor 1 activates SR protein-specific kinase 1 (SRPK1) which phosphorylates SRSF1. SRSF1 favours the splicing of the pro-angiogenic VEGF isoform (Nowak *et al.*, 2010). The anti-angiogenic VEGF-b is supressed by *Wilms tumour suppressor* (*WT1*) gene leading to the irregular angiogenesis and the production of Wilms tumours. The promoter of SRPK1 is directly bound by WT1, and mutant WT1 isoforms enhance SRPK1 expression causing hyperphosphorylation of SRSF1 and thus increasing VEGF expression alongside VEGF-b reduction (Amin *et al.*, 2011).

However aberrant alternative splicing can be an advantage to cancer cells as the protein produced may support disease development and progression. In the case of VEGF switching from the VEGF-165b to the VEGF 165 splice isoform increases angiogenesis, giving a tumour survival and growth advantage

(Woolard *et al.*, 2004). Identifying which splicing events produce these advantageous proteins is important but equally understanding how a splicing event is regulated can provide a potential diagnostic marker or therapeutic target (Le *et al.*, 2015). A comprehensive study, using high throughput reverse transcription PCR, of 600 cancer-associated genes found 41 alternative splicing events that were differentially expressed in breast cancer tissues compared to normal tissues. These changes were associated with increased cell survival and proliferation and some events could be stratified according to tumour grade meaning they had potential as biomarkers (Venables *et al.*, 2008).

Mutation in cell cycle control proteins can also contribute to malignant transformation due to aberrant cyclin-dependent kinase activity. One such example is cyclin D1 which can be alternatively spliced to produce two splice isoforms; cyclin D1a and cyclin D1b, in which intron 4 is retained and produces a truncated protein due the presence of a premature stop codon. Cyclin D1b is upregulated in prostate cancer (PCa) and its expression is influenced by two RBPs, Sam68 and SRSF1 (Olshavsky *et al.*, 2010; Paronetto *et al.*, 2010).

1.2 E26-transformation specific (ETS) family of transcription factors

1.2.1 Transcription factors and gene expression

Gene expression relies on the process of DNA transcription, in which DNA binding *trans*-factors better known as transcription factors play a regulatory role. By binding to DNA *cis*-elements such as enhancers (sequences that promote gene transcription), transcription factors can recruit co-factors and the transcriptional enzyme RNAPII and regulate the expression of the target gene (Ihn Lee and Young, 2013). A co-factor may activate or repress gene expression depending on whether a transcription factor regulates initiation or elongation of genes (Lelli, Slattery and Mann, 2012). In the case of transcription factors that regulate initiation co-activators are recruited with the P300 mediator complex (Malik and Roeder, 2010). Once RNAPII has begun transcribing the gene it will come to a pause 50 base pairs in, and whether it continues into elongation or terminates can be controlled by transcription factors (Adelman and Lis, 2012). The c-Myc transcription factor can stimulate positive transcription elongation factor b which restarts paused RNAPII via phosphorylation (Rahl *et al.*, 2010; Luo *et al.*, 2012).

Furthermore, gene expression is also controlled by the conformation of the chromatin which is itself regulated by protein complexes and histone modifications. Chromatin may be in a configuration specifically intended to repress expression of genes. For example, the polycomb complex which includes histones such as H3K27me3, associates with genes that are only activated at specific stages of differentiation and development (Orkin and

Hochedlinger, 2011). Remodelling of the chromatin may be required to allow for gene activation and this can involve enzymes such as histone methylases, acetylases (Bannister and Kouzarides, 2011) and remodelling complexes such as the SWItch/Sucrose Non-Fermentable (SWI/SNF) family to allow transcription factors to bind their intended DNA target (Hargreaves and Crabtree, 2011). The crosstalk between chromatin structure and transcription factors adds an additional layer of complexity to the regulation of transcription.

A gene can have a number of promoter elements with varying proximity to its target gene that can be bound by various transcription factors (Krivega and Dean, 2012). Each transcription factor can recognise and preferentially bind a set of binding motifs present in the promoter elements of a gene (Inukai, Kock and Bulyk, 2017). Moreover, every specific cell type will have a specific gene expression program dependent on the position, profile and number of transcription factors bound to the promoter and DNA sequence of a set of genes.

The basic domain structure of most transcription factors includes a DNA binding domain and additional domains that interact with co-activators or co-repressors (Bhagwat and Vakoc, 2015). Consequently, dysregulation of transcription factors can manifest both via improper DNA binding and altered protein interactions. Subsequently downstream signalling pathways may also alter transcription factor expression or may in themselves be disrupted by improper gene expression driven by a mutated transcription factor. Therefore, mutations in promoter sequences, co-factors and the genes that encode transcription

factors themselves can all contribute to the aberrant gene expression seen in diseases especially cancer.

1.2.2 Structure and binding specificity of ETS transcription factors

It must be acknowledged that many of the recognised oncogenes are actually transcription factors (Introna and Golay, 1999). The ETS transcription factor family is an example of oncogenic transcription factors involved in the pathogenesis of a number of human cancers. The ETS gene family is thought to have arisen from an ancestral gene duplication that pre-dates the emergence of chordates. Thus ETS members are present in both vertebrates and invertebrates (Seth *et al.*, 1992) and found in species across the metazoan phyla (Degnan *et al.*, 1993).

ETS-1 (E26-transformation specific 1) was the first member to be discovered and since then 30 members have been discovered in various species (Degnan *et al.*, 1993, Table 1). Classification of ETS proteins into subfamilies is also based on the conservation and homology in the C terminal region that contains the conserved and crucial DNA-binding ETS domain in addition to nuclear localisation sequences (Seth *et al.*, 1992; Laudet *et al.*, 1999). There are 28 human ETS family genes organised into 12 subfamilies and four classes (Class I-IV) based on the base at the +4 position of the consensus binding sequence for the subfamilies (Figure 1.3). In the case of class II members ETV6 and ETV7 have a preference for binding to A rather than T at this position, and therefore this class is divided into class IIa and class IIb. The subfamilies are *ELG*, *PDEF*, *SPI*, *TEL*, *ESE*, *ELF*, *PEA3*, *TCF*, *ETS*, *ERF*, *ERG* and *GABPA* (Figure 1.1A, Table 1, Hollenhorst et al., 2011).

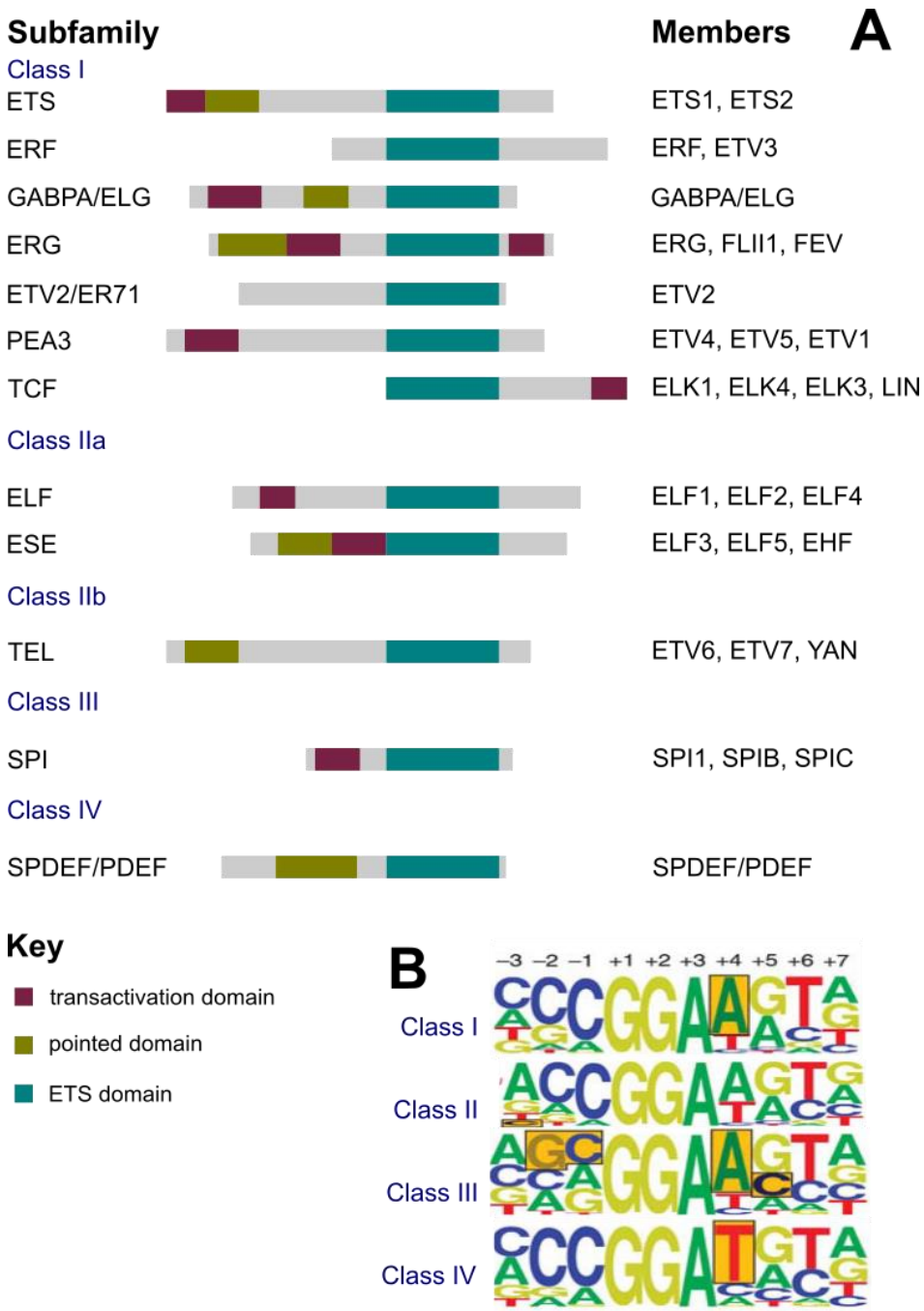


Figure 1.3 The ETS transcription factor family (adapted from Wei *et al.*, 2010 and Hollenhorst *et al.*, 2011)

A. There are 12 subfamilies in the ETS family in humans divided into four classes (listed on the left). A schematic representative of each one is shown with the position of the ETS domain (teal), and if present, the pointed domain (gold) and transactivation domain (purple) mapped. The members of each subfamily are listed on the right. **B.** The consensus binding sequence for each class of ETS transcription factor is defined by the base at position +4.

Gene Symbol	Full Name	Chromosome location
EHF	ETS homologous factor	11p13
ELF1	E74 like ETS transcription factor 1	13q14.11
ELF2	E74 like ETS transcription factor 2	4q31.1
ELF3	E74 like ETS transcription factor 3	1q32.1
ELF4	E74 like ETS transcription factor 4	Xq26.1
ELF5	E74 like ETS transcription factor 5	11p13
ELK1	ETS-like transcription factor 1	Xp11.23
ELK3	ETS-like transcription factor 3	12q23.1
ELK4	ETS-like transcription factor 4	1q32.1
ERF	ETS2 repressor factor	19q13.2
ERG	ETS related gene	21q22.2
ETS1	v-ets erythroblastosis virus E26 oncogene homolog 1	11q24.3
ETS2	v-ets erythroblastosis virus E26 oncogene homolog 2	21q22.2
ETV1	ETS variant 1	7p21.2
ETV2	ETS variant 2	19q13.12
ETV3	ETS variant 3	1q23.1
ETV3L	ETS variant 3 like	1q23.1
ETV4	ETS variant 4	17q21.31
ETV5	ETS variant 5	3q27.2
ETV6	ETS variant 6	12p13.2
ETV7	ETS variant 7	6p21.31
FEV	Fifth Ewing variant	2q35
FLI1	Friend leukaemia virus integration 1	11q24.3
GABPA	GA binding protein transcription factor subunit alpha	21q21.3
SPDEF	SAM pointed domain-containing ETS transcription factor	6p21.31
SPIB	Spleen focus forming virus proviral integration B	19q13.33
SPIC	Spleen focus forming virus proviral integration C	12q23.2
SPI1	Spleen focus forming virus proviral integration	11p11.2

Table 1 Gene names and locations of the ETS transcription family (adapted from Yates et al., 2017)

The gene symbol, full gene name and chromosome location of for the 28 members of the ETS transcription factor. The names and chromosome locations used are according to HUGO Gene Nomenclature Committee (HGNC) classification in the HGNC database - data accessed 18th August 2018.

The structure of the ETS domain has been resolved using high resolution X-ray crystallography (Suwa *et al.*, 2008) as well as NMR (Liang *et al.*, 1994; Werner *et al.*, 1997) and reveals an 85 amino acid helix-loop-helix structure of three alpha helices and four antiparallel beta sheets. The core recognition motif is GGA (A/T) (Karim *et al.*, 1990). The sequence that flanks this core motif determines the binding specificity of each ETS transcription factor allowing for activation of specific gene targets (Seth *et al.*, 1992; Donaldson *et al.*, 1996; Shore *et al.*, 1996). The two G residues interact with two conserved arginine residues found in the third alpha helix and a conserved tyrosine that interacts with the A base is also found in the same helix (Cooper *et al.*, 2015).

The class of an ETS transcription factor is based on the specificity of the binding motif sequence (Wei *et al.*, 2010). Class I which includes the ERG, TCF and ETS subfamilies is the largest and have a consensus sequence of ACCGGAAGT which was determined through *in vitro* binding experiments (Wang *et al.*, 1992; Wei *et al.*, 2010). The next largest group is Class II (TEL, ESE and ELF subfamilies) and the consensus sequence for these ETS proteins is CCCGGAAGT (Wang *et al.*, 1992). However as ETV6 and ETV7 have slightly different specificity to other class II members it has been suggested that class II be divided into class IIa and IIb (Wei *et al.*, 2010). Class III and Class IV contain one subfamily each, SPI and SPDEF respectively. Adenosine rich sequences upstream of the core motif are the preference of class III ETS transcription factors (Ray-Gallet *et al.*, 1995) whereas class IV favour a different core motif of GGAT(Oettgen *et al.*, 2000).

Some ETS proteins have a pointed domain which has a 4 or 5 alpha helical structure (Mackereth *et al.*, 2004). The domain has variable roles, some unique to each ETS protein, including protein-protein interactions that co-regulate transcriptional control and kinase interactions (Lacronique *et al.*, 1997; Seidel and Graves, 2002). Indeed dimerization of ETS proteins can be facilitated by both the ETS and POINTED domains, either as homo- or heterodimers (Sharrocks *et al.*, 1997; Carrère *et al.*, 1998). Other specific regions in transcription factors include nuclear localisation domains (Hoesel *et al.*, 2016), transcriptional activation domains (Siddique *et al.*, 1993; Vlaeminck-Guillem *et al.*, 2003), kinase phosphorylation sites (Seth *et al.*, 1992; Slupsky *et al.*, 1998; Seidel and Graves, 2002) and a B-box domain a conserved amino acid region which is facilitates interactions with serum response factors (Ling *et al.*, 1997).

1.2.3 Expression and biological function of ETS transcription factors

Generally ETS transcription factors are expressed in all tissues (Hollenhorst, Jones and Graves, 2004) and function as either transcriptional activators or repressors (Sharrocks *et al.*, 1997). It is estimated that ETS binding sites are present in 25% of human promoters and although not all of them are necessarily functional, this proposes that the ETS family is important for regulating gene expression (Hollenhorst, McIntosh and Graves, 2011). Post-translational modifications of ETS proteins as effectors of signalling cascades modulate the proteins activity and can dictate whether the ETS transcription factor acts as an activator or repressor when it binds a promoter. For example, an ERK/MAPK phosphorylation site in the B box of TCF facilitates the formation of a complex with serum response factor that binds to serum responses elements in the *Fos proto-oncogene, AP-1 transcription factor subunit* (*c-fos*) promoter (Gille, Sharrocks and Shaw, 1992). The signalling pathway may also recruit transcriptional co-regulatory proteins. Auto-inhibition of ETS1 occurs when the protein is phosphorylated but this can be reversed with cooperation with Runt-related transcription factor 1 (RUNX1) which relieves the inhibition at specific sites (Hollenhorst, McIntosh and Graves, 2011).

ETS transcription factors are important to developmental processes such as haematopoiesis and embryogenesis (Sharrocks *et al.*, 1997) as well as angiogenesis and vasculogenesis (Hollenhorst, McIntosh and Graves, 2011; Shah, Birdsey and Randi, 2016). Differentiation, proliferation, apoptosis, migration and cell matrix interactions have all been shown to involve ETS proteins as well (Oikawa and Yamada, 2003; Wei *et al.*, 2010). Some ETS

proteins are expressed exclusively or more highly in specific tissues, such as SP1 which is expressed in macrophages, B cells and myeloid lineage cells (Lloberas *et al.*, 1999).

1.3 The ERG transcription factor

1.3.1 Structure and isoforms

The *ERG* gene (ETS-related gene, also v-ets avian erythroblastosis virus E26 oncogene homolog) was discovered in 1987 while investigating human oncogene homologues that belong to the ETS (erythroblast transformation specific) family of transcription factors. The ETS domain is a highly conserved DNA-binding domain present in ERG and other ETS family members (Watson *et al.*, 1985). The 1987 study screened a human complementary DNA (cDNA) library derived from COLO320 cells (derived from colorectal adenocarcinoma) with a probe comprising the conserved ETS domain from the E26 virus known to cause erythroblastoid leukaemia in chicken (Reddy et al., 1987).

The *ERG* gene is located on chromosome 21 and is 300kb in size. When *ERG* was first discovered two mRNA transcripts were identified; these were named *erg 1* and *erg 2* producing a 41kDa and 52kDa polypeptide respectively (Rao et al. 1987). Two studies have sought to annotate *ERG* isoforms and now it is predicted that there are around 30 isoforms, although not all of them encode functional proteins, with 16 splice isoforms currently characterised (Owczarek et al. 2004; Zammarchi et al. 2013). The main open reading frame produces an mRNA transcript of 18 exons and encodes for a full length protein 486 amino acids long with a mass of 54.6kDa (Zammarchi et al. 2013; Owczarek et al. 2004).

The main isoform generating events in *ERG* are alternative polyadenylation sites, alternative promoter sites and the alternative splicing of two cassette exons; exons 4 and 7b, which are either included or excluded in most *ERG*

splice isoforms. There are also other cassette exons such as exon 7 but these are less commonly spliced (Zammarchi et al. 2013). It is not surprising that *ERG* is so thoroughly alternatively spliced considering its complex biological roles. The existence of so many isoforms can contribute to imparting a certain level of specificity as well as diversity to the transcriptional functions of ERG.

The ERG protein can form homodimers and heterodimers with FLI-1, ETS2 and SPI1 ETS transcription factors and this is aided by the pointed domain (Carrère *et al.*, 1998). In addition to the ETS and pointed domains, ERG also has two transcriptional activation domains - one located in between the pointedand ETS domain (AD) and another in the C-terminus of the protein (cAD). Both the AD and cAD encourage the binding of proteins to increase ERG transcriptional activation (Siddique *et al.*, 1993). Therefore, whereas the role of the ETS domain is to localise ERG, the AD is there to recruit and interact with transcriptional machinery.

Alternative splicing of exon 7b results in the inclusion or skipping of 24 amino acids in the AD and maintains a functional protein as the change is in-frame (Prasad *et al.*, 1994; Hagen *et al.*, 2014; Adamo and Ladomery, 2015). This change in amino acid sequence alters the tertiary structure of ERG and it is therefore possible that the alternative splicing of exon 7b significantly changes the transcriptional activity of ERG. The structure of the AD of ERG in relation to its exon structure is shown in Figure 1.5A.

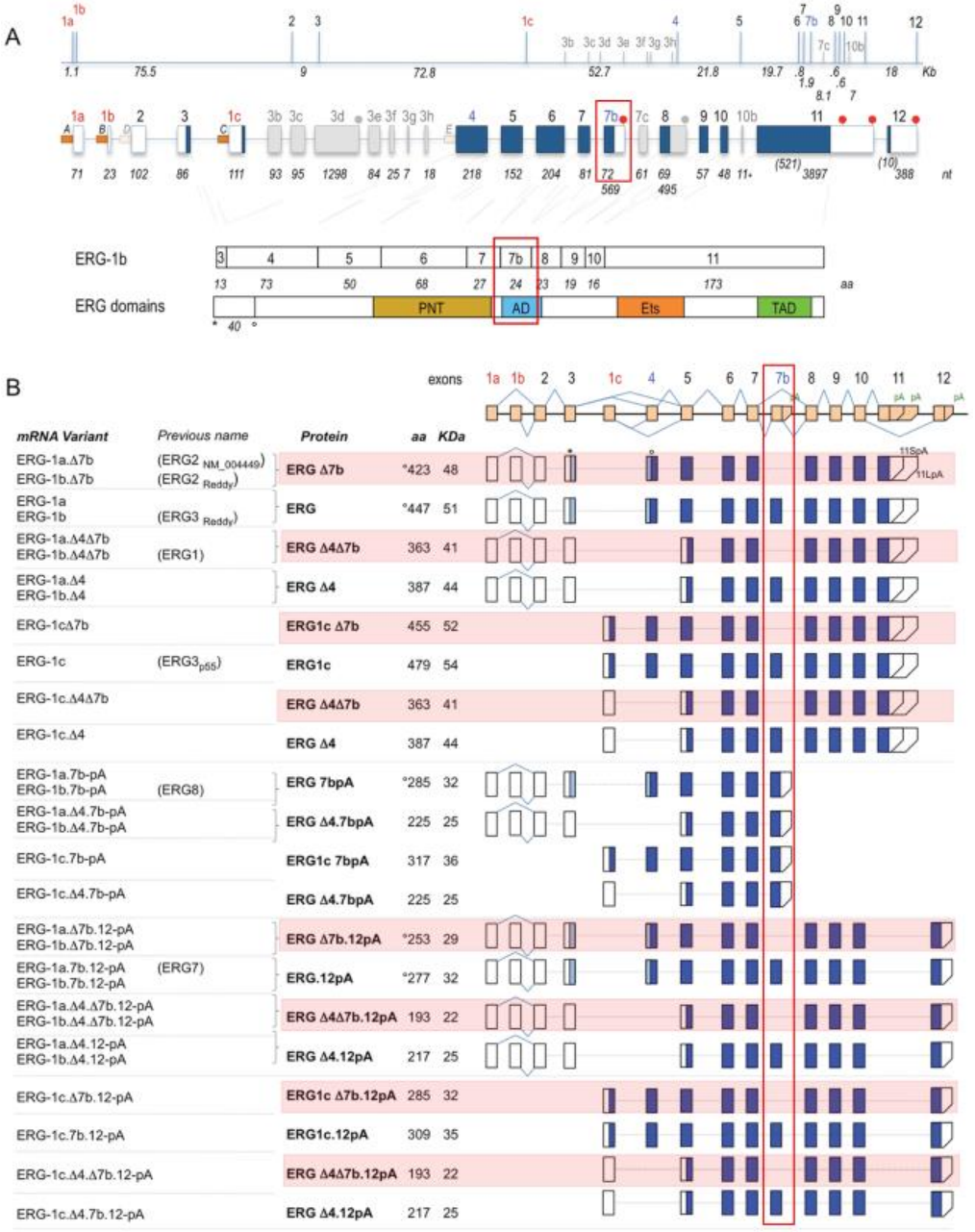


Figure 1.4 *ERG*: gene structure and isoforms (© Zammarchi et al., 2013)

A. Top: The ~300 Kb human *ERG* locus, roughly to scale with intron and exon (bars) sizes in Kb are shown. Red: first exons; blue: common alternative ones and grey uncommon ones. Middle: Exon structure, with exon sizes at the bottom. Blue boxes indicate the main predicted ORFs, white boxes the untranslated regions and grey the uncommon exons. Red circles indicate polyA sites. Bottom: alignment of the exons (7b in the red box) forming the main ORF (ERG-1b) with the protein's domains. Numbers indicate size in amino acids. PNT=pointed domain, AD=alternative domain, Ets=Ets domain, TAD=transactivation domain. Asterisk and circle indicate position of the first and second ATG. **B.** Human *ERG* main variants. The splicing pattern of exon 7b is shown in the red box and isoforms that exclude the exon are highlighted in red. Alignment of exons forming the 30 main RNA variants of human ERG. Blue indicates the ORF, light blue the additional region from the ATG in exon 3. For each variant, the proposed name is indicated next to previous nomenclature (if available).

1.3.2 ERG and Cancer

Cancer can be described as uncontrollable growth of cells and although this is true in part there are other key features that cells must develop to give them oncogenic properties. Loss of senescence, evasion of growth control signals and apoptosis, epithelial-mesenchymal transition (EMT) facilitating cell motility and angiogenesis are all additional attributes of cancerous development. Hanahan and Weinberg (2011) ascribe ten “biological capabilities” (originally six) to the stepwise progressive development of cancer. More recent evidence suggests that cell metabolism can be reprogrammed to better fuel tumorigenesis. Moreover, cells are able to evade detection and destruction by the immune system.

ERG contributes to cancer progression via the transcriptional regulation of its target genes and the profile of ERG target genes can transform with cancer development and progression. Altered signalling pathways can change post-translational modifications of ERG and therefore its transcriptional activity. The alternative splicing of *ERG* can produce an ERG protein with modified activity and this is regulated by a profile of splice factors, some of which may be alternatively spliced themselves in cancer. All of this can change cancer cell biology by excluding, and possibly introducing new ERG target genes.

1.3.2.1 Ewing's sarcoma

Ewing's sarcoma, commonly afflicting children and adolescents, is a group of bone cancers with neurogenic origin and is characterised by undifferentiated small round cells (Ozaki, 2015). It is named after Dr James Ewing who first described it in the early 20th century (Ewing, 1921). When present in the bone the main sites of malignancy are ribs, pelvis and femur however tumours can also develop in other sites such as the back and lower leg. Survival varies but worsens if metastases are present with a slightly reduced estimated 5 year survival of 58% in such cases compared to 68% overall (Ozaki, 2015). The pathogenesis of the malignancy has been linked to the formation of chromosomal rearrangements that result in the production of oncogenic fusion proteins that drive tumorigenic transformation in the tissue (Lessnick and Ladanyi, 2012). As the development of the disease is significantly linked to the chromosomal translocations present they pose a clear potential therapy target. As a result, much effort has been put into identifying and characterising the molecular landscape of Ewing sarcoma.

The presence of ETS gene translocations is critical for the tumorigenesis of Ewing sarcoma and is attributed to the combined gene regulation power of *EWS*, a transcriptional activator, and the DNA binding capability of the ETS domain. The EWS protein is produced from the *EWSR1* gene, a member of the TET protein family. The protein contains RNA binding domains, RGG regions, a zinc finger and in the N terminal region a glutamine-glycine-serine-tyrosine rich transactivation domain. This region is able to interact with splice factors, bind calmodulin preventing protein kinase C mediated phosphorylation and also

interacts with RNAPII and transcription factors (Bertolotti *et al.*, 1998; Cantile *et al.*, 2013). The protein is therefore thought to act as a bridge between gene transcription and splicing. EWS also has important roles as an RNA and nucleic acid binding protein facilitated by the RNA binding domains and the RGG domain that favour polyU and polyG nucleic acid sequences.

Several ETS transcription factors form fusion proteins with EWS, most commonly *FLI-1* (around 85% of occurrences) but *ERG* is the second most common fusion partner (Delattre *et al.*, 1992; Zucman *et al.*, 1993; Sorensen *et al.*, 1994). Detected consistently in Ewing patient tumours, cell lines and mice xenografts, the initial t(11;22) translocation was found in the late 1980s (Turc-Carel *et al.*, 1988). By 1992 *EWS-FLI1* was discovered following cloning of the translocation breakpoints (Zucman *et al.*, 1992). This is not so surprising as it is important to note that *FLI-1* and *ERG* share some significant homology especially in their ETS binding domains as they belong to the same ETS subfamily (Laudet *et al.*, 1999; Hollenhorst, McIntosh and Graves, 2011).

What does vary between the two fusions is the breakpoints at which the two chromosomes combine. The *EWS-ERG* translocation t(21;22) (q22;q12) has four potential breakpoints in both the *EWS* and *ERG* giving several fusion variants. The *EWS* breakpoints maintain the transactivation domain and are located in exons 7-10 (Lin *et al.*, 1999). In *ERG* the breakpoints are in exons 6-9 and all exclude the pointedpointed domain but maintain the functionally important ETS domain (Zucman *et al.*, 1993). In contrast the breakpoints in *FLI-1* are located in exons 4-7, and exclusion of exon 4 removes the pointed domain

(Lin *et al.*, 1999). Fusion variants that have been described for are *EWS-ERG 7/6*, *EWS-ERG 7/8*, *EWS-ERG 7/9* and *EWS-ERG 10/6* (Zucman *et al.*, 1993).

Although the *EWS-ERG* fusion has not been as well characterised as *EWS-FLI*, the oncogenic ability of *EWS-ERG* has been confirmed by several studies. The transcriptional activity of ERG is maintained and in combination with the conserved transactivation domain of EWS, it is thought the fusion protein functions as a dysfunctional transcription factor. The altered activation of downstream target genes contributes to the oncogenic activity of EWS-ERG in Ewing sarcoma. The structural similarities would support the hypothesis that both fusions have very similar functions. Indeed a study confirmed this assumption by comparing *EWS-FLI1* and *EWS-ERG* patients and found no significant differences in their clinical statistics, disease, progression or survival (Ginsberg *et al.*, 1999). The transformative potential of EWS-ERG was further evidenced by the colony forming ability of the fibroblast-derived cell line NIH3T3 although no tumours were made in a xenograft mouse model (Braunreiter *et al.*, 2006).

The fusion protein can bind promoters of genes altering cell signalling pathways that may aid oncogenic transformation and promotion. Reduction of transcription of transforming growth factor β type II receptor (TGF-β RII) abates sensitivity to TGF-β and increases tumorigenesis. TGF-β RII is supressed by EWS-ERG. The cancer microenvironment is reliant on the building and reorganisation of the extracellular matrix. The expression of two extracellular matrix proteins COL11A2 and lamin β-3 are regulated by EWS-ERG (Matsui *et al.*, 2003; Irifune *et al.*, 2005). In the case of COL11A2, promoter binding by

EWS-ERG also involves interaction with RNAPII (Matsui *et al.*, 2003). Lamin β-3 encodes for a chain in the lamin 5 glycoprotein which is involved in cell migration, differentiation and adhesion. ASO repression of EWS-ERG resulted in reduced transcription of lamin β-3 providing a possible target that may drive oncogenic cell development.

In addition another member of the TET protein family *FUS* (also called *TLS*) forms a fusion protein with *ERG* (Ichikawa *et al.*, 1994). This protein plays a role in leukaemia specifically acute myeloid leukaemia (AML) (Panagopoulos *et al.*, 1994). The *FUS-ERG* fusion first described in four Ewing tumour patients is exactly the same t(16;21)(p11;q22) translocation found in AML. The *FUS* gene is found of chromosome 16 and the protein has a similar structure to other TET family members with the presences of an RNA binding domain that facilitates regulation of gene transcription and RNA metabolism. The *FUS* breakpoint in exon 7 observed in these initial cases maintain the N terminal transactivation and are fused to exon 8 or 9 of *ERG*. However one case did have a breakpoint part way through exon 6 instead fused to *ERG* exon 9 and it was suggested that this may indicate a distinct oncogenic mechanism for *FUS-ERG* (Shing *et al.*, 2003). A case with an Ewing sarcoma-peripheral neuroectodermal tumour of the kidney had a *FUS-ERG* fusion where exons 1-5 of *FUS* were fused to *ERG* at exon 6 (Berg *et al.*, 2009). Overall the presence of *FUS-ERG* is rare and further attempt to identify it in small blue round cell tumours reported a prevalence of 8.2% of 85 cases assessed (Chen *et al.*, 2016).

Although the presence of *TET-ERG* fusions in Ewing diseases has been established their rarity means not as much classification has been performed.

Certainly the role of specific splice variants has not been established and therefore there is definite scope to increase our knowledge of ERG and ERG fusion as well their respective splice variants in the onset, development and progression of the Ewing family of malignancies.

1.3.2.2 Haematological cancers

A role for ERG in the regulation of normal haematopoiesis is well established, as it was found that homozygous mutant *ERG* mice did not survive past E13.5. Furthermore, one of the characteristics of heterozygous mice was decreased levels of less committed progenitor cells. Eventually ERG was shown to be critical to the regulation of haematopoietic stem cell maintenance (Loughran *et al.*, 2008). This is supported by the observation that endothelial stem cell differentiation is similarly dependent on ERG activity (Nikolova-Krstevski *et al.*, 2009). In addition to its regulation of haematopoiesis *ERG* is an oncogene in several haematological malignancies (Salek-Ardakani *et al.*, 2009; Adamo and Ladomery, 2015).

The aforementioned *FUS-ERG* fusion gene has been found to be expressed in AML and acute megakaryoblastic leukaemia (AMKL) patients and contribute to leukaemogenesis. AMKL is a subtype of AML, with more than 50% of blast cells of megakaryocyte lineage. The *FUS* promoter driven expression of *ERG* is able to lead to the transformation of NIH3T3 cells, increase cell proliferation and disrupt myeloid differentiation (Ichikawa *et al.*, 1994; Yi *et al.*, 1997; Pereira *et al.*, 1998; Pan *et al.*, 2008). Although initially it was thought that the presence of the fusion alone was sufficient to drive the development of leukaemia, it was later established that additional mutational events are required (Warner *et al.*, 2005). Mice with the *EWS-ERG* fusion can develop leukaemia however this mutation has not yet been reported in humans (Codrington *et al.*, 2005).

A recent study revealed that *FUS-ERG* fusion functions as a transcriptional repressor. In the presence of all-trans-retinoic acid, FUS-ERG binding to

promoters and enhancers is diminished. It was further found that the fusion protein interacts with and alters the activity of a heptad of transcription factors consisting of ERG, TAL1/SCL, GATA2, LYL1, LMO2, RUNX1 and FLI1 (Sotoca et al. 2015). Overexpression of ERG is one markers used for the classification of a specific complex AML karyotype (Marcucci et al. 2005). Therefore, the regulation of the transcriptional activity of ERG is an important contributor to AML and possibly other leukaemia.

Another translocation t(X;21)(q25–26;q22) fusing the N-terminus of the ETS transcription factor E74 Like ETS Transcription Factor 4 (*ELF4*) to the C-terminus of *ERG* was reported in an AML patient. Exon 2 of *ELF4* fuses to exon 2 of *ERG* mean the majority of the protein sequence in this fusion is attributed to ERG. Therefore it was suggested that its function would likely be similar to native ERG and as of yet no other AML cases have been reported to harbour this mutation (Moore *et al.*, 2006).

Individuals with trisomy 21 (Downs syndrome) have an increased risk of developing AMKL. This is thought to be in part because of the overexpression of genes such as ERG that play a significant part in megakaryoblast transformation on this chromosome due to its enhanced ploidy (Ng *et al.*, 2015; Tsuzuki, Taguchi and Seto, 2015). In mice ERG overexpression was shown to lead to aggressive AMKL (Salek-Ardakani *et al.*, 2009). This process is thought to involve other transcription factors including GATA1 (Xu *et al.*, 2003; Stankiewicz and Crispino, 2013) which cooperates with ERG to promote colony formation but also possibly RUNX1 (Elagib *et al.*, 2003; Salek-Ardakani *et al.*, 2009). Furthermore ETS2 has previously been reported to also contribute to

trisomy 21 associated AMKL and since different ETS transcription factors can work cooperatively this may be an additional way to enhance megakaryopoiesis and AMKL development (Sumarsono *et al.*, 1996; Stankiewicz and Crispino, 2009).

In addition to this ERG overexpression has been observed in T-cell acute lymphoblastic leukaemia (T-ALL) and was associated with poor relapse free survival (Baldus *et al.*, 2006). This overexpression profile has been linked to the leukaemia-specific and transient use of the +85 enhancer in the *ERG* promoter, and the transcription factor heptad associated with FUS-ERG was shown to bind as well (Thoms *et al.*, 2011). At a protein level several serine phosphorylation sites were found on ERG in both AML and T-ALL cells. However, one site S283 stood out as it was phosphorylated significantly more than other haematopoietic cells and it was shown to promote use of the +85 enhancer and increase cell proliferation. The phosphorylation at S283 is mediated by ERK which activates the RAS/MAPK signalling pathway and required a DEF domain located in exon 7b (Y. Huang *et al.*, 2016).

Not much is known about effect of alternative splicing on *ERG* in leukaemia. Variable use of the first exons (named exons 4 and 5 in this study) between AML and T-ALL patients has been shown with the former preferentially expressing isoforms using both exons and the later exon 5 only. Regulation of isoform expression was associated with CpG islands sites where methylation occurs on the promoter of the *ERG2* isoform particularly in T-ALL. Furthermore the study identified an isoform that lacked of exon 7b (called exon 12) in T-ALL patients, suggesting the use of specific signalling pathways associated with the

change in transactivation of ERG that this splicing event leads to (Bohne *et al.*, 2009).

More recently in a subtype of ALL, B-progenitor ALL (B-ALL) a novel *ERG* isoform was shown to be activated by the binding of the DUX4, a transcription factor commonly rearranged and overexpressed in B-ALL. The novel *ERG* isoform used exon 7b (called exon 6 in this study) as its first exon and inhibited the activity of native ERG (Zhang *et al.*, 2016).This proposes there is an ongoing need to establish a better understanding of the role of ERG in leukaemia and the study of alternative splicing of *ERG* particularly in the context leukaemia merits immediate attention.

1.3.2.3 Prostate cancer

The discovery of the *TMPRSS2-ERG* fusion gene, found in a remarkable proportion i.e. 50% of PCa cases, was reported by Tomlins et al. (2005). The *TMPRSS2-ERG* fusion gene occurs as a consequence of an interstitial deletion or translocation between *ERG* and the transmembrane protease serine 2 (*TMPRSS2*) gene, which is found in three megabases (Mb) upstream of *ERG* and encodes for an androgen responsive serine protease (Tomlins et al., 2005). The mutation combines the promoter of *TMPRSS2* and the coding region of *ERG* leading to the androgen responsive production of an ERG protein (Clark *et al.*, 2007).

The presence of the fusion gene has been associated with a poor prognosis, biochemical recurrence as indicated by levels of the PCa biomarker prostate specific antigen and decreased survival in patients (Wang *et al.*, 2006; Mehra *et al.*, 2007; Hu *et al.*, 2008; J. Wang *et al.*, 2008; Yoshimoto, Joshua and Cunha, 2008; Gopalan *et al.*, 2009). There is an abundance of evidence demonstrating that *ERG* promotes pro-tumorigenic processes and signalling pathways in PCa (King *et al.*, 2009; Bock *et al.*, 2013). It is important to note that *ERG* is not significantly expressed in normal prostate epithelial cells (Mohamed *et al.*, 2010). Inappropriate expression is observed frequently in prostatic intraepithelial neoplasia (PIN), a precursor tumour lesion that can develop into malignant PCa (Tomlins *et al.*, 2008; King *et al.*, 2009). There is evidence to suggest that ERG targets genes that increase the motility and invasiveness of PCa tumour cells therefore aiding the progression of the

disease (J. Wang *et al.*, 2008; Rickman *et al.*, 2010; Wu *et al.*, 2013; Tian *et al.*, 2014).

C-X-C Motif Chemokine Receptor 4 (CXCR4), a chemoreceptor whose chemokine is stromal cell derived factor 1 (SDF1), is transcriptionally activated by ERG and mediates increased cell invasion and metastasis specifically to bone tissue (Singareddy *et al.*, 2013). Invasiveness is also promoted via the involvement of ERG with matrix metalloproteases and WNT- signalling (Butticè *et al.*, 1996; Brase *et al.*, 2011; Wu *et al.*, 2013). Moreover ERG affects EMT by promoting the loss of VE-cadherin and epigenetic regulation of migration via histone deacetylase 6 (Mohamed *et al.*, 2011; Birdsey *et al.*, 2012).

In addition to this it has been suggested that the loss of phosphatase and tensin homologue (*PTEN*), which is associated with Protein kinase B (called *AKT*) overexpression, may play a role in the initiation of PCa. *PTEN* is a tumour suppressor gene that supresses *AKT* under normal physiology and is very frequently mutated in human cancers. *AKT* is part of a signalling pathway that promotes cell survival and proliferation (Carver *et al.*, 2009; Squire, 2009). Several studies suggest that the concomitant loss of *PTEN* and the presence of *TMPRSS2-ERG* are also associated with PCa progression, aggressive disease and a poor prognosis (Carver et al., 2009; Leinonen et al., 2013; Barbieri et al., 2013; Figure 2). This may be because ERG may repress *PTEN* when bound to its promoter (Adamo *et al.*, 2017).

Moreover, now *EWS* is thought to be essential to the role of ETS transcription factors in PCa tumorigenesis. Kedage *et al.* (2016) revealed a crucial role for EWS interaction with ERG in the tumorigenesis of PCa. This was shown to be

true when the N terminus of *EWS* was fused to *ERG* but a *EWS* interaction domain in the C terminus of *ERG* was also validated, which when mutated, reduced cell migration and colony formation ability. The interaction between the two proteins was shown to be required for transcriptional activity. Furthermore, when EWS was knocked down or the EWS interaction domain in ERG was mutated, there was a significant reduction in transcription of target genes. This data would suggest the EWS is a crucial co-activator of ETS transcription factors including ERG target genes (Kedage *et al.*, 2016).

Contrastingly other large cohort investigations have found no significant correlation between *TMPRRS2-ERG* and poor prognosis or disease progression. Gopalan et al. (2009) found that although *TMPRSS2-ERG* positive patients did overexpress *ERG,* the fusion gene was associated with low grade tumours and there was no association with outcome. They also suggested that associations with poor outcome may be due to general aneuploidy often found in cancer patients. Another study looked at the histological pathology of prostate tumour samples expressing *TMPRSS2-ERG* and found that samples positive for the rearrangement had a lower Gleason score in comparison to those without (Fine *et al.*, 2010). Therefore, there is some debate as to whether the gene rearrangement can be used as a diagnostic or prognostic marker.

It is important to note that the diversity of *TMPRSS2-ERG* variants including those due to alternative splicing means it is not always possible for all *ERG* isoforms to be expressed. Both *ERG* and *TMPRSS2* are alternatively spliced and accordingly there are various combinations that can occur between the two genes and consequently several *TMPRSS2-ERG* fusion variants exist. Indeed

recent estimates show that there are nearly 20 *TMPRSS2-ERG* variants (Clark *et al.*, 2007; Zammarchi, Boutsalis and Cartegni, 2013). The most common variant is a hybrid of exon 1 of *TMPRSS2* and exon 4 of *ERG* often referred to as T1:E4 (Clark *et al.*, 2007). The protein product of each variant is either type I ERG, a near full length product with the ETS domain, or the truncated protein lacking the ETS domain known as type II ERG (Clark *et al.*, 2007; Rastogi *et al.*, 2014).

Interestingly Hagen et al. (2011) examined the expression of exon 7 and 7b in PCa patient samples. The splice isoform ratios of exon 7 and exon 7b inclusion and exclusion were compared in benign tissues, T2 localised tumours and stage T3a and T3b cancer tumour and, advanced stage tumours. The study reported that both exon 7 and 7b are increasingly included as disease progresses (Hagen et al. 2014). As previously detailed, studies using prostate cancer cell lines identified more oncogenic behaviour in *ERG* splice variants that included this 72 base pair exon which encodes part of the AD (J. Wang *et al.*, 2008; Yin *et al.*, 2011) and may therefore encourage the transcription of pro-tumorigenic genes.

Therefore, the clinical manifestations associated with the presence of a fusion gene may not always be present in every patient as may be the case, for example, if the *TMPRSS2* promoter region is not activated or functional. Furthermore some *ERG* splice isoforms are non-coding and may not produce an ERG protein at all (Zammarchi, Boutsalis and Cartegni, 2013). This is why it is important to measure *ERG* expression in combination with clinical outcomes as this is a more reliable measure of fusion effect on PCa status.

1.3.3 Function of ERG

Compared to other ETS transcription factors ERG is highly expressed in endothelial cells, and significant research has revealed the several important roles that ERG supports in endothelial development and function (Baltzinger, Mager-Heckel and Remy, 1999; Hollenhorst, Jones and Graves, 2004; Nikolova-Krstevski *et al.*, 2009). As the mouse embryo develops, endothelial cells, haematopoietic and cartilage tissue all express ERG (Mohamed *et al.*, 2010). In humans other tissue that express ERG include chondrocytes (Iwamoto *et al.*, 2000), both B and T lymphocytes, myeloid cells and megakaryocytes (Rainis *et al.*, 2005). ERG is generally a nuclear protein but the ERG8 isoform was shown to localise to the cytosol (Hoesel *et al.*, 2016).

1.3.3.1 ERG as an activator and repressor of transcription

In an *in vivo* mouse model for haematopoiesis ERG was shown to be a positive regulator of the *RUNX1* and *GATA2* genes. The expression of both genes was downregulated when functional ERG was knocked down and ERG was observed to occupy enhancer sequences in the promoters of both genes in foetal liver where expression of these genes was high. This regulatory pathway was shown to be vital to maintaining haematopoietic stem cells and thus embryonic development (Taoudi *et al.*, 2011).

Vascular endothelial (VE)-cadherin regulates the integrity of the endothelial layer affecting cell growth and permeability of the cell layer. Data from a chromatin immunoprecipitation (ChIP) assay of the vascular endothelial (VE)-cadherin promoter demonstrated that ERG binds the promoter and further evidence from a transactivation assay showed increased transcriptional activity

at the promoter as a result of this binding (Birdsey *et al.*, 2008). Further research revealed that ERG-dependent regulation of VE-cadherin affects vascular stability, that is the control of endothelial cell cytoskeleton and junction integrity, and angiogenesis through the canonical Wingless-related integration site (Wnt) signalling pathway (Birdsey *et al.*, 2015).

Leukocyte recruitment and adhesion is important in inflammatory disease and has been shown to be influenced by the expression of intercellular adhesion molecule 1 (ICAM-1) and vascular cell adhesion molecule (VCAM) as well as the cytokine interleukin-8 (IL-8). ERG directly binds to the ICAM-1 promoter and represses gene transcription suggesting an anti-inflammatory role for ERG. Repressed expression of IL-8 and VCAM was also observed. This was shown to be the case *in vivo* using a murine model where inflammation was induced by causing swelling the paws of mice. Normally tumour necrosis factor α (TNFα) stimulates inflammation and upregulates ICAM-1 gene expression. However when ERG was overexpressed inflammation was reduced in a TNFα – dependent manner (Sperone *et al.*, 2010).

1.3.3.2 Function of ERG exon 7b isoforms

The importance of *ERG* exon 7b was suggested quite early in the genes discovery. The exon was one of the first alternative exons identified in the full length protein then called p55 (Prasad *et al.*, 1994) and the first study that investigated the role of the transactivation domains also carried out mutagenesis in this exon (Siddique *et al.*, 1993). Owczarek and colleagues (2004) were one of the first to map out main human *ERG* variants but this was followed by a more detailed study by Zammarchi and colleagues in 2013.

Although both studies identified the 72 bp exon, one called it exon 12 (Owczarek) whilst the other called it exon 7b (Zammarchi, Boutsalis and Cartegni, 2013). In this study the Zammarchi *et al.* naming system will be used. However, studies use variable naming systems and where a different one is used clarification will be made.

In chicken there are two main isoforms of ERG, chERG that when expressed includes all three exons in the 222 base pair (bp) variable region (these exon are exon 7 (81 bp), exon 7b (72 bp) and exon 8 (69 bp)). The other main isoform is C-1-1 which lacks exon 7. This variant has been cloned and characterised across several organisms (Baltzinger et al. 1999). In developing epiphyseal chondrocytes of chicks, the C-1-1 variant is preferentially expressed, and overexpression of this variant imparts stabilizing and anti-maturation abilities to the cells when compared to chERG (Iwamoto et al. 2000).

The function of the mouse isoform of ERG C-1-1 isoform was later examined. Transgenic mice which were made to conditionally express the human C-1-1 ERG variant displayed severe morphological abnormalities in skeletal development, although it is important to note that observations were made in E18.5 embryos due to the lethality of the ERG phenotype to mice at birth. The transgenic mice embryos were generally smaller with hypomineralised skeletons when compared to wild type embryo littermates. Transgenic embryos also showed poor cellular organisation, absent growth plates, abnormal chondrocyte morphology as well as decreased expression of gene markers associated with maturation (Pacifici et al. 2006). All of these findings reveal a crucial role for ERG in the development of the articular chondrocyte structures

and suggest that ERG may be a potential target for the development of new anti-arthritic intervention development therapies.

To better understand the biological role of the various splice isoforms of *TMPRSS2-ERG* expression a study by Wang et al. (2008) transiently expressed several *TMPRSS2-ERG* transcripts in PNT1a prostatic epithelial cells which do not normally express ERG. They particularly focussed on exon 7b (denoted 72bp for this in the study) and found that the VCaP cell line expressed the type III isoform of *TMPRSS2-ERG* (exon 1 of *TMPRSS2* bound fused to exon 4 of *ERG*) +/- exon 7b. Their finding showed that inclusion of exon 7b increased proliferation, increased invasive ability invasion (as assessed by the Matrigel assay), and increased motility (assessed using the wound healing assay). Although they were able to confirm that all isoforms can form heterodimers and homodimers, the isoforms including exon 7b bound significantly more protein suggesting that this the exon imparts an increased potential for exhibited stronger protein-protein interactions (Wang et al. 2008).

High Gleason score and recurrence of the tumour in PCa has been associated with high expression of prostate specific membrane antigen (PSMA). Similar to the TMPRSS2-ERG fusion protein, PSMA is found to be overexpressed in around 50% PCa cases. TMPRSS2-ERG fusions decrease PSMA expression at the mRNA level in VCaP and an inverse correlation between the expression of ERG fusion and PSMA was confirmed by increased expression of PSMA with RNAi ERG knockdown mediated by RNA interference. Type III and VI *TMPRSS2-ERG* fusions +/- exon 7b were transfected into LNCaP cells and mRNA levels were quantified using real time PCR. PSMA mRNA levels were

generally higher in the + exon 7b isoforms both in the presence and absence of the synthetic androgen R1881. However, PSMA luciferase activity was lower in type III + exon 7b compared to type III – exon 7b but higher in type VI + exon 7b compared to type VI – exon 7b, with the lowest overall activity found in type VI – exon 7b transfected cells (Yin et al. 2011). Thus, the study showed that the overexpression of specific TMPRSS2-ERG variants decreased PSMA levels, presenting new potential targets for treatment. This data suggests PSMA and ERG work cooperatively and broadens the potential targets and pathways that can be manipulated for treatment.

1.4 Aims of this Study

The aim of this study was to gain an understanding of the biological function of the *ERG* exon 7b splice isoforms and how the inclusion of this exon is regulated. The rationale for focusing on this cassette exon was based on the literature that indicates that its inclusion makes *ERG* more oncogenic. Exon 7b adds, in frame, 24 amino-acids to the transactivation domain a crucial part of ERG that influences its interactions and transcriptional activity. Presumably this splicing event leads to the selective expression of the full transactivation domain and is a specific mechanism for altering ERG activity. To explore this hypothesis, the project will used splice switching oligonucleotides to reduce exon 7b inclusion and therefore ablate the expression of the full transactivation domain isoforms of ERG. Upon successful achievement of exon skipping a series of functional assays were performed to characterise the oncogenic potential of the isoforms expressed in the cell model.

The factors that regulate oncogenic splicing events have been identified for several human genes and Furthermore a part of the project concentrated on identifying potential splice factors that play a regulatory role in the alternative splicing of exon 7b in *ERG*.

2 Materials and Methods

2.1 Cell Lines

MOLT4 (obtained from ECACC, Public Health England), CMK, TK6 and K562 (kindly provided by Dr Ruth Morse) cells were grown in RPMI media supplemented with 10% foetal bovine serum (FBS) and 4mM L-glutamine at 37°C, 5% CO2 in a humidified incubator. Cells were sub-cultured every 2-3 days.

MG63 cells (kindly provided by Dr Jason Mansell) were grown in DMEM supplemented with 10% FBS at 37°C, 5% CO2 in a humidified incubator and sub-cultured every 3- 4 days. Where serum starvation was required cells were cultured in DMEM:F12 supplemented with essential amino acids.

2.1.1 Cell counting and viability assessment

Cells were counted using a haemocytometer using the Trypan Blue exclusion assay to determine viability. A 10µL aliquot of cell suspension was diluted 1:1, 1:3, 1:5 or 1:10 depending on cell confluence with 0.4% Trypan Blue and loaded into the haemocytometer. The haemocytometer was observed on a light microscope using a x10 objective. Dead cells appeared blue as they took up the trypan blue dye whereas viable cells appeared bright and unstained. Viable cells and total cells were counted and viability was determined using the following equation:

$$\%\ viability = \frac{(number\ of\ viable\ cells)}{(total\ number\ of\ cells)} \times 100$$

2.2 Cell treatments

Phorbol myristate acetate (PMA) is a highly potent activator of protein kinase C. Changes in the downstream signalling pathways has been observed to increase the expression of several transcription factors. K562 and TK6 cells, which do not ordinarily express ERG, were treated with 15nM PMA for 72 hours to assess if the expression of ERG could be induced. All experiments also included a dimethyl sulfoxide (DMSO) control for which cells were treated with 0.5% DMSO as this was the maximum amount of DMSO ever used for each experiment.

2.3 Designing and using splice switching oligonucleotides

Splice switching oligonucleotides (SSOs) were designed, with assistance from Dr Lee Spraggon (Memorial Sloan Kettering, USA) complementary to the DNA sequence of the 3' and 5' splice site of the *ERG* gene as taken from the Ensembl Genome Browser (http://www.ensembl.org/index.html). All Vivo-morpholinos were manufactured by Gene Tools, LLC. An antisense morpholino was designed against the both the 5′ and 3′ splice site of *ERG* exon 7b. The antisense sequence of the ERG 7b 5′ss SSO (E7b5) 5′-UCCGGUCCAUGCUUUUGUGGGGACA -3′ and for ERG 7b 3′ss (E7b3) is 5′-AAGGAAAACAGACGTCCCCCACGUC-3′. The sequence for the control SSO (designed by Gene Tools, LLC), targeting position 705 in the beta-globin mRNA, is 5′-CCTCTTACCTCATTACAATTTATA-3′. Stocks of each Vivo-morpholino were made in sterile PBS at a concentration of 0.5 mM. Each SSOs has an octaguanidine dendrimer moiety to facilitate delivery for cellular uptake meaning no transfection reagent is required and the SSOs are added directly to cell medium.

2.4 RNA extraction

Total RNA was extracted using the Absolutely RNA Miniprep Kit (Agilent Limited) according to the manufacturer's protocol. Samples were DNase treated on-column using kit provided RNase-free DNase I.

For samples containing less than 500,000 cells RNA was extracted using the genesig easy DNA/RNA extraction kit (Primerdesign) and DNase treatment was performed with the Precision DNase kit (Primerdesign) according to manufacturer protocols.

RNA preparation for quantitative PCR was carried out using a phenol chloroform based method as follows. Cells were lysed in lysis buffer (0.1 M Tris pH 8.0, 5 mM EDTA pH 8.0, 0.1 M NaCl, 0.5% sodium dodecyl sulfate (SDS), 1% β 2-mercaptoethanol was added just before use) and homogenised using a 21-23 gauge needle and syringe. An equal volume of UltraPure™ Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v) (Thermofisher) was added to each sample. Samples were vortexed thoroughly and spun at maximum speed on a bench top centrifuge (13,000 x g) for 5 minutes. The aqueous phase was removed and put into a fresh tube. An equal volume of chloroform was added to the sample which was then vortexed and incubated at room temperature for 2-3 minutes. The samples were then centrifuged at 12,000 x g for 15 minutes at 4°C. This process was repeated.

Two volumes of ice-cold ethanol was added with a 1:10 dilution of 3M sodium acetate (pH 5.2) and the sample were placed in a -80°C freezer for a minimum of 30 minutes up to overnight or a -20°C freezer for a minimum of 1 hour up to

overnight to precipitate the RNA pellet. The samples were then centrifuged at 12,000 x g for 30 minutes at 4°C. The RNA pellet was washed in ice cold ethanol, excess ethanol was aspirated and then the pellet was left to air dry. Pellets were resuspended in nuclease-free water. To remove gDNA contamination an appropriate volume of 10x DNase reaction buffer and 2U of DNase I (New England Biolabs, NEB) was added to the sample which was then incubated at 37°C for 10 minutes. To deactivate the DNase EDTA to a final concentration of 5mM was added and the sample was incubated at 75°C for 10 minutes.

All RNA samples were quantified using a Nanodrop 2000 (Thermofisher) and aliquots were made prior to cDNA synthesis and storage at -80°C.

2.5 cDNA synthesis

Between 150-2000ng of total RNA was used to synthesise cDNA using the Precision nanoScript2 Reverse Transcription kit (Primerdesign) according to manufacturer's protocol with the following modifications. Samples that would be used for qPCR were spiked with 4ng of RNA sequence for the small RUBisCO subunit of *Arabidopsis thaliana* in order to normalise gene expression. To prime the reactions 0.5 µL each of oligo dT and random nonamer primers were used.

The qPCR spike was made by as follows. A cDNA plasmid clone for the small RUBisCO subunit of *Arabidopsis thaliana* (U13397, from The Arabidopsis Information Resource) was cultured in selective medium and a maxi preparation was made to yield adequate plasmid. The plasmid was linearised and sequenced to confirm its identity. The linearized plasmid (1µg) was in vitro

transcribed using 30U T3 RNA polymerase with 50U RNase inhibitor, rNTP mix (2 mM final concentration), 10 µL 5x transcription buffer made up to a 50 µL reaction volume with nuclease-free water. The reaction was incubated at 37°C overnight before the *in vitro* transcribed RNA was purified using the phenol: chloroform method described above. The RNA was DNase treated, further purified using the phenol: chloroform method, quantified using the Nanodrop 2000 and the final concentration was adjusted to make 10 µL aliquots with 1µg of RNA in each. These were stored at -80°C.

2.6 Standard PCR and Gel Electrophoresis

Hot Start Taq 2X master mix (NEB) was used for standard PCR reactions to determine gene expression. Reactions were set up at room temperature with 1-2µL undiluted cDNA template and had a final reaction volume of 25 µL. The sequences for the primers in the appendix. The annealing temperature was calculated using the NEB Tm calculator (available on https://tmcalculator.neb.com/#!/main). The final concentration for each primer in the reaction was 0.4µM. PCRs were run as follows: initial denaturation 95°C for 30 seconds, then 30-40 cycles of 95°C for 30 seconds, the annealing temperature for 1 minute, 68°C for 1 minute and a final extension at 68°C for 5 minutes. The PCR samples were stored a 4°C until they could be run on an agarose gel.

Agarose gels (1-2%) were made and stained with 5µL Midori Green Advance DNA stain (Geneflow) for every 100mL of TAE. Samples were loaded using purple 6X gel loading dye (no SDS) (NEB) and were run at 100V for 40-60 minutes. Splice isoform ratios were be determined by using the ImageStudio

Lite software (Licor, available from https://www.licor.com/bio/products/ software/ image_studio _lite/). The percent spliced in (PSI) was calculated by

$$PSI = \frac{signal\ for\ splice\ isoform\ of\ interest}{total\ signal\ for\ all\ splice\ isoforms} \times 1$$

2.7 Isolating and Sequencing PCR products

PCR bands were excised under UV light. The QIAquick Gel Extraction kit (Qiagen) was used to purify the PCR products for sequencing according to the manufacturer's protocol. Samples were sent to Eurofins Genomic Service for sequencing. The sequencing results for ERG can be found in the appendix.

2.8 Quantitative PCR

Quantitative PCR was used to measure gene expression. Template cDNA (pre-diluted as indicated in figures) was amplified using Precision FAST MasterMix with ROX (Primerdesign), 0.4 µM forward primer, 0.4 µM reverse primer and made up to 10µl using nuclease-free water. The cycling conditions comprised DNA denaturation for 3 minutes at 95°C and amplification repeated for 40 cycles at 95 °C for 15 seconds and 58-60 °C for 30 seconds. All PCR efficiencies were above 95%. Real-time PCR were performed using the StepOnePlus thermocycler (Applied Biosystems) and the results were analysed using StepOne Software version 2.3. The normalisation and quantification protocol can be found in the appendix.

2.9 RNAi high-throughput RT-PCR screen

The RNAi high-throughput RT-PCR screen was carried out by RNAomics Platform at University of Sherbrooke to determine the change in PSI as a result

of RNA binding protein knockdowns. This assay amplifies full length and exon 7b skipped *ERG* isoforms using cDNA from 56 RNA binding protein siRNA knockdowns that have been carried out in MCF-7 cells. A lipofectamine transfection reagent control (siCTRL) was included for each set of knockdowns. The analysis of alternative splicing was determined using endpoint PCR coupled with microcapillary electrophoresis. PSIs were then calculated and the change in PSI (∆PSI) was determined by the following formula:

$$\Delta\text{PSI} = \frac{PSI\ siRNA}{PSI\ siCTRL} \times 1$$

PSI and calculated ΔPSI values for all 56 siCTRL and siRNA MCF7 cell samples tested are included in the appendix. Any PSI values with no useable control or an error were not included in analyses.

2.10 Protein isolation and preparation

Cells were washed twice with ice-cold PBS, and whole cell protein lysates were prepared using an appropriate volume of RIPA buffer (10 mM Tris-Cl (pH 8.0), 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS and 140 mM NaCl) supplemented with protease inhibitor tablets (ThermoFisher). After a 30 minute incubation, with periodical vortexing, lysates were centrifuged for 15 min at 12,000 x g at 4°C for clarification. Samples were transferred to fresh tubes and quantified using the bicinchoninic acid (BCA) kit (Thermofisher) and a bovine serum albumin protein standard curve.

To prepare MOLT4 nuclear protein lysate using a method by Yu, Huang and Lung (2013). Briefly cells pellets were washed twice in PBS before addition of the subcellular fractionation buffer. Cells were lysed in the buffer and incubated

on ice for 30 minutes. This was followed by centrifugation and the leftover pellet (the nuclear fraction) was sonicated in a nuclear lysis buffer. The lysate was quantified using the BCA kit, aliquoted and stored at -80°C.

2.11 SDS-PAGE and Immunoblotting

Immunoblotting was carried out using 10 to 20µg of total cell lysate. Samples were added to 2x Laemelli Sample Buffer and heated for 5 minutes at 100 °C. Proteins were separated by hand cast SDS-polyacrylamide gel electrophoresis (PAGE) or TruPAGE™ Precast Gels and electro-transferred to polyvinylidene difluoride (PVDF) membrane (Thermofisher). The membranes were blocked for 1 hour at room temperature using 5% skimmed milk powder made in Tris-buffered saline with 0.1% Tween-20 (TBST) then probed with primary antibodies against target proteins diluted in TBS-T overnight at 4 °C. The following antibodies used in this study: anti-ERG (1:500-1000, ab92513), anti-beta-actin (1:10,000, ab8229), both antibodies were from Abcam. Following three TBST washes the membranes were then incubated with an appropriate HRP-linked IgG secondary antibody (1:2000, New England Biolabs) for 2 hours at room temperature. Following three TBST washes membranes were incubated in Luminata Forte Western HRP substrate (Millipore) for chemiluminescent detection for 2 minute prior to image acquisition using the Licor Odyssey Fc imaging system.

2.12 RNA Pull-Down

To determine proteins that associate with *ERG* exon 7b RNA an RNA pull-down experiment was performed using Pierce Magnetic RNA-Protein Pull-down kit (Thermofisher) following the manufacturers protocol. Firstly, 50 pmol synthetic

RNA oligonucleotide corresponding to the 3´ splice site of ERG exon 7b (gucuuuccuuuugucugcagGGGGUGCAGCUUUUAUUUUC, -20 to 20; ThermoFisher) was ligated with desthiobiotinylated cytidine bisphosphate at the 3' end and then purified by ethanol precipitation. Subsequently, labelled RNA was bound to streptavidin magnetic beads at room temperature followed by incubation with 100 µg MOLT4 nuclear extract at 4 °C for 2 hours with agitation. After incubation, the magnetic beads were washed three times with wash buffer (20 mM Tris (pH 7.5), 10 mM NaCl, 0.1% Tween-20) and the bound proteins were eluted using biotin elution buffer. Bound proteins were separated by SDS-PAGE and visualised by coomassie blue staining. Following imaging of gels, bands were excised using sterile scalpels and sent off for trypsin digestion prior to mass spectrometry analysis (Proteomics Facility, University of Bristol). The list of proteins identified using mass spectrometry can be found in the appendix.

2.13 siRNA knockdown of splice factors

RNAi was used to knockdown protein levels for candidate splice factors to establish if they have a role in regulating the expression of ERG exon 7b. MG63 cells (5 x 10^5 in 500 µL per well) were seeded in DMEM media without serum in a 12 well plate. For each well 20nM of siGENOME Human SMART pool siRNA solution (Dharmacon) was added to 200µL of DMEM without serum and mixed well before 6 µL of Lipofectamine RNAiMax transfection reagent (ThermoFisher) was added. All siRNA were a pool of four siRNA to increase efficacy of knockdown and a pool of non-targeting siRNA was included as a control. The solutions were vortexed for 10 seconds and incubated at room temperature for 10 minutes. These solution were then added to the cells drop

wise and the plate was gently swirled. After 4.5 hours of incubation at 37°C 1mL of complete DMEM media was added. After 24 hours cells were sub-cultured into T-25 flasks for a further 24 hours at which point RNA and protein were isolated from the cells.

2.14 Resazurin cell viability assay

The conversion of resazurin is carried out by enzymes in viable cells and can be measured colorimetrically. The activity of the metabolic enzymes within the cells allow for the assessment of both cell viability and proliferation. A 10mg/mL stock solution of resazurin was used to make a fresh working stock at 0.04mg/mL concentration every two weeks. All stocks were made in PBS. The resazurin reagent was added 1:10v/v to cells treated wells as well as to negative control wells (made up to the equivalent of the media volume) 2-4 hours prior to the end of the experimental incubation period in order to assess cell viability. The plates were then read at 570nm and 630nm in a plate reader (FLUOstar OPTIMA, BMG Labtech) to obtain readings for cell viability determination. The following equation was used to determine the percentage reduction of resazurin to resofurin.

$$\%\ \text{reduction of resazurin} = \frac{(\text{Eoxi630} \times \text{A570}) - (\text{Eoxi570} \times \text{A630})}{(\text{Ered570} \times \text{C630}) - (\text{Ered630} \times \text{C570})} \times 100$$

- Eoxi630= molar extinction coefficient (E) of oxidised resazurin reagent at 630nm = 34798
- A570 = absorbance of test wells at 570nm
- Eoxi570 = E of oxidised resazurin reagent at 570nm = 80586
- A630 = absorbance of test wells at 630nm

- Ered570 = E of reduced resazurin reagent at 570nm = 155677
- C630 = absorbance of negative control well (media, resazurin reagent, no cells) at 630nm
- Ered630 = E of reduced resazurin reagent at 630nm = 5494
- C570 = absorbance of negative control well at 570nm

The percentage viability was determined by correcting the other percentage reductions against the media only or PBS treated cells which were considered to be 100% viable.

2.15 Immunofluorescence analysis of Ki-67

Ki-67 is a nuclear protein that is expressed in all active stages of the cell cycle and can be used as a proliferation marker as high expression of the protein are found in proliferating cells. MG63 cells were treated with SSOs or PBS in flat bottomed 6 well plates for 48 hours and 72 hours. The media was removed and cells were washed three times in PBS. The cells were then fixed in 3.7% paraformaldehyde for 15 minutes at room temperature. The paraformaldehyde was discarded, and after three PBS washes, a solution of 0.1% Triton X-100 was applied for 5 minutes to permeabilise the cells. The cell layer was blocked in 3% FBS solution (in PBS) for 30 minutes at room temperature followed by incubation in a 1:200 dilution of Ki-67 primary antibody (Abcam, ab15580) made up in blocking solution for 2 hours. Three PBS washes were then followed by 1 hour incubation with the Alexa 568 fluorophore conjugated secondary antibody (Thermofisher, #A10042, 1:2000) made in 3% FBS in the dark, three PBS washes, a 2 minute incubation in 2 µg/mL Hoechst 33342 and storage in PBS. Images of six representative fields of view per treatment were acquired using

the using the 20x objective on the Eclipse 80i microscope (Nikon). An image using the Texas red fluorescent filter and the 4′,6-diamidino-2-phenylindole (DAPI) filter setting were taken for each field of view.

2.16 Caspase 3/7 assay

The process of apoptosis has two main pathways; the intrinsic and extrinsic pathway. These pathways converge at the activation of effector caspases which are a type of protease (Lamkanfi and Kanneganti, 2010). Measurement of the activity of caspase 3 and 7 allows reveals cells that are committed to the final events of the apoptotic process as the cleavage of substrates by caspase 3 and 7 initiates key structural and biochemical changes cell.

After seeding in 6 well plates, MG63 cells were treated with SSOs or PBS for 48 hours and 72 hours and 45 minutes prior to the end of the incubation period the Cell Event Caspase 3/7 reagent made in pre-warmed PBS was added to cells. Images of 6 fields of view per treatment were taken using the 20x objective on the Eclipse 80i microscope. An image using the green fluorescein isothiocyanate (FITC) fluorescent filter and the phase contrast setting were taken for each field of view.

2.17 Transwell invasion and migration assay

Tumour cells migration and invasion is a characteristic of a advancing disease in malignancy and in vitro investigation of a cells capacity for mobility is a useful indicator of oncogenicity. Invasion requires that a cells pass through a barrier which in the assay is a layer of basement membrane components. Cell migration is also an indicator of a cells motility a characteristic that contribute to its invasive ability.

For the invasion assay polyethylene terephthalate inserts (membrane pore size, 8 μm; Millipore) were coated with 50uL of Geltrex® Matrix (ThermoFisher) diluted 1:1 in serum-free medium for at least 2 hours and up to 24 hours to create an artificial basement membrane to facilitate cell invasion. Uncoated inserts were used for the migration assay. Inserts were set up in 24 well plates for both migration and invasion assays. MG63 cells were treated with SSOs or PBS for 24 hours and then were harvested. Subsequently 1×10^5 cells in 100 μL serum-free medium was added to the upper chamber of the insert with (invasion) or without (migration) Geltrex® Matrix and 600 μL of medium supplemented with 10% FBS was added to the lower chamber of a 24-per well. Following 24 hours of culture at 37 °C, the cells remaining on the upper membrane of the insert were removed and inserts were washed in PBS. After 15 minutes fixation in methanol and 2 minutes staining with haematoxylin the inserts were air dried prior to imaging of six representative fields of view using a camera attached to a light microscope (x20 objective). The number of cells adhering to the lower membrane of the inserts was counted. All experiments were carried out in triplicate.

2.18 Alkaline Phosphatase assay

Alkaline phosphatase activity increases during osteoblast maturation and can reliably be measured by the generation of p-nitrophenol (p-NP) from p-nitrophenylphosphate (p-NPP) under alkaline conditions. Cells were seeded into 96 well plates and starved in phenol red-free, serum-free DMEM:F12 medium overnight. The relevant wells were dosed with 3μM of SSOs or PBS and incubated for 24 hours. Subsequently a phosphatase resistant analogue of

lysophosphatidic acid (LPA), (3S)1-fluoro-3-hydroxy-4-(oleoyloxy)butyl-1-phosphonate (FHBP) and calcitriol, the active metabolite of vitamin D3 (1,25D), were added to the relevant wells as well as an additional dose of SSO to maintain the expression of *ERG* exon 7b skipped isoforms. Following another 24 hour incubation resazurin reagent was added to treated wells as well as to empty control wells 2-4 hours prior to the end of the incubation period in order to assess cell viability. The plates were then analysed to obtain readings for cell viability determination.

The resazurin reagent was removed and the monolayers were washed for 5 minutes in fresh phenol red-free DMEM:F12 media to remove the residual resazurin and its metabolites. Following this, the medium was removed and the monolayers lysed with 50 µL of lysis buffer (25mM sodium carbonate (pH 10.3) and 0.1% (v/v) Triton X- 100). After 2 minutes, each well was treated with 100 µL of 15mM p-NPP (di-Tris salt, Sigma, UK) in 250mM sodium carbonate (pH 10.3), 1mM MgCl2. The cell culture plates were then placed in an incubator at 37°C for 1 hour. After the incubation period the cell culture plates were transferred to the plate reader, and an ascending series of p-NP (10–500 µM) controls prepared in the incubation buffer were added to empty wells on the plate to enable quantification of product formation. The absorbance was read at 405 nm. Reported total alkaline phosphatase activity was corrected to the cell viability determined using the resazurin assay for each well.

2.19 Chromatin immunoprecipitation assay

Chromatin immunoprecipitation was used to determine if ERG interacts with the promoter of the alkaline phosphatase gene and was performed using the Imprint

Chromatin Immunoprecipitation Kit (Sigma, UK) according to the manufacturer's protocol. Briefly, $1x10^6$ MG63 cells were incubated for 10 minutes with 1% formaldehyde at room temperature to allow DNA to be cross-linked to the protein. Cells were then lysed and DNA sheared to produce fragments of about 1000bp using a sonicator six pulses for 15 seconds at 50% power output followed by incubation on ice for 60 seconds after each pulse. The DNA-protein mixture was incubated with in well of a cell culture plate coated with either 3µg of anti-ERG antibody (Santa Cruz, UK), 1µg RNA polymerase II antibody or 1µg non-specific IgG at room temperature for 3 hours. This was followed by six washes with the IP wash buffer. The cross-linked DNA-ERG complexes were released using Proteinase K at 65°C for 15 minutes. The DNA was cleaned up and eluted using GeneElute Binding Colum (Sigma, UK). PCR was then carried out using primers shown in the appendix for ALPL.

2.20 Xenograft mouse model and in vivo tumour growth

The xenograft mouse model was used as an *in vivo* model to measure the effects of a ERG SSOs on malignant growth. Two-month-old male nude mice (CD1; Charles River, USA) were housed under pathogen-free conditions. All animal operations were approved by the Animal Ethics Committee of University School of Exeter. Six mice were used for each experimental group and the experiment was repeated three times (a total of 18 mice per experimental group). MG63 cells were harvested, counted and resuspended in PBS.

To establish tumours mice MG63 cells (7×10^6) were injected subcutaneously with a 27- gauge needle under general anaesthesia (inhalation of isoflurane allowed to reach 3mm by 3mm in size. Once tumours reached this size

12.5mg/kg of SSO or PBS was administered by intraperitoneal injection into the mouse twice weekly. Health checks were performed daily, body weight was measured using scales and tumour sizes were measured using vernier callipers on dosing days. Once tumours reached 12mm in one direction, mice were culled under Schedule 1 procedure and tumours harvested. These experiments were conducted by Dr Sean Porazinski in collaboration with Dr Sebastian Oltean at the University of Exeter.

2.21 Bioinformatic analysis

2.21.1 Databases used for expression analysis

The Genotype-Tissue Expression (GTEx) Project was used to assess *ERG* expression across in human tissues. GTEx is supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this thesis were obtained from the GTEx Portal on 25 June 2018.

The Encyclopedia of DNA Elements (ENCODE) database was accessed via the UCSC Genome browser to assess *ERG* exon 7b in various human derived cell lines (Kent *et al.*, 2002; Rosenbloom *et al.*, 2012; Raney *et al.*, 2014).

2.21.2 Gene sequence annotation

Gene and amino acid sequences were obtained from GenBank (http://www.ncbi.nlm.nih.gov/genbank/) or Ensembl and annotated using Snap Gene Viewer (software can be downloaded from http://www.snapgene.com/products/snap

gene_viewer/).

2.21.3 Sequence alignments

Nucleotide and protein alignments of genes orthologous to the human *ERG* gene were carried out using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) or MUSCLE (MUltiple Sequence Comparison by Log- Expectation) (https://www.ebi.ac.uk/Tools/msa/muscle/) and further annotation was carried out using MEGA6 (downloaded from http://www.megasoftware.net/).

2.21.4 Prediction of splicing regulatory elements

To predict for the presence of potential exonic splicing enhancer sequences present in exon 7b exon and flanking intronic sequence ESEfinder 3.0 (http://krainer01.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home) and RESCUE-ESE (http://genes.mit.edu/burgelab/rescue-ese/) were used (Cartegni *et al.*, 2003; Fairbrother *et al.*, 2004; Smith *et al.*, 2006).

2.21.5 Splice factor binding prediction

Bioinformatic tools were used to predict potential splice factor binding sites in and around the *ERG* exon 7b sequence. This analysis was carried using Splice Aid 2 (http://193.206.120.249/splicing_tissue.html) and the Splice Aid F database (http://srv00.ibbe.cnr.it/SpliceAidF/).

2.22 Statistical Analysis

All statistical analysis was performed using the GraphPad Prism 7 software. The Shapiro-Wilk test, the kurtosis value, the skewness value and distribution graph for each treatment group was used to assess normality. At least one treatment group for all experiments did not have normal distribution meaning non-parametric tests were used. The Kruskal-Wallis (one-way analysis of variance

on ranks) non-parametric test was used to test for significant difference between groups. Multiple comparisons between different treatment groups were assessed using the Dunn's test. Significance was a p-value of ≤0.05.

3 Analysis of *ERG* exon 7b sequence and expression

3.1 Introduction

The aberrant expression of *ERG* in disease particularly PCa and haematological cancers is well established suggesting that the gene plays a key role in the development of these cancers. The overall expression patterns of ERG are well characterised and several databases and research papers provide information of both mRNA and protein expression levels in normal tissues (Iwamoto et al., 2000; Shon et al., 2015) and several cell lines (Reddy and Rao, 1991; Mohamed *et al.*, 2010).

The expansion and usability of RNA sequencing (RNA-seq) technology has opened up a new era of transcriptome analysis allowing us to establish the changes in expression of specific splice variants of a gene in human tissues and cell lines in different conditions. According to the Genotype-Tissue Expression (GTEx) project data *ERG* is expressed at moderate levels in most tissues with highest levels of expression in arterial tissue. High *ERG* expression has also been observed in peripheral blood mononuclear cells, adipocytes, breast, prostate, lung, thyroid and spleen (Figure 3.1). Within the developmental context ERG is expressed including in haematological (Loughran *et al.*, 2008) and cardiovascular progenitor tissues and cells (Nikolova-Krstevski *et al.*, 2009). Expression of *ERG* exon 7b specifically in human tissues is limited to arteries, breast, ovary, cervix and adipose tissue all at comparable levels (Figure 3.2). This corresponds to tissues in which overall *ERG* expression is greater and suggests that the inclusion of this exon is particularly important in these tissues.

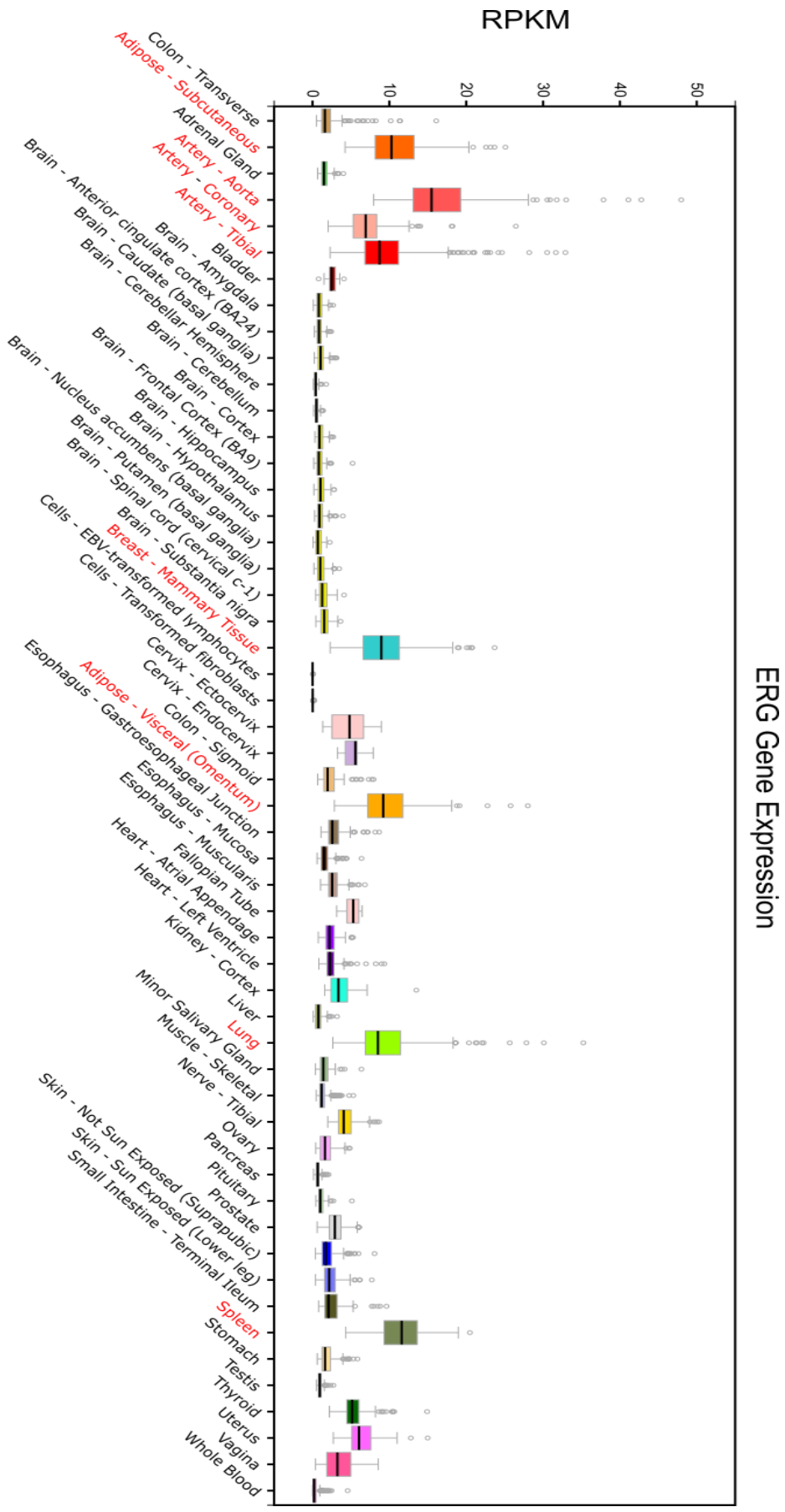

RPKM
0
10
20
30
40
50
ERG Gene Expression
Colon - Transverse
Adipose - Subcutaneous
Adrenal Gland
Artery - Aorta
Artery - Coronary
Artery - Tibial
Bladder
Brain - Amygdala
Brain - Anterior cingulate cortex (BA24)
Brain - Caudate (basal ganglia)
Brain - Cerebellar Hemisphere
Brain - Cerebellum
Brain - Cortex
Brain - Frontal Cortex (BA9)
Brain - Hippocampus
Brain - Hypothalamus
Brain - Nucleus accumbens (basal ganglia)
Brain - Putamen (basal ganglia)
Brain - Spinal cord (cervical c-1)
Brain - Substantia nigra
Breast - Mammary Tissue
Cells - EBV-transformed lymphocytes
Cells - Transformed fibroblasts
Cervix - Ectocervix
Cervix - Endocervix
Colon - Sigmoid
Adipose - Visceral (Omentum)
Esophagus - Gastroesophageal Junction
Esophagus - Mucosa
Esophagus - Muscularis
Fallopian Tube
Heart - Atrial Appendage
Heart - Left Ventricle
Kidney - Cortex
Liver
Lung
Minor Salivary Gland
Muscle - Skeletal
Nerve - Tibial
Ovary
Pancreas
Pituitary
Prostate
Skin - Not Sun Exposed (Suprapubic)
Skin - Sun Exposed (Lower leg)
Small Intestine - Terminal Ileum
Spleen
Stomach
Testis
Thyroid
Uterus
Vagina
Whole Blood

Figure 3.1 *ERG* expression in human tissues according to the Genotype-Tissue Expression project (GTEx)

Differential *ERG* expression across human tissues (x-axis). For each box the middle line is the median expression and outliers are shown as circles. The y-axis is reads per kilobase transcript per million (RPKM) and x-axis shows the different human tissues, those in red have the highest expression levels. The graph was obtained from the GTEx Browser (http://www.gtexportal.org/home/).

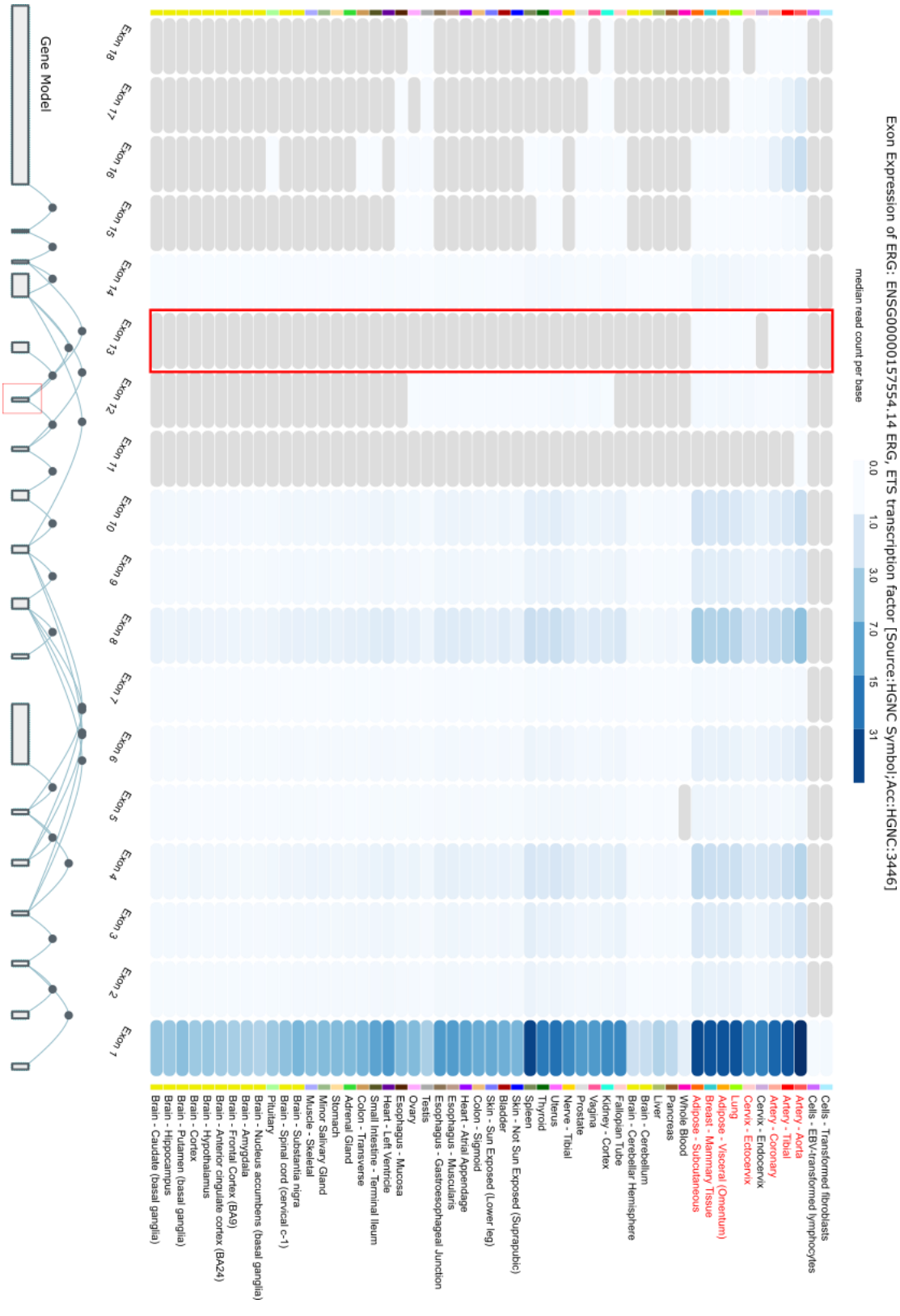

Exon Expression of ERG: ENSG00000157554.14 ERG, ETS transcription factor [Source:HGNC Symbol;Acc:HGNC:3446]
median read count per base
0.0
1.0
3.0
7.0
15
31
Exon 18
Exon 17
Exon 16
Exon 15
Exon 14
Exon 13
Exon 12
Exon 11
Exon 10
Exon 9
Exon 8
Exon 7
Exon 6
Exon 5
Exon 4
Exon 3
Exon 2
Exon 1
Gene Model
Cells - Transformed fibroblasts
Cells - EBV-transformed lymphocytes
Artery - Aorta
Artery - Tibial
Artery - Coronary
Cervix - Endocervix
Cervix - Ectocervix
Lung
Adipose - Visceral (Omentum)
Breast - Mammary Tissue
Adipose - Subcutaneous
Whole Blood
Pancreas
Liver
Brain - Cerebellum
Brain - Cerebellar Hemisphere
Fallopian Tube
Kidney - Cortex
Vagina
Prostate
Nerve - Tibial
Uterus
Thyroid
Spleen
Skin - Not Sun Exposed (Suprapubic)
Bladder
Skin - Sun Exposed (Lower leg)
Colon - Sigmoid
Heart - Atrial Appendage
Esophagus - Muscularis
Esophagus - Gastroesophageal Junction
Testis
Ovary
Esophagus - Mucosa
Heart - Left Ventricle
Small Intestine - Terminal Ileum
Colon - Transverse
Adrenal Gland
Stomach
Minor Salivary Gland
Muscle - Skeletal
Brain - Substantia nigra
Brain - Spinal cord (cervical c-1)
Pituitary
Brain - Nucleus accumbens (basal ganglia)
Brain - Amygdala
Brain - Frontal Cortex (BA9)
Brain - Anterior cingulate cortex (BA24)
Brain - Hypothalamus
Brain - Cortex
Brain - Putamen (basal ganglia)
Brain - Hippocampus
Brain - Caudate (basal ganglia)

Figure 3.2 *ERG* exon 7b expression in human tissues (GTEx)

Data from GTEx portal showing the mRNA expression of *ERG* exon 7b (exon 13 in this figure) is shown in the column highlighted in red. Each column represents an exon as shown in the *ERG* gene model at the bottom of the figure (exon 7b shown in the red box). The exon nomenclature used by GTEx varies to the one used in this thesis (Zammarchi, Boutsalis and Cartegni, 2013), however the exon size and position are equivalent. The colour gradient scale is the median read count per base where the more intense the blue colour the more reads per base and therefore the higher the expression. Grey indicates tissues for which no expression data is available. The human tissues (on the right hand side panel) in which exon 7b are expressed are shown in red.

There has been significant focus on the understanding the importance of the fusion genes that are formed by *ERG* namely *TMPRSS2-ERG* in PCa (Tomlins *et al.*, 2008), *FUS-ERG* in leukaemia (Panagopoulos *et al.*, 1994) and *EWS-ERG* in Ewing sarcoma (Sorensen *et al.*, 1994) but not much attention has been given the differential expression of *ERG* splice isoforms or splicing events (Hagen *et al.*, 2014; Hoesel *et al.*, 2016). Increasingly the significance of the generation of splice variants of several oncogenes via the process of alternative splicing has been shown to contribute to the development, maintenance and progression of the malignant phenotype. Therefore identification and characterisation of the expression of these splice isoforms is the foundation of better understanding their contribution to a specific disease and thus deciding their potential as a clinical marker.

3.1.1 Modification of alternative splicing using splice switching oligonucleotides (SSOs)

Splice switching oligonucleotides (SSOs) are a type antisense oligonucleotide (ASO); agents that utilise Watson-Crick base-pairing to inhibit gene expression (Aartsma-Rus and van Ommen, 2007). As RNA is transcribed into protein the use of ASO technology to modulate protein expression and production has been explored for some decades now as a potential therapeutic strategy. However, translation from bench top to clinic has been slow as issues with ASO chemistry, deliver and side effects have posed challenges. Encouragingly recent breakthroughs have been made. With improved ASO design and delivery solutions several clinical trials are ongoing and there is even a therapeutic ASO being used to treat patients.

The main challenges that come with using RNA oligonucleotides are their rapid degradation, insufficient uptake, poor biological distribution and poor target affinity (Shen and Corey, 2018). To overcome this, modifications to the nucleotide structure are made to the oligonucleotide to improve its pharmacokinetic properties. The non-bridging phosphodiester oxygen can be replaced by sulphur to make a phosphorothioate (PS) which is more resistant to nuclease digestion and increases binding to serum proteins facilitating systematic circulation and chances of successful uptake (Eckstein, 2014; Figure 3.3**Error! Reference source not found.**A).

Other common modifications are to the 2´ ribose where the 2´-hydroxyl group is replaced. Common substitutions are 2´-O-methyl, 2´-O-methoxyethyl and 2´-O-fluoro. As the 2´-hydrdoxyl is a site of nuclease attack the replacement of this moiety increases stability as well as increasing bonding affinity. The structural similarity of these 2´ modified ASOs to RNA means that RNase H does not cleave the hybridise molecules formed between the two and therefore these ASOs are popularly used to change alternative splicing (Prakash, 2011).

The SSOs used in this study are specifically antisense morpholino oligonucleotides. The structure of these ASOs mimics natural nucleic acids but have a 6-membered sugar ring in their backbone and non-charged phosphorodiamidate linkages instead of the phosphate linkages found in DNA and RNA molecules (Figure 3.3A). These structural modifications contribute to the biological stability, improved affinity and solubility of morpholinos (Morcos, 2007).

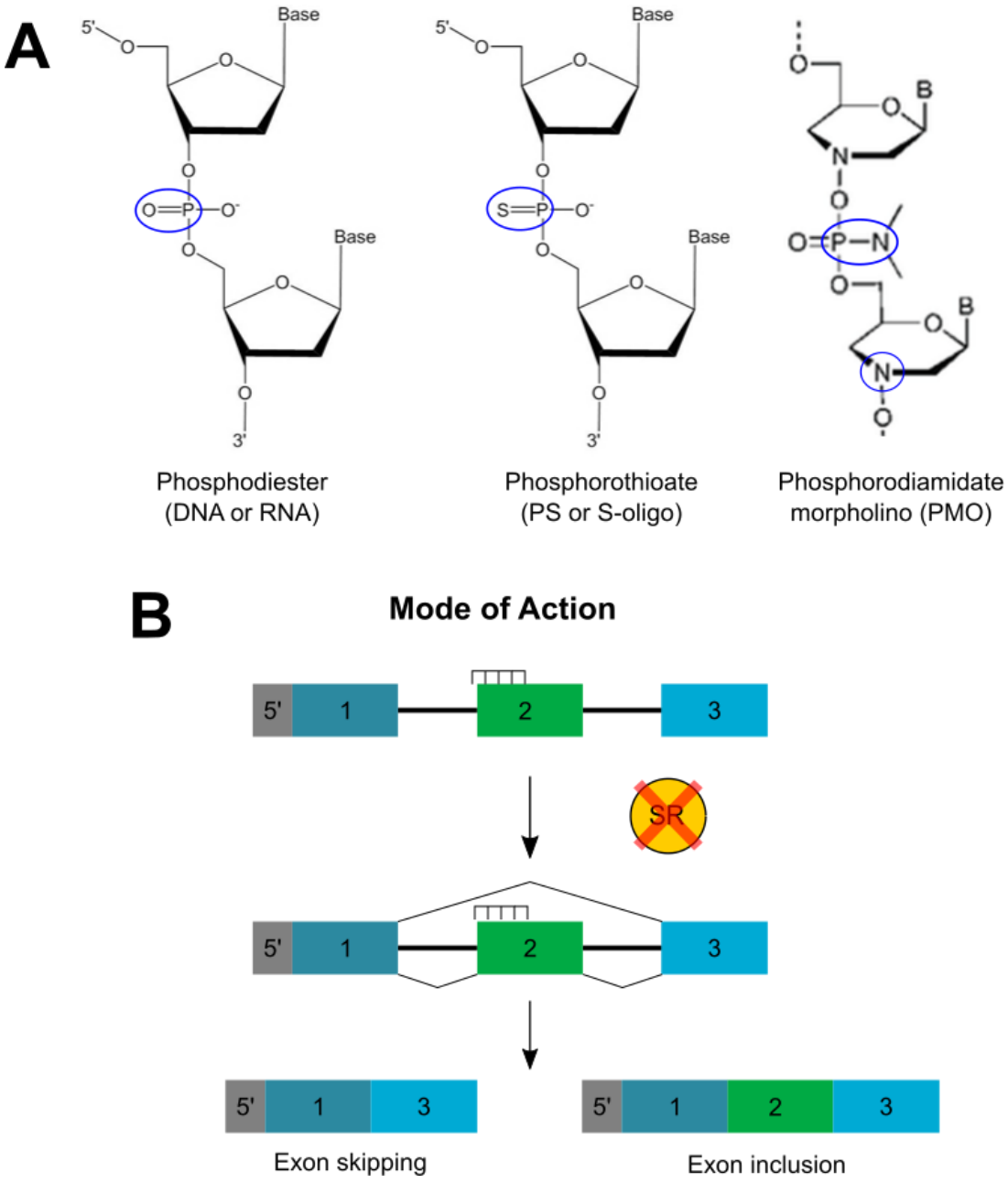


Figure 3.3 Antisense oligonucleotide chemistry and mode of action

A. The structure of a normal phosphodiester can be modified to produce ASO structures such as phosphorothioates and phosphodiamidate morpholinos. Sites of modification are circled in blue. **B.** SSOs are able to bind to intron-exon boundaries and block splice factors to cause exon skipping or inclusion.

On top of this the lack of charge means morpholinos do not significantly interact with proteins and assists with their easy delivery to biological models *in vitro* (Morcos, 2007). The method of delivery of ASOs does depend on the tissue and disease. Although PS ASOs can be delivered without a delivery moiety or reagent, high concentrations in the micromolar (μM) rather than nanomolar (nM) are required and the cost of such concentrations are quite high. In terms of formulation most ASOs can be made up in saline and indeed this is how they tend to be delivered in clinics (Shen and Corey, 2018). To counteract this transfection reagents can be to assist with delivery but these are associated with increased toxicity and are not translatable to the clinic. The morpholinos (Vivo-Morpholinos) used in this study have an octaguanidine dendrimer moiety that facilitates uptake and removes the requirement for transfection reagent facilitating in vivo delivery (Morcos, Li and Jiang, 2008). However, the recommended in vitro dosage is still relatively high at 1-10 μM as opposed to the ideal nM concentrations.

ASOs can be target nucleic acids using several mechanisms. Inhibition of translation via activation of RNase H activity or steric hindrance by microRNA or regulatory non-coding RNA inhibition, and modulation of alternative splicing via blockage of *cis*- or *trans*-acting elements (Aartsma-Rus and van Ommen, 2007; Schoch and Miller, 2017; Shen and Corey, 2018). SSOs work to encourage exon inclusion or exclusion by binding to pre-mRNA molecules and affecting splice factor recognition (Figure 3.3B). Development of an SSO requires identification of an appropriate disease linked splicing event, design and validation of several oligonucleotides near the intron-exon boundary to

identify the most efficient ones. Target sites can also be SRE such as ESEs and ISEs which have been shown to be able to significantly modulate the splicing of an exon (Aartsma-Rus and van Ommen, 2007).

Beta-thalassaemia was one of the first conditions to have SSOs used to restore normal phenotype (Suwanmanee *et al.*, 2002). Indeed much focus on therapeutic ASOs has been in the neurological conditions such as Duchenne muscular dystrophy (DMD) and famously spinal muscular atrophy (SMA) however interest in cancer treatment is beginning to increase (Bauman, Jearawiriyapaisarn and Kole, 2009). Mutations in the *survival motor neuron 1* (*SMN1*) gene are the cause of SMA. Although humans have a paralogue gene *survival motor neuron 2* (*SMN2*), the 11 nucleotides that vary in exon of the gene produce a truncated protein that is degraded. Therefore, patients are unable to produce sufficient SMA protein. Spinranza®, also called ISIS 396443, restores SMN level in patients by blocking an hnRNPA1/A2 ISS causing exon 7 inclusion in *SMN2* and leading to the synthesis of more functional protein. Intrathecal injections directly into the cerebrospinal fluid following lumbar puncture allows for optimum uptake into the target neuron sites (Prakash, 2017).

In terms of cancer, targets continue to be discovered and tested. Myeloid cell leukaemia-1 protein has two alternatively spliced isoforms, a long isoform that is anti-apoptotic and a shorter apoptotic isoform. ASOs were used in a skin basal cell carcinoma cell line to shift *Mcl-1* splicing to the short isoform and induce apoptosis (Shieh *et al.*, 2009). Some cancer related target proteins have been taken forward to clinical trials. The human oncogene eukaryotic translation

initiation factor 4E (eIF4E) is dysregulated in several cancers including colorectal cancer. The protein assists in the translation of cell survival genes. The ASO called ISIS183750, was initially used in cell lines and reduced eIF4E expression and proliferation. Success in the lab led to the ASOs against eIF4E being used in a clinical trial to treat patients with advanced colorectal cancer. The ASO was tested on its own and in combination with irinotecan chemotherapy (or its active component). Irinotecan works by interacting with DNA-topoisomerase I complexes disrupting DNA replication in S phase of cell cycle driving cells to apoptose. Both phase 1 and phase 2 patients had anti-proliferative profiles after both ASO and combination therapy. Samples collected from patients showed reduced eIF4E mRNA levels but protein remained unaffected. The research group suggested that the reason for these observations may be an increased binding of the ASO to the stroma or levels of eIF4E binding proteins. What did also clearly improve was survival with median progression-free survival of 1.9 months and median overall survival of 8.3 months based on an 11.4 month follow up period (Duffy *et al.*, 2016).

This results chapter aims to establish the evolutionary importance of exon 7b by analysing its sequence conservation in orthologues. In addition it will analyse the expression of exon 7b according to RNA sequencing data from the cancer genome atlas (TCGA) and the Encyclopedia of DNA Elements (ENCODE) consortium. Lastly it will demonstrate how the expression of exon 7b can be altered by splice switching oligonucleotides (SSO). These SSOs induce exon 7b skipping by blocking access to the splice sites for the exon.

3.2 *ERG* exon 7b is evolutionarily conserved in *ERG* orthologues

The conservation of ETS transcription factors in metazoa is well established (Degnan *et al.*, 1993). In the case of the ERG protein family predicted members have been identified in nematode and sea urchin species (Degnan *et al.*, 1993; Rizzo *et al.*, 2006). It is important to state at this point that most often conservation analysis was based on the identity with the ETS domain, though pointed domain conservation has also been analysed (Mackereth *et al.*, 2004).

With this in mind an analysis was carried out to determine the conservation of ERG exon 7b in order to determine when it may have emerged during evolution, as this would indicate when its presence became biologically significant. Initially an alignment of amino acid sequences of the main ERG isoforms identified in the NCBI database was performed to determine which isoforms include and exclude ERG exon 7b (Figure 3.4). It was found that all isoforms except ERG2, ERG5 and ERG8 included exon 7b. ERG6 and ERG7 encode for proteins that have a truncated C-termini lacking the ETS domain and therefore most likely lack DNA binding function. ERG4 and ERG5 had truncated N termini but the truncation does not affect the pointed domain.

To aid subsequent analysis predicted serine residues that are known to be phosphorylated were identified to establish which ones arose earliest, are best conserved and therefore of most potential functional importance. These serine residues are annotated with an asterisks in all alignments. Two of these residues are located in the transactivation domain with one in exon 7b. Within ERG isoforms two of the six serines are present across all isoforms analysed

here. For three residues the lack of presence of the relevant amino acids rather than substitutions is the reason why they are not present in all isoforms. One residue sees an S to R substitution in ERG7 although the alignment is not particularly strong for this part of the sequence.

Upon this analysis all subsequent alignments to human (*Homo sapiens*) were to ERG3 (termed hERG) which includes exon 7b and produced a protein 486 amino acids in length. Next an alignment of the amino acid sequence of hERG and various vertebrate species was performed to establish the conservation of exon 7b (Figure 3.5). Isoforms similar in size to hERG3 were chosen in the vertebrate species to aid comparison. Evidence suggests that ERG exon 7b is highly conserved at the amino-acid level in vertebrates as are 4 of the 6 serine residues identified, including the residues in the transactivation domain. In addition the docking site for ERK (DEF) domain, an ERK and MAPK docking site, located in exon 7b was highly conserved. The DEF domain consensus sequence is FXFP where X represents any amino acid residue (Jacobs *et al.*, 1999). This domain is located upstream of the serine residue in exon 7b and is thought to be important in facilitating its phosphorylation (Selvaraj, Kedage and Hollenhorst, 2015). In summary evidence points to ERG exon 7b amino-acids being strongly conserved throughout evolution.

With knowledge that ERG conservation has been established from humans to echinoderms we aimed to identify exon 7b in species within this phylum. Exon 7b like sequences were identified in two species, the purple sea urchin (*Strongylocentrotus purpuratus*) and crown-of-thorns starfish (*Acanthaster planci*). Although the serine in exon 7b was not conserved the DEF domain was,

establishing this as a potentially important site within exon 7b and the transactivation domain overall (Figure 3.5). Similarly the serine upstream of exon 7b was present in starfish but not sea urchin ERG. Therefore it has been established that exon 7b was present as far back as the echinoderm lineage.

Finally the overall conservation of ERG was investigated by scrutinizing the sequence conservation of the ETS domain in orthologous protein sequences present in the most primitive species. Figure 3.7 shows ERG-like proteins were found in a species as primitive as *Amphimedon queenslandica*, a marine sponge as well as two worm species (*Priapulus caudatus* and *Saccoglossus kowalevskii*) and a brachiopod (*Ligula anatina*). This provides further evidence of ERG conservation in metazoan possibly back to a common ancestor species. An *ERG* gene has not been reported in sponges and was identified in this analysis.

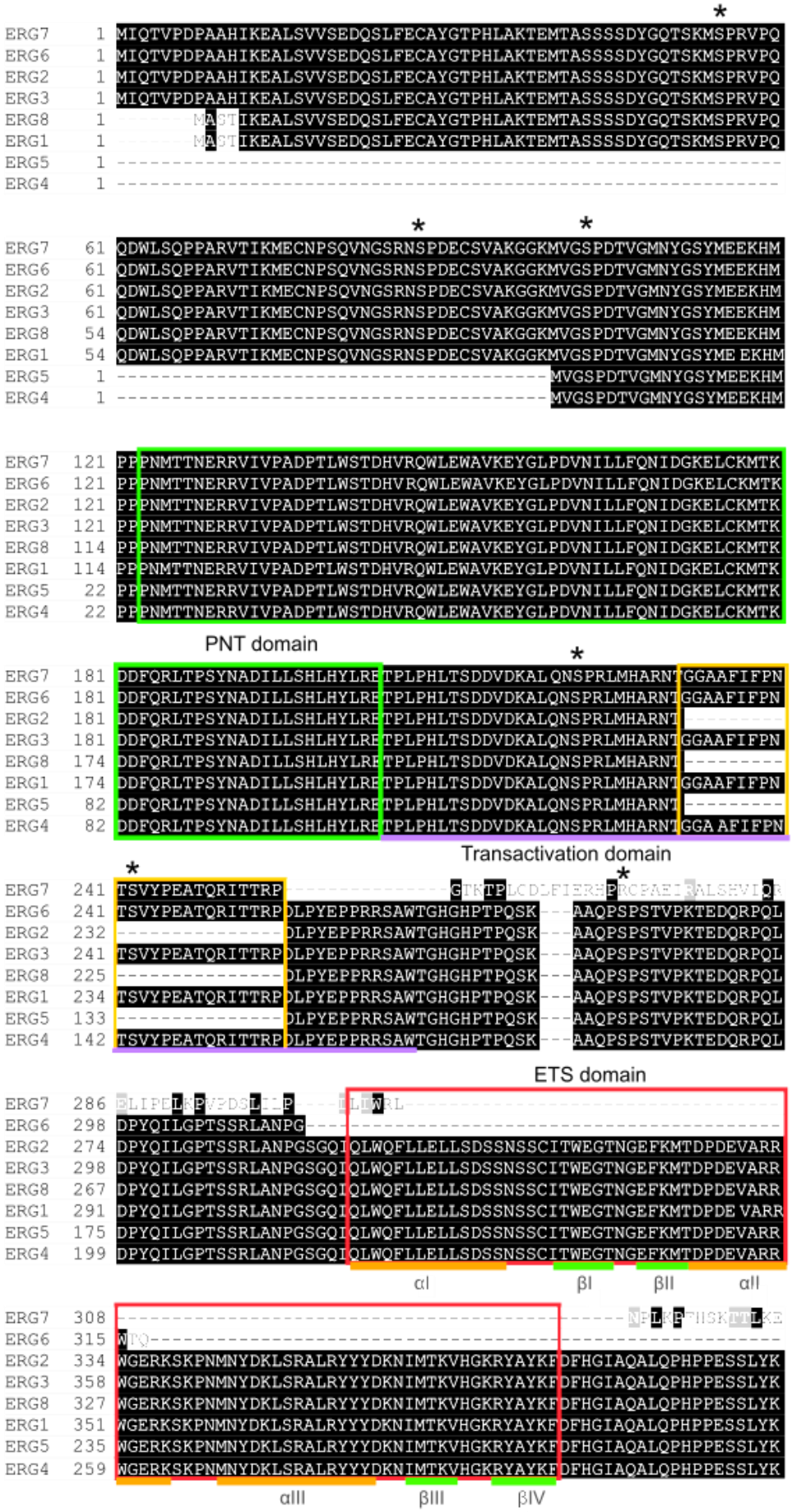

PNT domain
Transactivation domain
ETS domain
αI
βI
βII
αII
αIII
βIII
βIV

Figure 3.4 Amino acid alignment of human ERG protein isoforms

A MUSCLE amino acid alignment of the main human ERG isoforms. NCBI protein database accessions numbers are as follows: ERG1 NP_891548.1, ERG2 NP_004440.1, ERG3- NP_001129626.1, ERG4 NP_001129627.1, ERG5 NP_001230358.1, ERG6 NP_001230361.1, ERG7 NP_001278320.1 and ERG8 NP_001317954.1. Exon 7b (orange box) is excluded in three of the isoforms. Functional domains are indicated: pointed domain (green box), transactivation domain (underlined in purple) and ETS domain (red box). The sequences of the three alpha helices (underlined in orange) and four beta sheets (underlined in green) of the ETS domain are also labelled. Homologous regions with identical amino acid residues are box-shaded black and conserved amino acid substitutions in grey. Asterisks (*) are serine residues predicted to be phosphorylation site.

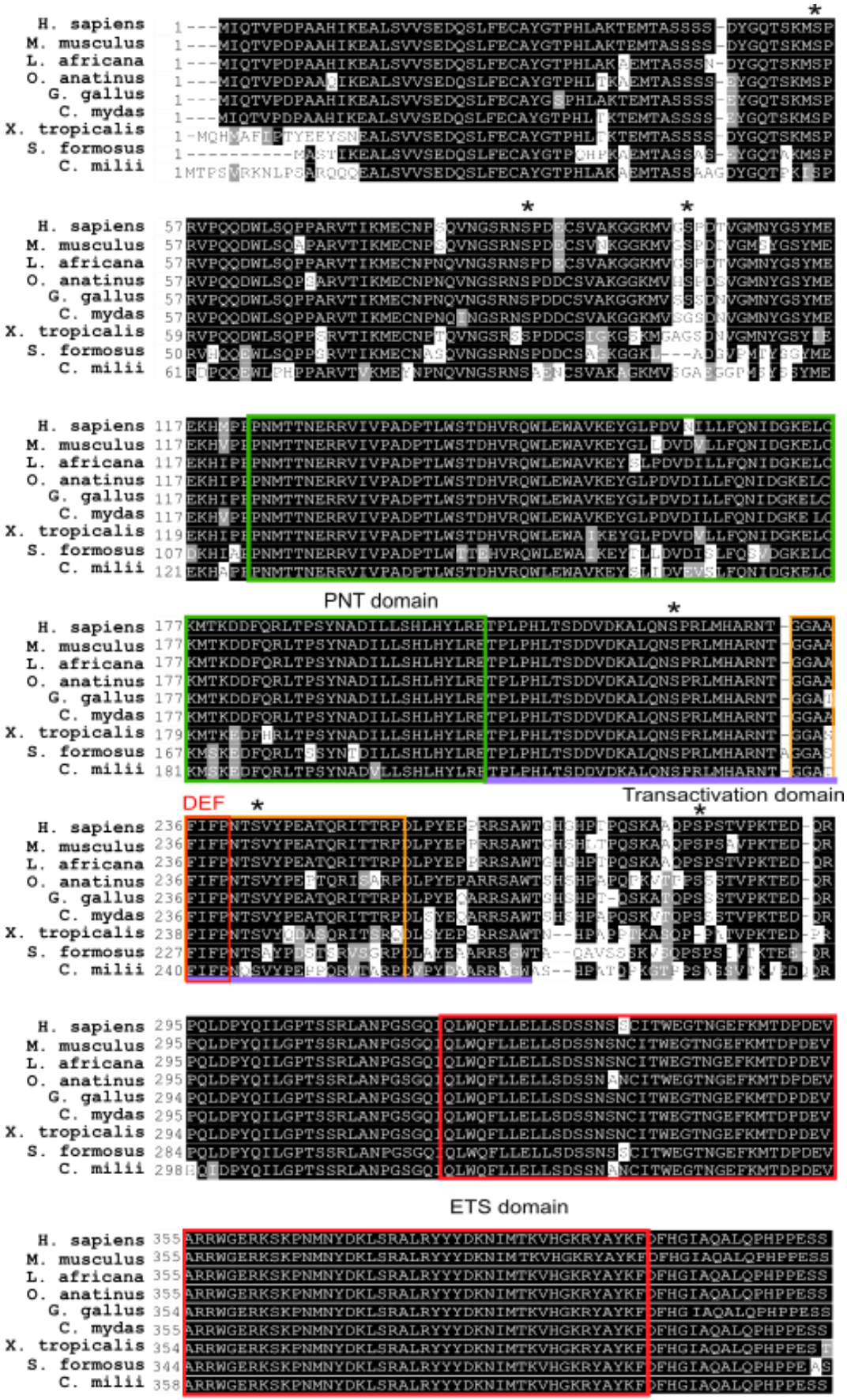

H. sapiens
M. musculus
L. africana
O. anatinus
G. gallus
C. mydas
X. tropicalis
S. formosus
C. milii
PNT domain
DEF
Transactivation domain
ETS domain

Figure 3.5 Amino acid alignment of ERG in representative vertebrate species

A MUSCLE amino acid alignment of ERG (corresponding to human ERG3] from different vertebrate species. NCBI protein database accessions numbers are as follows: *Homo sapiens* (human) NP_001230357.1, *Mus musculus* (mouse) XP_489769.1, *Loxodonta africana* (elephant) XP_010596953.1, Ornithorhynchus anatinus (platypus) XP_007665373.1 *Gallus gallus* (chicken) XP_015155738.1, *Chelonia mydas* (green sea turtle) XP_007064464.1, Xenopus tropicalis (clawed frog) XP_012812010.1, *Scleropages formosus* (Asian arowana fish) XP_018611073.1 and *Callorhinchus milii* (ghost shark) XP_007907661.1. Exon 7b (orange box) contained a DEF domain (labelled and highlighted by a red box) which suggests a MAPK/ERK docking site. Functional domains are indicated: pointed domain (green box), transactivation domain (underlined in purple) and ETS domain (red box). Homologous regions with identical amino acid residues are box-shaded black and conserved amino acid substitutions in grey. Asterisks (*) are serine residues predicted to be phosphorylation sites.

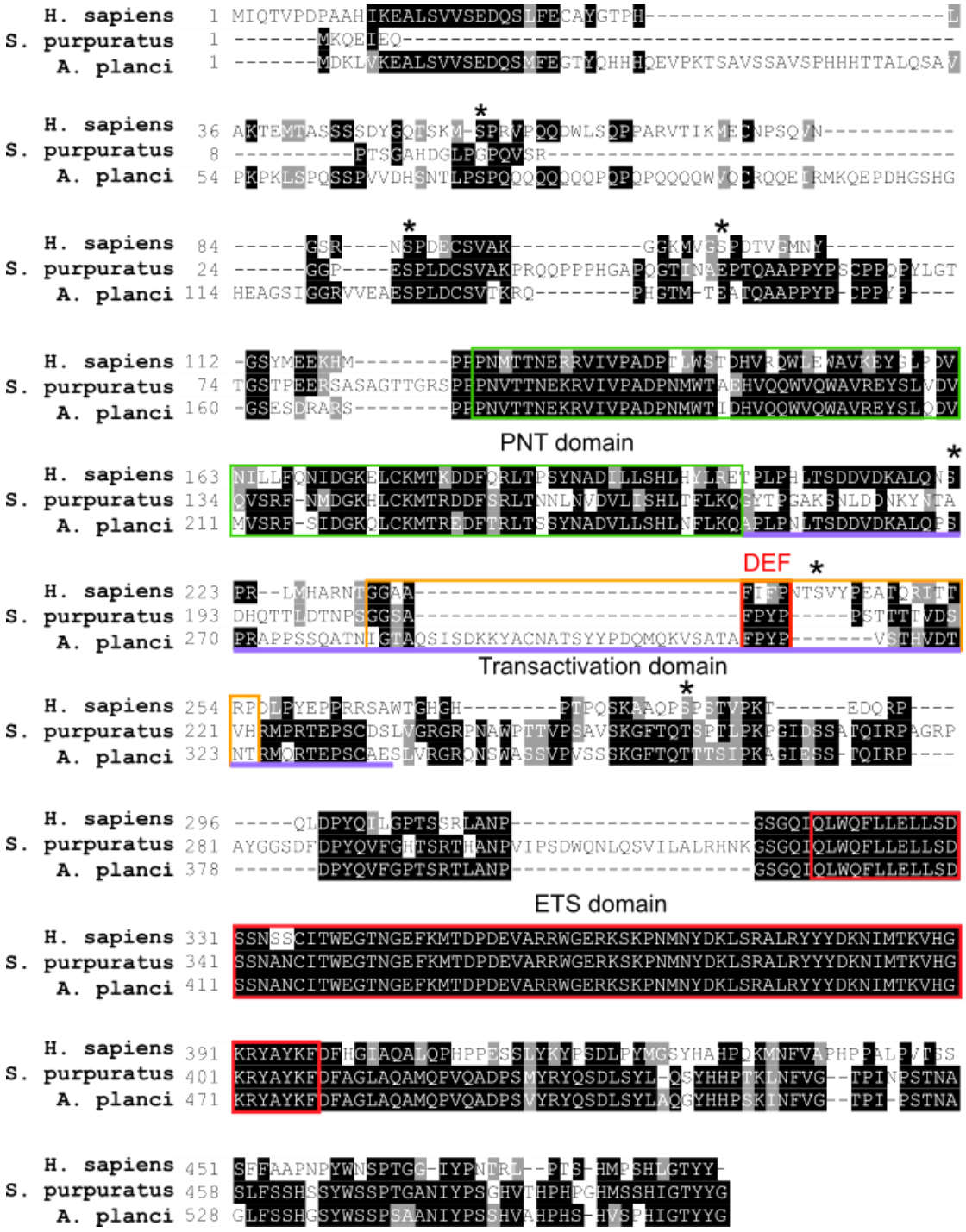

H. sapiens
S. purpuratus
A. planci
PNT domain
DEF
Transactivation domain
ETS domain

Figure 3.6 Amino acid alignment of human ERG and echinoderm ERG isoforms

A MUSCLE amino acid alignment of ERG isoforms from human versus two echinoderm species. NCBI protein database accessions numbers are as follows: *Homo sapiens* (human) NP_001230357.1, *Strongylocentrotus purpuratus* (purple sea urchin) and *Acanthaster planci* (crown-of-thorns starfish). Exon 7b (orange box) contained a DEF domain (labelled and highlighted by a red box) which is a MAPK/ERK docking site. Functional domains are indicated: pointed domain (green box), transactivation domain (underlined in purple) and ETS domain (red box). Homologous regions with identical amino acid residues are box-shaded black and conserved amino acid substitutions in grey. Asterisks (*) are serine residues predicted to be phosphorylation sites.

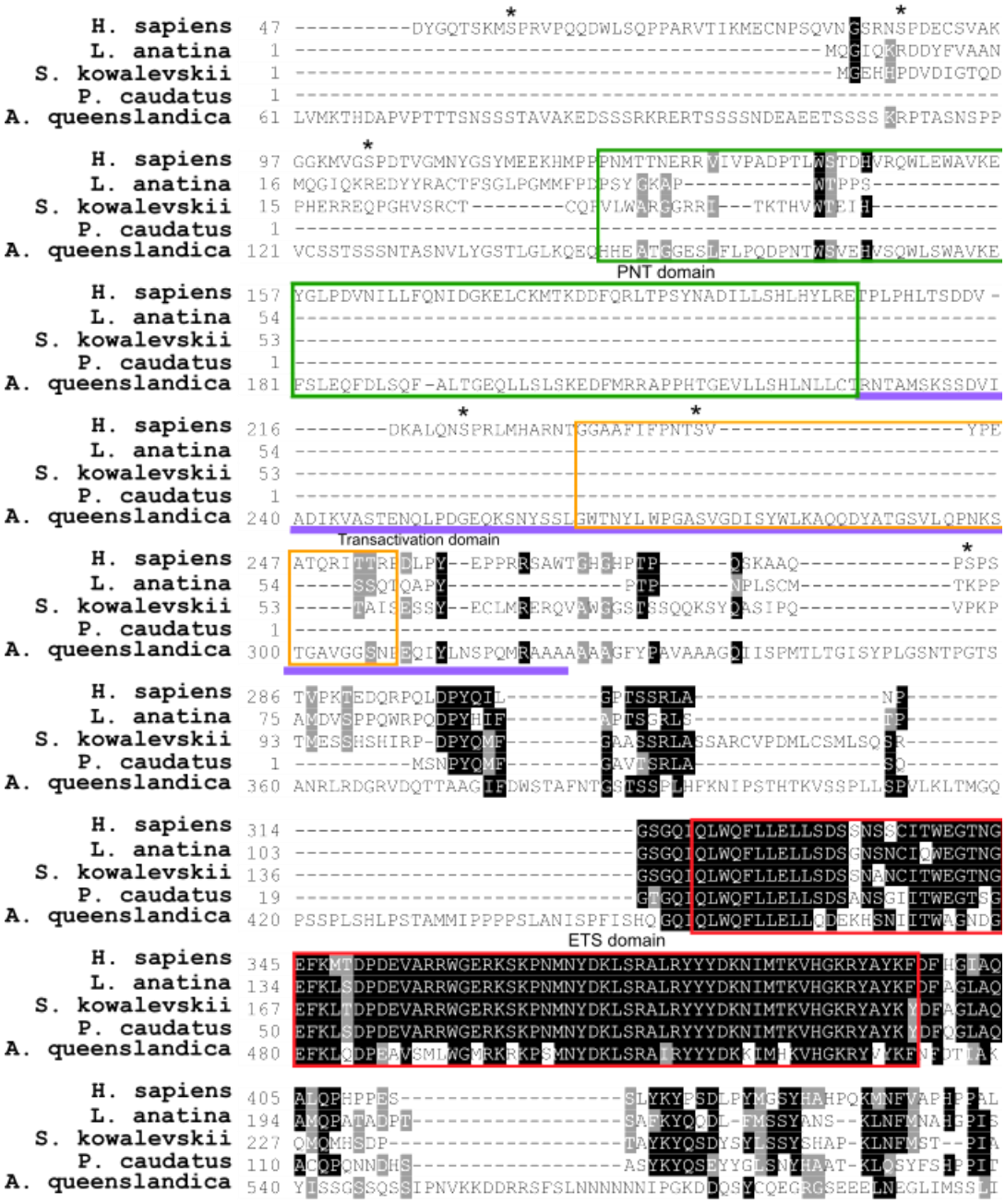

PNT domain
Transactivation domain
ETS domain

Figure 3.7 Amino acid alignment of ERG in primitive species

A MUSCLE amino acid alignment of human ERG versus ERG-like proteins in four metazoan species. NCBI protein database accessions numbers are as follows: *Homo sapiens* (human) NP_001230357.1, *Ligula anatine* (brachiopod) XP_013389107.1, *Priapulus caudatus* (worm) XP_014673650.1, *Saccoglossus kowalevskii* (acorn worm) XP_006822774.1 and *Amphimedon queenslandica* (marine demosponge) XP_019852323.1. Exon 7b (orange box) contained a DEF domain (labelled and highlighted by a red box) which is a MAPK/ERK docking site. Functional domains are indicated: pointed domain (green box), transactivation domain (underlined in purple) and ETS domain (red box). Homologous regions with identical amino acid residues are box-shaded black and conserved amino acid substitutions in grey. Asterisks (*) are serine residues predicted to be phosphorylation sites.

3.3 ERG exon 7b expression in prostate cancer

The Cancer Genome Atlas (TCGA) is a publically available dataset of comprehensive genomic analysis of 33 types of cancer. The collaborative project between the National Cancer Institute (NCI) and National Human genome Research Institute (NHGRI) uses tumour and normal matched tissues from over 10,000 patients.

Using data from TCGA, kindly provided by Dr Eduardo Eyras (Pompeu Fabra University & ICREA, Spain), the PSI of exon 7b in prostate adenocarcinoma (PRAD) patients was determined. PSI is a measure of the extent by which an exon is spliced in where 1 is complete inclusion and 0 complete exclusion. For the dataset analysed the cohort was made up of 243 samples and for each patient a normal tissue sample was taken alongside the tumour sample to allow for comparison. The tumours were given T scores which assess the size and extent of the tumour, where T2 is a tumour contained in the prostate, T3 a tumour that is extending out of the border of the prostate and T4 for tumours that have spread to other organs nearby.

The mean PSI of ERG exon 7b in tumour patients was 0.711 with a range of 0-1. Compared to normal tissue there is no significant difference in PSI in tumour (Figure 3.8A, p=0.233) or between the various tumour grades (Figure 3.8B, Figure 3.8p=0.4113). However due to the cohort size of the higher tumour grade group being smaller (5 patients for T4) the data may not be fully reflective of the population.

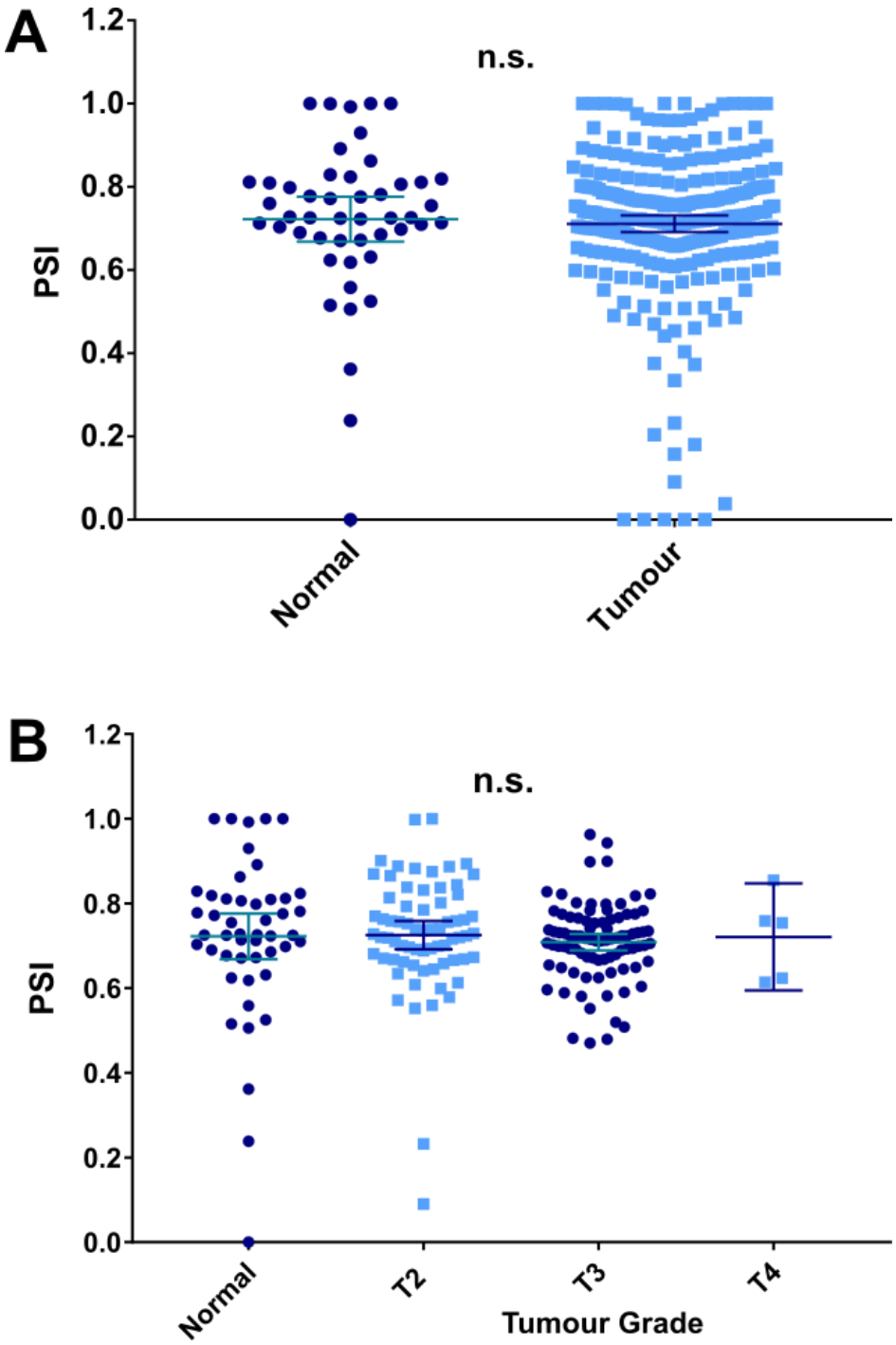


Figure 3.8 ERG exon 7b PSI values in prostate adenocarcinoma patients

A. PSI values for ERG exon 7b in normal and PRAD tumour tissue from the TCGA. **B.** PSI values for ERG exon 7b according to tumour grade. The middle line represents the mean PSI value and error bars show 95% confidence intervals (95%CI). Number of patient samples= 243

3.4 ERG exon 7b is alternatively spliced in several malignant cell lines

ERG is expressed in a number of haematological tissues but the rate of exon 7b inclusion has not been examined in detail. To determine expression of ERG exon 7b isoforms in a range of haematological cell lines a standard PCR was performed to determine exon 7b inclusion rates.

Of the seven cell lines tested four expressed ERG isoforms with and without exon 7b. K562 cells are chronic myeloid leukaemia derived cells and have previously been induced to express ERG after treatment with phorbol myristate acetate (PMA) as confirmed in this investigation (Rainis et al., 2005, Figure 3.9). When treated with PMA the lymphoblast like TK6 cell line did not express ERG. Most cell lines favoured exon 7b inclusion, as indicated by the PSI values, but an exception to this was the U937 cell line with a 0.43 PSI value suggesting almost equal ERG exon 7b isoform expression (Figure 3.9).

To determine if K562 expression could be detected by more sensitive means RNA-seq data from the ENCODE consortium database was analysed via the UCSC Genome browser. In addition to K562 two additional cells lines were chosen, HUVEC (human umbilical vein endothelial cells), K562 and human embryonic stem cells (hESC). HUVEC cells have been used to study ERG in previous studies (Yuan *et al.*, 2009; Birdsey *et al.*, 2012). In the 2005 study by Rainis *et al.* the expression of ERG in K562 was determined using PCR which is not as sensitive as RNA-seq. Expression of ERG exon 7b was considerably high in HUVEC and hESC cells but was not detected in K562 (Figure 3.10), consistent with findings shown in Figure 3.9.

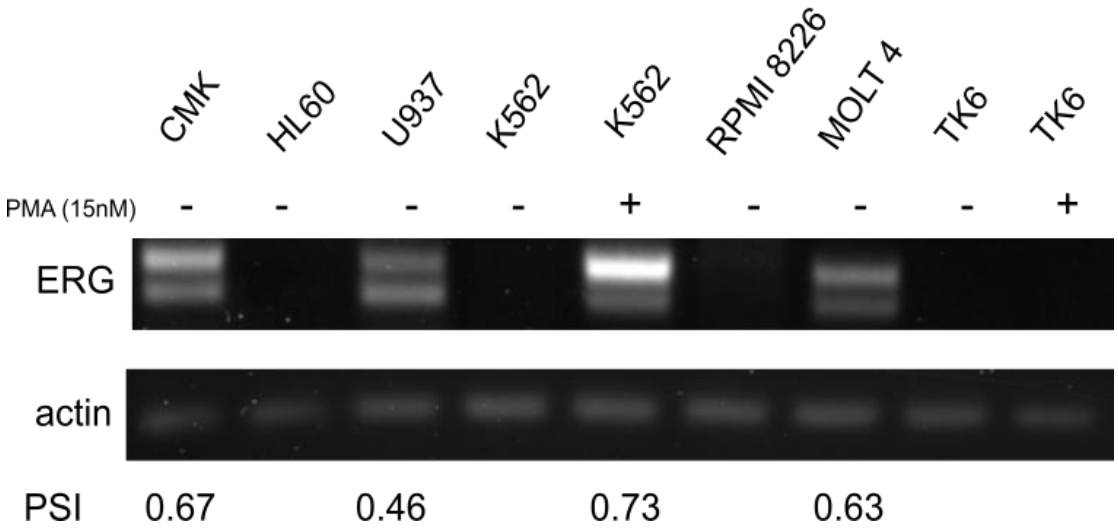


Figure 3.9 ERG exon 7b expression in haematological cell lines

Expression of ERG exon 7b was determined in CMK (acute megakaryocytic leukaemia), U937 (histiocytic lymphoma), HL-60 (acute myeloid leukaemia), K562, RPMI 8226 (multiple myeloma), MOLT4 (T-cell leukaemia) and TK6 (lymphoblast) human cell lines. Cell pellets were used for RNA isolation and cDNA synthesis. K562 and TK6 cells were treated with phorbol myristate acetate (PMA) to induce ERG expression. A standard PCR was carried out using primers that hybridise in the constitute exons 6 and 8 to detect skipping of exon 7b in mRNA. An actin cDNA control. PSI values indicate the degree of exon 7b skipping – lower values correlate with more skipping.

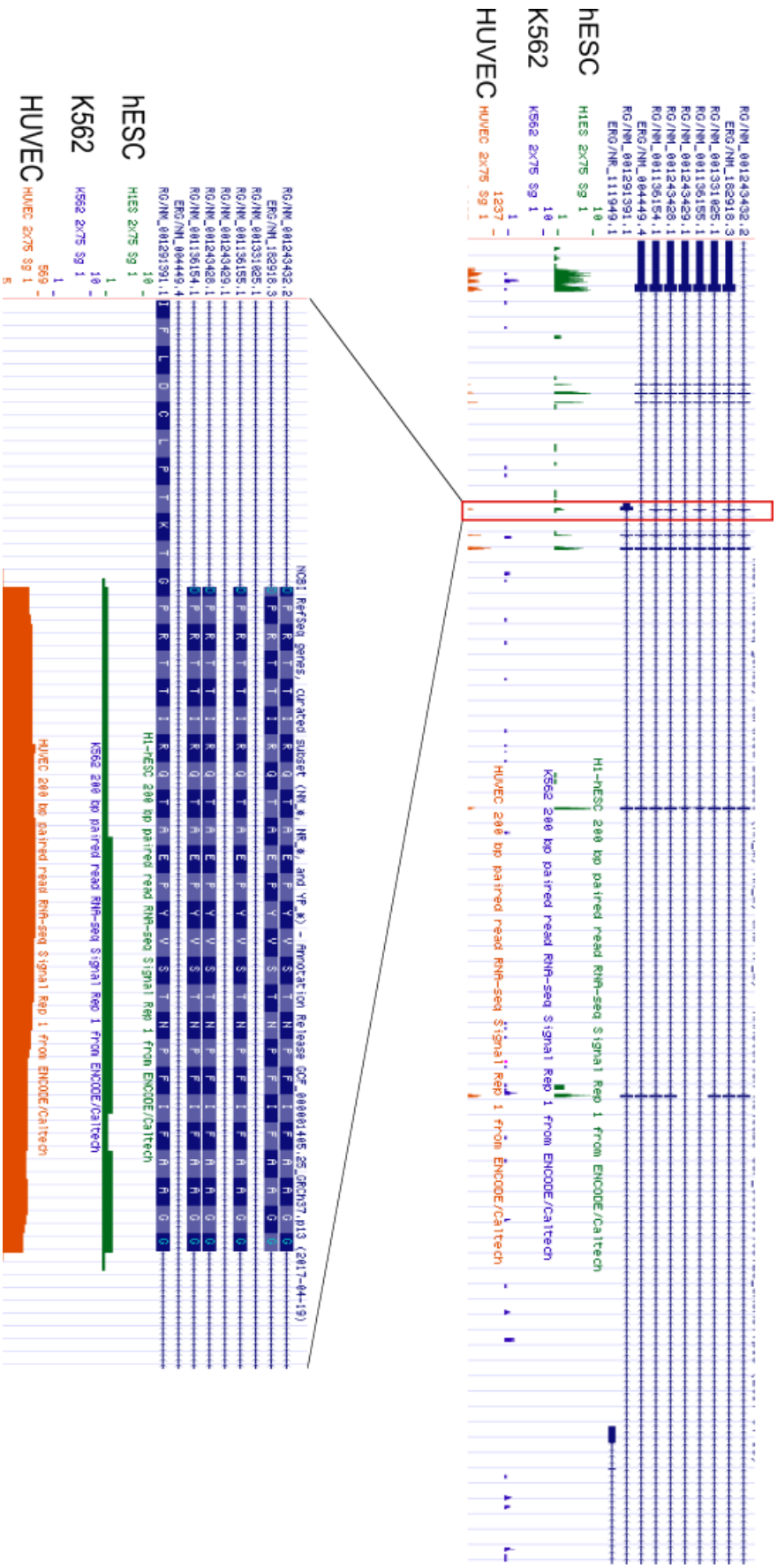

hESC
K562
HUVEC
RG/NM_001243432.2
ERG/NM_182918.3
RG/NM_001331825.1
RG/NM_001136155.1
RG/NM_001243429.1
RG/NM_001243428.1
RG/NM_001136154.1
ERG/NM_004449.4
RG/NM_001291391.1
ERG/NR_111949.1
H1-hESC 200 bp paired read RNA-seq Signal Rep 1 from ENCODE/Caltech
K562 200 bp paired read RNA-seq Signal Rep 1 from ENCODE/Caltech
HUVEC 200 bp paired read RNA-seq Signal Rep 1 from ENCODE/Caltech
NCBI RefSeq genes, curated subset (NM_* NR_* and YP_*) - Annotation Release GCF_000001405.25_GRCh37.p13 (2017-04-19)

Figure 3.10 ERG RNA expression in human cell lines

The top panel shows is a screenshot of UCSC Genome browser data for ERG. A density graph of signal enrichment for RNA sequencing data from the ENCODE project for hESC K562 and HUVEC cell lines shows the read density in reads per million (RPM) on y-axis and the nucleotide position on the x-axis. The main ERG isoforms of ERG are shown above the enrichment graphs, and include isoforms that do not express exon 7b. The bottom panel is a close up of the data for exon 7b with the exon sequence at the top and the signal for each base position in each cell line.

3.5 ERG exon 7b can be skipped with splice switching oligonucleotides (SSOs) targeted to the 5´ and 3´ splice sites

To facilitate the functional investigation of ERG exon 7b, detailed in the next chapter, two SSOs were designed with the assistance of Dr Lee Spraggon (Memorial Sloan Kettering, USA) against the splice sites of the exon in order to induce its skipping. The position and sequences of these SSOs in relation to the exon and its flanking introns are shown in Figure 3.11. In order to account for effects of SSO transfection a control SSO was included in all experiments. This SSO is designed to correct the mutation found at position 705 in the beta-globin mRNA present in human beta-thalassaemia patients' reticulocytes. A PBS control equivalent to the maximum volume of SSO transfected was also included.

The recommended range of SSO efficacy *in vitro* is 1-10 μM (GeneTools LLC, Oregon USA). Therefore with this in mind a titration of concentrations was carried out. Initially SSO were transfected into the MOLT4 cell line, a T-cell acute lymphoblastic leukaemia cell line. This cell line has previously been used to study the function of *ERG* (Thoms *et al.*, 2011) and as seen in Figure 3.9 has a predisposition for exon 7b inclusion.

Figure 3.12 A shows that the maximal dose of 10 μM is required to attain near complete exon 7b skipping. More effective of the two *ERG* SSOs is E7b3 with a PSI of 0.06 achieved at 10 μM as compared to the 0.34 PSI of E7b5. As expected the control SSO did not significantly alter exon 7b skipping.

As a dose of 10 μM is quite high an alternative cell line model was examined. Knowing that ERG is expressed in developing bone (Iwamoto et al., 2001, Iwamoto et al., 2000), the possibility that ERG may be expressed in the MG63 osteosarcoma cell line was explored. The cell line was found to express both exon 7b skipped and included isoforms of ERG and this is the first reporting of MG63 being used to study ERG. Therefore an SSO titration was also carried out in this cell line.

In contrast to the MOLT4 cell line much lower doses of ERG SSO were required to induce exon 7b skipping (Figure 3.12B). At 3 μM complete skipping of exon 7b was achieved with E7b3 SSO (PSI=<0) whereas near complete skipping of exon 7b was observed with E7b5 SSO (PSI=0.22). More impressively even at 1 μM the E7b3 SSO showed near complete exon 7b skipping (PSI=0.11). The dose of 3 μM is sufficient to cause skipping confirmed at the protein level after 72 and 96 hours of treatment (Figure 3.12C).

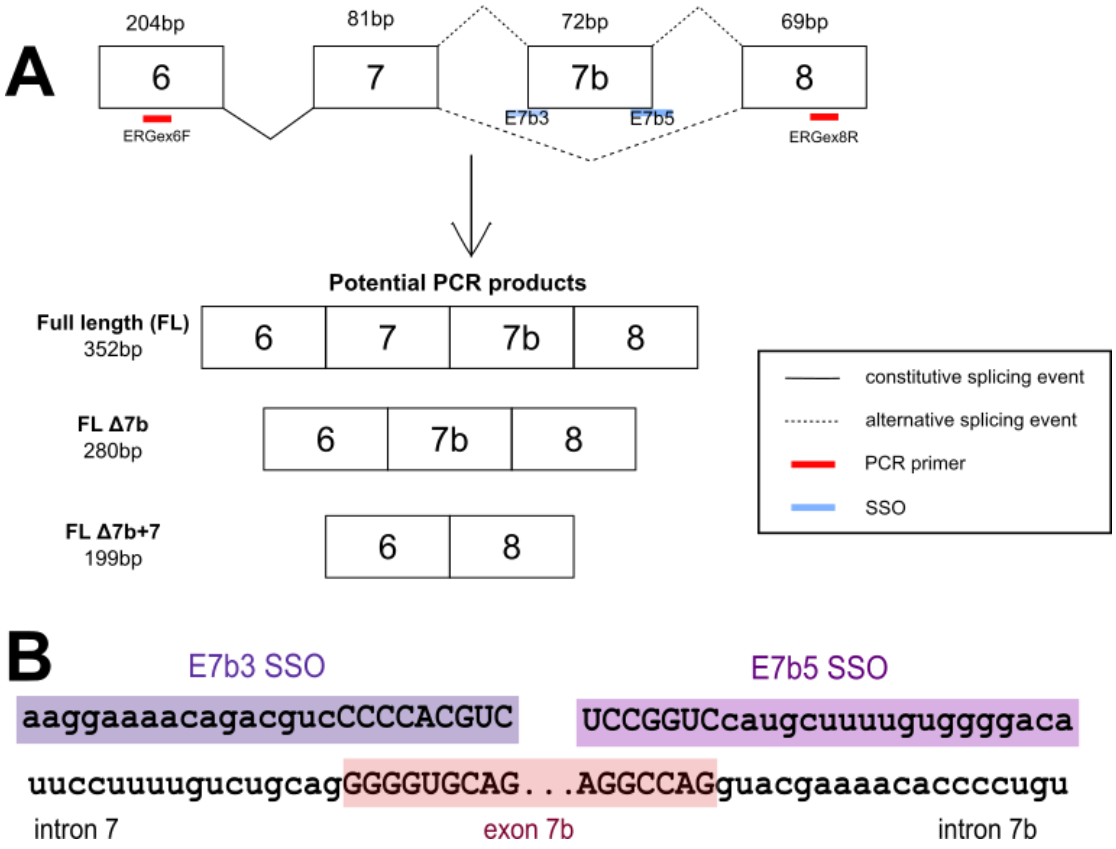


Figure 3.11 Schematic of ERG exon 7b splicing pattern and SSO sequences used in this study

A. Exon 6 to 8 of ERG and the sizes of each exon in base pairs (bp) are shown. The red bars show PCR primer positions and the light blue boxes the position of the SSOs. Three potential PCR products are produced, the full length (FL) product is 352bp long, the full length product with exon 7b skipped (FLΔ7b) is 280bp. The rarer skipping of both exon 7 and 7b (FLΔ7b+7) produces an even smaller 199bp PCR product. **B.** SSO sequences used in this study are in light purple and the boundaries of exon 7b in light pink. Intronic sequence is in lowercase and exon sequence in uppercase letters.

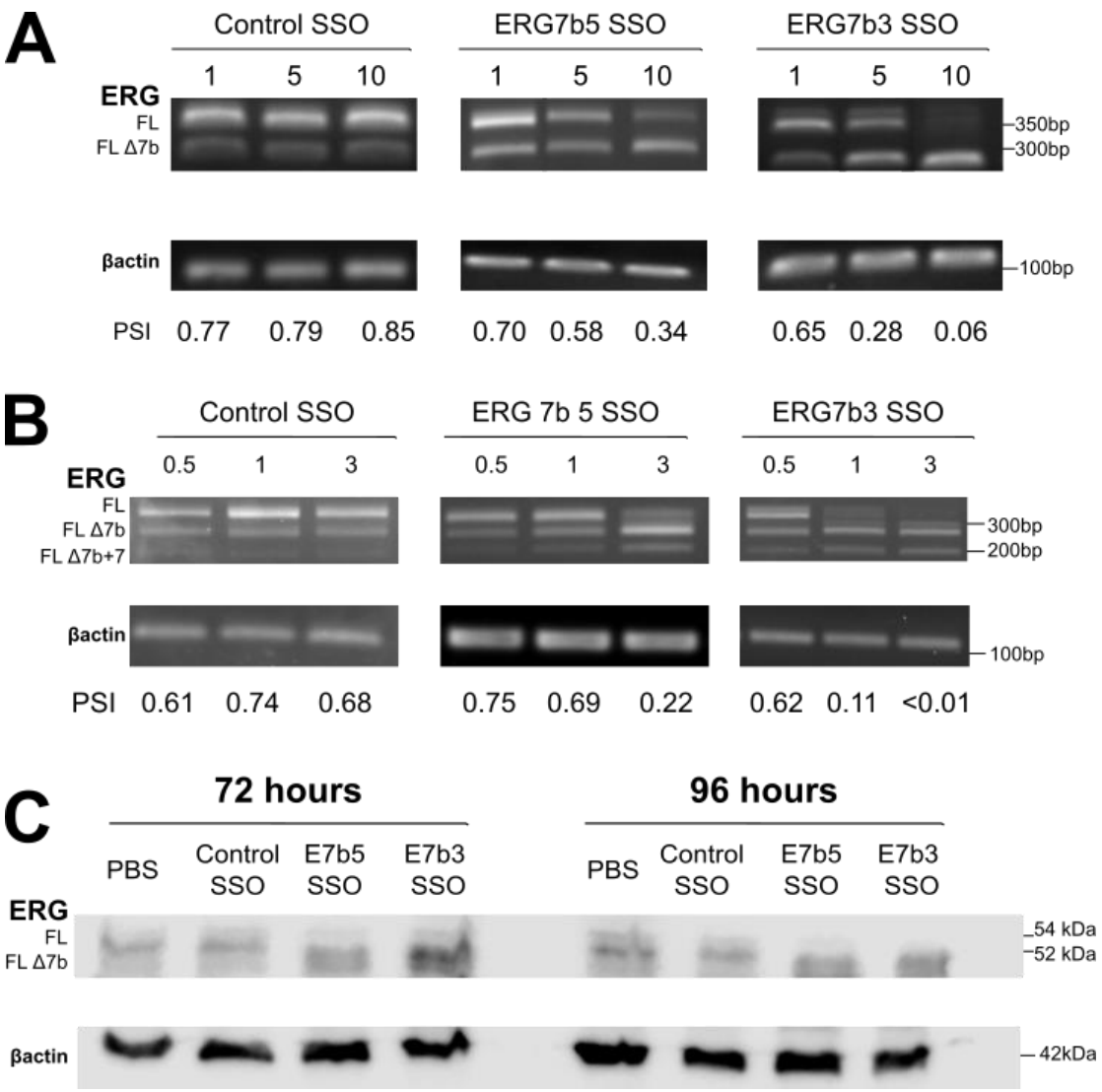


Figure 3.12 Titration of ERG exon 7b SSOs in MOLT4 and MG63 cells

A. MOLT4 cells were treated with control, E7b5 and E7b3 SSO at 1, 5 and 10 μM for 48 hours. The PCR show exon 7b skipping is detectable at all doses but is best achieved at the highest dose of 10 μM. **B.** MG63 cells were treated with control, E7b5 and E7b3 SSO at 0.5, 1 and 3 μM for 48 hours. The PCR show exon 7b skipping is detectable at 1 and 3 μM. The level of exon 7 b skipping with E7b3 is much higher even at the 1 μM dose. N=2 **C.** MG63 cells were treated with control, E7b5 and E7b3 SSO at 3 μM for 72 and 96 hours. An ERG western blot was performed using whole cell lysate extracted after treatment. Exon 7b encodes 24 amino acids which give a difference of 2kDa in protein size between full length and exon 7b excluded ERG proteins. Actin is included as a loading control. N= 3.

3.6 Summary of findings

A search of available isoforms on the NCBI GenBank database revealed that most ERG isoforms in humans include exon 7b with exception of ERG2, ERG5 and ERG8. Analysis established exon 7b conservation in echinoderms suggesting an evolutionarily conserved importance in ERG biological function. Exon 7b was one of the first alternative exons described along with exon 7 (Duterque-Coquillaud *et al.*, 1993). These two exon encode for part of the AD; this domain is known to be required for ERG's transcriptional activity and protein-protein interactions (Siddique *et al.*, 1993; Carrère *et al.*, 1998). The inclusion of the 24 amino-acids encoded by exon 7b is likely to modify ERG function in different developmental, physiological and disease contexts.

The presence of a DEF domain in exon 7b suggests that it is involved in ERK/MAPK signalling pathways. This domain may facilitate post-translational modification at serine and threonine residues. The observed conservation of the DEF domain and serine residues, particularly the flanking serine residues in metazoa in ERG suggests that these are similarly important to modulating ERG transcriptional activity. ERK phosphorylates ERG at serine 215 which is upstream to the DEF domain in exon 7 (Selvaraj, Kedage and Hollenhorst, 2015). Serine 215 is completely conserved in the metazoan sequences analysed in this study further suggesting that these two sites have important biological functions.

In general there seems to be a preference to include exon 7b in the majority of cell lines examined. Of the cell lines for which RNA-sequencing data is available HUVEC cells showed the highest levels of exon 7b expression. The main

exception was the U937 cell line in which exon 7b was predominantly skipped, contrary to a previous study (Rainis *et al.*, 2005) in which ERG was not seen to be expressed in U937.

Tumours for PRAD patients from the large TCGA cohort showed a shift to inclusion of *ERG* exon 7b as disease stage progressed confirming the findings of Hagen et al. (2014). Exon 7b encodes for part of the AD and truncation of this domain reduces transcriptional activity; therefore the conservation and inclusion of the exon may be important in modulating ERG's transcriptional activities. Increased inclusion in cancer may be linked to increasing the proportion of transcriptionally active ERG isoforms, contributing to disease progression.

By pushing splice isoform ratios in one direction (i.e. skipping or inclusion) the contribution of exon 7b to biological function can now be assessed. However to design SSOs with reasonable efficacy can be challenging as the sequences that are blocked may not modulate the splicing machinery adequately. In this study SSOs designed against the splice sites of *ERG* exon 7b successfully caused near complete exon skipping, detectable in the mRNA of MOLT4 and MG63 cell lines. Better efficacy was observed with the 3´ splice site targeting SSO suggesting that the design of this SSO comparatively better reduced the spliceosome's ability to include the exon. The SSO dose that induced exon skipping between the two cell lines differed with a higher dose required for the MOLT4 cells. This may be due to the MOLT4 cells being in suspension versus

the adherent MG63 cells. Furthermore the abundance of the mRNA in the two cell lines may vary as well.

In summary this chapter has demonstrated the importance of ERG exon 7b through its evolutionary conservation. SSOs are now available to reduce exon 7b inclusion, providing a valuable experimental tool to study the biological functions of this exon. The next chapter will describe a functional analysis of exon 7b, making use of these SSOs.

4 Function of ERG exon 7b splice isoforms

4.1 Introduction

As a transcription factor the function of ERG is to regulate gene expression. The profile of its target genes will vary according to the cell type that ERG is expressed in and is likely to be controlled by external stimuli as well. Furthermore the alternative splicing of ERG creates different protein isoforms with, presumably, different abilities to regulate transcription. The latter is particularly relevant in cancer where dysregulated signalling pathways are driven by the overexpression of oncogenes and tumour suppressor genes such as ERG.

Most studies that have looked at ERG isoform function in cancer have focussed on the *TMPRSS2-ERG* fusion variants. Overall the presence of these isoforms is hypothesised to drive the development and progression of prostate in an androgen responsive manner as the overexpression of the fusion gene leads to the overexpression of oncogenic ETS target genes such as *CXCR4* and *ADAM metallopeptidase with thrombospondin type 1 motif 1* (*ADAMST1*) (Seth and Watson, 2005; Carver *et al.*, 2009). *TMPRSS2-ERG* fusion isoforms can be type I (which encode full length ERG that has an ETS domain) or type II (encode ERG without and ETS domain) (Hu *et al.*, 2008). Of the two *ERG* exon 7b is included in type I variants only. Changes in motility, proliferation, invasion and apoptosis both in vitro and in vivo has been shown to be affected by the combination of 5´ fusion junction and coding exons of *TMPRSS2-ERG* variants and generally the isoforms that include exon 7b were more oncogenic (J. Wang *et al.*, 2008; Zammarchi, Boutsalis and Cartegni, 2013; Urbinati *et al.*, 2015).

More recently a study by Hoesel *et al.* (2016) studied native ERG isoform localisation and function focussing on the role of the nuclear localisation sequence (NLS), nuclear export sequence (NES) and ETS domain. Findings revealed that all isoforms could shuttle between nuclear and cytosolic compartments but the ERG8 isoform that lacked the NLS and ETS domain localised to the cytosol mainly whereas other isoforms were predominantly nuclear. In addition ERG8 may act as a transcriptional inhibitor of the active ERG isoforms studied (ERG1-4). Finally qRT-PCR was used to determine the differential expression of isoforms in four cancer cell lines vs non- cancer cells. Non-cancer cells had a greater ratio of active ERG isoforms to ERG8 compared to cancer cell lines.

With an established body of evidence that ERG splice isoforms have variable biological functions the aim of this chapter is to establish the biological function of the ERG exon 7b in a cancer cell line model. MG63 osteosarcoma cell line SSOs will be used to induce ERG exon 7b skipping after which cell growth, proliferation, apoptosis, invasion and migration will be assessed. In addition as a measure of cell differentiation *alkaline phosphatase, liver/bone/kidney* (*TNSALP*) activity will be determined and the expression of bone maturation genes will also be assessed.

4.2 SSO-induced ERG exon 7b skipping alters cell growth in MG63 cells

Initially the effect of SSO-induced exon 7b skipping on cell growth was assessed by monitoring cell numbers over time after SSO treatment. For each time point cells were seeded into a 12 well plate at 10^4 cells per well, and treated with 3μM of SSO or PBS as a control. Cells were re-dosed with SSO at the 96 hour time point. After 24, 48, 72, 96 and 168 hours cells were collected and counted using the trypan blue exclusion assay to determine the number of live cells. RNA and protein were then isolated from these samples to determine the degree of exon skipping at each time point. Three biological repeats were assessed. Cell growth was significantly inhibited in SSO treated cells (Figure 4.1). After 168h both E7b5 and E7b3 SSO treated cells exhibited significantly reduced growth when compared to the control SSO (Figure 4.1, $p<0.0001$, in both instances).

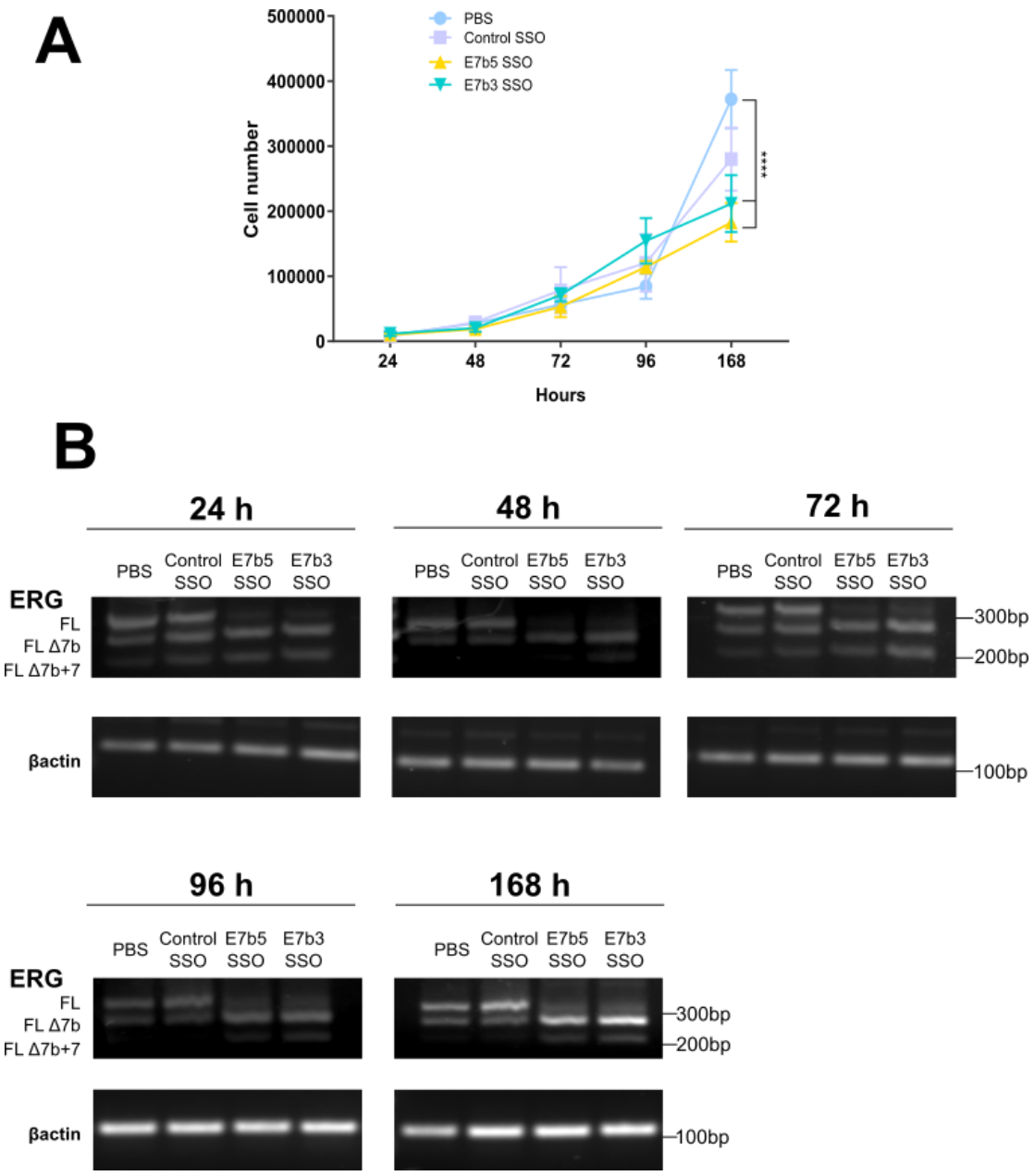


Figure 4.1 Growth curve of MG63 cells treated with ERG exon 7b SSO

MG63 cells were dosed at time zero and again at 96 hours with 3 μm of SSO. **A.** The Trypan blue assay was used to determine cell number at each time point. (**** p≤0.0001). Error bars show 95%CI. **B.** RNA was isolated in order to run PCR to confirm ERG exon 7b skipping at each time point. N=3

4.3 SSO-induced ERG exon 7b skipping alters cell proliferation in MG63 cells

A defining feature of cancer development is dysregulated proliferation (Hanahan & Weinberg, 2011). ERG is a regulator of key growth receptors and factors required for cellular proliferation (Stankiewicz and Crispino, 2013; Urbinati *et al.*, 2015). Initially a comparison of proliferation between tumour and normal tissue was determined using the proliferative index which was determined for PRAD patients from the TGCA database. This index was developed by assessing the expression of proliferation genes, shortlisted from microarray, gene ontology and gene expression studies across many tissues and developmental stages (Sandberg *et al.*, 2008). The values which range between 0 and 1 represent the proportion of genes in this gene set that are above median expression for each patient. Tumour tissue samples were shown to have a higher mean proliferation index of 0.8172 than normal control tissue of 0.6612 (Mann-Whitney, $p<0.0001$).

Therefore two assays were used to evaluate the effects of manipulating *ERG* exon 7b splicing on cell proliferation. The resazurin assay measures the reducing potential of a cell. The blue non-fluorescent resazurin is reduced to resofurin, which is pink and fluorescent by mitochondrial reductases and other metabolic enzymes such as NADPH dehydrogenase. These facilitate the reduction of resazurin by transferring electrons from NADPH + H^+. The colour change can be monitored either colorimetrically and the more proliferating cells are present the greater the reduction observed (Rampersad, 2012).

Cells were seeded into a 96 well plate (10^4/well) and treated with SSO (3μM) for 48, 72 and 96 hours. Two to four hours prior to the end of the time point the resazurin reagent was added. A microplate reader was used to measure the colorimetric changes compared to a control (of resazurin with no cells). The percentage difference in proliferation in comparison to the PBS control was determined. For each time point five biological repeats were assessed. After 48 hours of SSO treatment no significant changes in proliferation were detected (Figure 4.3). However 72 and 96 hours after SSO treatment significant changes were detected. For the 72 hour time point, E7b3 SSO treatment reduced cell proliferation ($p<0.0001$) more significantly than E7b5 SSO treated cells ($p=0.0485$) when compared to the control SSO (Figure 4.3C-D). Similar observations were made 96 hours post SSO (Figure 4.3E-F).

Another method to assess proliferation is by measuring the expression of Ki-67, a protein highly expressed in actively dividing cells. After SSO treatment cells were fixed, permeabilised and stained with primary and fluorescent secondary antibody in cell culture plates to permit detection using fluorescent microscopy. Counterstaining with DAPI allowed co-localisation analysis of Ki-67-positive foci. Three biological repeats were assessed.

Figure 4.4 shows after 72 hours of SSO treatment there is a reduction in Ki-67 positive foci in both E7b5 and E7b3 treated cells indicating a decrease in proliferation ($p=0.0012$ and $p=0.0036$, respectively). The observations demonstrate that the exclusion of exon 7b by SSOs reduces proliferation in MG63 cells.

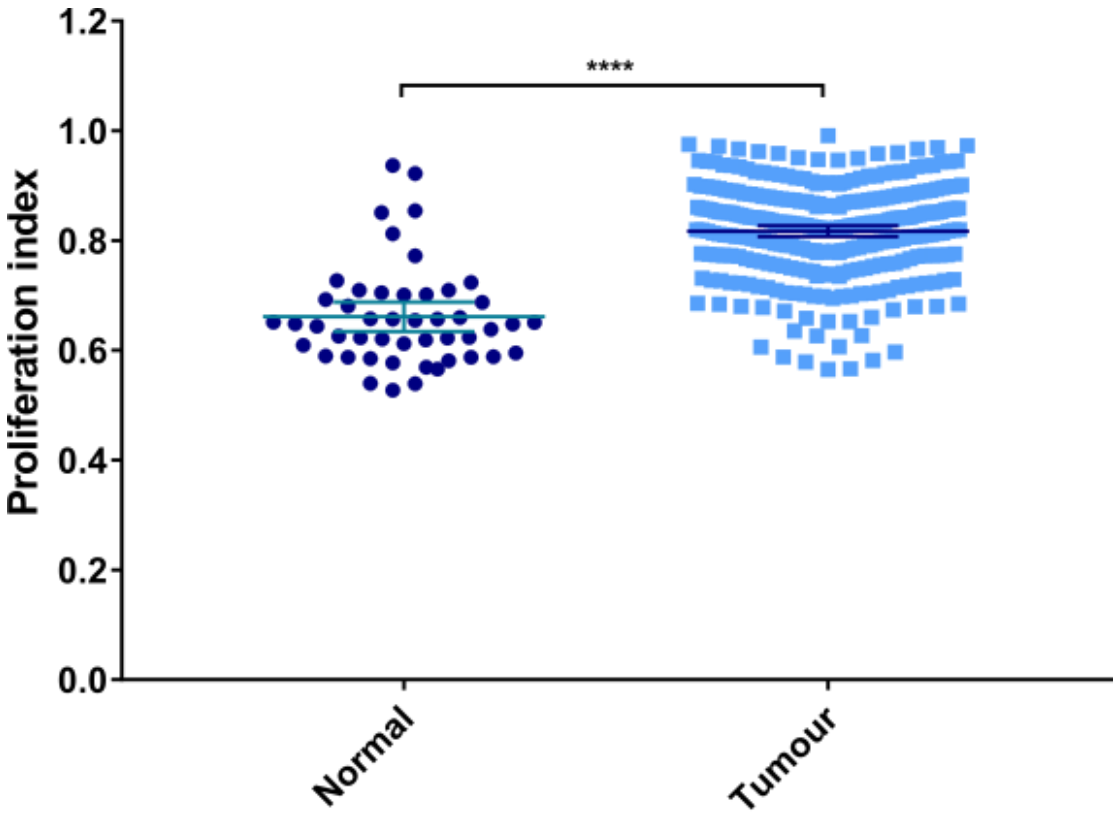


Figure 4.2 Proliferation index of PRAD patients from TCGA database

The proliferation index is a gene-expression based measure of proliferation and the values for normal and tumour tissue from the PRAD patients from the TCGA are shown. The middle line represents the mean and the error bars are 95%CI. (****p≤0.0001 Number of samples=243

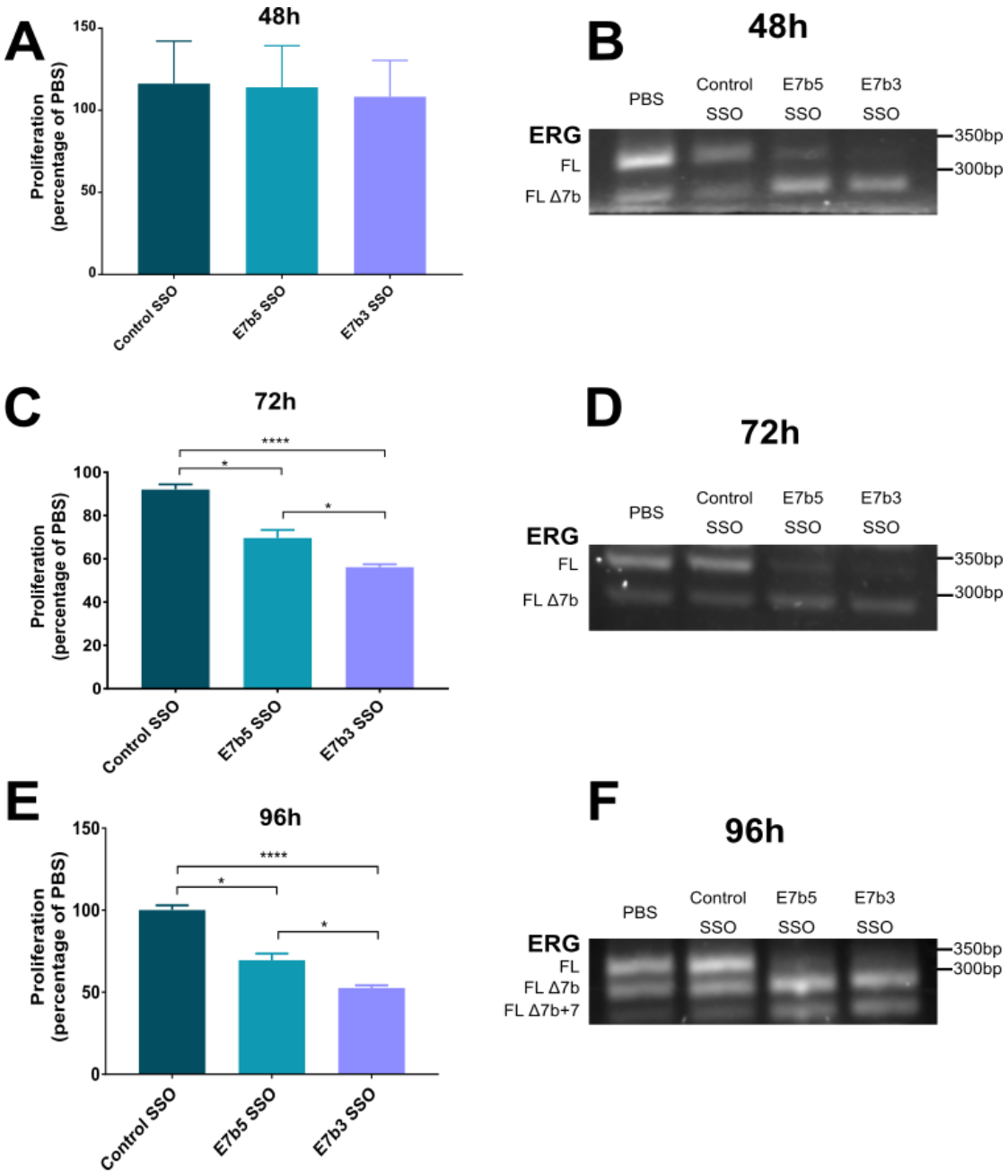


Figure 4.3 Proliferation of MG63 cells following SSO treatment

The resazurin assay was used to measure cell proliferation in MG63 cells treated with 3 μM SSOs for **A.** 48 hours **C.** 72 hours and **E.** 96 hours. The graphs shows the percentage difference of proliferation for each treatment in relation to the PBS control which is considered to be 100% proliferation. (*p≤0.05, **** p≤0.0001). Error bars show 95%CI. **B, D, F.** ERG exon 7b skipping was confirmed by PCR. N=5.

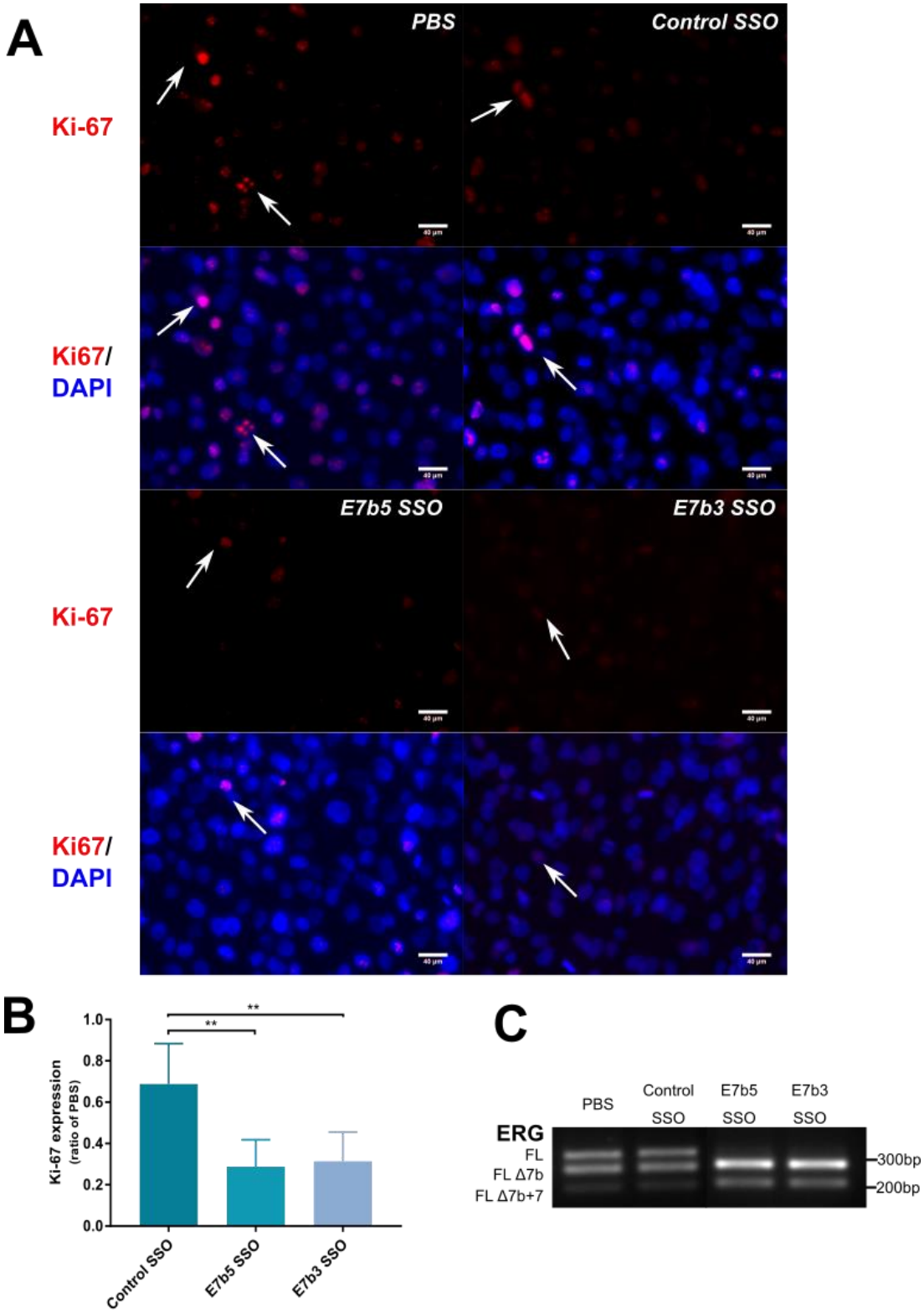

A
PBS
Control SSO
Ki-67
Ki67/
DAPI
E7b5 SSO
E7b3 SSO
Ki-67
Ki67/
DAPI
B
Ki-67 expression
(ratio of PBS)
1.0
0.8
0.6
0.4
0.2
0.0
**
**
Control SSO
E7b5 SSO
E7b3 SSO
C
PBS
Control
SSO
E7b5
SSO
E7b3
SSO
ERG
FL
FL Δ7b
FL Δ7b+7
300bp
200bp

Figure 4.4 Ki-67 expression in MG63 cells following 72 hours of SSO treatment

A. After 72 hours of treatment with 3μM SSO Ki-67 expression was determined using immunofluorescent microscopy to measure proliferation. Top panel: Ki-67 foci. Bottom panel: overlay with DAPI. Arrows indicate Ki-67 positive foci under the Texas red filter (top panel) and merged Texas red and DAPI filter (bottom panel). Scale bar = 40 μm. **B.** Graph represent the fold change in Ki-67 foci positive cells when compared to the PBS control. **p=0.01. Error bars show 95%CI. **C.** ERG exon 7b skipping was confirmed using PCR. N=3

4.4 SSO-induced ERG exon 7b skipping induces apoptosis in MG63 cells

The ability to undergo programmed cell death (apoptosis) is often inhibited in cancer cells. Some cancer treatments are based on the induction of apoptosis. Several genes that regulate apoptosis are ERG targets; for example, VE-cadherin (Birdsey *et al.*, 2008). VE-cadherin works via interactions with VEGF receptor 2 and β-catenin to prevent apoptosis by modulating the balance of pro- and anti-angiogenic factors in endothelial cells. The integrity of cell-cell junctions is degraded due to a reduction of ERG which interferes with signal transduction to VE-cadherin and other pathways and increasing pro-apoptotic factors such as Bcl-2 (Mallat and Tedgui, 2000).

Therefore apoptosis levels were investigated by assessing expression of caspase 3 and caspase 7 using the CellEvent Caspase 3/7 green detection reagent. The reagent labels apoptotic cells by staining their DNA with a fluorescent nucleic acid binding dye. A short peptide cleavage site sequence for caspase 3/7 is conjugated to the dye; in this state it remains non-fluorescent until it is cleaved by caspase 3/7 allowing it to bind DNA. The dye can then be detected using a FITC filter set.

Following 48 hours and 72 hours SSO treatment the caspase3/7 reagent was added 45 minutes prior to the end of the treatment time point. The experiment was carried out in triplicate. Images in six fields of view were taken to estimate the number of caspase3/7 positive cells. Results were expressed as a percentage of the total number of cells. At both time points three biological

repeats were assessed. The data presented in are fold change ratio of caspase 3/7 positive cells compared to the PBS control.

Both ERG exon 7b targeting SSOs induced significantly increased levels of caspase3/7 at both time points (Figure 4.5; Figure 4.6). Compared to the control SSO after 48 hours of treatment E7b3 SSO resulted in a 10.9 fold increase in apoptosis ($p<0.0001$) and at 72 hours of treatment the fold increase was even greater (182.8, $p<0.0001$). For the E7b5 SSO an increase in apoptosis was also observed when compared to control SSO after both 48 hours and 72 hours of treatment (4.8, $p=0.0323$ and 143.8,$p=0.0006$ respectively) although it was not as large an increase as obtained with the E7b3 SSO. The difference between the two treatments was statistically significant at the 48 hour time point ($p=0.0234$) but not at 72 hours.

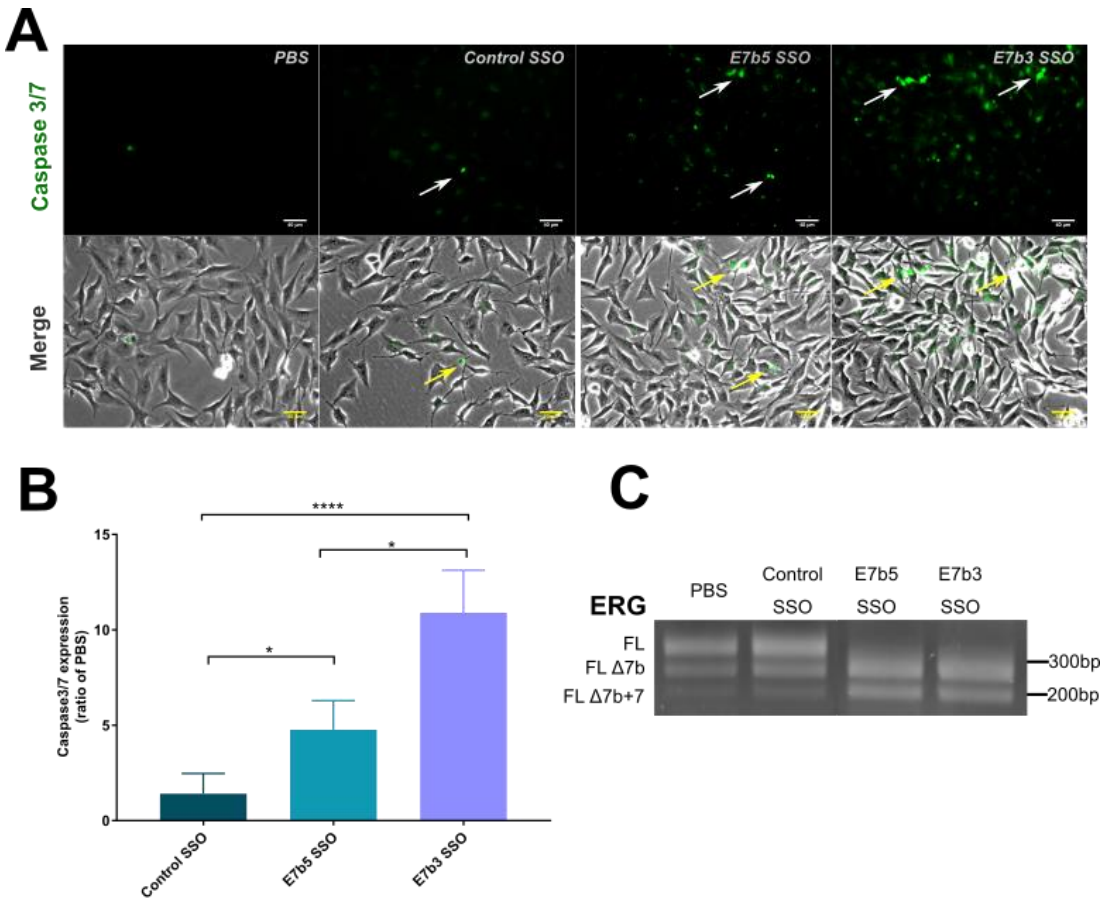


Figure 4.5 Caspase 3/7 expression following 48 hours of SSO treatment in MG63 cells

A. Caspase 3/7 positive foci were quantified using immunofluorescent microscopy to determine apoptosis in MG63 cells following treatment with 3μM SSO. Arrows identifying caspase 3/7 foci under FITC filters (top panel) and merged FITC and phase contrast images taken under a x20 objective. Scale bar = 40 μm. **B.** Graphs represent the fold change in caspase 3/7 positive foci quantified when compared to the PBS control. (***p≤0.001, ****p≤0.0001). Error bars show 95%CI. **C.** ERG exon 7b skipping was confirmed using PCR. N=3

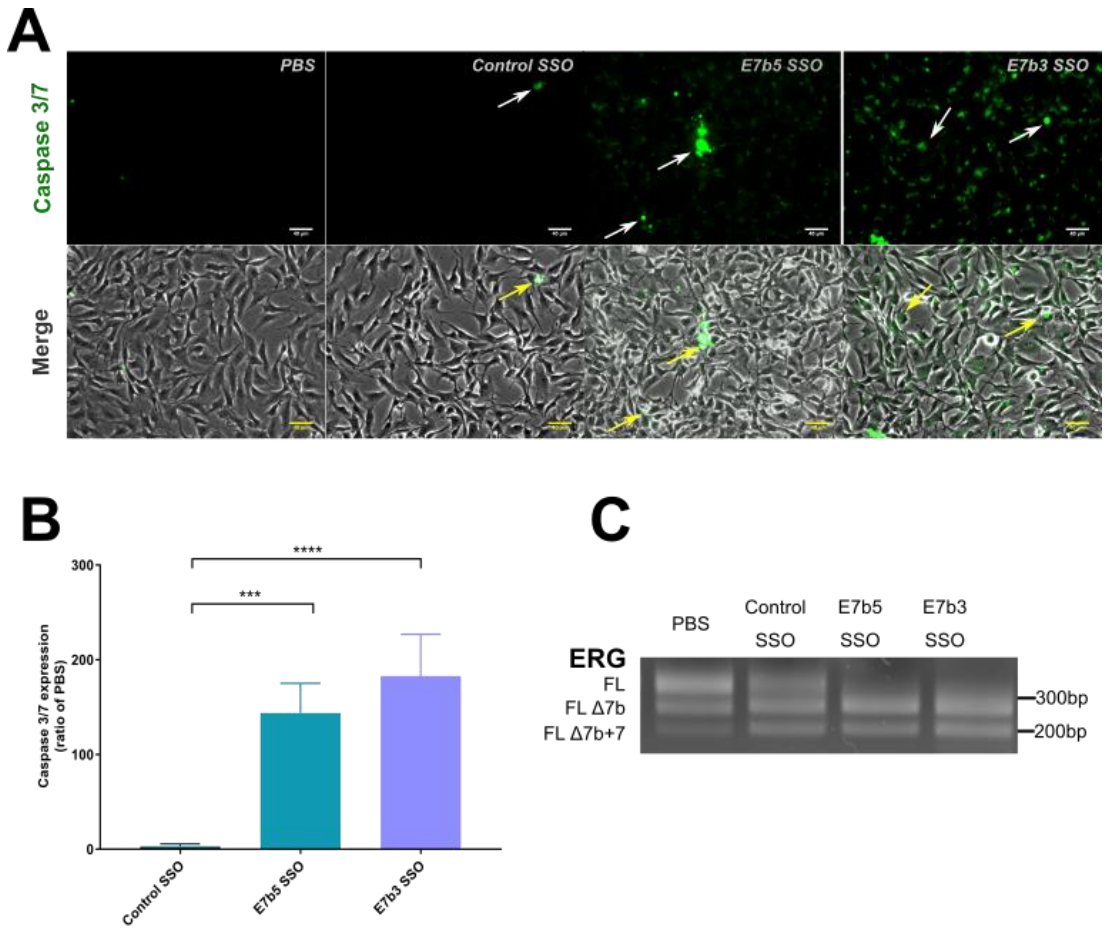


Figure 4.6 Caspase 3/7 expression following 72 hours of SSO treatment in MG63 cells

A. Caspase 3/7 positive foci were quantified using immunofluorescent microscopy to determine apoptosis in MG63 cells following treatment with 3μM SSO. Arrows identifying caspase 3/7 foci under FITC filters (top panel) and merged FITC and phase contrast images taken under a x20 objective. Scale bar = 40 μm. **B.** Graphs represent the fold change in caspase 3/7 positive foci quantified when compared to the PBS control. (***p≤0.001, ****p≤0.0001). Error bars show 95%CI. **C.** ERG exon 7b skipping was confirmed using PCR... N=3

4.5 SSO-induced ERG exon 7b skipping alters cell migration in MG63 cells

Migration was assessed using the transwell assay. Transwell inserts were set up in triplicate into the wells a 24 well plate. In this assay cells were pre-treated with SSO for 24 hours before addition to the inserts in serum free medium. Treatment was also carried out in serum free medium. A chemotactic gradient was created by addition of DMEM media with 10% FBS in the wells below the inserts. The transwell set up was incubated for 24 hours prior to collection, washing, fixing and staining of the inserts. Images of several fields of view of the bottom of the inserts were taken to assess cell number per insert. Three biological repeats were assessed. When compared to the control SSO treatment both E7b5 and E7b3 SSO treatment significantly reduced cell migration (Figure 4.7; $p=0.0022$ and $p=0.0062$, respectively). No significant change was seen between both ERG exon 7b-specific SSOs.

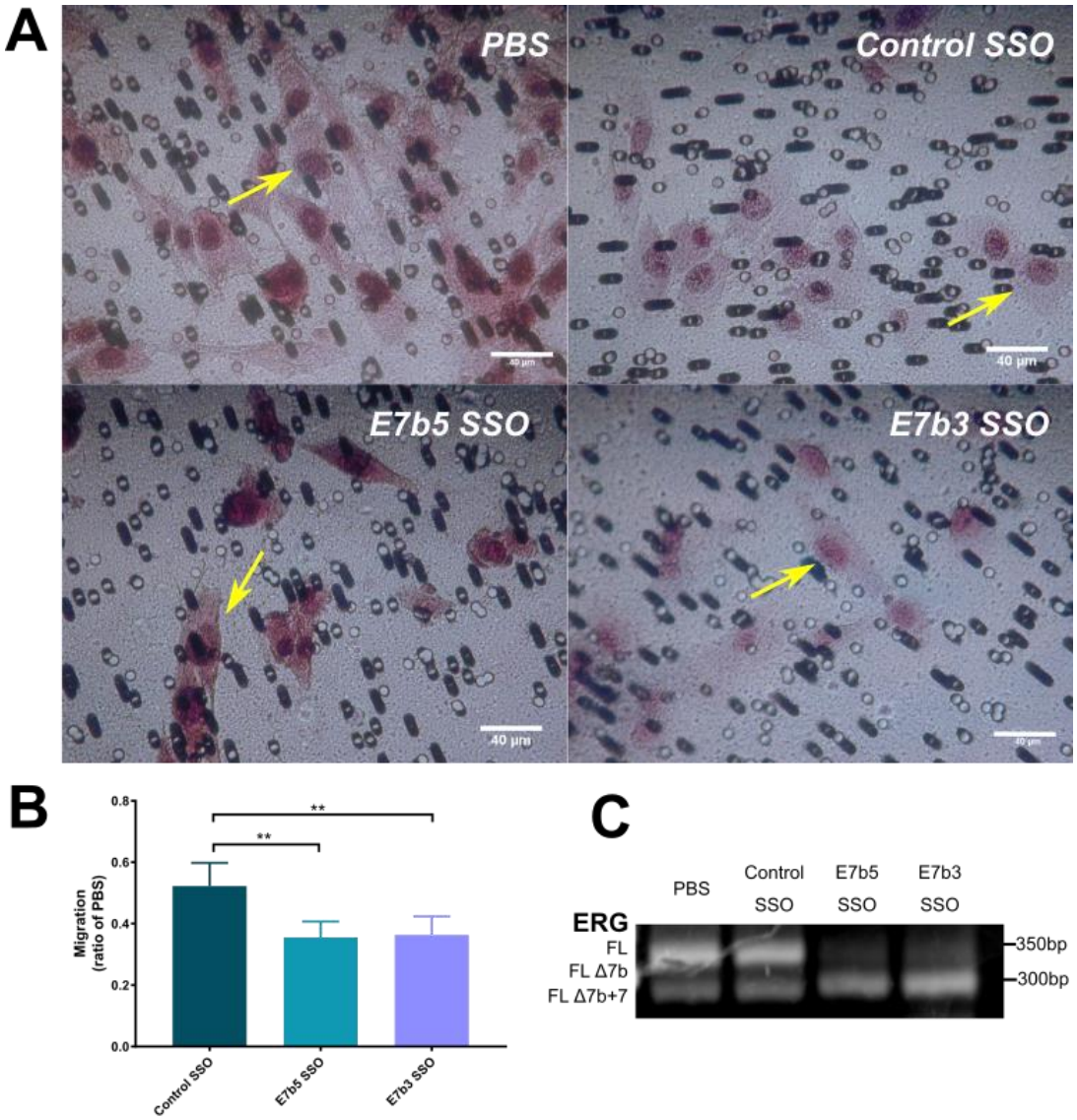


Figure 4.7 ERG exon 7b skipping causes reduction in MG63 cell migration

Following 24 hours of SSO treatment the transwell migration assay was carried out to measure migration after a total of 48 hours of SSO treatment. Cells were counted on images taken of the bottom of the filters using a x20 objective on a light microscope. **A.** The arrows indicate cells in representative images of the filters. The darker spots and transparent circles visible are pores in the filters. Scale bar= 40 µm. **B.** Graphs represent the fold change in cell numbers when compared to the PBS control. (**p≤0.01). Error bars show 95%CI. **C.** ERG exon 7b skipping was confirmed using PCR. N=3

4.6 SSO-induced ERG exon 7b skipping alters cell invasion in MG63 cell

The modification of the transwell inserts using Geltrex, a basement membrane matrix, to mimic a cell basement medium allowed for the investigation of the effect of exon 7b skipping on cell invasion. The inserts were pre-coated with Geltrex to form an even thick layer for the cells to invade. The rest of the experimental set up and image acquisition was the same as for the transwell migration assay. Three biological repeats were assessed. Figure 4.8 shows that treatment with E7b5 SSO significantly reduced cell invasion ($p<0.001$) and the mean fold change was lower for E7b3 SSO (0.35; $p<0.0001$).

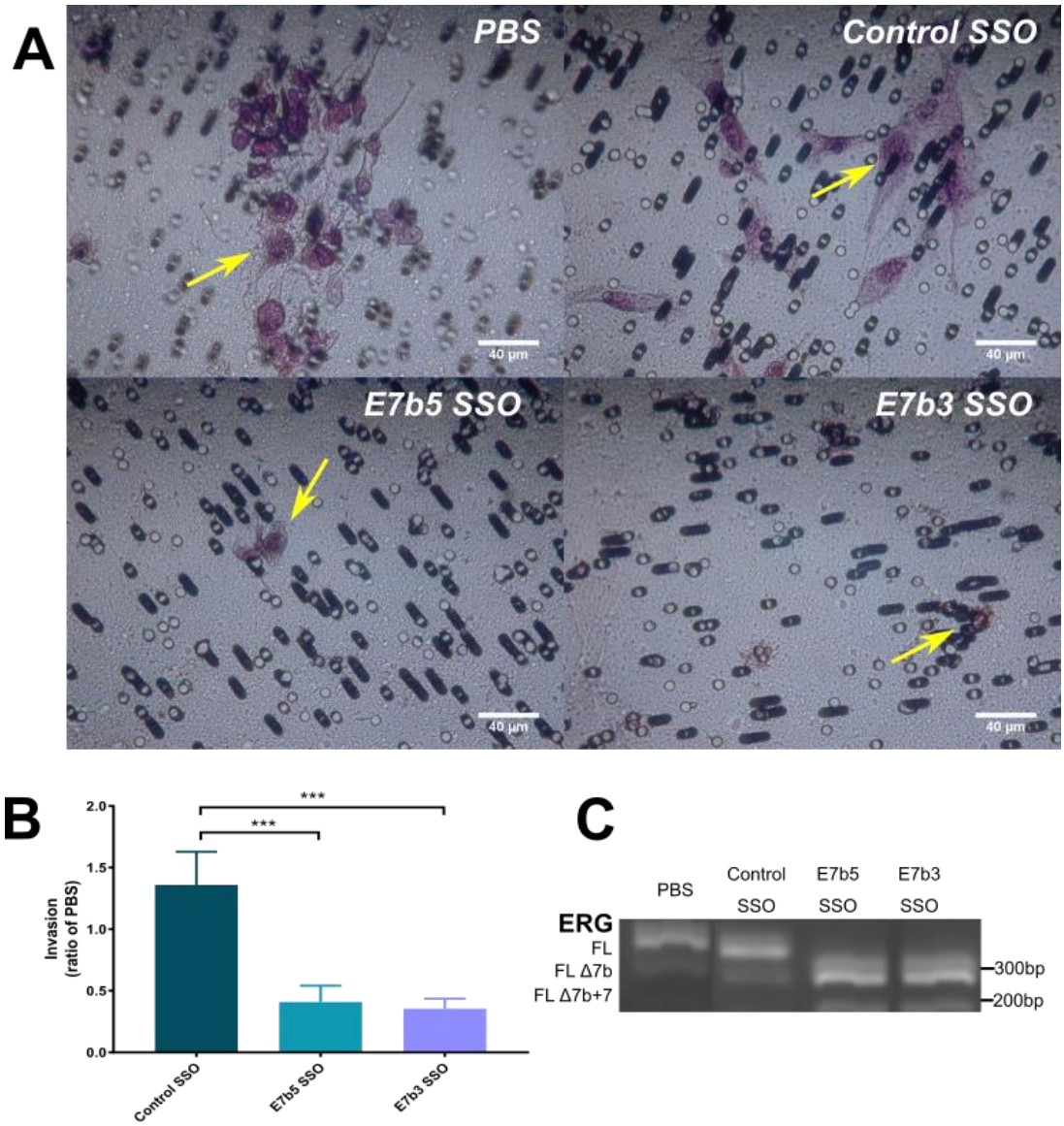


Figure 4.8 ERG exon 7b skipping causes reduction in MG63 cell invasion

Following 24 hours of SSO treatment the transwell invasion assay was carried out to measure migration after a total of 48 hours of SSO treatment. Cells were counted on images taken of the bottom of the filters using a x20 objective on a light microscope. **A.** The arrows indicate cells in representative images of the filters. The darker spots and transparent circles visible are pores in the filters. Scale bar= 40 µm. **B.** Graphs represent the fold change in cell numbers when compared to the PBS control. (***p≤0.001). Error bars show 95% CI. **C.** ERG exon 7b skipping was confirmed using PCR. N=3.

4.7 ERG exon 7b skipping reduces tumour growth in an MG63 xenograft mouse model

To investigate the *in vivo* efficiency of ERG exon 7b targeting SSO a mouse xenograft model was established. Mice were injected intraperitoneally with 12.5 mg/kg of control or E7b3 SSO which showed best efficacy *in vitro*, over a 63 day period. Tumour volume, tumour weight and the body weight of the mice was also measured over the time course of treatment. Six mice were used per treatment group and the experiment was repeated three times.

Although findings show a reduction of both tumour weight and volume, this was comparably similar between the control and E73b SSO and not statistically significant. The control SSO treated cells had an intermediate phenotype with a tumour volume smaller than the PBS tumours but greater than the E7b3 SSO treated tumours, suggesting that the chemistry of the SSOs may need to be revised to define if tumour growth inhibition observed is specific to exon 7b and not related to the toxicity of SSO chemistry. Body weight stayed fairly consistent between the treatment groups and over the time course of the experiment. It is important to stress that this data is preliminary but is proof of principle that indicates that the SSO may be effective in slowing down tumour growth *in vivo*.

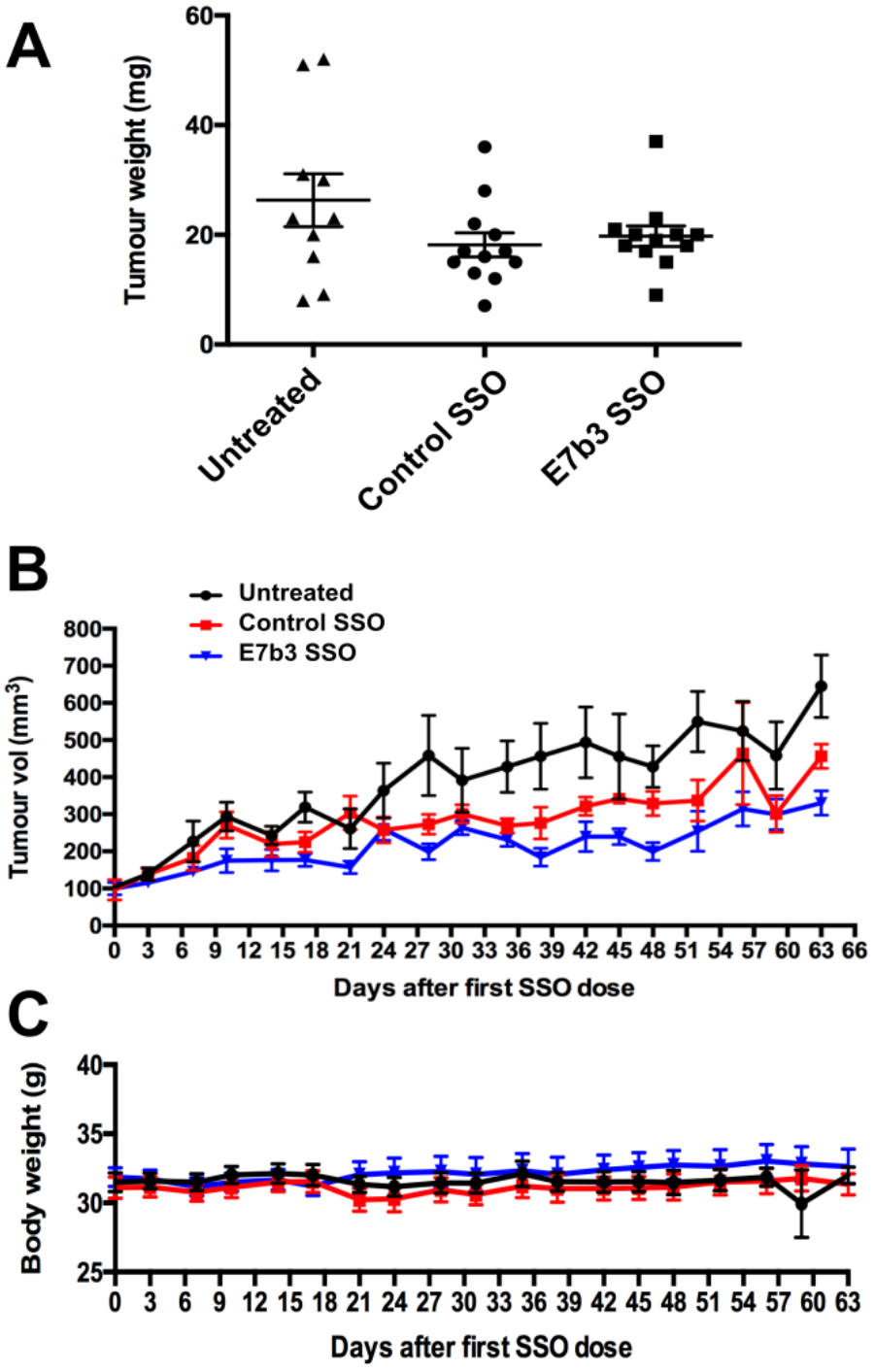


Figure 4.9 Growth of MG63 xenograft mouse model tumours treated with ERG exon 7b SSO

A. Tumour weights **B.** tumour volume and **C.** body weight of MG63 xenograft mouse model treated with 12.5mg/kg of control and E7b3 SSO for a 63 day period. An untreated PBS mouse cohort was also included. Mice were dosed at each time point seen on the graph.

4.8 SSO-induced ERG exon 7b exon skipping alters *TNSALP* activity in MG63 cells in a time-dependent manner

TNSALP expression can be used as a marker of cell differentiation in osteoblast and their derivative cells such as MG63 (Rutkovskiy, Stensløkken and Vaage, 2016). Co-treatment with the active vitamin D metabolite calcitriol (1,25D) and FHBP, a lysophosphatidic acid analogue, promotes cell differentiation (Gidley *et al.*, 2006). Therefore co-treatment with these reagents in the presence of SSO was carried out. Serum starved MG63 cells were treated with SSOs for 24 hours, 48 and 72 hours. In each case after SSO treatment cells were treated with 1,25D and FHBP and given an additional dose of SSO before being incubated for a further 24 hours. *TNSALP* expression was assessed based on the well-established method of measuring p-NPP conversion to p-NP resulting in a yellow colour that can be quantified by measuring absorbance at 405nm. The p-NP levels reported were corrected for using the percentage difference in cell viability with cell in media alone being used as a baseline to compare other treatments. For all three time points five biological repeats were assessed.

After a total 48 hours of SSO treatment the expected increase in p-NP production under co-stimulation of 1,25D and FHBP was observed and the addition of the control SSO did not lead to significant changes (Figure 4.10). In the presence of E7b3 SSO with the co-treatment this increase levels of p-NP was significantly lower when compared to the 1,25D and FHBP co-treatment ($p=0.0110$) and similarly control SSO with 1,25D and FHBP co-treatment ($p=0.0013$). The resazurin cell viability assay was confirmed changes observed with the TNSALP assay were not due to changes in cell number or health as

there were no significant changes in cell viability across the various treatments (Kruskal Wallis, p=0.0643).

In contrast after 72 hours of SSO treatment although the co-treatment with 1,25D and FHBP still increased TNSALP expression this was not observed at the same level in the presence of the SSOs including the control (Figure 4.11A). A significant reduction was observed for E7b3 SSO/co-treatment compared to 1,25D and FHBP co-treatment alone (p=0.0074) but not control SSO/co-treatment. Both E7b5 SSO/co-treatment (p=0.0406) and E7b3 SSO (p<0.0001) reduced TNSALP when compared to control SSO/co-treatment at 96 hours of SSO treatment (Figure 4.12 A) Even though the viability of E7b3 SSO treated cells was significantly reduced (Figure 4.11B and Figure 4.12B) all the p-NP concentrations reported were corrected for this effect.

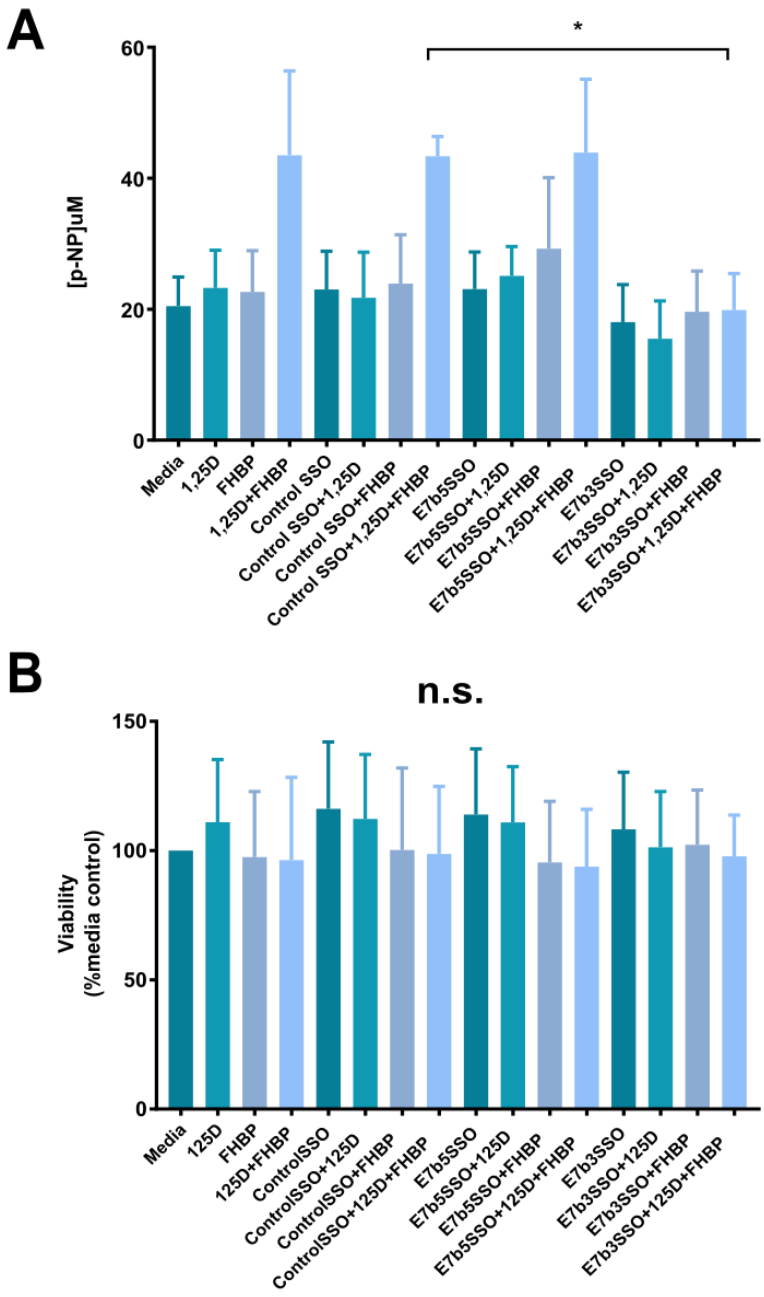


Figure 4.10 TNSALP expression is attenuated after 48 hours of SSO induced ERG exon 7b skipping in MG63 cells

A. TNSALP expression based on the concentration of p-NP in each sample was corrected for viability. MG63 cells were treated with SSOs for 24 hours prior to a second SSO dose as well as 1,25D and FHBP treatment for a further 24 hours. ERG SSO treatment were first compared to their respective equals in the presence of the control SSO. If significant change was observed an additional comparison was made to the non-SSO treated samples. (* p≤ 0.05). **B.** Cell viability measured relative to total cell viability in media (100% viable) determined using the resazurin assay. Statistical significance was determined against the cells treated in media. (n.s. not statistically significant). Error bars show 95% CI. N=5.

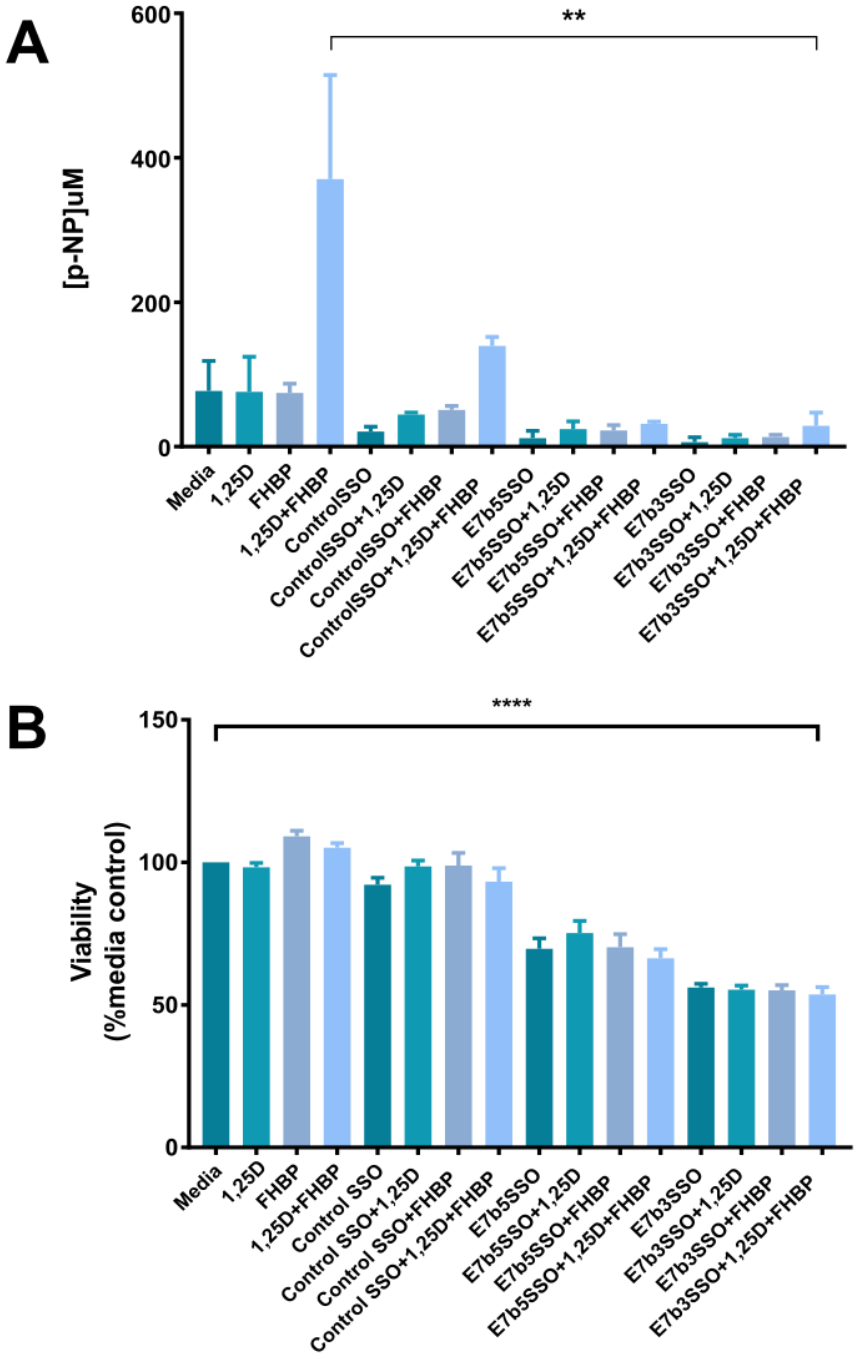


Figure 4.11 TNSALP expression is attenuated after 72 hours of SSO induced ERG exon 7b skipping in MG63 cells

A. TNSALP expression based on the concentration of p-NP in each sample corrected for viability. MG63 cells were treated with SSOs for 48h hours prior to a second SSO dose as well as 1,25D and FHBP treatment for a further 24 hours. ERG SSO treatment were first compared to their respective equals in the presence of the control SSO. If significant change was observed additional comparison was made to the non-SSO treated samples **B.** Cell viability against cells in media which was taken to be 100% viable and was determined using the resazurin assay. Statistical significance was determined against the cells treated in media. (*** p≤ 0.01, ****p≤0.0001). Error bars show 95% CI. N=5.

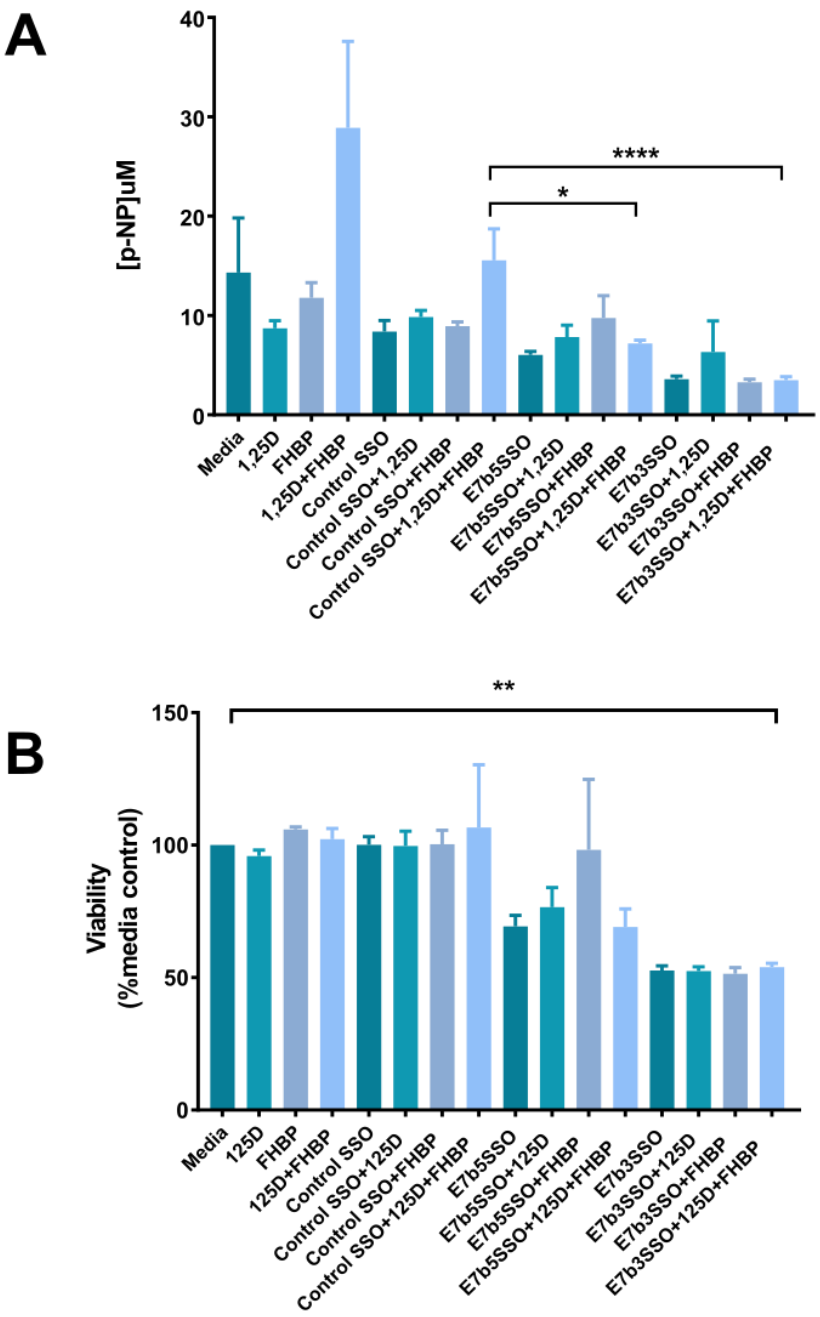


Figure 4.12 TNSALP expression is attenuated after 96 hours of SSO induced ERG exon 7b skipping in MG63 cells

A. TNSALP expression based on the concentration of p-NP in each sample corrected for viability. MG63 cells were treated with SSOs for 72 hours prior to a second SSO dose as well as 1,25D and FHBP treatment for a further 24 hours. ERG SSO treatment were first compared to their respective equals in the presence of the control SSO. If significant change was observed additional comparison was made to the non-SSO treated samples. (* p≤ 0.05, ** p≤ 0.01, ****p≤0.0001). **B.** Cell viability relative to cells in media which was taken to be 100% viable and was determined using the resazurin assay. Statistical significance was determined against the cells treated in media. (n.s. not statistically significant). Error bars show 95% CI. N=5.

4.9 SSO-induced ERG exon 7b skipping alters the expression of *TNSALP* and other bone maturation genes in MG63 cells

The expression of *runt-related transcription factor 2* (*RUNX2*), *osteopontin* (*OPN*) and *TNSALP* are all genes are involved in the biogenesis of osteoblast cells (Rutkovskiy, Stensløkken and Vaage, 2016). The expression of these genes was assessed in cells treated for 48 hours with the ERG7b3 SSO treatment which was chosen as the focus for this experiment as this is the more efficient SSO. RNA was collected from cells that had been treated with media, co-treatment of 1,25D and FHBP, co-treatment with control SSO and co-treatment with ERG7b3 SSO for 48 hours. In order to normalise the gene expression an exogenous spike made from the U13397 cDNA clone encoding the small RuBisCO subunit of *Arabidopsis thaliana* using *in vitro* transcription was added to each RNA sample prior to cDNA synthesis. The expression of the spike gene was also calculated and the mass of DNA obtained for each sample was normalised against this. Serial dilutions of known amounts of the U13397 plasmid DNA were used to produce standard curves in order to facilitate quantitation. Overall mRNA expression is expressed as a percentage of the total RNA used to make the cDNA pool.

Assessment of *TNSALP* expression showed significantly increased expression in the 1,25D and+ FHBP co-treated cells ($p=0.0024$) and in the presence of the control SSO ($p=<0.0001$) when compared to the media control (Figure 4.13A). The reduction of *TNSALP* mRNA levels in the presence of E7b3 SSO treatment was statistically significant when compared to control SSO levels ($p=0.0498$). These observations were in agreement with the TNSALP activity assay results.

Increased levels of *OPN* are observed in mature osteoblasts particularly during mineralisation as well as in proliferating cell in the pre-osteoblast stage. The highest levels of *RUNX2* are expressed in pre-mature osteoblasts and levels drop off as cells become more mature (Rutkovskiy, Stensløkken and Vaage, 2016, Figure 4.13D). Increased levels of *OPN* were observed following E7b3 treatment when compared to both media control and combined 1,25D and FHBP treatment ($p=<0.0001$ and $p=0.0005$, respectively, Figure 4.13B). When compared to both the media control and 1,25D and FHBP co-treatment increased levels of *RUNX2* expression were observed in the presence of E7b3 SSO ($p=0.027$ and $p=0.004$, respectively, Figure 4.13C).

To determine if the changes in *TNSALP* expression were a consequence of direct transcriptional control by ERG chromatin immunoprecipitation (ChIP) was carried out using MG63 cells. The ChIP analysis revealed an ERG binding site in the promoter of the *TNSALP* (-1114 to -85). Analysis of the *TNSALP* promoter using the Eukaryotic Promoter Database revealed the presence of a potential ERG binding site (AGAGGAAACG) within the region 1000bp upstream of the transcription start site one of them being within the region detected by the ChIP assay (Figure 4.14).

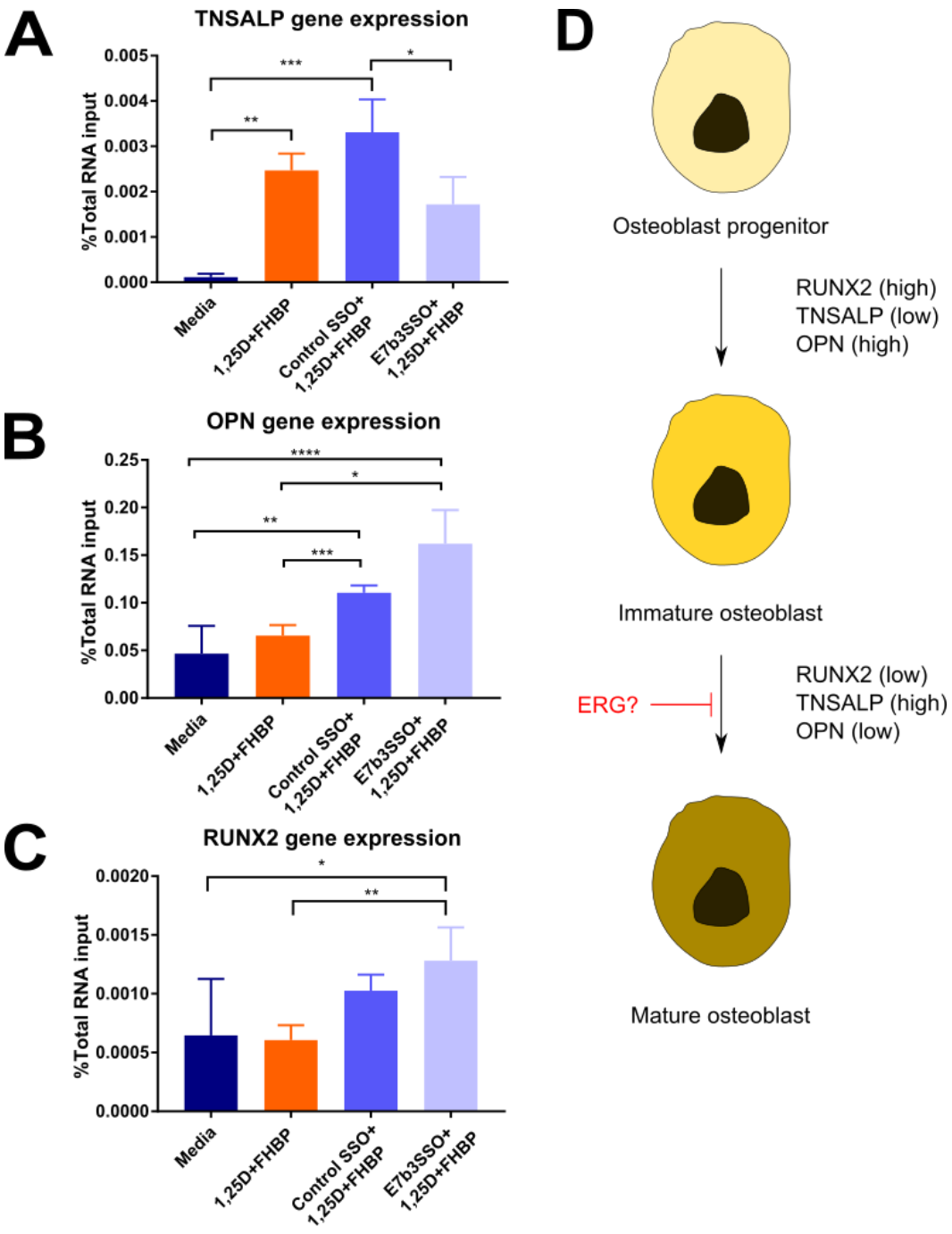


Figure 4.13 Expression of TNSALP, OPN and RUNX2 genes after 48h of 1,25D and FHBP co-treatment in the presence of SSOs.

Expression of **A.** TNSALP, **B.** OPN and **C.** RUNX2 is represented as a percentage of the total RNA used to make the cDNA library per sample. (* p≤ 0.05) ** p≤ 0.01, ***p≤0.001, ****p≤0.0001). Error bars show 95% CI. N=3. **D.** RUNX2 is important for osteoblast lineage commitment and maintains the expression of OPN. RUNX2 expression reduces in immature osteoblasts and TNSALP expression increases as osteoblast differentiate and mature.

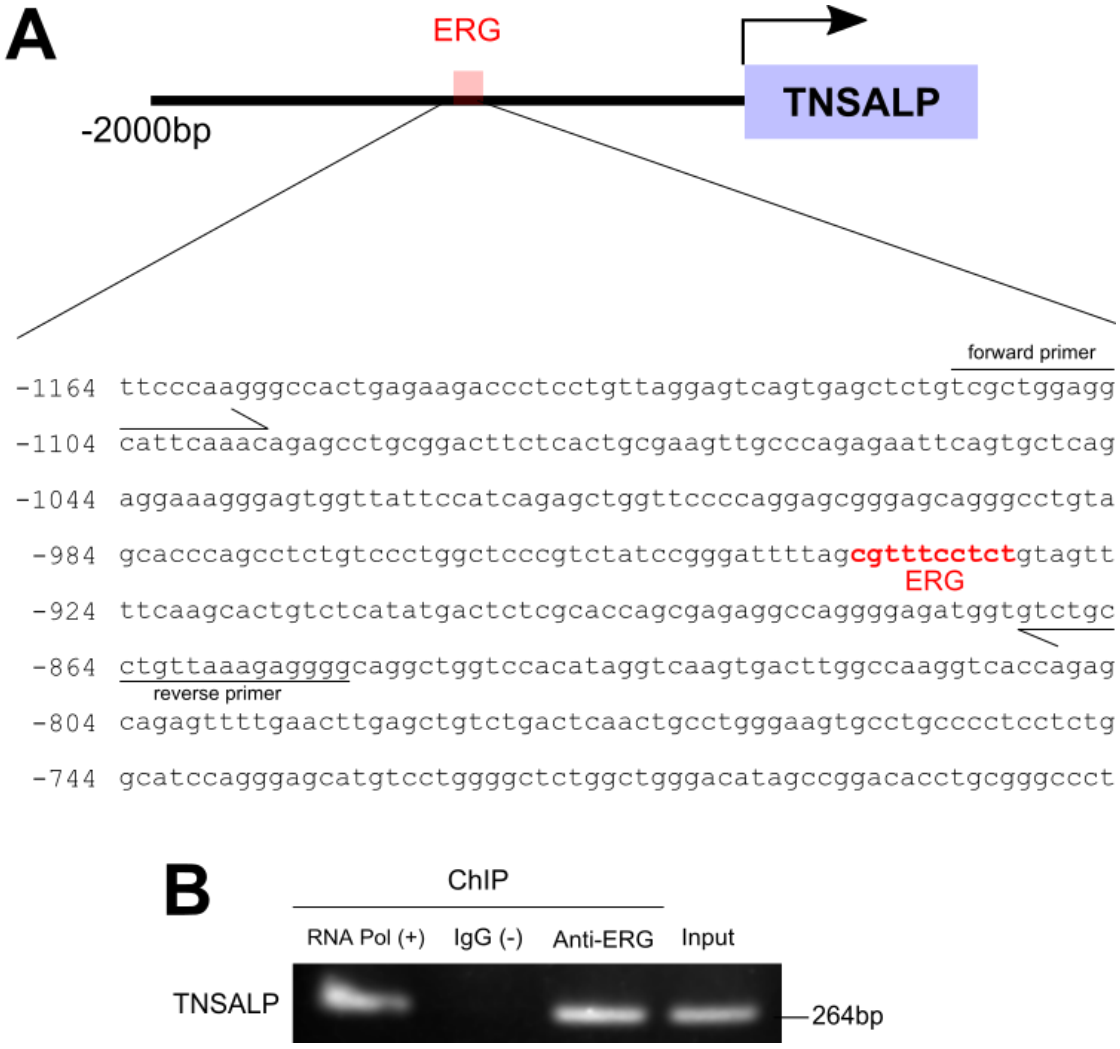


Figure 4.14 ERG binds to the promoter of *TNSALP*

A. Schematic of the TNSALP promoter showing the sequence of the predicted ERG binding site (-930 in red; note that the consensus is in the complementary sequence) and the primers designed to detect it. **B.** ChIP analysis using anti-ERG detected TNSALP promoter DNA. An RNA Pol positive control was included to confirm the gene was be transcribed, non-specific IgG negative control was included to prove specific antibody binding and the input lane confirms the presence of the sequence in the original DNA sample prior to immunoprecipitation. The PCR product size is shown.

4.10 Summary of findings

One of the hallmarks of cancer is uncontrolled cell proliferation. PRAD tumour samples showed more proliferative action compared to normal tissue controls. In addition ERG exon 7b skipping reduced cell growth and proliferation in MG63 cells in a time-dependent manner. Two other cancer phenotypes migration and invasion decreased in the presence of SSOs. These observations correlate with previous studies where inclusion of ERG exon 7b was associated with increased proliferation, invasion and migration (J. Wang *et al.*, 2008; Carver *et al.*, 2009; Wu *et al.*, 2013). Meanwhile the opposite was observed with apoptosis. Whereas ERG exon 7b inclusion is seen in cells with decreased apoptotic ability (J. Wang *et al.*, 2008) SSO-induced skipping of exon 7b increased apoptosis in a time-dependent manner. Generally phenotype changes were more pronounced with an increased period of SSO treatment (i.e. 48 to 72 hours). An explanation for this may be associated with more mRNA being translated into protein increasing protein activity and driving phenotype changes.

Cell proliferation and migration are both known to be regulated by RAS/ERK signalling, a cell signalling pathway of critical importance in cancer. ETS factors and activator protein 1 (AP1) work together to regulate transcription of RAS responsive genes by binding specific enhancer sequences (Hollenhorst, 2012). In addition homo- and heterodimer formation of ETS factors (Carrère *et al.*, 1998) and post-translational modification of ETS factors (Hollenhorst, McIntosh and Graves, 2011) mean that the binding target combinations are extensive. Skipping exon 7b potentially alters ERK phosphorylation by removing the DEF

domain and a serine residue that is a potential phosphorylation target. Furthermore a modification of ERG's AD will potentially alter its ability to dimerise or interact with other proteins.

Additionally it may be that the altered cell phenotype may be due to changes in target gene transcription levels, or even in the profile of target genes. In the case of cell migration and invasion a potential gene that may be undergoing altered transcription is *CXCR4* whose expression is regulated by ERG binding to its promoter. *CXCR4*, in association with *C-X-C motif chemokine ligand 12 (CXCL12)*, subsequently activates expression and secretion of matrix metalloproteinases which facilitate proteomic breakdown of the extracellular matrix contributing to cell migration and invasion in PCa cells (Chinni *et al.*, 2006; Singareddy *et al.*, 2013).

In terms of apoptosis, previous studies showed evidence that ERG and other ETS transcription factors are protective aiding cell survival (Birdsey *et al.*, 2012; Tan *et al.*, 2014). Inhibition of ERG in HUVEC cells increased apoptosis and this was shown to be mediated by VEGF/ VE-cadherin signalling suggesting ERG is required for cell survival. In angiogenesis the formation of new vessels requires a balance of pro- and anti-apoptotic factors. VE-cadherin associated with VEGF receptor 2 and other signalling pathway partners control levels of apoptotic proteins. VEGF receptor 2 activation by VEGF binding can activate Bcl proteins such as Bcl-2. Bcl-2 can activate caspases inducing apoptosis. VE-cadherin interactions with β-catenin also increases apoptosis. ERG is able to disrupt both VEGF and VE-cadherin apoptotic pathways (Birdsey *et al.*, 2008). Treatment-induced apoptosis was supressed in solid and haematopoietic

tumours with the *EWS-ERG* fusion, and knockdown of the fusion by RNAi resulted in cell death pointing to a protective role for aberrant ERG expression (Yi *et al.*, 1997). Therefore the observations made in this study suggest that in addition to complete knockdown skipping of *ERG* exon 7b is sufficient to counteract the protective capacity of ERG expression in cancer cells.

Within an *in vivo* MG63 xenograft model the ERG E7b3 SSO has been to have some efficacy reducing tumour volume and weight. Unfortunately the control SSO gave an intermediate effect. As the body weight of the mice was unaffected during the treatment it suggests that toxicity may not be the cause of this. Perhaps the moiety that facilitates transfection may be an issue, or perhaps there are off target effects of the control SSO. Further experimentation is needed but the data does suggest that the ERG exon 7b targeting SSO can be used *in vivo*. .

A new role for *ERG* in bone osteogenesis and *TNSALP* expression has been revealed. The multi-step process involved in osteoblast maturation from mesenchymal stem cell is carefully controlled by several transcriptional regulators that initiate, maintain and terminate the process (Rutkovskiy, Stensløkken and Vaage, 2016). Thus there was an attempt made to understand if ERG contributes to this process specifically through the transcriptional regulation of the *TNSALP* gene and possibly other genes required for osteoblast differentiation.

Addition of 1,25D and FHBP to the MG63 was used to establish a differentiation model in which increased TNSALP expression is observed. Expression of TNSALP was attenuated by ERG exon 7b skipping most especially when

induced by the 3´ splice site targeting SSO and this effect is time dependent. Two studies identified a reduction in alkaline phosphatase and mineralisation when the C-1-1 ERG variant, lacking exon 7, was expressed (Iwamoto *et al.*, 2000, 2001). This finding is consistent with my observations. As exon 7 also encodes for part of the AD it may be that ERG isoforms lacking exon 7b have a similar biological function in osteoblasts to C-1-1 in chondrocytes.

When mRNA expression was investigated a reduction in TNSALP was also observed, and ERG was shown to bind to the promoter of TNSALP using ChIP analysis. Therefore changes in TNSALP levels may be due to altered transcriptional activity of ERG induced by SSO mediated exon 7b skipping. In addition increased expression of *RUNX2* and *OPN* was observed in the presence of E7b3 SSO suggesting that ERG may be involved in regulation of the repression of these genes during osteoblast maturation. *RUNX2* is a master regulator of osteoblast maturation and the activity of the factor depends on association with other protein factors that may acts as co-repressors or co-activators (Komori, 2008). The expression of *OPN* is influenced by both RUNX2 and ERG as the promoter has binding site for RUNX2, ERG and other ETS transcription factors. As a result an increase in *RUNX2* expression does not necessarily correlate with the expression levels of the proteins target genes.

RUNX2 increases *OPN* expression near 3 fold but in combination with 1,25D increase in *OPN* expression was increased 8 fold as 1,25D enhances VDR and RUNX2 recruitment to the *OPN* promoter (Shen and Christakos, 2005). In PCa cells ERG was shown to increase *OPN* expression and there was a significant positive correlation between *TMPRSS2-ERG* and *OPN* expression in prostate

patient tumours (Flajollet *et al.*, 2011). Therefore although these genes were both found to be increased in agreement with an immature osteoblast phenotype effects by other ERG interactions may also be influencing their expression.

In summary, ERG exon 7b skipping induces anti-oncogenic features in the MG63 cell line model suggesting that the exon does contribute to tumorigenicity of malignancies with aberrant ERG expression. This finding was supported by observations of reduce tumour growth in an MG63 xenograft model. Furthermore ERG may be an important regulator of osteoblast maturation, as the induction of TNSALP was compromised by the skipping of exon 7b.

5 Regulation of *ERG* exon 7b splicing

5.1 Introduction

The regulation of alternative splicing of *ERG's* cassette exon 7b has not been investigated yet. Conservation analysis of the acceptor and donor site, branch point and polypyrimidine tract will potentially shed light on the functional importance of these key spliceosome assembly sites. Prediction of potential *cis*- and *trans*-elements will be coupled with *in vitro* validation of candidate splicing factors that may regulate *ERG* exon 7b splicing. Finally an RNA pull down of the exon 7b acceptor site will reveal splice factors that potentially regulate exon splicing at this site and, following siRNA knockdown of genes encoding the proteins detected in the pull down, the changes in PSI of exon 7b will be assessed.

5.2 Analysis of ERG exon 7b acceptor and donor splice sites

The specification of the 5′ (donor) and 3′ (acceptor) splice site is greatly important for recognising an exon and regulating its splicing. The strength of splice sites is greatly influenced by the surrounding sequence where the splicing machinery binds leading to the splicing of the exon. Two important sites that also influence the splicing of an exon are the branch point and polypyrimidine tract. Therefore a comparative analysis of the ERG exon 7b splice sites, branch point and polypyrimidine tract was performed by using the DNA sequences of human, mouse, chicken, frog and sea urchin species to produce a Clustal Omega alignment (Figure 5.1A).

Unsurprisingly the fairly degenerate splice site motifs were conserved across species as far back as sea urchin and more conservation was seen in exonic than in intronic sequences (Figure 5.1A). The conservation of two adenosine bases that may be potential branch points can be better seen in the sequence for the acceptor site and upstream intronic sequences. The polypyrimidine tract did appear to be particularly well conserved (Figure 5.1B). The sequence motif for the donor site graphically represents the good conservation of the consensus sequence based on the ERG donor site alignment (Figure 5.1C).

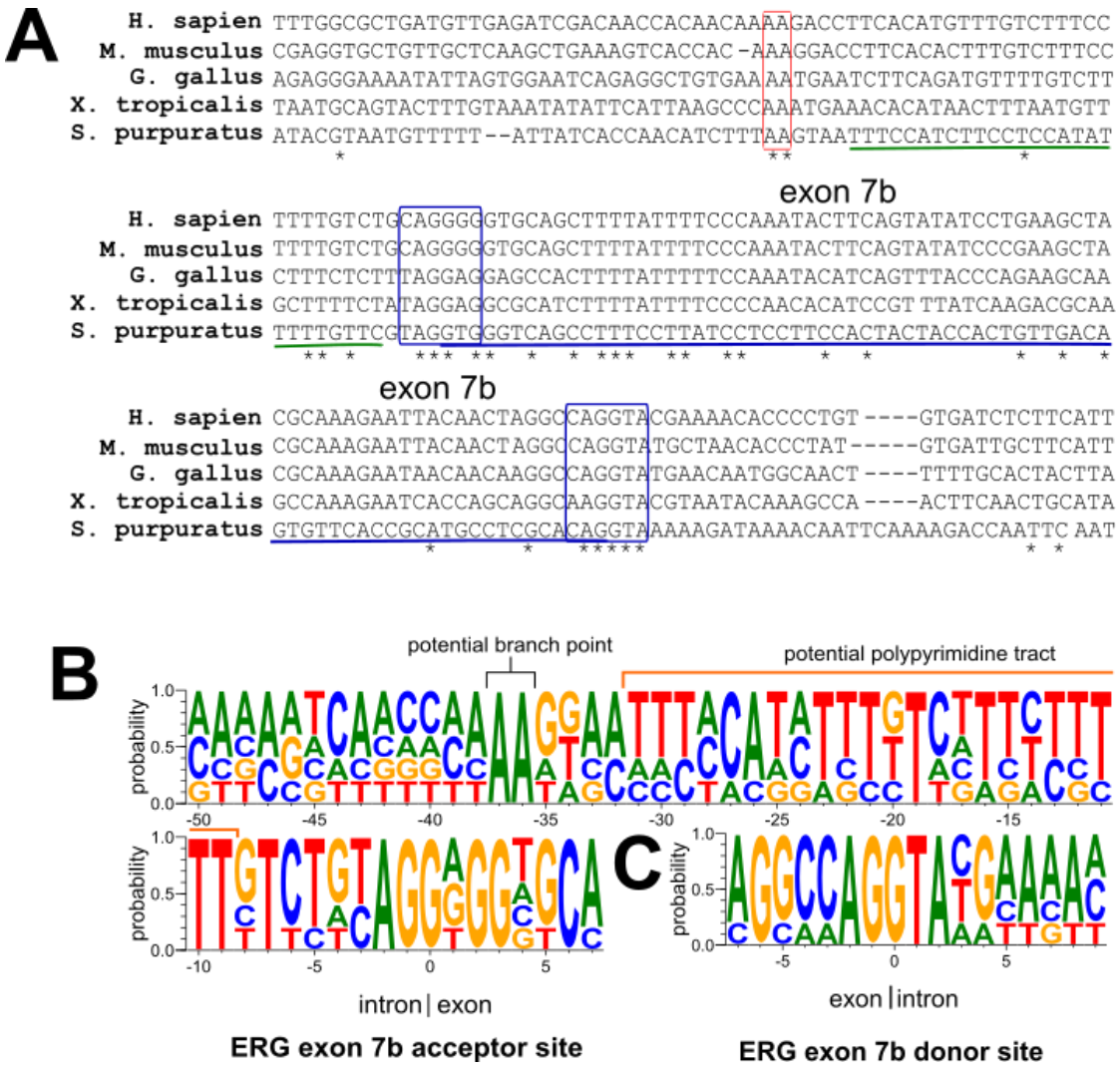


Figure 5.1 *ERG* exon 7b splice sites compared across metazoan species

A. Clustal Omega alignment of the DNA sequence of ERG exon 7b and it flanking introns was carried out. Exon 7b is underlined in blue and the splice sites sequences are shown in blue boxes. The potential polypyrimidine tract is underlined in green and predicted branch points in the red box. Asterisks (*) are conserved bases. Ensembl and EnsemblMetazoa gene ID are as follows: *Homo sapiens* ENSG00000157554 (Human), *Mus musculus* ENSMUSG00000040732 (Mouse), *Gallus gallus* ENSGALG00000016058 (chicken), *Xenopus tropicalis* ENSXETG00000023717 (frog) and *Strongylocentrotus purpuratus* SPU_018483 (purple sea urchin). **B.** The sequence logo for the acceptor site, potential polypyrimidine tract (orange bracket) and branch point of ERG exon 7b produced from the sequences used in the alignment showing the probability of having a specific nucleotide (y-axis) at that position (x-axis). **C.** The sequence consensus for the donor site of ERG produced from the sequences used in the alignment showing the probability of having a specific nucleotide (y-axis) at that position (x-axis).

5.3 Prediction of *cis*-acting splicing features that regulate exon 7b splicing

ESEs, short nucleotides sequences found with high frequency in alternatively spliced exons, are capable of regulating exon inclusion by facilitating the binding of splice factors, for example SR proteins. ESS sequences in contrast facilitate exon skipping. Regulatory sequences are also present in adjacent intronic sequences (ISEs and ISSs). The binding of SR proteins and other splice factors can cooperate with the core spliceosome machinery or override the effects of active ISSs or ESSs. To predict the effect that potential ESEs have on ERG exon 7b splicing two web server programs were sued to identify potential ESEs; RESCUE-ESE and ESEfinder (Cartegni *et al.*, 2003; Fairbrother *et al.*, 2004; Smith *et al.*, 2006).

RESCUE-ESE is an online server program that can find human hexamer ESEs that fulfil two main criteria; significant enrichment in exons compared to introns and the frequency is greater in exons with weaker splice sites. These criteria produced a list of 238 human ESEs but has been expanded to predict ESEs found in mouse, zebrafish and pufferfish species (Fairbrother *et al.*, 2004). Therefore the sequences for human and mouse acceptor and donor sites were analysed initially but did not identify any ESE sequences. Next the sequence for exon 7b was analysed. Human and mouse *ERG* exon 7b share 98.6% identity with one nucleotide variation T to C shown in Figure 5.2. Of the potential ESE predicted, three in human and two in mouse were unique and the sequences were found near the nucleotide variation between the two species (Figure 5.2 A).

ESEfinder was then used to predict which SR proteins may bind to human *ERG* exon 7b and surrounding its splice sites based on weight matrices for the motifs of four human SR proteins; SRSF1, SRSF2, SRSF5 and SRSF6. These matrices are generated from the frequency of the motif in alignments with winner sequences of SELEX experiments (Cartegni *et al.*, 2003). A threshold can be user-determined or server-determined and high scores are those that exceed it. Server-defined thresholds were used for this analysis.

Across 15 motifs all four SR proteins were predicted to bind to the sequence and a cluster of eight SR motifs were found at the acceptor site compared to four at the donor site (Figure 5.2C). SRSF2 and SRSF6 were predicted to bind in humans at the nucleotide position that varies to mice (T>C). The highest scores for each of the individual SR proteins were determined at the splice sites. The motifs used in ESEfinder are longer than those for RESCUE-ESE but with that in mind overlap of ESE sequences was present in this region (Figure 5.2A and Figure 5.2C).

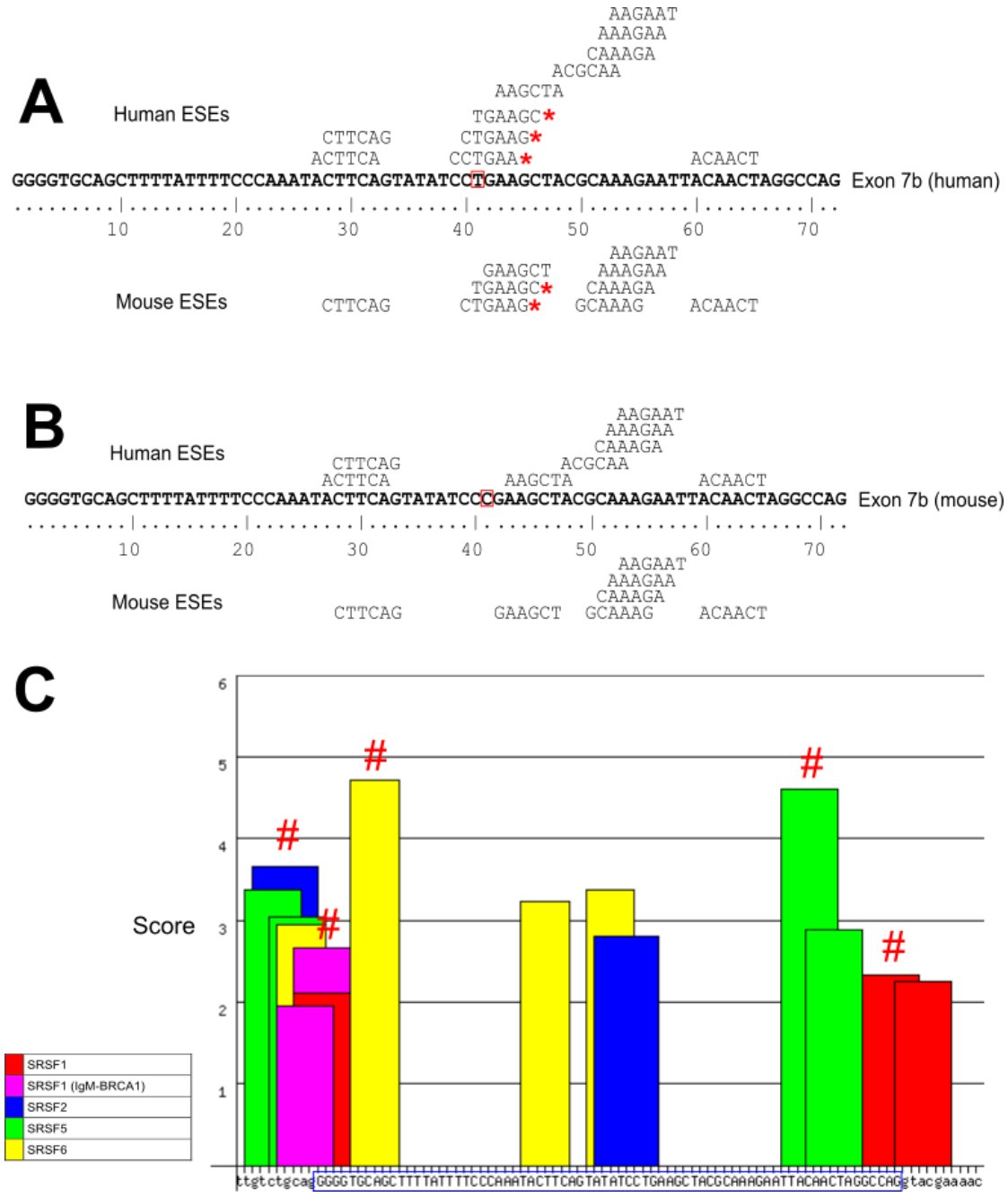


Figure 5.2 ESE prediction analysis for *ERG* exon 7b

RESCUE-ESE was used to predict human (top of exon 7b sequence) and mouse (bottom of exon 7b sequence) ESEs that may be present in the **A.** human and **B.** mouse sequence of ERG exon 7b. Red asterisks (*) denote ESEs that are unique and the red box shows the position of variation in the nucleotide sequence (T>C). C. SRSF1, SRSF2, SRSF5 and SRSF6 are all predicted to bind to human ERG exon 7b and its flanking introns. The red hash tag (#) denote the highest score for each individual SR protein and the blue box demarcates exon 7b, intronic sequence is in lowercase.

5.4 Prediction of potential *trans*-acting splice factors that regulate exon 7b splicing

In silico prediction of where splice factors and RNA binding proteins (RBPs) might bind can be informative of potential regulators for a specific splicing event. Using data from the RNAcompete assay RBPs that preferentially bind to ERG exon 7b and its flanking intron sequences was analysed. RNAcompete is an *in vitro* method utilising binding competitions for RBPs in the presence of a pool of RNAs in molar excess (Ray *et al.*, 2009). Microarray and computational methods can then be used to assess the affinity selected recovered proteins to determine the sequence preference of a specific RBP (Ray *et al.*, 2013).

Data matching RNAcompete motifs to the sequence of *ERG* exon 7b and 200bp of its flanking intronic sequences (provided by Dr Eduardo Eyras) revealed many RBPs that potentially bind. Most RBPs predicted to bind with higher motif matching scores were found in the upstream intron sequence and hnRNPA1, hnRNPA2B1, matrin3 and splice factor 3b subunit 4 (SF3B4) had the highest scores in this region. Of those only matrin3 was predicted to bind within 50bp of the acceptor site. ELAV like RNA binding protein 1 (ELAVL1) and poly RC binding protein (PCBP) both had high scores and were predicted to bind exonic sequence. ELAVL1 was also highly predicted to bind the downstream intron and had the most number of motif matches in the entire sequence analysed (Figure 5.3).

Spearman's correlations for SF and RBP that may potentially regulate exon 7b splicing were calculated by comparing normal and tumour samples from the PRAD TCGA cohort. For RBPs only those that are differentially expressed

between tumour and normal sample were used in this calculation whereas all SF were analysed regardless of differential expression.

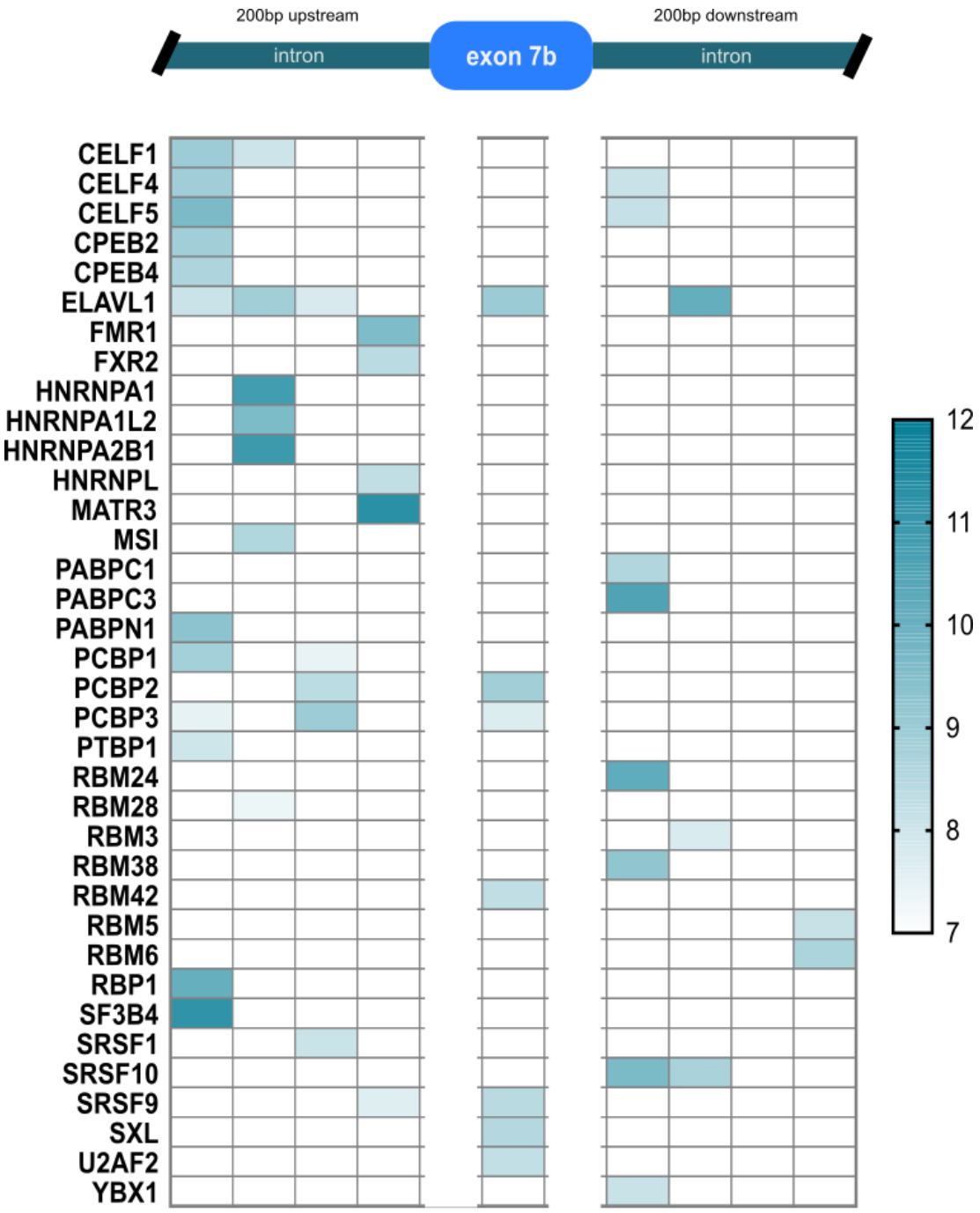


Figure 5.3 RBPs predicted to bind to ERG exon 7b and its flanking introns

Heat map showing motif matching scores (blue scale) of RBPs predicted to bind to ERG exon 7b and 200bp of flanking intronic sequence based on RNAcompete matrices. Each column represents 50bp bins except for the exon column which represents the entire 72bp of exon 7b. A total of 36 RBPs are presented.

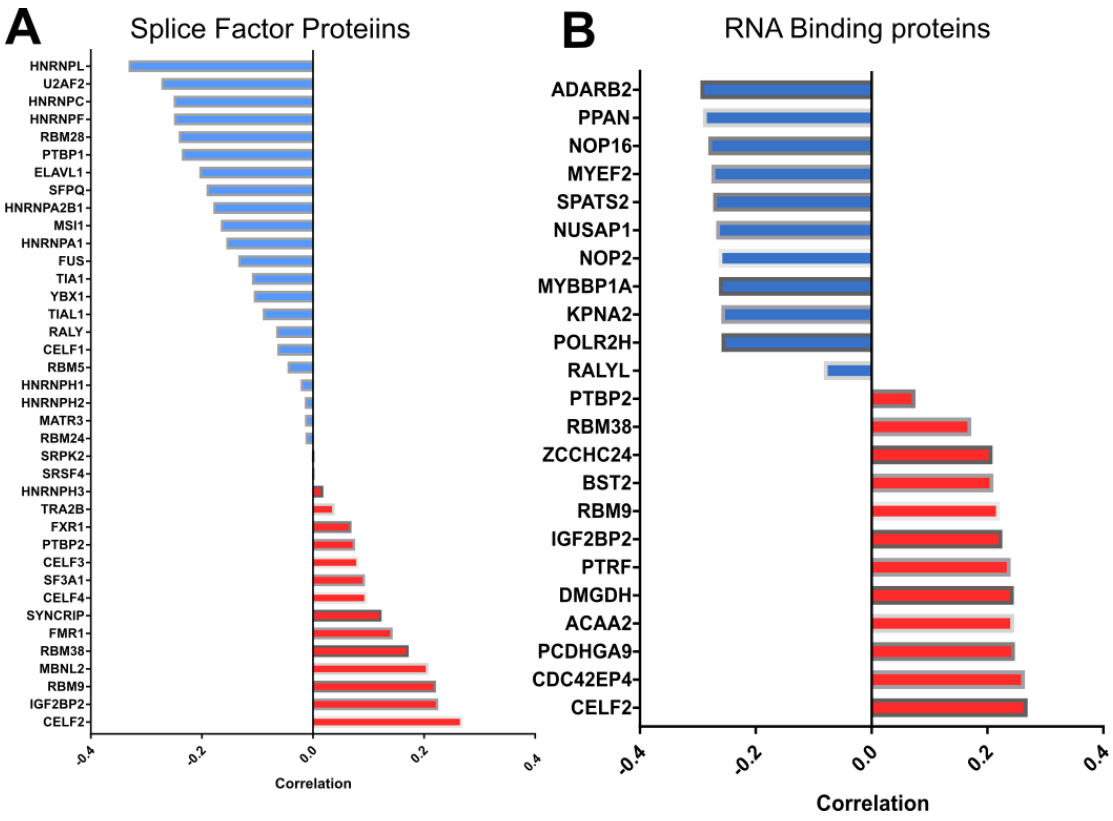


Figure 5.4 Predicted SF and RBP regulators of *ERG* exon 7b alternative splicing

The Spearman's correlation **A.** SF proteins and **B.** RBPs that have high correlation with the PSI exon skipping values of ERG exon 7b. Only RBPs that are differentially expressed between prostate tumour and normal samples were used in this calculation whereas all SFs were analysed regardless of differential expression. The red indicates splice factors that generally upregulate exon inclusion and blue are those that downregulate exon inclusion.

5.5 The knockdown of several RBPs significantly changes *ERG* exon 7b splicing in MCF7 cells

In collaboration with the University of Sherbrooke RNAomics Platform in Sherbrooke, Canada, alternative splicing using endpoint PCR coupled with microcapillary electrophoresis (ASPCR) was used to investigate how exon 7b inclusion might be regulated. A targeted RNAi screen of 56 RNA-binding proteins (RBPs) was coupled with high-throughput RT-PCR to identify RBPs that when knocked down would affect *ERG* alternative exon 7b inclusion. The RNAi screen used the human breast carcinoma cell line MCF7 and ERG exon 7b inclusion was determined for each of the 56 RBPs, some of which are known splicing regulators.

The results of this screen were represented in the heat map that expressed the change in PSI (∆PSI) in the knockdown MCF7 cells compared to a transfection control. It was established that ERG exon 7b splicing is affected by the knockdown of a number of RBPs and using a ∆PSI >10% a total of 21 RBPs were identified which had the greatest impact on the ∆PSI of *ERG* exon 7b were specifically identified (RBP names in red). Upon depletion of musashi RNA binding protein 2 (MSI2), RNA binding protein with serine rich domain 1 (RNPS1), cyclin dependent kinase (CDK11), SFRS4, SNW domain containing 1 (SNW1), transactivation response RNA-binding protein 1 (TARBP1), SFRS8, HNRNPL and steroid receptor RNA activator 1 (SR-A1) an increase in exon 7b inclusion was observed. In contrast, the knockdown of twelve RBPs (SF3A1, SF3A2, KHDRBS3, HEAB, PRPF8, RALY, RBM9, BRCA1, RNPC2, SF3B4, SFRS12 and U2AF1) resulted in decreased exon 7b inclusion (Figure 5.5A).

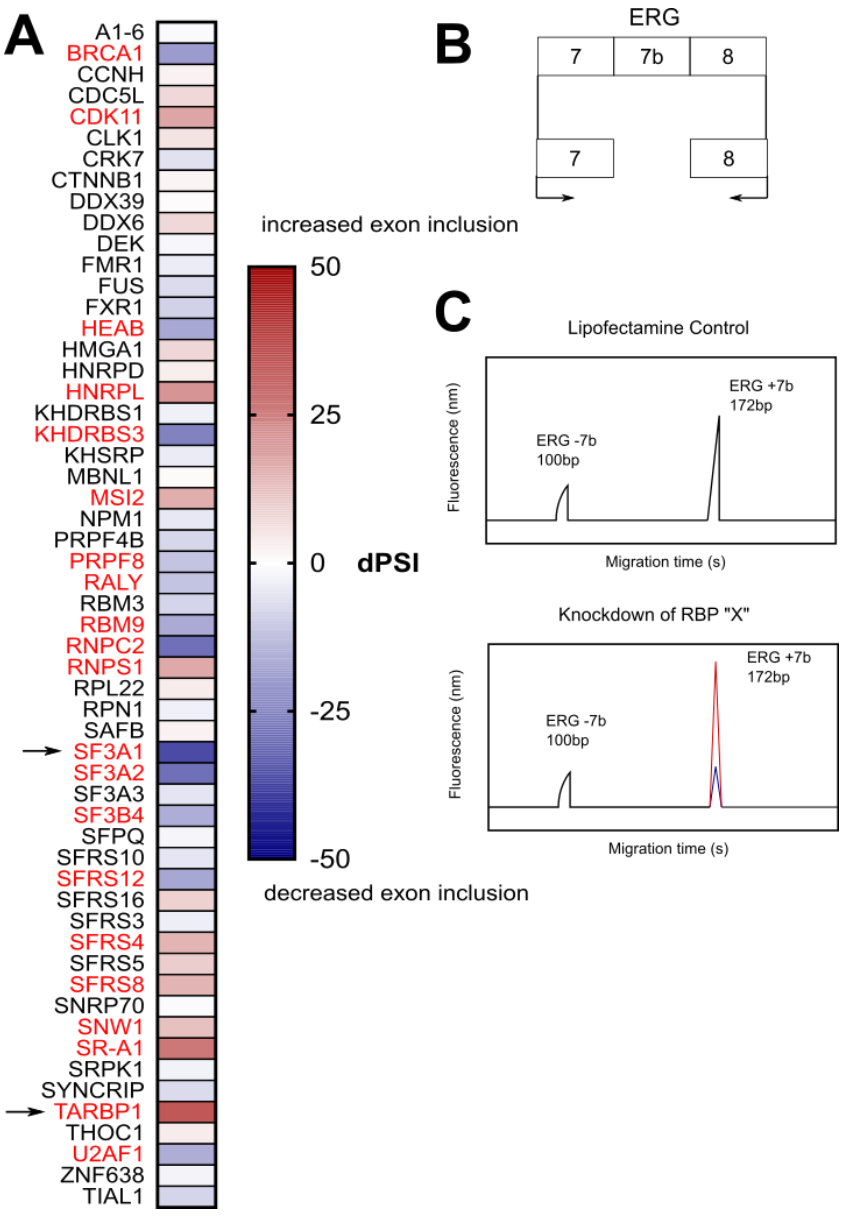


Figure 5.5 RNAi knockdown of RBPs change ERG exon 7b inclusion in MCF-7 cells

Alternative splicing using endpoint PCR coupled with microcapillary electrophoresis (ASPCR) revealed several RBPs that regulate ERG exon 7b alternative splicing. High-throughput PCR was performed to amplify +/-exon 7b isoforms in 56 MCF7 cell lines, each of which had a RBP knocked down. Expression of the isoforms is quantified by capillary electrophoresis of the PCR reactions. **A.** Heatmap showing the difference in percent splicing index (□PSI) of the ERG exon 7b ratios (exon 2 inclusion) of different RBP knockdown MCF7 cell lines that showed □PSI for the splicing event. The top results with the most apparent changes in the ERG exon 7b exclusion and inclusion levels from the screen are represented in red. **B.** A schematic representation of the splicing event. **C.** Primers located in exon 7 and exon 8 were used to amplify the ERG -7b (100bp) and ERG+7b (172bp) isoforms. Changes in the peak height of the 172bp product represent increased (red) or decreased (blue) exon inclusion.

5.6 Several proteins interact with the 3´ splice site of ERG exon 7b exon *in vitro*

An RNA pull down assay was carried out to identify potential candidate *trans*-acting factors that may regulate exon 7b splicing. A synthetic RNA oligo including the acceptor (3´) splice site of exon 7b and surrounding sequence (-20 to 20) was designed and ordered. The RNA oligonucleotide was attached to magnetic streptavidin beads prior to incubation with nuclear extract from MOLT4 cells. Proteins that bound were eluted and in addition to the original nuclear lysate were separated on an SDS-PAGE gel. Gels were stained using coomassie blue to allow for visualisation of protein bands and were imaged. Further analysis of the protein bands from the RNA oligonucleotide bound sample as compared to a no RNA lysate only control was carried out using mass spectrometry (Proteomics Facility, University of Bristol).

The oligo sequence was also analysed using SpliceAid2, which allows for cell line specific prediction of RBP that may bind to the query sequence. Figure 5.6A shows the results of this analysis based on the MOLT4 cell line as the lysate used for the pull down was from this cell line. The results of the mass spectrometry revealed numerous proteins were pulled down by the exon 7b splice site oligonucleotide. Figure 5.6B highlights the proteins that were particularly enriched in comparison to the no RNA control sample. Splicing factor proline and glutamine rich (SFPQ), SF3A2, hnRNPL, hnRNPD and SRSF5 were also identified in the ASPCR assay described in section 5.5. Peptide spectrum matches (PSM) ratios, a measure of the peptides that were more enriched in the exon 7b pull down sample compared to the no RNA

control, were calculated for all the proteins identified and Table 2 shows the values for the proteins highlighted in Figure 5.6 B.

Interestingly U2AF65, a key component of the spliceosome that binds the pyrimidine tract near the acceptor site, was pulled down (Figure 5.6 B) and had a relatively high PSM ratio of 6.82. CELF1 and family with sequence similarity 98 member B (FAM98B) had very high PSM ratios but were not predicted to bind to the RNA oligo by *in silico* analysis. Of the proteins predicted to specifically bind to the RNA oligo sequence by SpliceAid2 (Figure 5.6A) SRSF5, hnRNPF and PTBP1 were the three with the highest PSM ratios of 2.8, 4.14 and 5.4 respectively (Table 2) and all three had corresponding bands on the RNA pull down gel although confirmation using western blot was not possible due to elution sample size limitations.

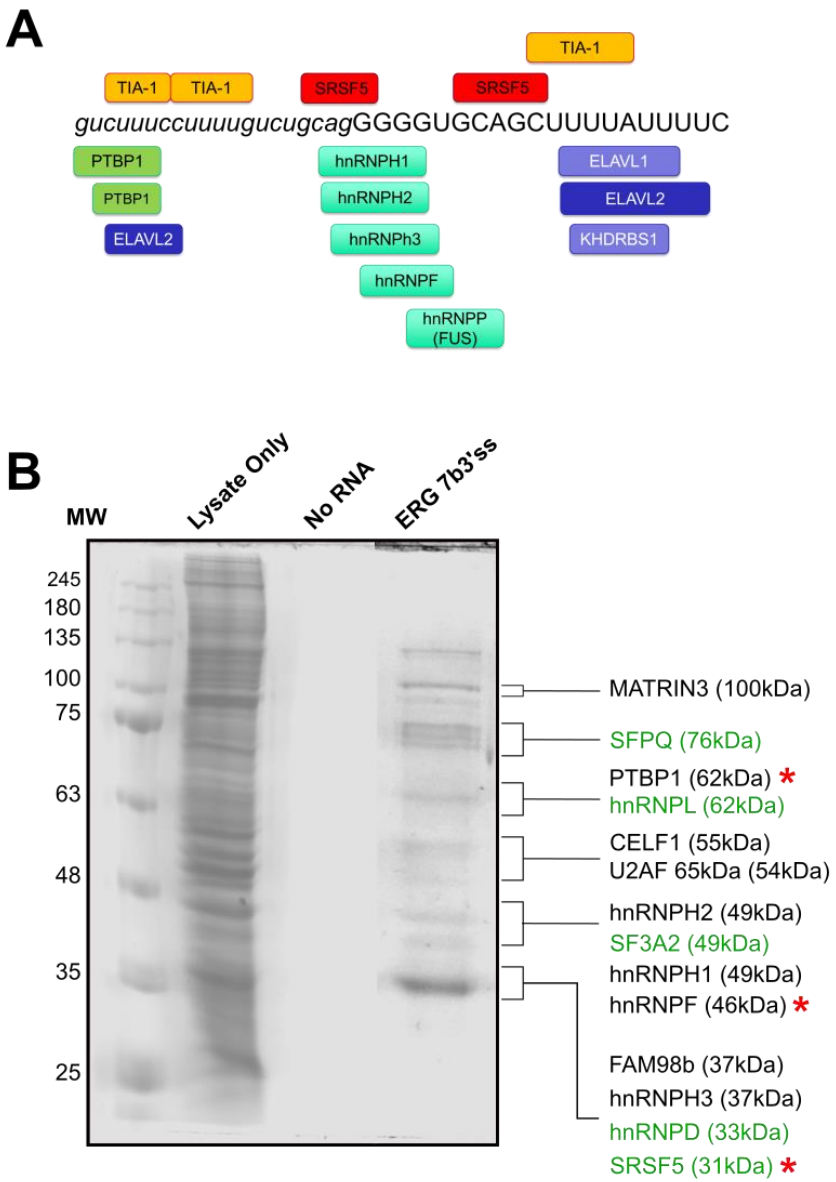


Figure 5.6 RNA pull down of ERG exon 7b acceptor site and surrounding sequence

An RNA pull down of a synthetic RNA oligonucleotide of the acceptor splice site of ERG exon 7b was performed using nuclear protein lysate from MOLT4 cells. **A.** SpliceAid2 was used to predict which RBP may bind to the exon 7b acceptor splice site RNA oligo; exon 7b sequence is in uppercase and intron in lowercase. The length of each RBP box corresponds to the predicted motif. **B.** SDS-PAGE gel stained with Coomassie blue shows the results of the RNA pull down. A no RNA and lysate only control were run alongside the acceptor site oligo (ERG7b 3´ss). The molecular weight markers are shown on the left hand side. Proteins, identified in the mass spectrometry analysis, which have molecular weights that correspond with bands in the ERG7b 3´ss lane are shown. Proteins in green were also identified in the ASPCR assay in Figure 5.5A. Red asterisks (*) represent the three proteins predicted to bind and had the highest PSM ratios. N=1

Protein identification	PSM ratio
hnRNPH1	1.80
U2AF65	6.82
CUGBP ELAV-like family member 1 (CELF1)	24.00
SF3A2	3.00
Matrin-3	6.03
PTBP1	5.40
hnRNPD	2.90
SFPQ	1.48
hnRNPL	7.00
FAM98B	39.00
SRSF5	2.80
hnRNPF	4.14
hnRNPH3	2.70
hnRNPH2	1.88

Table 2 RBPs from mass spectrometry of RNA pull down of exon 7b acceptor site

The RBPs predicted to bind to the RNA oligonucleotide sequence, with a molecular weight that corresponds to the RNA pull down gel results and peptide spectrum matches (PSM) ratios greater than 1 are shown. The PSM ratio compares the PSM of the no RNA control and exon 7b RNA pull down sample.

5.7 Knockdown of PTBP1, SRSF5 and hnRNPF did not significantly alter ERG exon 7b splicing

Results from the RNA pull down established a substantial list of potential RBPs that may regulate ERG exon 7b splicing. From this list three RBPs were knocked down to assess their potential in role in the regulation of exon 7b inclusion; SRSF5, PTBP1 and hnRNPF. The choice of RBPs was based on bioinformatic prediction, presence in the MCF7 ASPCR assay and mass spectrometry data from the RNA pull down assay.

Firstly all three factors were predicted to bind to the synthetic RNA oligonucleotide used in the RNA pull down assay (Figure 5.6 B). SRSF5 was found in both the MCF7 ASPCR assay and RNA pull down assay (Figure 5.5A and Figure 5.6B). The correlation analysis in prostate tumours predicted hnRNPF and PTBP1 as a splice factors that down regulates exon 7b inclusion (Figure 5.4A). PTBP1 is known to bind to polypyrimidine tract and it RRM interacts with Matrin3 (Coelho *et al.*, 2015). Matrin3 was pulled down (Figure 5.6B), had a good PSM ratio (Table 2) and was predicted by the correlation analysis in prostate tumours (Figure 5.4A) and RNAcompete prediction (Figure 5.3). However it was not predicted to bind to the ERG7b 3´ss RNA oligo and so PTBP1 was chosen to explore the possibility that it may be working with Matrin3 to regulate ERG exon 7b alternative splicing. A list of RBPs from the RNA pull down that were found in significant abundance in the ERG sample compared to control, and a table comparing the prediction of RBPs in this study can be found in the appendix.

RNAi was used to knockdown SRSF5, hnRNPF and PTBP1. MG63 cells were transfected with control and RBP targeting siRNA for 48 hours. Changes in ERG exon 7b alternative splicing were assessed using standard PCR. Knockdown of all three RBPs did not cause significant changes in *ERG* exon 7b inclusion as confirmed by the PSI values which were similar between all controls and knockdowns (Figure 5.7A). Knockdown for SRSF5, hnRNPF and PTBP1 was confirmed at the mRNA level using qPCR (Figure 5.7B).

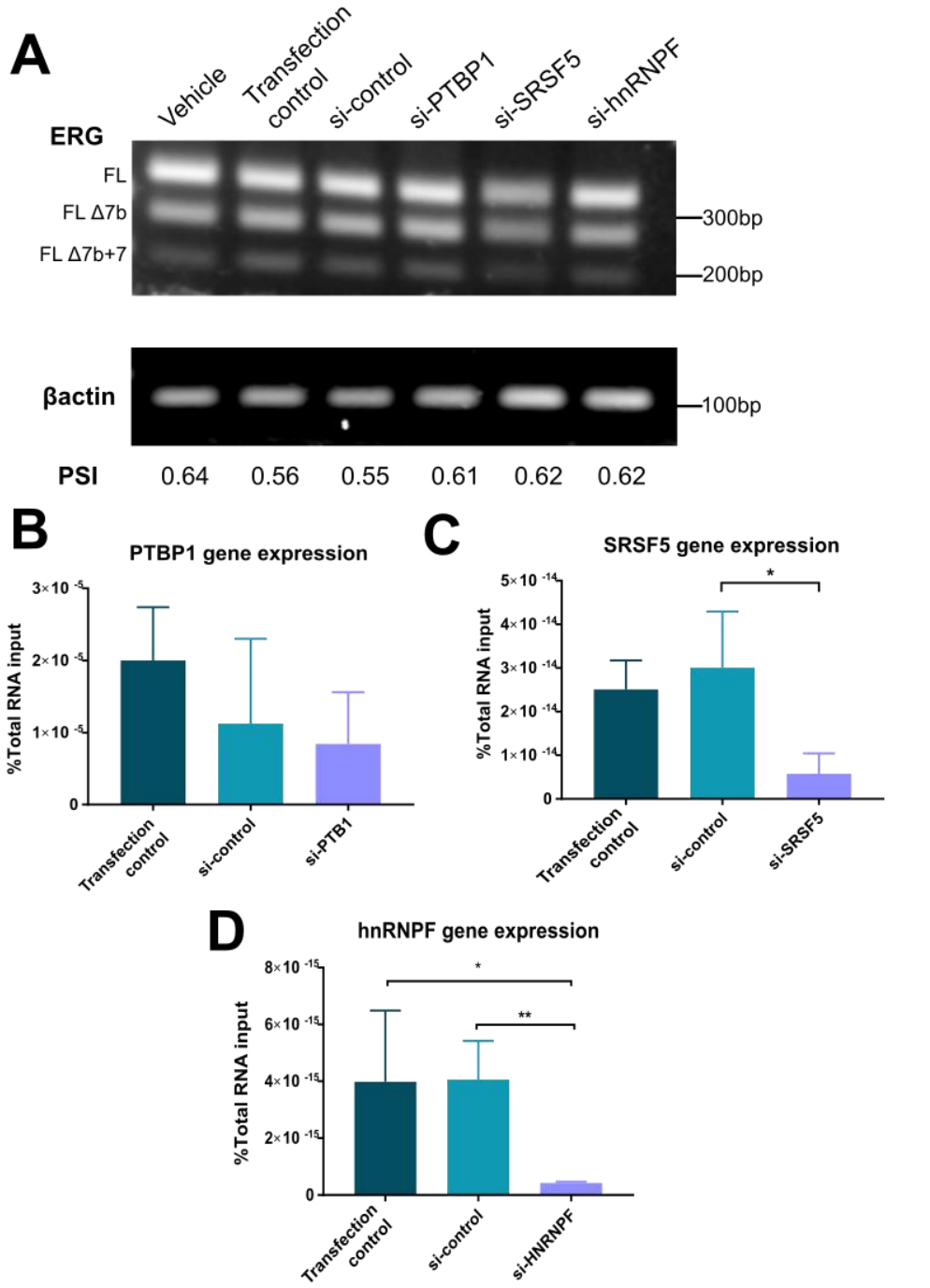


Figure 5.7 RNAi knockdown of SRSF5, hnRNPF and PTBP1 does not change ERG exon 7b alternative splicing

RNAi was used to knockdown SRSF5, hnRNPF and PTBP1. Pools of siRNA targeting each RBP and a control siRNA pool was transfected into MG63 cells alongside a vehicle control and transfection control for 48 hours. **A**. Standard PCR of ERG exon 7b splicing following RNAi knockdown of PTBP1, SRSF5 and hnRNPF and the actin control. PSI values are shown under the corresponding lane. **B.** Knockdown of PTBP1, SRSF5 and hnRNPF using qPCR showing mRNA expression. Error bars show 95%CI. (*p≤0.05 and **p≤0.01). N=3

5.8 Summary of findings

The SSOs designed to induce ERG exon 7b skipping in this study targeted the exon donor and acceptor sites. As successful skipping was achieved using both SSOs the conservation of the splice sites and their flanking sequences was examined as sequences that are conserved may play an important role in regulating the splicing of the exon. The human sequences were compared to other metazoan species for this analysis. The acceptor and donor splice site consensus sequences and branch point of *ERG* exon 7b are well conserved across metazoan species. However the polypyrimidine tract is less conserved. The conservation of the acceptor and donor site is expected and has been reported before (Abril, Castelo and Guigó, 2005). The lack of conservation in the polypyrimidine tract suggests that proteins binding here may not be as important to regulation of the splicing of exon 7b.

Analysis to identify SREs, SFs and RBPs revealed several *cis*- and *trans*-acting elements were predicted to regulate *ERG* exon 7b splicing. Most predicted ESEs were conserved between human and mouse with ESEs unique to humans found near a single nucleotide variation between the two species (T>C) in the ERG exon 7b sequence. This is unsurprising as the change in sequence would create potential for organism specific splicing regulation. Several SR proteins were predicted to bind to the exon 7b sequence and most cluster around the splice sites where there is greater sequence conservation. Of particular interest was that all four SR proteins were predicted to bind to the 3´ splice site which when targeted with the

SSOs showed the most exon skipping. It may be that this splice site is more critical to determining the inclusion exon 7b.

Prediction of splicing regulators is useful in providing potential targets for experimental validation as *ERG* exon 7b splicing regulators. RNACompete prediction provided several potential proteins, with those with the highest scores mainly binding in the intronic sequence both upstream and downstream of exon 7b. Matrin3, HNRNPA1, HNRNPA2B1, SF3B4 were predicted to bind in the upstream intron and ELAVL1, which was predicted to bind in the downstream intron, were among the highest scoring RBPs. Spearman's correlations of SFs and RBPs that may regulate ERG exon 7b from the TCGA PRAD cohort also ascertained some potential regulators namely HNRNPL and CUGBP ELAV-like family member 2 (CELF2).

Among splicing regulatory proteins (SFs and RBPs) able to modify *ERG* exon 7b inclusion the most significant changes seen in the ASPCR experiment in breast cancer cells (MCF7) were with knockdown of TARBP1 and HNRNPL which increased ERG exon 7b inclusion. Knockdown of splice factor 3 subunit 1 (SF3A1) and breast cancer 1 (BRCA1) decreased exon 7b inclusion. There is an added complication when considering the role of these splicing regulating proteins when one takes account of mutations that alter the splicing of the proteins themselves. Therefore there is a complex layering and interplay between mutations in *trans*- and *cis*-acting elements and how any changes in them may contribute to the pathogenesis related to any one splicing event that may be deemed significant. For instance that BRCA1 is an important regulatory factor in MCF7 cells in this study is not

unusual as overexpression and mutation in the protein is often seen in breast cancers (Vissac *et al.*, 2002). However BRCA1 could be of less, equal or no importance in regulating ERG splicing in another cell line model, *in vivo* or human cancer patients.

An RNA pull-down of *ERG* exon 7b acceptor site identified several splicing regulators. Choosing which proteins may regulate exon 7b splicing from the pool of potential regulators from the pull down and other predictive data lead to a choice of three initial candidates. The knockdown of PTBP1, HNRNPF and SRSF5 using siRNA did not significantly change exon 7b splicing suggesting they may not be relevant to *ERG* exon 7b splicing regulation. Further experimentation can be carried out from the data provided in this study.

Although the details of the regulation of ERG exon 7b have not been determined definitively yet, several splice regulating proteins have been identified with some *in vitro* validated in an ASPCR assay of MCF7 cells. Further analysis could be carried out by knocking down other potential candidate splice factors identified in the RNA pull down. In addition the list of proteins that are predicted to bind and that also appear in the knockdown experiments could be used to inform decisions on future potential candidate regulators of exon 7b inclusion.

6 Discussion and Future Work

The aim of this study was to understand the function and regulation of inclusion of ERG exon 7b. The changes in the splicing of this exon were hypothesised to significantly alter the transcriptional function of ERG by changing the size and nature of the AD. The inclusion of exon 7b has been shown to increase in PCa (Hagen *et al.*, 2014) and contribute to malignant cell behaviour (Wang et al., 2008). Therefore further investigation of the expression of ERG exon 7b splice isoforms is of interest as it may help understand oncogenic processes. If this cassette exon significantly alters the function of ERG then the regulation of the cassette exon may go astray in cancer. As well as the ERG splice isoform itself, regulators of exon 7b splicing could themselves be targeted therapeutically.

6.1 Evolutionary conservation of ERG and its exon 7b

As a starting point it was deemed important to understand and determine the evolutionary origins of exon 7b.The functional importance of ERG sequence was suggested by its conservation as far back as the metazoa with *ERG* being identified in what the authors of the study called an “unknown nematode” species (Degnan *et al.*, 1993). Soon after this was reported a putative *ERG* gene was identified in the *Nereis diversicolor* species of worms (Lelibvre-Chotteau *et al.*, 1994). Previous conservation studies for *ERG* and other ETS factors had focused on the ETS and pointed domains. However a similar approach was adapted to investigate both ERG exon 7b and ERG conservation.

This study has shown *ERG* exon 7b conservation in echinoderms including a sea urchin species. Previously the ETS domain was reported to be conserved in sea urchin species (Rizzo *et al.*, 2006); in this study AD conservation was also demonstrated, including several serine residues that may be potential sites for post translational modification. ERG exon 7b encodes for part of the AD and within the AD itself are two serine residues one in exon 7 and the other in exon 7b located downstream of the conserved DEF domain, an ERK docking site. Mutational changes in the AD have been shown to change transcriptional activity (Siddique *et al.*, 1993). The DEF domain located in exon 7b is crucial to anchoring ERK and facilitates ERG phosphorylation at the aforementioned serine residues, most especially the serine in exon 7b (Huang et al., 2016), activating transcription of gene targets required for cellular processes such as migration (Kedage *et al.*, 2017). Therefore the conservation of this exon may be

driven by the requirement of this phosphorylation site for appropriate signal transduction.

Also the conservation of ERG has been shown to go even further back to the porifera phylum as ERG was identified in *A. queenslandica* sponge species. It is thought that sponges and metazoa diverged from an ancient common ancestor some hundred million years ago (Müller *et al.*, 2001). The presence of ERG in a sponge points towards its possible presence and importance to the common ancestor for animals and the development of multicellular life. Therefore this study provides further evidence that ERG, including exon 7b, is among the most conserved metazoan proto-oncogenes, and this implies an essential role for the gene in development of a wide range of organisms across the taxa.

6.2 Role of ERG exon 7b in oncogenic processes

Having established the conservation of ERG exon 7b, the extent of its inclusion was investigated in several cell lines both *in vitro* and *in silico.* Overall exon 7b inclusion was favoured in human cancer cell lines tested suggesting that the ERG transcriptional activity in these cells may be mainly attributed to exon 7b inclusion isoforms. RNA-seq data also showed a high level of ERG exon expression especially in the HUVEC cells. ERG is vital for regulating signalling pathways in endothelial cells (Birdsey *et al.*, 2008; Yuan *et al.*, 2009; Dryden *et al.*, 2012), and so high expression of exon 7b in this context further points to it possible functional importance.

SSOs were designed against the two splice sites to induce exon 7b skippingand caused exon 7b skipping in two cell lines, MOLT4 and MG63 showing that the SSOs are effective in various tissue types. Initially the in vitro cell experiments were going to be conducted in both VCaP prostate cancer and MOLT4 cell lines in this study and this is why prostate cancer data is included. However the SSO *in vitro* cell biology work was conducted by Dr Sean Porazinski and this in addition to the data from this study is published (Jumbe *et al.*, 2019).

Going forward the MG63 cell line was used as the experimental model as the dose used in MOLT4 cell lines was significantly higher than in MG63. Therefore the assumption was that lower doses would result in less off target effects. Moreover working with an adherent cell line presented practical advantages. Interestingly ERG expression and function in MG63 cells had not previously been reported; therefore better understanding of ERG in this context.

An additional amplicon could be detected in the ERG PCR using cDNA from MG63 cells that represented the exclusion of both exons 7 and 7b. Due to project time constraints this investigation only focused on exon 7b. Future studies looking at removal of the entire AD or specifically studying exon 7 instead of exon 7b should be carried out using this cell line. Although exon 7 exclusion has been reported it is less frequent and normally in minor isoforms which use less commonly used promoters (Zammarchi, Boutsalis and Cartegni, 2013) perhaps explaining why exon 7 skipping was not detected in most other cell lines tested.

The importance of the AD has been shown as mutation or elimination of the domain reduces ERG transcriptional activity (Siddique *et al.*, 1993). Furthermore ERG protein activity is modulated by its phosphorylation status. The expression of the full length AD facilitates the phosphorylation of ERG by ERK via the DEF domain located in exon 7b in collaboration with the serine residue (S215) in the upstream exon 7 which in turn facilitates the phosphorylation of S96 (Selvaraj *et al.*, 2015). If exon 7b is excluded S96 is not phosphorylated and therefore downstream targets and signal transduction pathways activated and repressed by ERG will vary when if exon 7b is not included. Kedage *et al.* (2017) demonstrated how phosphorylation of S96 modulates the expression level of ERG target genes and enhances cell migration in RPWE cells. In addition hypothetically the serine residue found in the protein sequence for exon 7b may be a site for further phosphorylation by other kinases.

A preference for exon 7b inclusion in several malignant cell lines and PRAD patients was observed and suggest that the full AD, with exon 7b amino-acids included, encodes an ERG isoform with a higher oncogenic potential. The data presented in this study demonstrates SSO induced ERG exon 7b skipping decreased cell migration, invasion and proliferation but increased apoptosis. This was in agreement with the hypothesis that ERG exon inclusion increases oncogenic potential, based on previous literature (Wang et al., 2008; Hagen et al., 2014). All of these observations lead me to hypothesise that full transcriptional activity may be favoured in a number of biological contexts including cancer.

In particular Wang et al. (2008) found that TMPRSS2-ERG variants that included exon 7b increased proliferation and invasion and attributed this to potential protein folding and dimerization changes. As the splicing of exon 7b alters the polypeptide sequence it would be expected to result in a varied tertiary structure in the protein. It has been suggested that the proximity of the AD to both the pointed and ETS domain which are primarily involved in homo and hetero dimerization interactions may be required to facilitate the process (Carrère *et al.*, 1998).

Previous studies that investigated the role of ERG in proliferation (Thoms *et al.*, 2011), invasion (Singareddy *et al.*, 2013), migration (Deplus *et al.*, 2017) and apoptosis (Birdsey *et al.*, 2008), and examined either native ERG, *TMPRSS2-ERG*, other fusion variants (Irifune *et al.*, 2005; Sotoca *et al.*, 2015) or a combination. The present study did not assess whether or not MG63 cells

express any ERG fusion variants; however all ERG exon 7b containing isoforms would have been targeted by the SSOs.

There is a potential that these isoforms may be potential biomarkers which may become a key part of cancer diagnosis, treatment and in determining prognosis. Currently ERG has been detected and monitored in PCa patients using urine-based and immunological testing methods (Yao *et al.*, 2014). The expression of *TMPRSS2-ERG* variants +7b is associated with a poorer outcome in most cases with a 58-80% chance of recurrence within 5 years (St. John *et al.*, 2012). This suggest that clinical picture of a patient may be informed by the *ERG* or *TMPRSS2-ERG* status. However an interplay exists between native ERG variants and fusion variants that influences the tumorigenic potential of ERG (Mani *et al.*, 2011), and so diagnostic tests should include the measurement of ERG alternative splicing too.

Zammarchi *et al.* (2013) observed a specific expression profile of native ERG splice isoforms was identified in PCa cells compared to normal tissue and this led to the suggestion of a positive feedback loop with *TMPRSS2-ERG* variants that can circumvent androgen dependence. Of the three native *ERG* promoters, a switch occurs where promoter B replaces promoter C as the principle promoter in both from PCa samples and cell lines (Zammarchi, Boutsalis and Cartegni, 2013). Therefore the ERG splice isoforms expressed are changed. As promoter elements for the *MYC* proto-oncogene, BHLH Transcription Factor (c-Myc), a downstream target gene of ERG elements are present on promoter B, it may be that ERG transcription can be maintained independent of the

androgen dependent *TMPRSS2-ERG* through a feed-forward loop between ERG and c-Myc (Zammarchi, Boutsalis and Cartegni, 2013).

In addition the expression of *ERG* splice isoforms is complicated by the presence of the fusion with *TMPRSS2* in prostate cancer. Changes in the androgen responsive element that mediates the androgen response of *TMPRSS2* would also contribute to androgen driven ERG expression. Reduced binding to the androgen responsive element of *TMPRSS2* was identified and shown to be caused by a single nucleotide polymorphism in the gene sequence (Clinckemalie *et al.*, 2013). Therefore fusion genes with this mutation may not express ERG at comparative levels potentially reducing the clinical significance of the presence of the fusion.

Similar analyses would need to be carried out in cancer cases involving *FUS-ERG* and *EWS-ERG* fusion variants but what these studies highlight is the complication of understanding the clinical significance and contribution of native *ERG* and ERG fusion splice isoforms to cancer pathogenesis. Therefore although exon 7b inclusion generally seems to be an indicator of poor clinical outcome additional, detailed analysis of the expression profile of splice isoforms in different cancers needs to be carried out to validate the significance of exon 7b as a clinical biomarker.

6.3 Therapeutic potential of SSOs

To gain an indication of the *in vivo* efficacy and potential of the SSOs an MG63 xenograft model was treated over a two month course period. Tumour volume and weight were both reduced by ERG E7b3 SSO over the time course. However the control SSO seemed to have undesirable effects as tumour weight and volume were both reduced with this treatment as well although not to the same extent as the ERG SSO. This may be a consequence of the toxicity of the chemistry of the vivo-morpholinos in this model, though mouse body weight was not affected over the treatment course between groups suggesting minimal toxicity at the systemic level. It may be that within the tumour milieu the potency of the SSO is altered by pharmacokinetic complications.

Clinical trials in humans using SSOs are underway especially for neuromuscular diseases with the first ever treatment developed for spinal muscular atrophy the FDA approved ASO Spinranza® which has greatly increasing patient survival (Prakash, 2017). With regards to cancer ASO against eIF4E have been used in a clinical trial to treat advanced colorectal cancer. In patients treatment with the ASO alone and in combination with chemotherapy showed increased stabilisation of disease and survival but suggestions were made that efficacy was inhibited by stromal binding of the ASO (Duffy *et al.*, 2016). Therefore the dose received by each mouse may be equal but the proportion of this that would be pharmaceutically active could vary.

One of the benefits of vivo-morpholinos is that the same ASO can be used in cell culture and a variety of animal models as demonstrated by this investigation. Finally it must be acknowledged that this cohort was limited in size

and only a first pilot study and therefore further work is required to finesse the efficacy of SSOs *in vivo*. In light of all this encouragingly the SSOs have be proven to have potential for efficient inhibition of tumour growth in an *in vivo* context.

6.4 Role of exon 7b in osteoblast maturation

TNSALP is used as a clinical marker in prostate cancer. Many prostate cancer patients with advanced disease develop metastases in their bones (Kamiya *et al.*, 2012). The metastases are usually osteoblastic in nature, which is bone producing rather that osteoclastic which is bone destructive (Brown and Sim, 2010). These lesions mimic *de novo* bone cancer and exhibit high levels of TNSALP which can be detected in the serum (Kamiya *et al.*, 2012). Individuals with high TNSALP levels have a poor progressive-free survival and overall survival rate (Li *et al.*, 2018). This study investigated the relationship between TNSALP and ERG, two important prostate cancer biomarkers.

The findings of this study have revealed an unexpected role for ERG in osteoblast maturation. The treatment of differentiating osteoblasts with SSOs attenuated the process which was measured in terms of TNSALP induction. *TNSALP* mRNA expression was downregulated and the expression of two important osteoblast genesis regulators, *RUNX2* and *OPN* was upregulated. *TNSALP* is expressed at higher levels in more mature osteoblast cells. The reduction of activity and expression in the presence of SSO suggested that ERG may regulate its transcription.

The initial characterisation of the *TNSALP* promoter revealed the presence of potential ERG binding sites in the *TNSALP* promoter and interaction of ERG on this promoter was confirmed using ChIP analysis. There are also potential binding sites for the ETS transcription factor *SP1* (Terao *et al.*, 1990). Inhibition of SP1 promoter binding by Mithramycin A in the presence of 1,25D/FHBP co-treatment reduced the levels of TNSALP (Mansell *et al.*, 2016) suggesting that

the *SP1* promoter binding sites may have a biological function. Therefore there is potential to investigate if an *ERG* isoform-specific interaction with the *TNSALP* promoter is required to activate *TNSALP* transcription. Alternatively it is also possible that an interaction occurs between ERG isoforms and SP1, thereby activating *TNSALP* transcription.

The observations may also suggest that ERG plays a role in bone mineralization regulation via interaction with 1,25D and LPA sensitive pathways. The expression of TNSALP is upregulated in the presence of 1,25D (Bonewald *et al.*, 1992). Indeed *TNSALP* mRNA stability was shown to be increased by the addition of 1,25D to MG63 cells (Orimo and Shimada, 2006). However the exact mechanism of action remains unexplained. More recently the vitamin D receptor (VDR) was identified as a synergistic partner of the *TMPRSS2-ERG* gene fusion in PCa cell lines inducing the expression of CYP24A1, the enzyme responsible for the inactivation of 1,25D. The study carried out by Kim *et al.* (2014) expressed type VI ERG + exon 7b in LNCaP cells in a doxycycline-inducible dose-dependent manner and found that intracellular 1,25D levels were markedly lowered. In the case of this study it could be that ERG isoforms regulate TNSALP expression post-transcriptionally as well through interaction with the VDR induced by 1,25D.

Wnt and Notch signalling pathways that regulate osteoblast differentiation are known to be targets of ETS transcription factors and more specifically ERG (Wu *et al.*, 2013; Fish *et al.*, 2017). Wnt signalling is a well-studied regulator of RUNX2. After stimulation at the cell surface via Wnt receptor a signal is transduced that leads to the translocation of beta-catenin to the nucleus where

it binds to TCF and LEF1 forming a complex (Wu *et al.*, 2013). This complex is a transcriptional regulator and in osteoblasts activates the expression of RUNX2 and other key osteogenic genes. ERG stabilizes beta-catenin through interaction with Frizzled-4, a Wnt receptor (Birdsey *et al.*, 2015), and also interacts with LEF1 (Wu *et al.*, 2013). Interestingly LPA has been shown to also upregulate beta-catenin dependent Wnt target gene expression via aggregation of beta 2 integrins in ovarian cancer cells (Burkhalter *et al.*, 2015). Therefore increased levels of *RUNX2* may be mediated by ERG isoform specific regulation as well as LPA mediated effects on Wnt signalling.

Finally that the changes in osteoblast differentiation are triggered by the SSO-induced skipping of ERG exon 7b provides an example of how the altered transcriptional activity of ERG significantly changes the biological function of the transcription factor. More work is required to specify which downstream targets and signalling pathways are being affected by this splicing event. However these data support the hypothesis that changes in the structure of the ERG AD caused by exon 7b skipping modulates cell functions as a result of changes in transcriptional transactivation.

6.5 Regulation of exon 7b inclusion

Another aim of this project was to identify splicing regulators of ERG exon 7b. Aberrant regulation of alternative splicing is clearly associated with human disease including cancer. As ERG is an oncogenic transcription factor is not unreasonable to assume that changes in the alternative splicing of exon 7b may be driving increased inclusion that is observed in cancer cell lines and in patients. Using both *in silico* and *in vitro* methods a series of proteins were identified that potentially regulate exon 7b inclusion. However one of the main challenges to assess the biological relevance of these predictions was the *in vitro* validation of predicted regulators of alternative splicing.

The ASPCR assay allowed for validation of some the predicted proteins that are present in the repertoire i*n vitro* in MCF7 cells. Several splice factors were also identified in the RNA pull down of the 3´ end of ERG exon 7b. However the splicing of a gene is not only species specific but may be tissue specific in some cases. Therefore although some SFs and RBPs were found in the RNA pull down, ASPCR and prediction results that different cell lines were used as the basis for these analyses may negate the relevance of some of these proteins as ERG exon 7b splicing regulators. This is because splice factors may regulate splicing in a tissue specific manner (Lee and Donald C Rio, 2015).

SREs are relatively degenerate consensus sequences around 10 nucleotides in length and as such many splice factors may bind to them (Keren, Lev-Maor and Ast, 2010). These RBPs also possess a certain amount of redundancy depending on the structure of their RNA binding domain, which in turn is dependent on how the RBPs themselves are spliced (Hallegger, Llorian and

Smith, 2010). Further the activation of the spliceosome also relies on external ques such as the epigenetics, hormones and other small molecules (Luco *et al.*, 2011; Roy, Haupt and Griffiths, 2013).

This may have been the reason as to why no changes were observed in exon 7b splicing following siRNA knockdowns of PTBP1, HNRNPF and SRSF5. Of the three candidate proteins SRSF5 was detected in the pull down, ASPCR and in silico prediction whereas the other two were not detected in ASPCR assay. However the observations may have been affected by 1) experimental limitations (siRNA dose and treatment period) 2) mutations in these splice factors in the cell line derived nuclear extract preventing them from regulating exon 7b or 3) missing co-factors not present in the nuclear extract. Not only could the assay be repeated under different experimental parameters but different splice factors could also be targeted. Also to appropriately understand the splicing code for the splicing of ERG 7b would require further interrogation of possible combinations, possibly using a high throughput approach.

As the most effective SSO targeted the 3´ splice site and this suggests that this splice site may be more influential in driving the splicing of the exon and this is why the RNA pull down focussed on this splice site. Analysis of ESEs also predicted that all four SRSF proteins analysed bind to this site at least once. However it would be wise to also conduct analysis of the 5´ splice site of exon 7b, as well as intra-exonic sequences. Analysis actually predicted and highly scored many SFs and RBPs binding to the intronic sequence flanking exon 7b. Therefore although this study provides a good foundation for exon 7b inclusion

regulation analysis more work is required to fully understand how the event is regulated.

6.6 Main conclusions

In conclusion, this study aimed to assess the contribution of ERG exon 7b to the oncogenic function of the transcription factor and elucidate the regulation of the alternative splicing of the exon. The conservation of ERG exon 7b in echinoderms points to its functional importance. Cancer cell lines and PRAD patients favour expression of ERG splice isoforms that include exon 7b. Cell biology analysis revealed SSO induced exon 7b skipping reduced in cancer hallmarks such as reduced proliferation, cell motility and cell survival. ERG exon 7b specific changes in osteoblast differentiation partly mediated by ERG binding the TNSALP promoter is a clear example of exon 7b specific biological change. Therefore this study supports an oncogenic function for the full length AD of ERG which is more greatly expressed in cancer, and these results are in agreement with the observations and suggestions made in other studies (Wang et al., 2008; Hagen et al., 2014). In addition this study provides experimental tools in the form of the MG63 cell line as a new model in which to investigate ERG as well as validated SSOs that cause exon 7b skipping.

The ERG targeting SSOs were able to reduce tumour growth in a mouse xenograft model. However size reduction was also observed with the control SSOs and therefore further *in vivo* work would require a new control SSO be designed. An alternative scrambled SSO was trialled but showed no reduced effect compared to the intron targeting SSO. Therefore further careful design of the SSO structure itself may be necessary going forward to ensure that chemistry toxicity is not masking sequence specific effects *in vivo*.

This study recognised a need to investigate the regulation of alternative splicing in ERG and other ETS transcription factors, particularly specific splicing event that may contribute to a disease phenotype. This study has begun to understand the splice regulators that may contribute to regulation of ERG exon 7b inclusion but more investigation must be performed to discern the specificities of the regulation of this splicing event.

This study has highlighted the importance of the transactivation domain of ERG as a key site of regulation for the activity of the transcription factor. Truncating the domain produces a functional protein that has less oncogenic activity as supported by the data in this study. Therefore manipulation of the splicing of exon 7b is a feasible and potential therapeutic target to reduce disease burden in ERG related cancers.

6.7 Future Work

6.7.1 Further study of the role of exon 7 in conjunction with exon 7b

As both exon 7 and exon 7b encode for the AD study of the contribution of each individual exon and both exon to changing ERG function could be of interest. The approach for this could be two fold. SSOs could designed to cause exon 7 skipping and skipping of both exons. Similarly gene editing methods such as clustered regularly interspaced short palindromic repeats (CRISPR) CRISPR associate protein 9 (Cas9) could be used. CRISPR-Cas 9 uses specifically designed guide RNA to direct the Cas9 bacterial nuclease to the intended target site where the enzyme makes an incision in the genomic DNA. Formation of a double strand break leads to its repair via the non-homologous end joining (NHEJ) or homology-directed repair (HDR) pathway. It is at this point that users are able to apply their genomic manipulation to generate the desired modification (Gaj, Gersbach and Barbas, 2013; Ran *et al.*, 2013; Munoz *et al.*, 2014).

ERG variants lacking exon 7 and 7b separately and together could be generated in a cell line that does not normally express the mutation using CRISPR-Cas9. Upon successful generation the model cell line, cell biology assays as used in this study could assess the effect of that each genetic modification has *in vitro*. Further the exons could be removed in an ERG expressing cancer cell line to investigate if biological changes are cell line, cancer or tissue specific. These genome editing experiments would complement the SSO work for which cell biology analysis would also be carried out. The future use of the SSO would require the redesign of the SSO to reduce off target effects. Both standard and

scrambled controls were tested in the course of this study therefore alternatives would need to be used. An SSO with 5 or more base pairs mismatched to the target SSO could be an option. Similarly the chemistry of the vivo moiety on the morpholino could require be modification. Lastly these manipulations would allow for more powerful *in vivo* experiments in which the selected model cancer cells, now genetically modified, would be injected orthotopically into mouse either directly into the prostate or an alternative site.

It would also be of interest to understand how the exclusion of exon 7 and 7b affect the RNA secondary structure of the *ERG* transcript and of the protein structure of ERG. It is known the RNA conformation can contribute to the regulation of splicing and the presence of structures such as hairpins and G4 may drive splicing of these particular exons. Furthermore resolution of the protein structure could be informative of the binding interactions that occur within this domain. To study RNA structure chemical mapping and nuclear magnetic resonance could be used whereas X-ray crystallography could be used to study protein structure.

To gain a better insight into transcriptional changes caused by alternative splicing qPCR of recognised ERG targets could be carried out alongside traditional transcriptional assays such as the renilla-luciferase assay. This would reveal promoters for which transcription is directly affected as well as target genes for which ERG alternative splicing may indirectly influence expression. In addition mutagenesis of phosphorylation sites in exon 7b (and perhaps exon 7) and the DEF domain prior to transcriptional investigation would

inform on the specific influence that these sites contribute to the changes seen in ERG function.

6.7.2 Further investigation of the regulation of ERG exon 7b regulation

The data in this study with regards to ERG exon 7b regulation provides a robust resource from which to continue investigation. A number of approaches could be employed to gain further insight. Firstly mutagenesis of predicted ESEs could be carried out and the effect on exon 7b splicing discerned using PCR initially. ISE prediction could be carried out and these could also be investigated particularly in light of the RNACompete data that showed that a number of RBPs may be binding in the intronic sequence flanking exon 7b.

Several ERG minigenes with mutations in the ESEs and ISEs would be designed and transfected into a cell line that does not normally express ERG. If the role of a specific splice factor were to be investigated alongside then an expression vector for the protein could be co-transfected.

Certainly further refinement of the experimental conditions for the siRNA knockdowns of PTBP1, HNRNPF and SRSF5 could be carried out, and other proteins pulldown by at the acceptor site could be investigated as well. The RNA pull down experiment could be repeated based on results from the ESE and ISE mutagenesis screen to identify which proteins may be binding at these sites.

The use of PCR to assess the extent of alternative splicing is simple, convenient and in this case allows for semi-quantitative determination of PSI. However a more sensitive method may reveal more subtle changes that may still be of relevance. An alternative method could be the use of carefully designed qRT-PCR primers especially those that could detect the different splice variants with various combinations of exon 7b and exon 7.

6.7.3 The role of ERG in osteogenesis of osteoblasts

SSO and genome editing approaches could be used to investigate the role of ERG in osteoblast maturation as well. Not only would alternative splicing of the AD be of interest but also knockdown and overexpression experiments could inform on the role of the complete profile of ERG splice variants. Coupled with the 1, 25D/FHBP co-treatment changes in TNSALP activity could be assessed when different exon 7b and 7 variants are expressed. Study of the role of ERG exon 7 and 7b in bone development would also be of interest as currently this is not really understood.

Conditioned media could be collected from the various treatments and the presence of several osteogenic factors measured using biochemical assays. Drivers of osteogenesis that can be detected using radioimmunoassays and enzyme-linked immunoabsorbent assays include osteocalcin, osteoprotegrin, VEGF and transforming growth factor beta 1(TGF-β1). In addition to the MG63 cell line model experiments could be carried out in primary osteoblasts.

Gene expression studies could be carried out for more bone marker genes and this could be complemented with western blot analysis to assess changes at the protein level as well. Transcriptional assays could inform on which osteoblast genes may be directly regulated by ERG. Based on the findings of this study with regards to the *TNSALP* promoter it would be of interest to understand if ERG binds other promoters of key osteoblast genes.

Finally knowing that ERG is an important Wnt signalling regulator investigation of the role of the transcription factor in this pathway should also be examined. Wnt signalling is a key regulator of bone homeostasis maintaining the balance

between formation and resorption of bone. PCR and western blot analysis of downstream Wnt signalling components could be performed initially then inhibition of any proteins that are seen to change could be carried out to see if changes can be detected.

7 References

Aartsma-Rus, A. and van Ommen, G.-J. B. (2007) Antisense-mediated exon skipping: a versatile tool with therapeutic and research applications. *RNA*.Cold Spring Harbor Laboratory Press. 13 (10), pp. 1609–1624. doi: 10.1261/rna.653607.

Abril, J. F., Castelo, R. and Guigó, R. (2005) Comparison of splice sites in mammals and chicken. *Genome research*.Cold Spring Harbor Laboratory Press. 15 (1), pp. 111–9. doi: 10.1101/gr.3108805.

Adamo, P., Porazinski, S., Rajatileka, S., Jumbe, S., Hagen, R., Cheung, M., Wilson, I. and Ladomery, M. (2017) The oncogenic transcription factor ERG represses the transcription of the tumour suppressor gene PTEN in prostate cancer cells. *Oncology Letters*. 14 (5), pp. 5605–5610. doi: 10.3892/ol.2017.6841.

Adamo, P. and Ladomery, M. R. (2015) The oncogene ERG: a key factor in prostate cancer. *Oncogene*.Macmillan Publishers Limited. 35 (4), pp. 403–414. Available at: http://dx.doi.org/10.1038/onc.2015.109.

Adelman, K. and Lis, J. T. (2012) Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature Reviews Genetics*. 13 (10), pp. 720–731. doi: 10.1038/nrg3293.

Amin, E. M. … Ladomery, M. R. (2011) WT1 mutants reveal SRPK1 to be a downstream angiogenesis target by altering VEGF splicing. *Cancer cell*. 20 (6), pp. 768–80. doi: 10.1016/j.ccr.2011.10.016.

Baldus, C. D., Burmeister, T., Martus, P., Schwartz, S., Gokbuget, N.,

Bloomfield, C. D., Hoelzer, D., Thiel, E. and Hofmann, W. K. (2006) High expression of the ETS transcription factor ERG predicts adverse outcome in acute T-lymphoblastic leukemia in adults. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*.United States. 24 (29), pp. 4714–4720. doi: 10.1200/JCO.2006.06.1580.

Baltzinger, M., Mager-Heckel, A.-M. and Remy, P. (1999) Xl erg: Expression pattern and overexpression during development plead for a role in endothelial cell differentiation. *Developmental Dynamics*.John Wiley & Sons, Inc. 216 (4/5), pp. 420–433. doi: 10.1002/(SICI)1097-0177(199912)216:4/5<420::AID-DVDY10>3.0.CO;2-C.

Bannister, A. J. and Kouzarides, T. (2011) Regulation of chromatin by histone modifications. *Cell Research*. 21 (3), pp. 381–395. doi: 10.1038/cr.2011.22.

Barberan-Soler, S., Medina, P., Estella, J., Williams, J. and Zahler, A. M. (2011) Co-regulation of alternative splicing by diverse splicing factors in Caenorhabditis elegans. *Nucleic acids research*.Oxford University Press. 39 (2), pp. 666–74. doi: 10.1093/nar/gkq767.

Barbieri, C. E., Bangma, C. H., Bjartell, A., Catto, J. W. F., Culig, Z., Grönberg, H., Luo, J., Visakorpi, T. and Rubin, M. A. (2013) The mutational landscape of prostate cancer. *European urology*. 64 (4), pp. 567–76. doi: 10.1016/j.eururo.2013.05.029.

Bauman, J., Jearawiriyapaisarn, N. and Kole, R. (2009) Therapeutic Potential of Splice-Switching Oligonucleotides. *Oligonucleotides*. 19 (1), pp. 1–13. doi:

10.1089/oli.2008.0161.

Berg, T., Kalsaas, A.-H., Buechner, J. and Busund, L.-T. (2009) Ewing sarcoma–peripheral neuroectodermal tumor of the kidney with a FUS–ERG fusion transcript. *Cancer Genetics and Cytogenetics*. 194 (1), pp. 53–57. doi: 10.1016/j.cancergencyto.2009.06.002.

Berget, S. M. (1995) Exon recognition in vertebrate splicing. *The Journal of biological chemistry*. 270 (6), pp. 2411–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7852296 (Accessed: 5 July 2018).

Berget, S. M., Moore, C. and Sharp, P. A. (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America*.National Academy of Sciences. 74 (8), pp. 3171–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/269380 (Accessed: 5 July 2018).

Berget, S. M., Berk, A. J., Harrison, T. and Sharp, P. A. (1978) Spliced segments at the 5' termini of adenovirus-2 late mRNA: a role for heterogeneous nuclear RNA in mammalian cells. *Cold Spring Harbor symposia on quantitative biology*. 42 Pt 1 , pp. 523–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/277361 (Accessed: 5 July 2018).

Bertolotti, A., Melot, T., Acker, J., Vigneron, M., Delattre, O. and Tora, L. (1998) EWS, but not EWS-FLI-1, is associated with both TFIID and RNA polymerase II: interactions between two members of the TET family, EWS and hTAFII68, and subunits of TFIID and RNA polymerase II complexes. *Molecular and cellular*

biology. 18 (3), pp. 1489–97. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9488465 (Accessed: 9 August 2018).

Bhagwat, A. S. and Vakoc, C. R. (2015) Targeting Transcription Factors in Cancer. *Trends in cancer*.NIH Public Access. 1 (1), pp. 53–65. doi: 10.1016/j.trecan.2015.07.001.

Birdsey, G. M., Dryden, N. H., Shah, a. V., Hannah, R., Hall, M. D., Haskard, D. O., Parsons, M., Mason, J. C., Zvelebil, M., Gottgens, B., Ridley, a. J. and Randi, a. M. (2012) The transcription factor Erg regulates expression of histone deacetylase 6 and multiple pathways involved in endothelial cell migration and angiogenesis. *Blood*. 119 (3), pp. 894–903. doi: 10.1182/blood-2011-04-350025.

Birdsey, G. M., Dryden, N. H., Amsellem, V., Gebhardt, F., Sahnan, K., Haskard, D. O., Dejana, E., Mason, J. C. and Randi, A. M. (2008) Transcription factor erg regulates angiogenesis and endothelial apoptosis through VE-cadherin. *Blood*. 111 (7), pp. 3498–3506. doi: 10.1182/blood-2007-08-105346.

Birdsey, G. M., Shah, A. V., Dufton, N., Reynolds, L. E., Osuna Almagro, L., Yang, Y., Aspalter, I. M., Khan, S. T., Mason, J. C., Dejana, E., Göttgens, B., Hodivala-Dilke, K., Gerhardt, H., Adams, R. H. and Randi, A. M. (2015) The Endothelial Transcription Factor ERG Promotes Vascular Stability and Growth through Wnt/β-Catenin Signaling. *Developmental Cell*.Elsevier. 32 (1), pp. 82–96. doi: 10.1016/j.devcel.2014.11.016.

Black, D. L. (2003) Mechanisms of Alternative Pre-Messenger RNA Splicing.

Annual Review of Biochemistry. 72 (1), pp. 291–336. doi: 10.1146/annurev.biochem.72.121801.161720.

Bock, J., Mochmann, L. H., Schlee, C., Farhadi-Sartangi, N., Göllner, S., Müller-Tidow, C. and Baldus, C. D. (2013) ERG transcriptional networks in primary acute leukemia cells implicate a role for ERG in deregulated kinase signaling. *PloS one*. 8 (1), p. e52872. doi: 10.1371/journal.pone.0052872.

Bohne, A., Schlee, C., Mossner, M., Thibaut, J., Heesch, S., Thiel, E., Hofmann, W. K. and Baldus, C. D. (2009) Epigenetic control of differential expression of specific ERG isoforms in acute T-lymphoblastic leukemia. *Leukemia Research*. 33 (6), pp. 817–822. doi: 10.1016/j.leukres.2008.11.012.

Bonewald, L. F., Kester, M. B., Schwartz, Z., Swain, L. D., Khare, A., Johnson, T. L., Leach, R. J. and Boyan, B. D. (1992) Effects of combining transforming growth factor beta and 1,25-dihydroxyvitamin D3 on differentiation of a human osteosarcoma (MG-63). *The Journal of biological chemistry*. 267 (13), pp. 8943–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1577731 (Accessed: 5 March 2018).

Bothwell, A. L. M., Ballardsli, D. W., Philbrick, W. M., Lindwalls, G., Maher, S. E., Bridgetts, M. M., Jamison, S. F. and Garcia-Blanco, M. A. (1991) Murine polypyrimidine tract binding protein. Purification, cloning, and mapping of the RNA binding domain. *Journal of Biological Chemistry*. 266 (36), pp. 24657–24663. Available at: http://www.jbc.org/content/266/36/24657.full.pdf (Accessed: 11 July 2018).

Brase, J. C., Johannes, M., Mannsperger, H., Falth, M., Metzger, J., Kacprzyk, L. a, Andrasiuk, T., Gade, S., Meister, M., Sirma, H., Sauter, G., Simon, R., Schlomm, T., Beissbarth, T., Korf, U., Kuner, R. and Sultmann, H. (2011) TMPRSS2-ERG -specific transcriptional modulation is associated with prostate cancer biomarkers and TGF-beta signaling. *BMC Cancer*.BioMed Central Ltd. 11 (1), p. 507. doi: 10.1186/1471-2407-11-507.

Braunreiter, C. L., Hancock, J. D., Coffin, C., Boucher, K. M. and Lessnick, S. L. (2006) Expression of EWS-ETS Fusions in NIH3T3 Cells Reveals Significant Differences to Ewing's Sarcoma. *Cell Cycle*. 5 (23), pp. 2753–2759. doi: 10.4161/cc.5.23.3505.

Brow, D. A. (2002) Allosteric Cascade of Spliceosome Activation. *Annual Review of Genetics*. 36 (1), pp. 333–360. doi: 10.1146/annurev.genet.36.043002.091635.

Brown, J. E. and Sim, S. (2010) Evolving role of bone biomarkers in castration-resistant prostate cancer. *Neoplasia (New York, N.Y.)*.Neoplasia Press. 12 (9), pp. 685–96. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20824045 (Accessed: 12 January 2019).

Burkhalter, R. J., Westfall, S. D., Liu, Y. and Stack, M. S. (2015) Lysophosphatidic Acid Initiates Epithelial to Mesenchymal Transition and Induces β-Catenin-mediated Transcription in Epithelial Ovarian Carcinoma. *The Journal of biological chemistry*.American Society for Biochemistry and Molecular Biology. 290 (36), pp. 22143–54. doi: 10.1074/jbc.M115.641092.

Butticè, G., Duterque-Coquillaud, M., Basuyaux, J. P., Carrère, S., Kurkinen, M. and Stéhelin, D. (1996) Erg, an Ets-family member, differentially regulates human collagenase1 (MMP1) and stromelysin1 (MMP3) gene expression by physically interacting with the Fos/Jun complex. *Oncogene*. 13 (11), pp. 2297–306. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8957070 (Accessed: 12 October 2015).

Cantile, M., Marra, L., Franco, R., Ascierto, P., Liguori, G., De Chiara, A. and Botti, G. (2013) Molecular detection and targeting of EWSR1 fusion transcripts in soft tissue tumors. *Medical oncology (Northwood, London, England)*.Springer. 30 (1), p. 412. doi: 10.1007/s12032-012-0412-8.

Carrère, S., Verger, a, Flourens, a, Stehelin, D. and Duterque-Coquillaud, M. (1998) Erg proteins, transcription factors of the Ets family, form homo, heterodimers and ternary complexes via two distinct domains. *Oncogene*. 16 (25), pp. 3261–3268.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q. and Krainer, A. R. (2003) ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic acids research*.Oxford University Press. 31 (13), pp. 3568–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12824367 (Accessed: 22 August 2018).

Carver, B. S., Tran, J., Gopalan, A., Chen, Z., Shaikh, S., Carracedo, A., Alimonti, A., Nardella, C., Varmeh, S., Scardino, P. T., Cordon-Cardo, C., Gerald, W. and Pandolfi, P. P. (2009) Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate. *Nature genetics*. 41 (5), pp. 619–624. doi: 10.1038/ng.370.

Chen, S., Deniz, K., Sung, Y.-S., Zhang, L., Dry, S. and Antonescu, C. R. (2016) Ewing sarcoma with ERG gene rearrangements: A molecular study focusing on the prevalence of FUS-ERG and common pitfalls in detecting EWSR1-ERG fusions by FISH. *Genes, chromosomes & cancer*.NIH Public Access. 55 (4), pp. 340–9. doi: 10.1002/gcc.22336.

Cheng, C. and Sharp, P. A. (2006) Regulation of CD44 alternative splicing by SRm160 and its potential role in tumor cell invasion. *Molecular and cellular biology*.American Society for Microbiology (ASM). 26 (1), pp. 362–70. doi: 10.1128/MCB.26.1.362-370.2006.

Chinni, S. R., Sivalogan, S., Dong, Z., Filho, J. C. T., Deng, X., Bonfil, R. D. and Cher, M. L. (2006) CXCL12/CXCR4 signaling activates Akt-1 and MMP-9 expression in prostate cancer cells: The role of bone microenvironment-associated CXCL12. *The Prostate*. 66 (1), pp. 32–48. doi: 10.1002/pros.20318.

Chow, L. T., Gelinas, R. E., Broker, T. R. and Roberts, R. J. (1977) An amazing sequence arrangement at the 5’ ends of adenovirus 2 messenger RNA. *Cell*. 12 (1), pp. 1–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/902310 (Accessed: 5 July 2018).

Clark, J., Merson, S., Jhavar, S., Flohr, P., Edwards, S., Foster, C. S., Eeles, R., Martin, F. L., Phillips, D. H., Crundwell, M., Christmas, T., Thompson, A., Fisher, C., Kovacs, G. and Cooper, C. S. (2007) Diversity of TMPRSS2-ERG fusion transcripts in the human prostate. *Oncogene*. 26 , pp. 2667–2673. doi: 10.1038/sj.onc.1210070.

Clinckemalie, L., Spans, L., Dubois, V., Laurent, M., Helsen, C., Joniau, S. and Claessens, F. (2013) Androgen Regulation of the TMPRSS2 Gene and the Effect of a SNP in an Androgen Response Element. *Molecular Endocrinology*.Oxford University Press. 27 (12), pp. 2028–2040. doi: 10.1210/me.2013-1098.

Codrington, R., Pannell, R., Forster, A., Drynan, L. F., Daser, A., Lobato, N., Metzler, M. and Rabbitts, T. H. (2005) The Ews-ERG Fusion Protein Can Initiate Neoplasia from Lineage-Committed Haematopoietic Cells. *PLoS Biology*.Edited by C. Marshall.Public Library of Science. 3 (8), p. e242. doi: 10.1371/journal.pbio.0030242.

Coelho, M. B., Attig, J., Bellora, N., Konig, J., Hallegger, M., Kayikci, M., Eyras, E., Ule, J. and Smith, C. W. (2015) Nuclear matrix protein Matrin3 regulates alternative splicing and forms overlapping regulatory networks with PTB. *The EMBO Journal*.EMBO Press. 34 (5), pp. 653–668. doi: 10.15252/embj.201489852.

Cooper, C. D. O., Newman, J. A., Aitkenhead, H., Allerston, C. K. and Gileadi, O. (2015) Structures of the Ets Protein DNA-binding Domains of Transcription Factors Etv1, Etv4, Etv5, and Fev. *The Journal of biological chemistry*.American Society for Biochemistry and Molecular Biology. 290 (22), pp. 13692–709. doi: 10.1074/jbc.M115.646737.

Degnan, B. M., Degnan, S. M., Naganuma, T. and Morse, D. E. (1993) The ets multigene family is conserved throughout the Metazoa. *Nucleic Acids Research*. 21 (15), pp. 3479–3484. Available at:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC331448/pdf/nar00064-0143.pdf (Accessed: 14 May 2018).

Delattre, O., Zucman, J., Plougastel, B., Desmaze, C., Melot, T., Peter, M., Kovar, H., Joubert, I., de Jong, P., Rouleau, G., Aurias, A. and Thomas, G. (1992) Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature*. 359 (6391), pp. 162–165. doi: 10.1038/359162a0.

Deplus, R., Delliaux, C., Marchand, N., Flourens, A., Vanpouille, N., Leroy, X., de Launoit, Y. and Duterque-Coquillaud, M. (2017) TMPRSS2-ERG fusion promotes prostate cancer metastases in bone. *Oncotarget*. 8 (7), pp. 11827–11840. doi: 10.18632/oncotarget.14399.

Desmet, F.-O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M. and Béroud, C. (2009) Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Research*.Oxford University Press. 37 (9), pp. e67–e67. doi: 10.1093/nar/gkp215.

Donaldson, L. W., Petersen, J. M., Graves, B. J. and McIntosh, L. P. (1996) Solution structure of the ETS domain from murine Ets-1: a winged helix-turn-helix DNA binding motif. *The EMBO journal*. 15 (1), pp. 125–134. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC449924/pdf/emboj00001-0127.pdf.

Dryden, N. H., Sperone, A., Martin-Almedina, S., Hannah, R. L., Birdsey, G. M., Khan, S. T., Layhadi, J. A., Mason, J. C., Haskard, D. O., Göttgens, B. and

Randi, A. M. (2012) The Transcription Factor Erg Controls Endothelial Cell Quiescence by Repressing Activity of Nuclear Factor (NF)-κB p65. *Journal of Biological Chemistry* . 287 (15), pp. 12331–12342. doi: 10.1074/jbc.M112.346791.

Duffy, A. G. … Greten, T. F. (2016) Modulation of tumor eIF4E by antisense inhibition: A phase I/II translational clinical trial of ISIS 183750-an antisense oligonucleotide against eIF4E-in combination with irinotecan in solid tumors and irinotecan-refractory colorectal cancer. *International Journal of Cancer*. 139 (7), pp. 1648–1657. doi: 10.1002/ijc.30199.

Duterque-Coquillaud, M., Niel, C., Plaza, S. and Stehelin, D. (1993) New human erg isoforms generated by alternative splicing are transcriptional activators. *Oncogene*. 8 , pp. 1865–1873.

Eckstein, F. (2014) Phosphorothioates, Essential Components of Therapeutic Oligonucleotides. *Nucleic Acid Therapeutics*. 24 (6), pp. 374–387. doi: 10.1089/nat.2014.0506.

Elagib, K. E., Racke, F. K., Mogass, M., Khetawat, R., Delehanty, L. L. and Goldfarb, A. N. (2003) RUNX1 and GATA-1 coexpression and cooperation in megakaryocytic differentiation. *Blood*.American Society of Hematology. 101 (11), pp. 4333–41. doi: 10.1182/blood-2002-09-2708.

Ewing, J. (1921) Diffuse endothelioma of bone. *New York Pathological Society*. 21 , pp. 17–24.

Fairbrother, W. G., Yeo, G. W., Yeh, R., Goldstein, P., Mawson, M., Sharp, P.

A. and Burge, C. B. (2004) RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. *Nucleic Acids Research*. 32 (Web server issue), pp. W187–W190. doi: 10.1093/nar/gkh393.

Fine, S. W., Gopalan, A., Leversha, M. a, Al-Ahmadie, H. a, Tickoo, S. K., Zhou, Q., Satagopan, J. M., Scardino, P. T., Gerald, W. L. and Reuter, V. E. (2010) TMPRSS2-ERG gene fusion is associated with low Gleason scores and not with high-grade morphological features. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*.Nature Publishing Group. 23 (10), pp. 1325–33. doi: 10.1038/modpathol.2010.120.

Fish, J. E., Cantu Gutierrez, M., Dang, L. T., Khyzha, N., Chen, Z., Veitch, S., Cheng, H. S., Khor, M., Antounians, L., Njock, M.-S., Boudreau, E., Herman, A. M., Rhyner, A. M., Ruiz, O. E., Eisenhoffer, G. T., Medina-Rivera, A., Wilson, M. D. and Wythe, J. D. (2017) Dynamic regulation of VEGF-inducible genes by an ERK/ERG/p300 transcriptional network. *Development (Cambridge, England)*.Oxford University Press for The Company of Biologists Limited. 144 (13), pp. 2428–2444. doi: 10.1242/dev.146050.

Flajollet, S., Tian, T. V., Flourens, A., Tomavo, N., Villers, A., Bonnelye, E., Aubert, S., Leroy, X. and Duterque-Coquillaud, M. (2011) Abnormal Expression of the ERG Transcription Factor in Prostate Cancer Cells Activates Osteopontin. *Molecular Cancer Research*. 9 (7), pp. 914–924. doi: 10.1158/1541-7786.MCR-10-0537.

Fu, X.-D. and Maniatis, T. (1990) Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature*. 343 (6257), pp.

437–441. doi: 10.1038/343437a0.

Gaj, T., Gersbach, C. a and Barbas, C. F. (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in biotechnology*.Elsevier Ltd. 31 (7), pp. 397–405. doi: 10.1016/j.tibtech.2013.04.004.

Ge, H. and Manley, J. L. (1990) A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. *Cell*. 62 (1), pp. 25–34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2163768 (Accessed: 11 July 2018).

Gidley, J., Openshaw, S., Pring, E. T., Sale, S. and Mansell, J. P. (2006) Lysophosphatidic acid cooperates with 1α,25(OH)2D3 in stimulating human MG63 osteoblast maturation. *Prostaglandins & Other Lipid Mediators*. 80 (1–2), pp. 46–61. doi: 10.1016/j.prostaglandins.2006.04.001.

Gille, H., Sharrocks, A. D. and Shaw, P. E. (1992) Phosphorylation of transcription factor p62TCF by MAP kinase stimulates ternary complex formation at c-fos promoter. *Nature*. 358 (6385), pp. 414–417. doi: 10.1038/358414a0.

Ginsberg, J. P., de Alava, E., Ladanyi, M., Wexler, L. H., Kovar, H., Paulussen, M., Zoubek, A., Dockhorn-Dworniczak, B., Juergens, H., Wunder, J. S., Andrulis, I. L., Malik, R., Sorensen, P. H., Womer, R. B. and Barr, F. G. (1999) EWS-FLI1 and EWS-ERG gene fusions are associated with similar clinical phenotypes in Ewing's sarcoma. *Journal of clinical oncology : official journal of*

the American Society of Clinical Oncology.American Society of Clinical Oncology. 17 (6), pp. 1809–14. doi: 10.1200/JCO.1999.17.6.1809.

Gopalan, A., Leversha, M. a, Satagopan, J. M., Zhou, Q., Al-Ahmadie, H. a, Fine, S. W., Eastham, J. a, Scardino, P. T., Scher, H. I., Tickoo, S. K., Reuter, V. E. and Gerald, W. L. (2009) TMPRSS2-ERG gene fusion is not associated with outcome in patients treated by prostatectomy. *Cancer research*. 69 (4), pp. 1400–6. doi: 10.1158/0008-5472.CAN-08-2467.

Graveley, B. R. (2000) Sorting out the complexity of SR protein functions. *RNA (New York, N.Y.)*. 6 (9), pp. 1197–211. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10999598 (Accessed: 11 July 2018).

Graveley, B. R. (2005) Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell*.NIH Public Access. 123 (1), pp. 65–73. doi: 10.1016/j.cell.2005.07.028.

Hagen, R., Adamo, P., Karamat, S., Oxley, J., Aning, J., Gilliatt, D., Persad, R., Ladomery, M. R. and Rhodes, A. (2014) Quantitative analysis of ERG expression and its splice isoforms in formalin-fixed paraffin-embedded prostate cancer samples: Association with seminal vesicle invasion and biochemical recurrence. *American journal of clinical pathology*. 142 (4), pp. 533–540. Available at: http://ajcp.ascpjournals.org/content/142/4/533.short (Accessed: 17 November 2014).

Hallegger, M., Llorian, M. and Smith, C. W. J. (2010) Alternative splicing: global insights. *The FEBS journal*. 277 (4), pp. 856–66. doi: 10.1111/j.1742-

4658.2009.07521.x.

Hanahan, D. and Weinberg, R. A. (2011) Hallmarks of cancer: The next generation. *Cell*. pp. 646–674.

Hargreaves, D. C. and Crabtree, G. R. (2011) ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Research*. 21 (3), pp. 396–420. doi: 10.1038/cr.2011.32.

Hertel, K. J. (2008) Combinatorial Control of Exon Recognition. *Journal of Biological Chemistry*. 283 (3), pp. 1211–1215. doi: 10.1074/jbc.R700035200.

Hicks, M. J., Yang, C.-R., Kotlajich, M. V and Hertel, K. J. (2006) Linking Splicing to Pol II Transcription Stabilizes Pre-mRNAs and Influences Splicing Patterns. *PLoS Biology*.Edited by M. Wickens.Public Library of Science. 4 (6), p. e147. doi: 10.1371/journal.pbio.0040147.

Hoesel, B., Malkani, N., Hochreiter, B., Basílio, J., Sughra, K., Ilyas, M. and Schmid, J. A. (2016) Sequence-function correlations and dynamics of ERG isoforms. ERG8 is the black sheep of the family. *Biochimica et biophysica acta*. 1863 (2), pp. 205–18. doi: 10.1016/j.bbamcr.2015.10.023.

Hollenhorst, P. C. (2012) RAS/ERK pathway transcriptional regulation through ETS/AP-1 binding sites. *Small GTPases*.Taylor & Francis. 3 (3), pp. 154–8. doi: 10.4161/sgtp.19630.

Hollenhorst, P. C., Jones, D. A. and Graves, B. J. (2004) Expression profiles frame the promoter specificity dilemma of the ETS family of transcription factors. *Nucleic Acids Research*. 32 (18), pp. 5693–5702. doi: 10.1093/nar/gkh906.

Hollenhorst, P. C., McIntosh, L. P. and Graves, B. J. (2011) Genomic and Biochemical Insights into the Specificity of ETS Transcription Factors. *Annual Review of Biochemistry*. Annual Reviews . 80 (1), pp. 437–471. doi: 10.1146/annurev.biochem.79.081507.103945.

Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R. and Zamir, A. (1965) Structure of a Ribonucleic Acid. *Science*. 147 (3664), p. 1462 LP-1465. Available at: http://science.sciencemag.org/content/147/3664/1462.abstract.

House, A. E. and Lynch, K. W. (2006) An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition. *Nature Structural & Molecular Biology*. 13 (10), pp. 937–944. doi: 10.1038/nsmb1149.

Hu, Y., Dobi, A., Sreenath, T., Cook, C., Tadase, A. Y., Ravindranath, L., Cullen, J., Furusato, B., Chen, Y., Thangapazham, R. L., Mohamed, A., Sun, C., Sesterhenn, I. a, McLeod, D. G., Petrovics, G. and Srivastava, S. (2008) Delineation of TMPRSS2-ERG splice variants in prostate cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 14 (15), pp. 4719–25. doi: 10.1158/1078-0432.CCR-08-0531.

Huang, Y., Thoms, J. A. I., Tursky, M. L., Knezevic, K., Beck, D., Chandrakanthan, V., Suryani, S., Olivier, J., Boulton, A., Glaros, E. N., Thomas, S. R., Lock, R. B., MacKenzie, K. L., Bushweller, J. H., Wong, J. W. H. and Pimanda, J. E. (2016) MAPK/ERK2 phosphorylates ERG at serine 283 in leukemic cells and promotes stem cell signatures and cell proliferation. *Leukemia*. 30 (7), pp. 1552–1561. doi: 10.1038/leu.2016.55.

Huang, Y., Thoms, J. A. I., Tursky, M. L., Knezevic, K., Beck, D., Chandrakanthan, V., Suryani, S., Olivier, J., Boulton, A., Glaros, E. N., Thomas, S. R., Lock, R. B., MacKenzie, K. L., Bushweller, J. H., Wong, J. W. H. and Pimanda, J. E. (2016) MAPK/ERK2 phosphorylates ERG at serine 283 in leukemic cells and promotes stem cell signatures and cell proliferation. *Leukemia*.Nature Publishing Group. 30 (7), pp. 1552–1561. doi: 10.1038/leu.2016.55.

Ichikawa, H., Shimizu, K., Hayashi, Y. and Ohki, M. (1994) An RNA-binding Protein Gene, TLS/FUS, Is Fused to ERG in Human Myeloid Leukemia with t(16;21) Chromosomal Translocation. *Cancer Research*. 54 (11), pp. 2865–2868. Available at: http://cancerres.aacrjournals.org/content/54/11/2865.abstract.

Ihn Lee, T. and Young, R. A. (2013) Transcriptional Regulation and Its Misregulation in Disease. *Cell*.Taatjes. 152 , pp. 1237–1251. doi: 10.1016/j.cell.2013.02.014.

Introna, M. and Golay, J. (1999) *How can oncogenic transcription factors cause cancer: a critical review of the myb story*. *Leukemia*. Available at: http://www.stockton-press.co.uk/leu (Accessed: 10 August 2018).

Inukai, S., Kock, K. H. and Bulyk, M. L. (2017) Transcription factor-DNA binding: beyond binding site motifs. *Current opinion in genetics & development*.NIH Public Access. 43 , pp. 110–119. doi: 10.1016/j.gde.2017.02.007.

Irifune, H., Nishimori, H., Watanabe, G., Yoshida, K., Ikeda, T., Matsui, C.,

Morohashi, M., Kawaguchi, S., Nagoya, S., Wada, T., Yamashita, T., Nakamura, Y. and Tokino, T. (2005) Aberrant laminin β3 isoforms downstream of EWS-ETS fusion genes in ewing family tumors. *Cancer Biology and Therapy*. 4 (4), pp. 449–455. doi: 1623 [pii].

Iwamoto, M., Higuchi, Y., Koyama, E., Enomoto-Iwamoto, M., Kurisu, K., Yeh, H., Abrams, W. R., Rosenbloom, J. and Pacifici, M. (2000) Transcription Factor Erg Variants and Functional Diversification of Chondrocytes during Limb Long Bone Development. *The Journal of Cell Biology*. 150 (1). Available at: http://jcb.rupress.org/content/150/1/27.long (Accessed: 3 May 2017).

Iwamoto, M., Higuchi, Y., Enomoto-Iwamoto, M., Kurisu, K., Koyama, E., Yeh, H., Rosenbloom, J. and Pacifici, M. (2001) The role of ERG (ets related gene) in cartilage development. *Osteoarthritis and Cartilage*.W.B. Saunders. 9 , pp. S41–S47. doi: 10.1053/JOCA.2001.0443.

Jacobs, D., Glossip, D., Xing, H., Muslin, A. J. and Kornfeld, K. (1999) Multiple docking sites on substrate proteins form a modular system that mediates recognition by ERK MAP kinase. *Genes & development*.Cold Spring Harbor Laboratory Press. 13 (2), pp. 163–75. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9925641 (Accessed: 20 August 2018).

Jakubauskiene, E., Vilys, L., Makino, Y., Poellinger, L. and Kanopka, A. (2015) Increased Serine-Arginine (SR) Protein Phosphorylation Changes Pre-mRNA Splicing in Hypoxia. *Journal of Biological Chemistry*. 290 (29), pp. 18079–18089. doi: 10.1074/jbc.M115.639690.

St. John, J., Powell, K., Conley-lacomb, M. K. and Chinni, S. R. (2012) TMPRSS2-ERG Fusion Gene Expression in Prostate Tumor Cells and Its Clinical and Biological Significance in Prostate Cancer Progression. *Journal of cancer science & therapy*. 4 (4), pp. 94–101. doi: 10.4172/1948-5956.1000.

Jumbe, S. L., Porazinski, S. R., Oltean, S., Mansell, J. P., Vahabi, B., Wilson, I. D. and Ladomery, M. R. (2019) The Evolutionarily Conserved Cassette Exon 7b Drives ERG's Oncogenic Properties. *Translational Oncology*.Elsevier Inc. 12 (1), pp. 134–142. doi: 10.1016/j.tranon.2018.09.001.

Kamiya, N., Suzuki, H., Endo, T., Yano, M., Naoi, M., Nishimi, D., Kawamura, K., Imamoto, T. and Ichikawa, T. (2012) Clinical usefulness of bone markers in prostate cancer with bone metastasis. *International Journal of Urology*.John Wiley & Sons, Ltd (10.1111). 19 (11), pp. 968–979. doi: 10.1111/j.1442-2042.2012.03098.x.

Karim, F. D., Urness, L. D., Thummel, C. S., Klemsz, M. J., McKercher, S. R., Celada, A., Van Beveren, C., Maki, R. A., Gunther, C. V and Nye, J. A. (1990) The ETS-domain: a new DNA-binding motif that recognizes a purine-rich core DNA sequence. *Genes & development*.Cold Spring Harbor Laboratory Press. 4 (9), pp. 1451–3. doi: 10.1101/GAD.4.9.1451.

Kedage, V., Selvaraj, N., Nicholas, T. R., Budka, J. A., Plotnik, J. P., Jerde, T. J. and Hollenhorst, P. C. (2016) An Interaction with Ewing's Sarcoma Breakpoint Protein EWS Defines a Specific Oncogenic Mechanism of ETS Factors Rearranged in Prostate Cancer. *Cell Reports*.Cell Press. 17 (5), pp. 1289–1301. doi: 10.1016/J.CELREP.2016.10.001.

Kedage, V., Strittmatter, B. G., Dausinas, P. B. and Hollenhorst, P. C. (2017) Phosphorylation of the oncogenic transcription factor ERG in prostate cells dissociates polycomb repressive complex 2, allowing target gene activation. *The Journal of biological chemistry*.American Society for Biochemistry and Molecular Biology. 292 (42), pp. 17225–17235. doi: 10.1074/jbc.M117.796458.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome research*.Cold Spring Harbor Laboratory Press. 12 (6), pp. 996–1006. doi: 10.1101/gr.229102.

Keren, H., Lev-Maor, G. and Ast, G. (2010) Alternative splicing and evolution: diversification, exon definition and function. *Nature Reviews Genetics*. 11 (5), pp. 345–355. doi: 10.1038/nrg2776.

King, J. C., Xu, J., Wongvipat, J., Hieronymus, H., Carver, B. S., Leung, D. H., Taylor, B. S., Sander, C., Cardiff, R. D., Couto, S. S., Gerald, W. L. and Sawyers, C. L. (2009) Cooperativity of TMPRSS2-ERG with PI3-kinase pathway activation in prostate oncogenesis. *Nature genetics*. 41 (5), pp. 524–6. doi: 10.1038/ng.371.

Komori, T. (2008) Regulation of bone development and maintenance by Runx2. *Frontiers in Bioscience*. 13 , p. 898. doi: 10.2741/2730.

Koushika, S. P., Lisbin, M. J. and White, K. (1996) ELAV, a Drosophila neuron-specific protein, mediates the generation of an alternatively spliced neural protein isoform. *Current Biology*.Cell Press. 6 (12), pp. 1634–1641. doi:

10.1016/S0960-9822(02)70787-2.

Krainer, A. R., Conway, G. C. and Kozak, D. (1990) The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. *Cell*. 62 (1), pp. 35–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2364434 (Accessed: 11 July 2018).

Krivega, I. and Dean, A. (2012) Enhancer and promoter interactions—long distance calls. *Current Opinion in Genetics & Development*. 22 (2), pp. 79–85. doi: 10.1016/j.gde.2011.11.001.

Kuroyanagi, H., Ohno, G., Mitani, S. and Hagiwara, M. (2007) The Fox-1 family and SUP-12 coordinately regulate tissue-specific alternative splicing in vivo. *Molecular and cellular biology*.American Society for Microbiology (ASM). 27 (24), pp. 8612–21. doi: 10.1128/MCB.01508-07.

de la Mata, M., Alonso, C. R., Kadener, S., Fededa, J. P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D. and Kornblihtt, A. R. (2003) A slow RNA polymerase II affects alternative splicing in vivo. *Molecular cell*. 12 (2), pp. 525–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14536091 (Accessed: 11 July 2018).

Lacronique, V., Boureux, A., Valle, V. D., Poirel, H., Quang, C. T., Mauchauffé, M., Berthou, C., Lessard, M., Berger, R., Ghysdael, J. and Bernard, O. A. (1997) A TEL-JAK2 fusion protein with constitutive kinase activity in human leukemia. *Science (New York, N.Y.)*. 278 (5341), pp. 1309–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9360930 (Accessed: 3 July 2018).

Lamkanfi, M. and Kanneganti, T.-D. (2010) Caspase-7: a protease involved in

apoptosis and inflammation. *The international journal of biochemistry & cell biology*.NIH Public Access. 42 (1), pp. 21–4. doi: 10.1016/j.biocel.2009.09.013.

Laudet, V., Hänni, C., Stéhelin, D. and Duterque-Coquillaud, M. (1999) Molecular phylogeny of the ETS gene family. *Oncogene*.Nature Publishing Group. 18 (6), pp. 1351–9. doi: 10.1038/sj.onc.1202444.

Le, K., Prabhakar, B. S., Hong, W. and Li, L. (2015) Alternative splicing as a biomarker and potential target for drug discovery. *Acta pharmacologica Sinica*.Nature Publishing Group. 36 (10), pp. 1212–8. doi: 10.1038/aps.2015.43.

Lee, Y. and Rio, D. C. (2015) Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annual review of biochemistry*.NIH Public Access. 84 , p. 291. doi: 10.1146/ANNUREV-BIOCHEM-060614-034316.

Lee, Y. and Rio, D. C. (2015) Mechanisms and Regulation of Alternative Pre-mRNA Splicing HHS Public Access. *Annu Rev Biochem*. 84 , pp. 291–323. doi: 10.1146/annurev-biochem-060614-034316.

Leinonen, K. a., Saramaki, O. R., Furusato, B., Kimura, T., Takahashi, H., Egawa, S., Suzuki, H., Keiger, K., Ho Hahm, S., Isaacs, W. B., Tolonen, T. T., Stenman, U.-H., Tammela, T. L. J., Nykter, M., Bova, G. S. and Visakorpi, T. (2013) Loss of PTEN Is Associated with Aggressive Behavior in ERG-Positive Prostate Cancer. *Cancer Epidemiology Biomarkers & Prevention*. 22 (12), pp. 2333–2344. doi: 10.1158/1055-9965.EPI-13-0333-T.

Lelibvre-Chotteau, A., Laudetb, V., Flourensb, A., B&ueb, A., Leprinceb, D. and

Fontainea, F. (1994) *Identification of two ets related genes in a marine worm, the polychaete annelid Nereis diversicolor*. *FEBS Letters*. Available at: https://febs.onlinelibrary.wiley.com/doi/pdf/10.1016/0014-5793%2894%2901091-9 (Accessed: 23 August 2018).

Lelli, K. M., Slattery, M. and Mann, R. S. (2012) Disentangling the Many Layers of Eukaryotic Transcriptional Regulation. *Annual Review of Genetics*. 46 (1), pp. 43–68. doi: 10.1146/annurev-genet-110711-155437.

Lessnick, S. L. and Ladanyi, M. (2012) Molecular pathogenesis of Ewing sarcoma: new therapeutic and transcriptional targets. *Annual review of pathology*.NIH Public Access. 7 , pp. 145–59. doi: 10.1146/annurev-pathol-011110-130237.

Li, D., Lv, H., Hao, X., Hu, B. and Song, Y. (2018) Prognostic value of serum alkaline phosphatase in the survival of prostate cancer: evidence from a meta-analysis. *Cancer management and research*.Dove Press. 10 , pp. 3125–3139. doi: 10.2147/CMAR.S174237.

Liang, H., Mao, X., Olejniczak, E. T., Nettesheim, D. G., Yu, L., Meadows, R. P., Thompson, C. B. and Fesik, S. W. (1994) Solution structure of the ets domain of Fli-1 when bound to DNA. *Nature structural biology*. 1 (12), pp. 871–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7773776 (Accessed: 2 July 2018).

Lin, P. P., Brody, R. I., Hamelin, A. C., Bradner, J. E., Healey, J. H. and Ladanyi, M. (1999) Differential transactivation by alternative EWS-FLI1 fusion proteins

correlates with clinical heterogeneity in Ewing's sarcoma. *Cancer research*.American Association for Cancer Research. 59 (7), pp. 1428–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10197607 (Accessed: 9 August 2018).

Ling, Y., Lakey, J. H., Roberts, C. E. and Sharrocks, A. D. (1997) Molecular characterization of the B-box protein-protein interaction motif of the ETS-domain transcription factor Elk-1. *The EMBO Journal*.Wiley-Blackwell. 16 (9), pp. 2431–2440. doi: 10.1093/emboj/16.9.2431.

Lloberas, J., Soler, C., Celada, A., Lloberas, J., Soler, C. and Celada, A. (1999) The key role of PU.1/SPI-1 in B cells, myeloid cells and macrophages. *Immunology today*.Elsevier. 20 (4), pp. 184–9. doi: 10.1016/S0167-5699(99)01442-5.

Lobanov, A. V, Turanov, A. A., Hatfield, D. L. and Gladyshev, V. N. (2010) Dual functions of codons in the genetic code. *Critical reviews in biochemistry and molecular biology*.NIH Public Access. 45 (4), pp. 257–65. doi: 10.3109/10409231003786094.

Loomis, R. J., Naoe, Y., Parker, J. B., Savic, V., Bozovsky, M. R., Macfarlan, T., Manley, J. L. and Chakravarti, D. (2009) Chromatin Binding of SRp20 and ASF/SF2 and Dissociation from Mitotic Chromosomes Is Modulated by Histone H3 Serine 10 Phosphorylation. *Molecular Cell*. 33 (4), pp. 450–461. doi: 10.1016/j.molcel.2009.02.003.

Lou, H., Yang, Y., Cote, G. J., Berget, S. M. and Gagel, R. F. (1995) An intron

enhancer containing a 5' splice site sequence in the human calcitonin/calcitonin gene-related peptide gene. *Molecular and cellular biology*. 15 (12), pp. 7135–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8524281 (Accessed: 11 July 2018).

Loughran, S. J., Kruse, E. A., Hacking, D. F., de Graaf, C. A., Hyland, C. D., Willson, T. A., Henley, K. J., Ellis, S., Voss, A. K., Metcalf, D., Hilton, D. J., Alexander, W. S. and Kile, B. T. (2008) The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. *Nature immunology*. 9 (7), pp. 810–9. doi: 10.1038/ni.1617.

Luco, R. F., Pan, Q., Tominaga, K., Blencowe, B. J., Pereira-Smith, O. M. and Misteli, T. (2010) Regulation of Alternative Splicing by Histone Modifications. *Science*. 327 (5968), pp. 996–1000. doi: 10.1126/science.1184208.

Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R. and Misteli, T. (2011) Epigenetics in alternative pre-mRNA splicing. *Cell*. 144 (1), pp. 16–26. doi: 10.1016/j.cell.2010.11.056.

Luo, M.-J., Zhou, Z., Magni, K., Christoforides, C., Rappsilber, J., Mann, M. and Reed, R. (2001) Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature*. 413 (6856), pp. 644–647. doi: 10.1038/35098106.

Luo, Z., Lin, C., Guest, E., Garrett, A. S., Mohaghegh, N., Swanson, S., Marshall, S., Florens, L., Washburn, M. P. and Shilatifard, A. (2012) The Super Elongation Complex Family of RNA Polymerase II Elongation Factors: Gene

Target Specificity and Transcriptional Output. *Molecular and Cellular Biology*. 32 (13), pp. 2608–2617. doi: 10.1128/MCB.00182-12.

Mackereth, C. D., Schärpf, M., Gentile, L. N., MacIntosh, S. E., Slupsky, C. M. and McIntosh, L. P. (2004) Diversity in Structure and Function of the Ets Family PNT Domains. *Journal of Molecular Biology*. 342 (4), pp. 1249–1264. doi: 10.1016/j.jmb.2004.07.094.

Malik, S. and Roeder, R. G. (2010) The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nature reviews. Genetics*.NIH Public Access. 11 (11), pp. 761–72. doi: 10.1038/nrg2901.

Mallat, Z. and Tedgui, A. (2000) Apoptosis in the vasculature: mechanisms and functional importance. *British journal of pharmacology*.Wiley-Blackwell. 130 (5), pp. 947–62. doi: 10.1038/sj.bjp.0703407.

Mani, R.-S., Iyer, M. K., Cao, Q., Brenner, J. C., Wang, L., Ghosh, A., Cao, X., Lonigro, R. J., Tomlins, S. a, Varambally, S. and Chinnaiyan, A. M. (2011) TMPRSS2-ERG-mediated feed-forward regulation of wild-type ERG in human prostate cancers. *Cancer research*. 71 (16), pp. 5387–92. doi: 10.1158/0008-5472.CAN-11-0876.

Mansell, J. P., Cooke, M., Read, M., Rudd, H., Shiel, A. I., Wilkins, K. and Manso, M. (2016) Chitinase 3-like 1 expression by human (MG63) osteoblasts in response to lysophosphatidic acid and 1,25-dihydroxyvitamin D3. *Biochimie*.Elsevier Ltd. 128–129 , pp. 193–200. doi: 10.1016/j.biochi.2016.08.011.

Matsui, Y., Chansky, H. A., Barahmand-Pour, F., Zielinska-Kwiatkowska, A., Tsumaki, N., Myoui, A., Yoshikawa, H., Yang, L. and Eyre, D. R. (2003) *COL11A2* Collagen Gene Transcription Is Differentially Regulated by EWS/ERG Sarcoma Fusion Protein and Wild-type ERG. *Journal of Biological Chemistry*. 278 (13), pp. 11369–11375. doi: 10.1074/jbc.M300164200.

Mehra, R., Tomlins, S. a, Shen, R., Nadeem, O., Wang, L., Wei, J. T., Pienta, K. J., Ghosh, D., Rubin, M. a, Chinnaiyan, A. M. and Shah, R. B. (2007) Comprehensive assessment of TMPRSS2 and ETS family gene aberrations in clinically localized prostate cancer. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*. 20 (5), pp. 538–44. doi: 10.1038/modpathol.3800769.

Mohamed, A. A., Tan, S. H., Mikhalkevich, N., Ponniah, S., Vasioukhin, V., Bieberich, C. J., Sesterhenn, I. A., Dobi, A., Srivastava, S. and Sreenath, T. L. (2010) Ets family protein, erg expression in developing and adult mouse tissues by a highly specific monoclonal antibody. *Journal of Cancer*. 1 , pp. 197–208.

Mohamed, A., Tan, S. H., Sun, C., Shaheduzzaman, S., Hu, Y., Petrovics, G., Chen, Y., Sesterhenn, I. a., Li, H., Sreenath, T., McLeod, D. G., Dobi, A. and Srivastava, S. (2011) ERG oncogene modulates prostaglandin signaling in prostate cancer cells. *Cancer Biology and Therapy*. 11 (4), pp. 410–417. doi: 10.4161/cbt.11.4.14180.

Moore, S. D. P., Offor, O., Ferry, J. A., Amrein, P. C., Morton, C. C. and Dal Cin, P. (2006) ELF4 is fused to ERG in a case of acute myeloid leukemia with a t(X;21)(q25–26;q22). *Leukemia Research*.Pergamon. 30 (8), pp. 1037–1042.

doi: 10.1016/J.LEUKRES.2005.10.014.

Morcos, P. A. (2007) Achieving targeted and quantifiable alteration of mRNA splicing with Morpholino oligos. *Biochemical and Biophysical Research Communications*. 358 , pp. 521–527. doi: 10.1016/j.bbrc.2007.04.172.

Morcos, P. A., Li, Y. and Jiang, S. (2008) Vivo-Morpholinos: A non-peptide transporter delivers Morpholinos into a wide array of mouse tissues. *BioTechniques*. 45 (6), pp. 613–623. doi: 10.2144/000113005.

Motta-Mena, L. B., Heyd, F. and Lynch, K. W. (2010) Context-Dependent Regulatory Mechanism of the Splicing Factor hnRNP L. *Molecular Cell*. 37 (2), pp. 223–234. doi: 10.1016/j.molcel.2009.12.027.

Müller, W. E., Schröder, H. C., Skorokhod, A., Bünz, C., Müller, I. M. and Grebenjuk, V. A. (2001) Contribution of sponge genes to unravel the genome of the hypothetical ancestor of Metazoa (Urmetazoa). *Gene*. 276 (1–2), pp. 161–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11591483 (Accessed: 23 August 2018).

Munoz, I. M., Szyniarowski, P., Toth, R., Rouse, J. and Lachaud, C. (2014) Improved Genome Editing in Human Cell Lines Using the CRISPR Method. *PloS one*. 9 (10), p. e109752. doi: 10.1371/journal.pone.0109752.

Naro, C. and Sette, C. (2013) Phosphorylation-mediated regulation of alternative splicing in cancer. *International Journal of Cell Biology*. (Article ID 151839), p. 15 pages. doi: 10.1155/2013/151839.

Ng, A. P., Hu, Y., Metcalf, D., Hyland, C. D., Ierino, H., Phipson, B., Wu, D.,

Baldwin, T. M., Kauppi, M., Kiu, H., Di Rago, L., Hilton, D. J., Smyth, G. K. and Alexander, W. S. (2015) Early Lineage Priming by Trisomy of Erg Leads to Myeloproliferation in a Down Syndrome Model. *PLOS Genetics*.Edited by H. L. Grimes.Public Library of Science. 11 (5), p. e1005211. doi: 10.1371/journal.pgen.1005211.

Nikolova-Krstevski, V., Yuan, L., Le Bras, A., Vijayaraj, P., Kondo, M., Gebauer, I., Bhasin, M., Carman, C. V and Oettgen, P. (2009) ERG is required for the differentiation of embryonic stem cells along the endothelial lineage. *BMC developmental biology*. 9 , p. 72. doi: 10.1186/1471-213X-9-72.

Nowak, D. G., Amin, E. M., Rennel, E. S., Hoareau-Aveilla, C., Gammons, M., Damodoran, G., Hagiwara, M., Harper, S. J., Woolard, J., Ladomery, M. R. and Bates, D. O. (2010) Regulation of vascular endothelial growth factor (VEGF) splicing from pro-angiogenic to anti-angiogenic isoforms: a novel therapeutic strategy for angiogenesis. *The Journal of biological chemistry*.American Society for Biochemistry and Molecular Biology. 285 (8), pp. 5532–40. doi: 10.1074/jbc.M109.074930.

Oettgen, P., Finger, E., Sun, Z., Akbarali, Y., Thamrongsak, U., Boltax, J., Grall, F., Dube, A., Weiss, A., Brown, L., Quinn, G., Kas, K., Endress, G., Kunsch, C. and Libermann, T. A. (2000) PDEF, a novel prostate epithelium-specific ets transcription factor, interacts with the androgen receptor and activates prostate-specific antigen gene expression. *The Journal of biological chemistry*.American Society for Biochemistry and Molecular Biology. 275 (2), pp. 1216–25. doi: 10.1074/JBC.275.2.1216.

Oikawa, T. and Yamada, T. (2003) Molecular biology of the Ets family of transcription factors. *Gene*. 303 , pp. 11–34. doi: 10.1016/S0378-1119(02)01156-3.

Olshavsky, N. A., Comstock, C. E. S., Schiewer, M. J., Augello, M. A., Hyslop, T., Sette, C., Zhang, J., Parysek, L. M. and Knudsen, K. E. (2010) Identification of ASF/SF2 as a Critical, Allele-Specific Effector of the Cyclin D1b Oncogene. *Cancer Research*. 70 (10), pp. 3975–3984. doi: 10.1158/0008-5472.CAN-09-3468.

Orimo, H. and Shimada, T. (2006) Posttranscriptional modulation of the human tissue–nonspecific alkaline phosphatase gene expression by 1,25-dihydroxyvitamin D3 in MG-63 osteoblastic osteosarcoma cells. *Nutrition Research*.Elsevier. 26 (5), pp. 227–234. doi: 10.1016/J.NUTRES.2006.05.004.

Orkin, S. H. and Hochedlinger, K. (2011) Chromatin Connections to Pluripotency and Cellular Reprogramming. *Cell*. 145 (6), pp. 835–850. doi: 10.1016/j.cell.2011.05.019.

Ozaki, T. (2015) Diagnosis and treatment of Ewing sarcoma of the bone: a review article. *Journal of Orthopaedic Science*.Elsevier. 20 (2), pp. 250–263. doi: 10.1007/S00776-014-0687-Z.

Pan, J., Zou, J., Wu, D. Y., Roberson, R. S., Hennings, L. J., Ma, X., Yared, M., Blackburn, M. L., Chansky, H. A. and Yang, L. (2008) TLS-ERG Leukemia Fusion Protein Deregulates Cyclin-Dependent Kinase 1 and Blocks Terminal Differentiation of Myeloid Progenitor Cells. *Molecular Cancer Research*. 6 (5),

pp. 862–872. doi: 10.1158/1541-7786.MCR-07-2070.

Panagopoulos, I., Aman, P., Fioretos, T., Hoglund, M., Johansson, B., Mandahl, N., Heim, S., Behrendtz, M. and Mitelman, F. (1994) Fusion of the FUS gene with ERG in acute myeloid leukemia with t(16;21)(p11;q22). *Genes, chromosomes & cancer*.UNITED STATES. 11 (4), pp. 256–262.

Park, E., Pan, Z., Zhang, Z., Lin, L. and Xing, Y. (2018) The Expanding Landscape of Alternative Splicing Variation in Human Populations. *The American Journal of Human Genetics*.Cell Press. 102 (1), pp. 11–26. doi: 10.1016/J.AJHG.2017.11.002.

Paronetto, M. P., Cappellari, M., Busa, R., Pedrotti, S., Vitali, R., Comstock, C., Hyslop, T., Knudsen, K. E. and Sette, C. (2010) Alternative Splicing of the Cyclin D1 Proto-Oncogene Is Regulated by the RNA-Binding Protein Sam68. *Cancer Research*. 70 (1), pp. 229–239. doi: 10.1158/0008-5472.CAN-09-2788.

Penalva, L. O. F. and Sánchez, L. (2003) RNA binding protein sex-lethal (Sxl) and control of Drosophila sex determination and dosage compensation. *Microbiology and molecular biology reviews : MMBR*.American Society for Microbiology. 67 (3), p. 343–59, table of contents. doi: 10.1128/MMBR.67.3.343-359.2003.

Pereira, D. S., Dorrell, C., Ito, C. Y., Gan, O. I., Murdoch, B., Rao, V. N., Zou, J. P., Reddy, E. S. and Dick, J. E. (1998) Retroviral transduction of TLS-ERG initiates a leukemogenic program in normal human hematopoietic cells. *Proceedings of the National Academy of Sciences of the United States of*

America. 95 (14), pp. 8239–44. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9653171 (Accessed: 11 August 2018).

Prakash, T. P. (2011) An Overview of Sugar-Modified Oligonucleotides for Antisense Therapeutics. *Chemistry & Biodiversity*. 8 (9), pp. 1616–1641. doi: 10.1002/cbdv.201100081.

Prakash, V. (2017) Spinraza—a rare disease success story. *Gene Therapy*.Nature Publishing Group. 24 (9), pp. 497–497. doi: 10.1038/gt.2017.59.

Prasad, D. D., Rao, V. N., Lee, L. and Reddy, E. S. (1994) Differentially spliced erg-3 product functions as a transcriptional activator. *Oncogene*. 9 , pp. 669–673.

Query, C. C., Moore, M. J. and Sharp, P. A. (1994) Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes & development*. 8 (5), pp. 587–97. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7926752 (Accessed: 5 July 2018).

Rahl, P. B., Lin, C. Y., Seila, A. C., Flynn, R. A., McCuine, S., Burge, C. B., Sharp, P. A. and Young, R. A. (2010) c-Myc regulates transcriptional pause release. *Cell*.NIH Public Access. 141 (3), pp. 432–45. doi: 10.1016/j.cell.2010.03.030.

Rainis, L., Toki, T., Pimanda, J. E., Rosenthal, E., Machol, K., Strehl, S., Göttgens, B., Ito, E. and Izraeli, S. (2005) The proto-oncogene ERG in megakaryoblastic leukemias. *Cancer research*. 65 (17), pp. 7596–602. doi:

10.1158/0008-5472.CAN-05-0147.

Rampersad, S. N. (2012) Multiple applications of Alamar Blue as an indicator of metabolic function and cellular health in cell viability bioassays. *Sensors (Basel, Switzerland)*.Multidisciplinary Digital Publishing Institute (MDPI). 12 (9), pp. 12347–60. doi: 10.3390/s120912347.

Ran, F., Hsu, P., Wright, J. and Agarwala, V. (2013) Genome engineering using the CRISPR-Cas9 system. *Nature protocols*. 8 (11), pp. 2281–2308. doi: 10.1038/nprot.2013.143.Genome.

Raney, B. J., Dreszer, T. R., Barber, G. P., Clawson, H., Fujita, P. A., Wang, T., Nguyen, N., Paten, B., Zweig, A. S., Karolchik, D. and Kent, W. J. (2014) Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics*.Oxford University Press. 30 (7), pp. 1003–1005. doi: 10.1093/bioinformatics/btt637.

Rastogi, A., Tan, S., Mohamed, A. and Chen, Y. (2014) Functional antagonism of TMPRSS2-ERG splice variants in prostate cancer. *Genes & Cancer*. 5 (July), pp. 9–12. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4162137/ (Accessed: 17 November 2014).

Ray-Gallet, D., Mao, C., Tavitian, A. and Moreau-Gachelin, F. (1995) DNA binding specificities of Spi-1/PU.1 and Spi-B transcription factors and identification of a Spi-1/Spi-B binding site in the c-fes/c-fps promoter. *Oncogene*. 11 (2), pp. 303–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7624145 (Accessed: 2 July 2018).

Ray, D., Kazan, H., Chan, E. T., Peña Castillo, L., Chaudhry, S., Talukder, S., Blencowe, B. J., Morris, Q. and Hughes, T. R. (2009) Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nature biotechnology*. 27 (7), pp. 667–70. doi: 10.1038/nbt.1550.

Ray, D. … Hughes, T. R. (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature*.NIH Public Access. 499 (7457), pp. 172–7. doi: 10.1038/nature12311.

Reddy, E. S. and Rao, V. N. (1991) erg, an ets-related gene, codes for sequence-specific transcriptional activators. *Oncogene*. 6 , pp. 2285–2289.

Reddy, E. S., Rao, V. N. and Papas, T. S. (1987) The erg gene : A human gene related to the ets oncogene. *Proceedings of the National Academy of Sciences of the United States of America*. 84 (17), pp. 6131–6135.

Reed, R. (1996) Initial splice-site recognition and pairing during pre-mRNA splicing. *Current opinion in genetics & development*. 6 (2), pp. 215–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8722179 (Accessed: 5 July 2018).

Rickman, D. S., Chen, Y. B., Banerjee, S., Pan, Y., Yu, J., Vuong, T., Perner, S., Lafargue, C. J., Mertz, K. D., Setlur, S. R., Sircar, K., Chinnaiyan, a M., Bismar, T. a, Rubin, M. a and Demichelis, F. (2010) ERG cooperates with androgen receptor in regulating trefoil factor 3 in prostate cancer disease progression. *Neoplasia*. 12 (12), pp. 1031–1040. doi: 10.1593/neo.10866.

Rizzo, F., Fernandez-Serra, M., Squarzoni, P., Archimandritis, A. and Arnone,

M. I. (2006) Identification and developmental expression of the ets gene family in the sea urchin (Strongylocentrotus purpuratus). *Developmental Biology*.Academic Press. 300 (1), pp. 35–48. doi: 10.1016/J.YDBIO.2006.08.012.

Rosenbloom, K. R. … Kent, W. J. (2012) ENCODE Data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Research*.Oxford University Press. 41 (D1), pp. D56–D63. doi: 10.1093/nar/gks1172.

Roy, B., Haupt, L. and Griffiths, L. (2013) Review: Alternative Splicing (AS) of Genes As An Approach for Generating Protein Complexity. *Current genomics*. 14 , pp. 182–194. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3664468/ (Accessed: 17 November 2014).

Rutkovskiy, A., Stensløkken, K.-O. and Vaage, I. J. (2016) Osteoblast Differentiation at a Glance. *Medical science monitor basic research*.International Scientific Literature, Inc. 22 , pp. 95–106. doi: 10.12659/MSMBR.901142.

Salek-Ardakani, S., Smooha, G., de Boer, J., Sebire, N. J., Morrow, M., Rainis, L., Lee, S., Williams, O., Izraeli, S. and Brady, H. J. M. (2009) ERG is a megakaryocytic oncogene. *Cancer research*. 69 (11), pp. 4665–73. doi: 10.1158/0008-5472.CAN-09-0075.

Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A. and Burge, C. B. (2008) Proliferating Cells Express mRNAs with Shortened 3’ Untranslated Regions and

Fewer MicroRNA Target Sites. *Science*. 320 (5883), pp. 1643–1647. doi: 10.1126/science.1155390.

Schoch, K. M. and Miller, T. M. (2017) Antisense Oligonucleotides: Translation from Mouse Models to Human Neurodegenerative Diseases. doi: 10.1016/j.neuron.2017.04.010.

Schwer, B. and Shuman, S. (1996) Conditional inactivation of mRNA capping enzyme affects yeast pre-mRNA splicing in vivo. *RNA (New York, N.Y.)*.Cold Spring Harbor Laboratory Press. 2 (6), pp. 574–83. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8718686 (Accessed: 5 July 2018).

Seidel, J. J. and Graves, B. J. (2002) An ERK2 docking site in the Pointed domain distinguishes a subset of ETS transcription factors. *Genes & development*. 16 (1), pp. 127–37. doi: 10.1101/gad.950902.

Selvaraj, N., Kedage, V. and Hollenhorst, P. C. (2015) Comparison of MAPK specificity across the ETS transcription factor family identifies a high-affinity ERK interaction required for ERG function in prostate cells. *Cell communication and signaling : CCS*. 13 (1), p. 12. doi: 10.1186/s12964-015-0089-7.

Séraphin, B. (1995) Sm and Sm- like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *The EMBO Journal*.Wiley-Blackwell. 14 (9), pp. 2089–2098. doi: 10.1002/J.1460-2075.1995.TB07200.X.

Seth, A., Ascione, R., Fisher, R. J., Mavrothalassitis, G. J., Bhat, N. K. and Papas, T. S. (1992) The ets gene family. *Trends Biochem. Sci.,*. 17 (May), pp.

251–256. doi: 10.1016/0968-0004(92)90404-W.

Seth, A. and Watson, D. K. (2005) ETS transcription factors and their emerging roles in human cancer. doi: 10.1016/j.ejca.2005.08.013.

Shah, A. V., Birdsey, G. M. and Randi, A. M. (2016) Regulation of endothelial homeostasis, vascular development and angiogenesis by the transcription factor ERG. *Vascular Pharmacology*. doi: 10.1016/j.vph.2016.05.003.

Shapiro, M. B. and Senapathy, P. (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic acids research*. 15 (17), pp. 7155–74. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3658675 (Accessed: 5 July 2018).

Sharma, S., Maris, C., Allain, F. H.-T. and Black, D. L. (2011) U1 snRNA Directly Interacts with Polypyrimidine Tract-Binding Protein during Splicing Repression. *Molecular Cell*. 41 (5), pp. 579–588. doi: 10.1016/j.molcel.2011.02.012.

Sharrocks, A. D., Brown, A. L., Ling, Y. and Yates, P. R. (1997) The ETS-domain transcription factor family. *The International Journal of Biochemistry & Cell Biology*. 29 (12), pp. 1371–1387. doi: 10.1016/S1357-2725(97)00086-1.

Shen, Q. and Christakos, S. (2005) The vitamin D receptor, Runx2, and the notch signaling pathway cooperate in the transcriptional regulation of osteopontin. *Journal of Biological Chemistry*. 280 (49), pp. 40589–40598. doi: 10.1074/jbc.M504166200.

Shen, X. and Corey, D. R. (2018) Chemistry, mechanism and clinical status of antisense oligonucleotides and duplex RNAs. *Nucleic Acids Research*.Oxford

University Press. 46 (4), pp. 1584–1600. doi: 10.1093/nar/gkx1239.

Shieh, J.-J., Liu, K.-T., Huang, S.-W., Chen, Y.-J. and Hsieh, T.-Y. (2009) Modification of alternative splicing of Mcl-1 pre-mRNA using antisense morpholino oligonucleotides induces apoptosis in basal cell carcinoma cells. *The Journal of investigative dermatology*. 129 , pp. 2497–2506. doi: 10.1038/jid.2009.83.

Shing, D. C., McMullan, D. J., Roberts, P., Smith, K., Chin, S.-F., Nicholson, J., Tillman, R. M., Ramani, P., Cullinane, C. and Coleman, N. (2003) FUS/ERG Gene Fusions in Ewing's Tumors. *Cancer Research*. 63 (15), p. 4568 LP-4576. Available at: http://cancerres.aacrjournals.org/content/63/15/4568.abstract.

Shon, W., Folpe, a. L. and Fritchie, K. J. (2015) ERG expression in chondrogenic bone and soft tissue tumours. *Journal of Clinical Pathology*. 68 (2), pp. 125–129. doi: 10.1136/jclinpath-2014-202601.

Shore, P., Whitmarsh, A. J., Bhaskaran, R., Davis, R. J., Waltho, J. P. and Sharrocks, A. D. (1996) Determinants of DNA-binding specificity of ETS-domain transcription factors. *Molecular and Cellular Biology* . 16 (7), pp. 3338–3349. doi: 10.1128/MCB.16.7.3338.

Siddique, H. R., Rao, V. N., Lee, L. and Reddy, E. S. (1993) Characterization of the DNA binding and transcriptional activation domains of the erg protein. *Oncogene*. 8 (7), pp. 1751–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8510921 (Accessed: 9 October 2015).

Singareddy, R., Semaan, L., Conley-Lacomb, M. K., St John, J., Powell, K., Iyer,

M., Smith, D., Heilbrun, L. K., Shi, D., Sakr, W., Cher, M. L. and Chinni, S. R. (2013) Transcriptional regulation of CXCR4 in prostate cancer: significance of TMPRSS2-ERG fusions. *Molecular cancer research : MCR*. 11 (11), pp. 1349–61. doi: 10.1158/1541-7786.MCR-12-0705.

Slupsky, C. M., Gentile, L. N., Donaldson, L. W., Mackereth, C. D., Seidel, J. J., Graves, B. J. and McIntosh, L. P. (1998) Structure of the Ets-1 pointed domain and mitogen-activated protein kinase phosphorylation site. *Proceedings of the National Academy of Sciences of the United States of America*. 95 (October), pp. 12129–12134. doi: 10.1073/pnas.95.21.12129.

Smith, P. J., Zhang, C., Wang, J., Chew, S. L., Zhang, M. Q. and Krainer, A. R. (2006) An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Human Molecular Genetics*. 15 (16), pp. 2490–2508. doi: 10.1093/hmg/ddl171.

Sorensen, P. H. B., Lessnick, S. L., Lopez-Terrada, D., Liu, X. F., Triche, T. J. and Denny, C. T. (1994) A second Ewing's sarcoma translocation, t(21;22), fuses the EWS gene to another ETS–family transcription factor, ERG. *Nature Genetics*.Nature Publishing Group. 6 (2), pp. 146–151. doi: 10.1038/ng0294-146.

Sotoca, A. M., Prange, K. H. M., Reijnders, B., Mandoli, A., Nguyen, L. N., Stunnenberg, H. G. and Martens, J. H. A. (2015) The oncofusion protein FUS-ERG targets key hematopoietic regulators and modulates the all-trans retinoic acid signaling pathway in t(16;21) acute myeloid leukemia. *Oncogene*.Macmillan Publishers Limited. Available at:

http://dx.doi.org/10.1038/onc.2015.261.

Spellman, R., Rideau, A., Matlin, A., Gooding, C., Robinson, F., McGlincy, N., Grellscheid, S. N., Southby, J., Wollerton, M. and Smith, C. W. J. (2005) Regulation of alternative splicing by PTB and associated factors. *Biochemical Society Transactions*. 33 (3), pp. 457–460. doi: 10.1042/BST0330457.

Sperone, A., Dryden, N., Birdsey, G. M., Madden, L. E., Evans, P., Mason, J. C., Boyle, J., Paleolog, E., Haskard, D. O. and Randi, A. M. (2010) The transcription factor Erg represses ICAM-1 expression and vascular inflammation. *Atherosclerosis*. 213 (1), p. e17. doi: 10.1016/j.atherosclerosis.2010.08.039.

Squire, J. (2009) TMPRSS2-ERG and PTEN loss in prostate cancer. *Nature genetics*. 41 (5), pp. 509–510. Available at: http://www.nature.com/ng/journal/v41/n5/abs/ng0509-509.html (Accessed: 17 November 2014).

Staley, J. P. and Woolford, J. L. (2009) Assembly of ribosomes and spliceosomes: complex ribonucleoprotein machines. *Current Opinion in Cell Biology*. 21 (1), pp. 109–118. doi: 10.1016/j.ceb.2009.01.003.

Stankiewicz, M. J. and Crispino, J. D. (2009) ETS2 and ERG promote megakaryopoiesis and synergize with alterations in GATA-1 to immortalize hematopoietic progenitor cells. *Blood*.American Society of Hematology. 113 (14), pp. 3337–47. doi: 10.1182/blood-2008-08-174813.

Stankiewicz, M. J. and Crispino, J. D. (2013) AKT collaborates with ERG and

Gata1s to dysregulate megakaryopoiesis and promote AMKL. *Leukemia*.Macmillan Publishers Limited. 27 (6), pp. 1339–47. doi: 10.1038/leu.2013.33.

Sumarsono, S. H., Wilson, T. J., Tymms, M. J., Venter, D. J., Corrick, C. M., Kola, R., Lahoud, M. H., Papas, T. S., Seth, A. and Kola, I. (1996) Down's syndrome-like skeletal abnormalities in Ets2 transgenic mice. *Nature*. 379 (6565), pp. 534–7. doi: 10.1038/379534a0.

Sun, H. and Chasin, L. A. (2000) Multiple splicing defects in an intronic false exon. *Molecular and cellular biology*. 20 (17), pp. 6414–25. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10938119 (Accessed: 5 July 2018).

Suwa, Y., Nakamura, T., Toma, S., Ikemizu, S., Kai, H. and Yamagata, Y. (2008) Preparation, crystallization and preliminary X-ray diffraction analysis of the DNA-binding domain of the Ets transcription factor in complex with target DNA. *Acta crystallographica. Section F, Structural biology and crystallization communications*.International Union of Crystallography. 64 (Pt 3), pp. 171–4. doi: 10.1107/S1744309108002662.

Suwanmanee, T., Sierakowska, H., Lacerra, G., Svasti, S., Kirby, S., Walsh, C. E., Fucharoen, S. and Kole, R. (2002) Restoration of human beta-globin gene expression in murine and human IVS2-654 thalassemic erythroid cells by free uptake of antisense oligonucleotides. *Molecular pharmacology*. 62 (3), pp. 545–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12181431 (Accessed: 31 July 2018).

Tan, S. H., Furusato, B., Fang, X., He, F., Mohamed, A. A., Griner, N. B., Sood, K., Saxena, S., Katta, S., Young, D., Chen, Y., Sreenath, T., Petrovics, G., Dobi, A., McLeod, D. G., Sesterhenn, I. A., Saxena, S. and Srivastava, S. (2014) Evaluation of ERG responsive proteome in prostate cancer. *Prostate*. 74 , pp. 70–89. doi: 10.1002/pros.22731.

Taoudi, S., Bee, T., Hilton, A., Knezevic, K., Scott, J., Willson, T. A., Collin, C., Thomas, T., Voss, A. K., Kile, B. T., Alexander, W. S., Pimanda, J. E. and Hilton, D. J. (2011) ERG dependence distinguishes developmental control of hematopoietic stem cell maintenance from hematopoietic specification. *Genes & Development*. 25 (3), pp. 251–262. doi: 10.1101/gad.2009211.

Terao, M., Studer, M., Gianní, M. and Garattini, E. (1990) Isolation and characterization of the mouse liver/bone/kidney-type alkaline phosphatase gene. *The Biochemical journal*.Portland Press Ltd. 268 (3), pp. 641–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2363702 (Accessed: 5 March 2018).

Thoms, J. A. I., Birger, Y., Foster, S., Knezevic, K., Kirschenbaum, Y., Chandrakanthan, V., Jonquieres, G., Spensberger, D., Wong, J. W., Oram, S. H., Kinston, S. J., Groner, Y., Lock, R., MacKenzie, K. L., Göttgens, B., Izraeli, S. and Pimanda, J. E. (2011) ERG promotes T-acute lymphoblastic leukemia and is transcriptionally regulated in leukemic cells by a stem cell enhancer. *Blood*.American Society of Hematology. 117 (26), pp. 7079–89. doi: 10.1182/blood-2010-12-317990.

Tian, T. V, Tomavo, N., Huot, L., Flourens, A., Bonnelye, E., Flajollet, S., Hot, D., Leroy, X., de Launoit, Y. and Duterque-Coquillaud, M. (2014) Identification

of novel TMPRSS2:ERG mechanisms in prostate cancer metastasis: involvement of MMP9 and PLXNA2. *Oncogene*.Macmillan Publishers Limited. 33 (17), pp. 2204–14. doi: 10.1038/onc.2013.176.

Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcárcel, J. and Guigó, R. (2009) Nucleosome positioning as a determinant of exon recognition. *Nature Structural & Molecular Biology*. 16 (9), pp. 996–1001. doi: 10.1038/nsmb.1658.

Tomlins, S. A., Laxman, B., Varambally, S., Cao, X., Yu, J., Helgeson, B. E., Cao, Q., Prensner, J. R., Rubin, M. A., Shah, R. B., Mehra, R. and Chinnaiyan, A. M. (2008) Role of the TMPRSS2-ERG gene fusion in prostate cancer. *Neoplasia (New York, N.Y.)*. 10 (2), pp. 177–88. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2244693&tool=pmcentrez&rendertype=abstract (Accessed: 12 February 2015).

Tomlins, S. a, Rhodes, D. R., Perner, S., Dhanasekaran, S. M., Mehra, R., Sun, X.-W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., Lee, C., Montie, J. E., Shah, R. B., Pienta, K. J., Rubin, M. a and Chinnaiyan, A. M. (2005) Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science (New York, N.Y.)*. 310 (5748), pp. 644–8. doi: 10.1126/science.1117679.

Tsuzuki, S., Taguchi, O. and Seto, M. (2015) Promotion and maintenance of leukemia by ERG. 117 (14), pp. 3858–3869. doi: 10.1182/blood-2010-11-320515.The.

Turc-Carel, C., Aurias, A., Mugneret, F., Lizard, S., Sidaner, I., Volk, C., Thiery, J. P., Olschwang, S., Philip, I. and Berger, M. P. (1988) Chromosomes in Ewing's sarcoma. I. An evaluation of 85 cases of remarkable consistency of t(11;22)(q24;q12). *Cancer genetics and cytogenetics*. 32 (2), pp. 229–38. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3163261 (Accessed: 9 August 2018).

Urbinati, G., Ali, H. M., Rousseau, Q., Chapuis, H., Desmaële, D., Couvreur, P. and Massaad-Massade, L. (2015) Antineoplastic effects of siRNA against TMPRSS2-ERG junction oncogene in prostate cancer. *PLoS ONE*. 10 (5). doi: 10.1371/journal.pone.0125277.

Venables, J. P. … Elela, S. A. (2008) Identification of Alternative Splicing Markers for Breast Cancer. *Cancer Research*. 68 (22), pp. 9525–9531. doi: 10.1158/0008-5472.CAN-08-1769.

Venter, J. C. … Zhu, X. (2001) The Sequence of the Human Genome. *Science*. 291 (5507), pp. 1304–1351. doi: 10.1126/science.1058040.

Vissac, C., Peffault De Latour, M., Communal, Y., Bignon, Y.-J. and Bernard-Gallon, D. J. (2002) Expression of BRCA1 and BRCA2 in different tumor cell lines with various growth status. *Clinica Chimica Acta*.Elsevier. 320 (1–2), pp. 101–110. doi: 10.1016/S0009-8981(02)00055-4.

Vlaeminck-Guillem, V., Vanacker, J.-M., Verger, A., Tomavo, N., Stehelin, D., Laudet, V. and Duterque-Coquillaud, M. (2003) Mutual repression of transcriptional activation between the ETS-related factor ERG and estrogen

receptor. *Oncogene*. 22 (November), pp. 8072–8084. doi: 10.1038/sj.onc.1207094.

Wang, C. Y., Petryniak, B., Ho, I. C., Thompson, C. B. and Leiden, J. M. (1992) Evolutionarily conserved Ets family members display distinct DNA binding specificities. *Journal of Experimental Medicine*. 175 (5). Available at: http://jem.rupress.org/content/175/5/1391 (Accessed: 3 May 2017).

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P. and Burge, C. B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*.NIH Public Access. 456 (7221), pp. 470–6. doi: 10.1038/nature07509.

Wang, J., Cai, Y., Ren, C. and Ittmann, M. (2006) Expression of variant TMPRSS2/ERG fusion messenger RNAs is associated with aggressive prostate cancer. *Cancer research*. 66 (17), pp. 8347–51. doi: 10.1158/0008-5472.CAN-06-1966.

Wang, J., Cai, Y., Yu, W., Ren, C., Spencer, D. M. and Ittmann, M. (2008) Pleiotropic biological activities of alternatively spliced TMPRSS2/ERG fusion gene transcripts. *Cancer research*. 68 (20), pp. 8516–8524. doi: 10.1158/0008-5472.CAN-08-1147.

Wang, Z. and Burge, C. B. (2008) Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA*. 14 (5), pp. 802–813. doi: 10.1261/rna.876308.

Warner, J. K., Wang, J. C. Y., Takenaka, K., Doulatov, S., McKenzie, J. L.,

Harrington, L. and Dick, J. E. (2005) Direct evidence for cooperating genetic events in the leukemic transformation of normal human hematopoietic cells. *Leukemia*. 19 (10), pp. 1794–1805. doi: 10.1038/sj.leu.2403917.

Watson, D. K., McWilliams-Smith, M. J., Nunn, M. F., Duesberg, P. H., O'Brien, S. J. and Papas, T. S. (1985) The ets sequence from the transforming gene of avian erythroblastosis virus, E26, has unique domains on human chromosomes 11 and 21: both loci are transcriptionally active. *Proceedings of the National Academy of Sciences of the United States of America*. 82 (21), pp. 7294–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=391330&tool=pmcentrez&rendertype=abstract.

Wei, G. … Taipale, J. (2010) Genome- wide analysis of ETS- family DNA-binding in vitro and in vivo. *The EMBO Journal*. 29 (13), p. 2147 LP-2160. Available at: http://emboj.embopress.org/content/29/13/2147.abstract.

Weldon, C., Dacanay, J. G., Gokhale, V., Boddupally, P. V. L., Behm-Ansmant, I., Burley, G. A., Branlant, C., Hurley, L. H., Dominguez, C. and Eperon, I. C. (2018) Specific G-quadruplex ligands modulate the alternative splicing of Bcl-X. *Nucleic Acids Research*.Oxford University Press. 46 (2), pp. 886–896. doi: 10.1093/nar/gkx1122.

Werner, M. H., Clore, G. M., Fisher, C. L., Fisher, R. J., Trinh, L., Shiloach, J. and Gronenborn, A. M. (1997) Correction of the NMR structure of the ETS1/DNA complex. *Journal of biomolecular NMR*. 10 (4), pp. 317–28. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9460239 (Accessed: 2 July

2018).

Woolard, J. … Bates, D. O. (2004) VEGF $_{165}$ b, an Inhibitory Vascular Endothelial Growth Factor Splice Variant. *Cancer Research*. 64 (21), pp. 7822–7835. doi: 10.1158/0008-5472.CAN-04-0934.

Wu, L., Zhao, J. C., Kim, J., Jin, H.-J., Wang, C.-Y. and Yu, J. (2013) ERG is a critical regulator of Wnt/LEF1 signaling in prostate cancer. *Cancer research*. 73 (19), pp. 6068–79. doi: 10.1158/0008-5472.CAN-13-0882.

Xu, G. … Ito, E. (2003) Frequent mutations in the GATA-1 gene in the transient myeloproliferative disorder of Down syndrome. *Blood*.American Society of Hematology. 102 (8), pp. 2960–8. doi: 10.1182/blood-2003-02-0390.

Yao, Y., Wang, H., Li, B. and Tang, Y. (2014) Evaluation of the TMPRSS2:ERG fusion for the detection of prostate cancer: a systematic review and meta-analysis. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*. 35 (3), pp. 2157–66. doi: 10.1007/s13277-013-1286-x.

Yates, B., Braschi, B., Gray, K. A., Seal, R. L., Tweedie, S. and Bruford, E. A. (2017) Genenames.org: the HGNC and VGNC resources in 2017. *Nucleic Acids Research*. 45 (D1), pp. D619–D625. doi: 10.1093/nar/gkw1033.

Yeo, G. and Burge, C. B. (2004) Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. *Journal of Computational Biology*. 11 (2–3), pp. 377–394. doi: 10.1089/1066527041410418.

Yi, H. K., Fujimura, Y., Ouchida, M., Prasad, D. D. K., Rao, V. N. and Reddy, E.

S. P. (1997) Inhibition of apoptosis by normal and aberrant Fli-1 and erg proteins involved in human solid tumors and leukemias. *Oncogene*. 14 (11), pp. 1259–1268.

Yin, L., Rao, P., Elson, P., Wang, J., Ittmann, M. and Heston, W. D. W. (2011) Role of TMPRSS2-ERG gene fusion in negative regulation of PSMA expression. *PloS one*.Edited by J. A. Bauer.Public Library of Science. 6 (6), p. e21319. doi: 10.1371/journal.pone.0021319.

Yoshimoto, M., Joshua, A. and Cunha, I. (2008) Absence of TMPRSS2: ERG fusions and PTEN losses in prostate cancer is associated with a favorable outcome. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*. 21 (12), pp. 1451–1460. doi: 10.1038/modpathol.2008.96.

Yu, Y., Maroney, P. A., Denker, J. A., Zhang, X. H.-F., Dybkov, O., Lührmann, R., Jankowsky, E., Chasin, L. A. and Nilsen, T. W. (2008) Dynamic regulation of alternative splicing by silencers that modulate 5' splice site competition. *Cell*.NIH Public Access. 135 (7), pp. 1224–36. doi: 10.1016/j.cell.2008.10.046.

Yuan, L., Nikolova-Krstevski, V., Zhan, Y., Kondo, M., Bhasin, M., Varghese, L., Yano, K., Carman, C. V, Aird, W. C. and Oettgen, P. (2009) Antiinflammatory effects of the ETS factor ERG in endothelial cells are mediated through transcriptional repression of the interleukin-8 gene. *Circulation research*.United States. 104 (9), pp. 1049–1057. doi: 10.1161/CIRCRESAHA.108.190751.

Zammarchi, F., Boutsalis, G. and Cartegni, L. (2013) 5' UTR control of native

ERG and of Tmprss2:ERG variants activity in prostate cancer. *PloS one*. 8 (3), p. e49721. doi: 10.1371/journal.pone.0049721.

Zhang, J. and Manley, J. L. (2013) Misregulation of pre-mRNA alternative splicing in cancer. *Cancer discovery*.NIH Public Access. 3 (11), pp. 1228–37. doi: 10.1158/2159-8290.CD-13-0253.

Zhang, J. … Mullighan, C. G. (2016) Deregulation of DUX4 and ERG in acute lymphoblastic leukemia. *Nature Genetics*. 48 (12), pp. 1481–1489. doi: 10.5588/ijtld.16.0716.Isoniazid.

Zhang, M. Q. (1998) Statistical features of human exons and their flanking regions. *Human molecular genetics*. 7 (5), pp. 919–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9536098 (Accessed: 5 July 2018).

Zucman, J., Melot, T., Desmaze, C., Ghysdael, J., Plougastel, B., Peter, M., Zucker, J. M., Triche, T. J., Sheer, D., Turc-Carel, C. and al., et (1993) Combinatorial generation of variable fusion proteins in the Ewing family of tumours. *The EMBO journal*.European Molecular Biology Organization. 12 (12), pp. 4481–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8223458 (Accessed: 9 August 2018).

Zucman, J., Delattre, O., Desmaze, C., Plougastel, B., Joubert, I., Melot, T., Peter, M., De Jong, P., Rouleau, G., Aurias, A. and Thomas, G. (1992) Cloning and characterization of the Ewing's sarcoma and peripheral neuroepithelioma t(11;22) translocation breakpoints. *Genes, Chromosomes and Cancer*.Wiley-Blackwell. 5 (4), pp. 271–277. doi: 10.1002/gcc.2870050402.

8 Appendices

Sequences of primers used in this study 262

Sequencing results for PCR products in this this study 263

Normalisation and quantification procedure for qPCR 264

Proteins identified in mass spectrometry of RNA pull down 268

Summary of splicing regulatory proteins identified in this study 273

Publications arising from this thesis 276

Sequences of primers used in this study

Primer Name	Sequence (5′ to 3′)
ERGex6F	GAATATGGCCTTCCAGACGTCAAC
ERGex8R	CGGTCCAGGCTGATCTCCTG
BactinF	CCTGGCACCCAGCACAAT
BactinR	GCCGATCCACACGGAGTACT
	For ChIP
ALPLF	TCGCTGGAGGCATTCAAAC
ALPLR	CCCCTCTTTAACAGGCAGAC
	For qPCR
qRUNX2 ex9 F	CCAACCCACGAATGCACTATC
qRUNX2 ex9 R	TAGTGAGTGGTGGCGGACATAC
qALP F	CGTCGATTGCATCTCTGGGC
qALP R	GTCTCTTGCGCTTGGTCTCG
qOPN F	TGGCTAAACCCTGACCCATCT
qOPN R	TCATTGGTTTCTTCAGAGGACACA
qHNRNPF F	GGTTGAGAACAAAAACATGC
qHNRNPF R	ATGGTAATGAAATGTCCACG
qSRSF5 F	AGATTTCATGAGACAAGCTG
qSRSF5 R	CTCCCATTTATTTCCTTTCCAG
qPTBP1F	AGCGCGTGAAGATCCTGT TC
qPTBP1R	CAGGGGTGAGTTGCCGTAG

Sequencing results for PCR products in this this study

> ERG top

ATGATTACGAATCGATGGGAGGACTGTGCAAGATGACCAAGGACGACTTCCAGAGGCT
CACCCCCAGCTACAACGCCGACATCCTTCTCTCACATCTCCACTACCTCAGAGAGACT
CCTCTTCCACATTTGACTTCAGATGATGTTGATAAAGCCTTACAAAACTCTCCACGGT
TAATGCATGCTAGAAACACAGGGGGTGCAGCTTTTATTTTCCCAAATACTTCAGGATA
TCCTGAAGGTACGCAAAGAAACTTACTAGGCCAGATTTACCATATGAGCCCCCCTGGA
GATCTACCAGGACGGGTCGCGGCCACCGTTCCATAATACTATATATAACATATCTAAC
ATTTTGGACTAGAATTAAGTAGCTGATTTTCTTTTGCCACGTTTCCTGCCATAACTTC
ACCGAGTGTGGGTGTCGCATCTGAAGTCACAAACTAGTTTACATGTTACCAATATAGA
CTTTATATTCTCAATGTCTAGCCCAGTTCCTACCCTACTCTAACTGTTGCATTCTCTG
CAGGTTTCTCGAATGCAAGTGCGGGGAAAGTTACAAAACACGTCAGCAACCTCGGACC
AAAAATTCTGAGATCGGAGACATACTTGTACATGACACTCTTGCCATAACCTCTTTTC
TGGACTGTTGCAGAGGTGTAGTAGGAAGCAGACACTCTGTAGCCATCAAGATCTTCTT
ATGCAGCACCTGT

> ERG middle

ATGATACGAATCGATGGGAGGACTGTGCAAGATGACCAAGGACGACTTCCAGAGGCTC
ACCCCCAGCTACAACGCCGACATCCTTCTCTCACATCTCCACTACCTCAGAGAGACTC
CTCTTCCACATTTGACTTCAGATGATGTTGATAAAGCCTTACAAAACTCTCCACGGTT
AATGCATGCTAGAAACACAGATTTACCATATGAGCCCCCCAGGAGATCAGCCTGGACC
GGTCACGGCCACCCAGCCTCGTGTCCGCTCTGACCTAGTGCCTTCATTGATTCCCCCG
CTACCCCAAGCACATGAACCAGCACTACTTCTTACGCGGTCCCTGCCCGAAAGCAACA
TCAACGAGCATCTTCCTTCTTCTATGACGACGGCAACTACCAGACCAGCTCCGAAGTG
AAGTTCTAACTCCACACCGATCCGAACCGAATTGAGATGAATCGCATCCATCTCCGAG
AGGATCGTCACGTTCTGCGGCACATCCTGTAGTACCAGTACAACGTCCTCTCGTTCTA
TATCATAGGCGACAAGTAGAACATCGGCGTCTACGTGCAACATCCGCATCCGCCACAA
CGTCGAGGACGGG

> ERG bottom

ATTGCATTCCGACTCGATGGGAGGAACTGTGCAAGATGACCAAGGACGACTTCCAGAG
GCTCACCCCCAGCTACAACGCCGACATCCTTCTCTCACATCTCCACTACCTCAGAGAG
AATTTACCATATGAGCCCCCCAGGAGATCAGCCTGGACCGGTCACGGCCACCCCACGC
CCCAGTCGAAAGCTGCTCAACCATCTCCTTCCACAGTGCCCAAAACTGAAGACCAGCG
TCCTCAGTTAGATCCTTATCAGATTCTTGGACCAACAAGTAGCCGCCTTGCAAATCCA
G

Exon 6 Exon 7 Exon 7b Exon 8

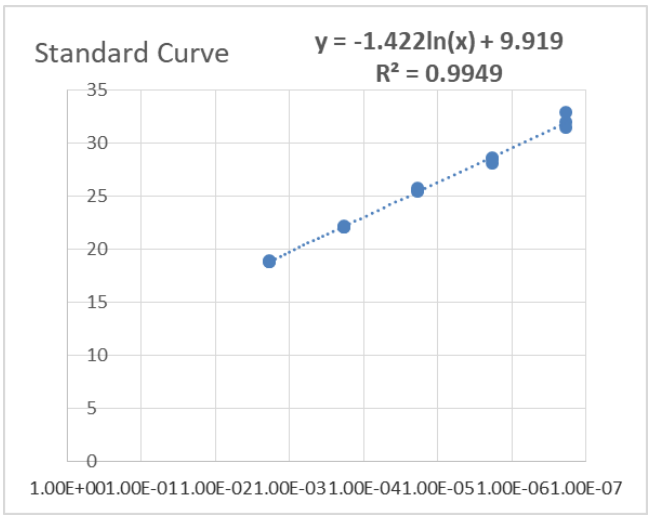
Underlined bases are miscalled by sequencing reaction

Normalisation and quantification procedure for qPCR

As stated in the qPCR method a spike was run for each sample to facilitate normalization. Linearized DNA for the same spike sequence was used to create a minimum five point 10-fold dilution standard curve included on each plate run. The following method describes how qPCR data was normalized and quantified.

1. A standard curve was created of the mass of cDNA standard in nanogram (ng) (x-axis) and Ct (y-axis). The Ct values for each well was used to derive the mass of cDNA amplified in each sample.

Amount of standard (ng)	Ct
1.86E-03	18.91226
1.86E-03	18.82157
1.86E-03	18.85822
1.86E-04	22.19721
1.86E-04	22.0679
1.86E-04	22.05155
1.86E-05	25.7451
1.86E-05	25.4576
1.86E-05	25.57102
1.86E-06	28.13166
1.86E-06	28.5323
1.86E-06	28.564
1.86E-07	31.42349
1.86E-07	31.94827
1.86E-07	32.89028



SPIKE				SRSF5		
Sample	Ct	Amount of standard (ng)		Sample	Ct	Amount of standard (ng)
untreated	16.62271	0.008966312		untreated	22.95869	0.00010412
untreated	16.53816	0.009515609		untreated	22.91039	0.00010772
untreated	16.62316	0.00896345		untreated	22.9103	0.00010772
transfection	16.22703	0.011842902		transfection	22.46821	0.00014701
transfection	16.25094	0.011645463		transfection	22.45071	0.00014883
transfection	16.39106	0.010552664		transfection	22.46825	0.000147
control sIRNA	17.10127	0.006404059		control sIRNA	22.13436	0.00018591
control sIRNA	17.07617	0.006518123		control sIRNA	22.11804	0.00018805
control sIRNA	17.08081	0.006496886		control sIRNA	21.95165	0.0002114
PTBP1 10	16.22092	0.011893907		PTBP1 10	22.4272	0.00015131
PTBP1 10	16.23546	0.011772929		PTBP1 10	22.37702	0.00015674
PTBP1 10	16.35033	0.010859274		PTBP1 10	22.45064	0.00014883
PTBP1 20	17.70487	0.004188979		PTBP1 20	23.10477	9.3955E-05
PTBP1 20	17.70528	0.004187788		PTBP1 20	22.95594	0.00010432

2. As each sample was spiked with an equal mass of spike RNA the expression of the spike could be used to normalize expression across samples. A multiplication factor was derived for each sample using the equation below.

$$Multiplicator\ factor\ = \frac{\text{largest mass of all spike cDNA samples}}{\text{mass of spike cDNA in the sample}}$$

3. To work out amount of gene expression. The mass of gene cDNA for each sample was normalized using the multiplication factor and converted to grams by multiplying the mass by 10^{-9}.

$$normalized\ mass\ of\ cDNA$$
$$= amount\ of\ cDNA\ in\ ng\ \times\ multiplication\ factor\ \times 10^{-9}$$

4. The number of amplicons at the Ct value was calculated. Avogadro's number = 6.02214 x 10^{23}.

$$\text{number of amplicons at the Ct value} = \frac{(normalized\ mass\ of\ cDNA\ in\ g \times avogadro's\ number)}{amplicon\ size}$$

5. The molecular weight of the mRNA transcript was calculated. The largest transcript size for each gene was used for the calculation. The average molecular weight of an RNA nucleotide 320.5 and the addition of 159 takes into account the molecular weight of a 5' triphosphate.

$$\text{molecular weight of the mRNA transcript}$$
$$= (\text{mRNA transcript}\ \times 320.5) + 159$$

6. The number of gene transcripts at Ct value was calculated. The normalized mass of the cDNA was divided by two to account for the single stranded nature of RNA vs cDNA.

$$\text{number of gene transcripts at Ct value} = \frac{(normalised\ mass\ of\ cDNA\ in\ g\ \div 2)}{molecular\ weight\ of\ mRNA\ transcript} \times Avogadro's number$$

7. The total mass of gene transcript in the sample was calculated

$$\text{mass of gene transcript in the sample} = \frac{\text{number of gene transcripts at Ct value}}{Avogadro's number} \times molecular\ weight\ of\ mRNA\ transcript$$

8. The expression of each gene is reported as a percentage of the total RNA used for the original cDNA synthesis reaction.

$$\%\ total\ RNA\ input\ (for\ gene\ of\ interest) = \frac{\text{mass of gene transcript in the sample}}{mass\ of\ RNA\ used\ for\ cDNA\ synthesis} \times 100$$

An example is shown of the next page to illustrate the method. This is how the expression of SRSF5 could be calculated in an untreated sample. Note that the experimetns were carried out in with three technical repeats. What is shown below is one of the three technical repeats. The method below starts at step 3

Step 3

A	B	C	D
Sample	**Amount of DNA (ng)**	**Multiplication factor**	**Normalised amount of DNA (g)**
untreated	0.00010412	1.4	=B3*C3*(10^-9)

Step 4 and 5

E	F	J	K
Amplicon size	**Amplicon number at Ct**	**Transcript size**	**MW of SRSF5 transcript**
141	=((D3*(6.02214*10^23))/(E3))	2815	=(J3*320.5)+159

Step 6, 7 and 8

L	M	N
Transcript number at Ct	**Mass of transcript (g)**	**Percentage of total RNA**
=(((D3)/2)/K3)*(6.02214*10^23)	=((L3)/(6.02214*10^23))*K3	=(M3/(200*10^-9))*100

PSI and ΔPSI values for high throughput RNAi experiment

Gene	PSI
Lipo MCF7 tr100	0.567
DEK MCF7 tr100	0.548
HMGA1 MCF7 tr100	0.645
PRPF8 MCF7 tr100	0.449
Lipo MCF7 tr102	0.524
CCNH MCF7 tr102	0.549
CRK7 MCF7 tr102	0.464
SFRS16 MCF7 tr102	0.612
Lipo MCF7 tr104	0.599
KHDRBS1 MCF7 tr104	0.569
Lipo MCF7 tr105	0.639
CDC5L MCF7 tr105	0.714
KHDRBS3 MCF7 tr105	0.392
MBNL1 MCF7 tr105	0.648
Lipo MCF7 tr108	0.656
ZNF638 MCF7 tr108	0.631
Lipo MCF7 tr110	0.563
CDK11 MCF7 tr110	0.739
CTNNB1 MCF7 tr110	0.584
Lipo MCF7 tr111	0.532
DDX6 MCF7 tr111	0.608
THOC1 MCF7 tr111	0.570
Lipo MCF7 tr115	0.589
FUS MCF7 tr115	0.517
FXR1 MCF7 tr115	0.498
PRPF4B MCF7 tr115	0.509
SAFB MCF7 tr115	0.614
SF3B4 MCF7 tr115	0.428
Lipo MCF7 tr123	0.389
SFRS4 MCF7 tr123	0.536
SR-A1 MCF7 tr123	0.652
Lipo MCF7 tr125	0.525
BRCA1 MCF7 tr125	0.329
Lipo MCF7 tr127	0.517
HEAB MCF7 tr127	0.347
MSI2 MCF7 tr127	0.675
Lipo MCF7 tr129	0.657
NPM1 MCF7 tr129	0.610
RPL22 MCF7 tr129	0.695
RPN1 MCF7 tr129	0.627
Lipo MCF7 tr44	0.497
TIAL1 MCF7 tr44	0.410

Gene	ΔPSI
A1-6	-1.0
BRCA1	-19.7
CCNH	2.5
CDC5L	7.5
CDK11	17.6
CLK1	5.3
CRK7	-6.0
CTNNB1	2.1
DDX39	0.6
DDX6	7.6
DEK	-1.9
FMR1	-3.9
FUS	-7.2
FXR1	-9.2
HEAB	-17.1
HMGA1	7.8
HNRPD	3.3
HNRPL	20.8
KHDRBS1	-3.0
KHDRBS3	-24.7
KHSRP	-4.0
MBNL1	0.9
MSI2	15.7
NPM1	-4.7
PRPF4B	-8.0
PRPF8	-11.8
RALY	-11.8
RBM3	-8.6
RBM9	-16.8
RNPC2	-28.1
RNPS1	16.7
RPL22	3.8
RPN1	-3.0
SAFB	2.5
SF3A1	-35.3
SF3A2	-28.2
SF3A3	-5.3
SF3B4	-16.1
SFPQ	-2.0
SFRS10	-5.1
SFRS12	-17.3
SFRS16	8.8

Lipo MCF7 tr68	0.488
SFRS3 MCF7 tr68	0.454
SFRS5 MCF7 tr68	0.586
Lipo MCF7 tr75	0.515
SNRP70 MCF7 tr75	0.506
U2AF1 MCF7 tr75	0.352
Lipo MCF7 tr78	0.600
RBM9 MCF7 tr78	0.432
Lipo MCF7 tr82	0.569
CLK1 MCF7 tr82	0.622
RNPS1 MCF7 tr82	0.736
Lipo MCF7 tr83	0.509
SF3A3 MCF7 tr83	0.456
Lipo MCF7 tr84	0.535
A1-6 MCF7 tr84	0.525
Lipo MCF7 tr86	0.579
DDX39 MCF7 tr86	0.585
KHSRP MCF7 tr86	0.539
Lipo MCF7 tr88	0.609
RBM3 MCF7 tr88	0.523
RNPC2 MCF7 tr88	0.328
SFRS12 MCF7 tr88	0.436
Lipo MCF7 tr89	0.439
FMR1 MCF7 tr89	0.400
SFRS8 MCF7 tr89	0.586
TARBP1 MCF7 tr89	0.765
Lipo MCF7 tr95	0.582
RALY MCF7 tr95	0.465
SFPQ MCF7 tr95	0.563
SNW1 MCF7 tr95	0.705
SRPK1 MCF7 tr95	0.556
Lipo MCF7 tr96	0.443
HNRPD MCF7 tr96	0.476
Lipo MCF7 tr97	0.577
SF3A1 MCF7 tr97	0.225
SF3A2 MCF7 tr97	0.296
Lipo MCF7 tr99	0.560
HNRPL MCF7 tr99	0.769
SFRS10 MCF7 tr99	0.509
SYNCRIP MCF7 tr99	0.490

SFRS3	-3.4
SFRS4	14.7
SFRS5	9.8
SFRS8	14.6
SNRP70	-0.9
SNW1	12.3
SR-A1	26.4
SRPK1	-2.7
SYNCRIP	-7.1
TARBP1	32.6
THOC1	3.8
U2AF1	-16.2
ZNF638	-2.4
TIAL1	-8.6

Proteins identified in mass spectrometry of RNA pull down

Description	Score A2	Coverage A2	# Peptides A2	# PSM A2	Score B2	Coverage B2	# Peptides B2	# PSM B2	MW [kDa]	#PSM ratio
Nucleolin OS=Homo sapiens GN=NCL PE=1 SV=3 - [NUCL_HUMAN]	128.86	32.68	25	43	563.91	54.37	40	204	76.6	4.74
Heterogeneous nuclear ribonucleoprotein M OS=Homo sapiens GN=HNRNPM PE=1 SV=3 - [HNRPM_HUMAN]	199.96	45.34	30	77	562.35	71.37	52	218	77.5	2.83
Heterogeneous nuclear ribonucleoprotein K, isoform CRA_d OS=Homo sapiens GN=HNRPK PE=4 SV=1 - [A0A024R228_HUMAN]	134.90	51.84	21	55	551.41	64.79	30	202	51.2	3.67
Polypyrimidine tract-binding protein 1 OS=Homo sapiens GN=PTBP1 PE=1 SV=1 - [A0A0U1RRM4_HUMAN]	91.11	37.41	14	30	531.30	58.67	26	162	62.4	5.40
Matrin-3 OS=Homo sapiens GN=MATR3 PE=1 SV=1 - [A8MXP9_HUMAN]	88.83	29.50	21	34	472.74	69.39	60	205	99.9	6.03
Poly(rC)-binding protein 1 OS=Homo sapiens GN=PCBP1 PE=1 SV=2 - [PCBP1_HUMAN]	87.97	56.74	13	32	318.32	71.91	16	94	37.5	2.94
Heterogeneous nuclear ribonucleoprotein H OS=Homo sapiens GN=HNRNPH1 PE=1 SV=4 - [HNRH1_HUMAN]	135.05	42.09	14	56	284.61	50.78	16	101	49.2	1.80
Heterogeneous nuclear ribonucleoprotein F OS=Homo sapiens GN=HNRNPF PE=1 SV=3 - [HNRPF_HUMAN]	60.35	26.02	6	22	269.34	51.08	15	91	45.6	4.14
Heterogeneous nuclear ribonucleoprotein M (Fragment) OS=Homo sapiens GN=HNRNPM PE=1 SV=1 - [M0QZM1_HUMAN]	103.41	47.26	14	41	268.50	67.10	21	103	40.0	2.51
Heterogeneous nuclear ribonucleoprotein A1 OS=Homo sapiens GN=HNRNPA1 PE=1 SV=2 - [F8W6I7_HUMAN]	111.09	63.84	19	42	252.94	64.82	21	102	33.1	2.43
Heterogeneous nuclear ribonucleoprotein D (AU-rich element RNA binding protein 1, 37kDa), isoform CRA_e OS=Homo sapiens GN=HNRPD PE=4 SV=1 - [A0A024RDF4_HUMAN]	65.98	43.46	16	31	240.74	52.29	21	90	32.8	2.90
Splicing factor U2AF 65 kDa subunit OS=Homo sapiens GN=U2AF2 PE=1 SV=4 - [U2AF2_HUMAN]	35.14	25.47	7	11	220.50	46.53	15	75	53.5	6.82
HNRNPL protein (Fragment) OS=Homo sapiens GN=HNRNPL PE=2 SV=2 - [Q6NTA2_HUMAN]	31.56	20.80	7	9	181.43	51.05	19	63	61.9	7.00
Poly(U)-binding-splicing factor PUF60 (Fragment) OS=Homo sapiens GN=PUF60 PE=1 SV=1 - [A0A0J9YVP6_HUMAN]	34.94	24.72	8	13	172.49	50.94	21	56	57.4	4.31
Heterogeneous nuclear ribonucleoproteins A2/B1 OS=Homo sapiens GN=HNRNPA2B1 PE=1 SV=2 - [ROA2_HUMAN]	106.27	55.81	18	40	159.43	57.51	21	61	37.4	1.53
Splicing factor 1, isoform CRA_a OS=Homo sapiens GN=SF1 PE=4 SV=1 - [A0A024R566_HUMAN]	32.61	21.19	8	11	158.04	35.73	15	62	61.8	5.64
Heterogeneous nuclear ribonucleoprotein H3 OS=Homo sapiens GN=HNRNPH3 PE=1 SV=2 - [HNRH3_HUMAN]	42.65	33.24	8	20	152.54	53.76	12	54	36.9	2.70
Heterogeneous nuclear ribonucleoprotein H2 OS=Homo sapiens GN=HNRNPH2 PE=1 SV=1 - [HNRH2_HUMAN]	66.22	28.51	10	33	137.68	40.53	14	62	49.2	1.88
Protein FAM98B OS=Homo sapiens GN=FAM98B PE=1 SV=1 - [FA98B_HUMAN]	1.91	3.94	1	1	124.91	46.06	11	39	37.2	39.00

Heterogeneous nuclear ribonucleoprotein D-like OS=Homo sapiens GN=HNRNPDL PE=1 SV=1 - [A0A087WUK2_HUMAN]	34.78	26.45	8	14	108.25	31.96	14	40	40.0	2.86
Mitochondrial antiviral-signaling protein OS=Homo sapiens GN=MAVS PE=1 SV=2 - [MAVS_HUMAN]					107.07	35.37	10	32	56.5	32.00
G-rich sequence factor 1 OS=Homo sapiens GN=GRSF1 PE=1 SV=3 - [GRSF1_HUMAN]	2.82	4.38	2	2	100.37	45.21	16	35	53.1	17.50
Heterogeneous nuclear ribonucleoprotein A0 OS=Homo sapiens GN=HNRNPA0 PE=1 SV=1 - [ROA0_HUMAN]	36.30	25.25	7	12	99.19	42.95	11	34	30.8	2.83
Splicing factor, proline- and glutamine-rich OS=Homo sapiens GN=SFPQ PE=1 SV=2 - [SFPQ_HUMAN]	56.27	29.99	17	25	93.71	28.43	17	37	76.1	1.48
Heterogeneous nuclear ribonucleoprotein A/B OS=Homo sapiens GN=HNRNPAB PE=1 SV=1 - [D6R9P3_HUMAN]	42.65	34.64	10	16	85.55	36.43	11	35	30.3	2.19
Serine/arginine-rich splicing factor 1 OS=Homo sapiens GN=SRSF1 PE=1 SV=2 - [SRSF1_HUMAN]	23.92	31.85	8	14	85.40	47.18	13	37	27.7	2.64
Splicing factor U2AF 35 kDa subunit-like protein OS=Homo sapiens GN=U2AF1L5 PE=3 SV=1 - [U2AF5_HUMAN]	10.03	16.25	3	4	84.21	47.08	8	31	27.9	7.75
RNA-binding protein 10 OS=Homo sapiens GN=RBM10 PE=1 SV=3 - [RBM10_HUMAN]	1.86	1.29	1	1	73.99	19.25	16	31	103.5	31.00
Ribonucleoprotein PTB-binding 1 OS=Homo sapiens GN=RAVER1 PE=1 SV=1 - [A0A087WZ13_HUMAN]	4.38	5.68	2	3	73.62	32.61	15	30	77.8	10.00
RNA-binding protein 12 OS=Homo sapiens GN=RBM12 PE=1 SV=1 - [RBM12_HUMAN]					70.73	19.85	17	35	97.3	35.00
Cap-specific mRNA (nucleoside-2'-O-)-methyltransferase 1 OS=Homo sapiens GN=CMTR1 PE=1 SV=1 - [CMTR1_HUMAN]					67.54	18.92	15	28	95.3	28.00
Pre-mRNA-splicing factor RBM22 OS=Homo sapiens GN=RBM22 PE=1 SV=1 - [RBM22_HUMAN]	1.92	2.62	1	1	64.25	28.57	11	28	46.9	28.00
CUG triplet repeat, RNA binding protein 1, isoform CRA_c OS=Homo sapiens GN=CELF1 PE=1 SV=1 - [G5EA30_HUMAN]					63.63	23.54	13	24	55.1	24.00
Polypyrimidine tract-binding protein 3 OS=Homo sapiens GN=PTBP3 PE=1 SV=2 - [PTBP3_HUMAN]	1.72	1.09	1	1	59.99	23.91	12	28	59.7	28.00
Serine/arginine-rich splicing factor 3 OS=Homo sapiens GN=SRSF3 PE=1 SV=1 - [SRSF3_HUMAN]	25.91	31.10	5	11	50.67	39.63	7	21	19.3	1.91
Serine/arginine-rich splicing factor 5 OS=Homo sapiens GN=SRSF5 PE=1 SV=1 - [SRSF5_HUMAN]	9.01	8.46	2	5	38.23	25.00	6	14	31.2	2.80
Serine/arginine-rich splicing factor 9 OS=Homo sapiens GN=SRSF9 PE=1 SV=1 - [SRSF9_HUMAN]					26.64	40.27	11	16	25.5	16.00
Polypyrimidine tract-binding protein 2 OS=Homo sapiens GN=PTBP2 PE=1 SV=1 - [PTBP2_HUMAN]					26.50	14.31	7	14	57.5	14.00
RNA-binding protein 42 OS=Homo sapiens GN=RBM42 PE=1 SV=1 - [K7EQ03_HUMAN]					18.71	16.43	6	7	44.6	7.00
Probable RNA-binding protein 19 OS=Homo sapiens GN=RBM19 PE=1 SV=3 - [RBM19_HUMAN]					16.03	6.46	5	6	107.3	6.00

RNA-binding protein 5 OS=Homo sapiens GN=RBM5 PE=1 SV=2 - [RBM5_HUMAN]					10.23	6.01	4	4	92.1	4.00
Splicing factor 3A subunit 2 OS=Homo sapiens GN=SF3A2 PE=1 SV=2 - [SF3A2_HUMAN]					8.08	3.88	1	3	49.2	3.00
Cold-inducible RNA-binding protein OS=Homo sapiens GN=CIRBP PE=1 SV=1 - [CIRBP_HUMAN]					5.78	12.79	2	2	18.6	2.00
RNA-binding protein 38 OS=Homo sapiens GN=RBM38 PE=1 SV=1 - [F6VZ39_HUMAN]					4.58	21.14	2	3	13.6	3.00
RNA-binding protein fox-1 homolog 2 (Fragment) OS=Homo sapiens GN=RBFOX2 PE=1 SV=1 - [A0A0G2JSB3_HUMAN]					4.57	16.67	3	3	20.2	3.00
Nucleoporin 153kDa, isoform CRA_a OS=Homo sapiens GN=NUP153 PE=4 SV=1 - [A0A024QZW7_HUMAN]					3.42	1.02	1	1	153.9	1.00
RNA-binding protein MEX3B OS=Homo sapiens GN=MEX3B PE=1 SV=1 - [MEX3B_HUMAN]					2.54	5.10	2	4	58.8	4.00

Summary of splicing regulatory proteins identified in this study

Protein	Mass Spectrometry	MCF assay	RNA Compete Motif	SF correlation	RBP correlation	Splice Aid2
A1-6		Y				
BRCA1		Y				
CCNH		Y				
CDC5L		Y				
CDK11		Y				
CELF1	Y		Y			Y
CELF2			Y	Y	Y	Y
CELF4			Y	Y		
CELF5			Y	Y	Y	
CIRBP	Y					
CLK1		Y				
CMTR1	Y					
CPEB2			Y			
CPEB4			Y			
CRK7		Y				
CTNNB1		Y				
DDX39		Y				
DEK		Y				
ELAVL1			Y	Y		
ENOX1			Y	Y		
FAM98B	Y					
FMR1		Y	Y			
FUS		Y		Y		Y
FXR1		Y	Y	Y		
GRSF1	Y					
HEAB		Y				
HMGA1		Y				
HNRNPA0	Y					
HNRNPA1	Y		Y			Y
HNRNPA2B1	Y		Y	Y		
HNRNPAB	Y					
HNRNPD	Y	Y				
HNRNPDL	Y					
HNRNPF	Y					Y
HNRNPH1	Y					Y
HNRNPH2	Y					Y
HNRNPH3	Y					Y
HNRNPK	Y					Y
HNRNPL	Y	Y	Y	Y		

HNRNPM	Y					
KHDRBS1		Y				Y
KHDRBS3		Y		Y		Y
KHSRP		Y		Y		Y
MATRIN3	Y		Y	Y		
MAVS	Y					
MBNL1		Y			Y	Y
MEX3B	Y					
MSI2		Y				
NPM1		Y				
NUP153	Y					
PABP			Y			
PCBP1	Y		Y			
PCBP2			Y			
PCBP3			Y		Y	
PRPF4B		Y				
PRPF8		Y				
PTBP1	Y		Y			Y
PTBP2	Y			Y	Y	
PTBP3	Y					
PUF60	Y					
RALY		Y		Y	Y	
RAVER	Y					
RBFOX2	Y					
RBM10	Y					
RBM12	Y					
RBM19	Y					
RBM22	Y					
RBM24			Y			
RBM28			Y	Y		
RBM3		Y	Y			
RBM38	Y			Y	Y	
RBM42	Y		Y			
RBM5	Y		Y	Y		
RBM6			Y			
RBM9		Y		Y	Y	
RBMS1			Y			
RBP1			Y			
RNPC2		Y				
RNPS1		Y				
RPL22		Y				
RPN1		Y				
SAFB		Y				

SF1	Y			Y		
SF3A1		Y		Y		
SF3A2	Y	Y				
SF3A3		Y				
SF3B4		Y	Y			
SFPQ	Y	Y		Y		
SNRP70		Y		Y		
SNW1		Y				
SR-A1		Y				
SRPK1		Y				
SRSF1	Y			Y		
SRSF10		Y	Y			
SRSF12		Y				
SRSF16		Y				
SRSF3	Y	Y				Y
SRSF4		Y				
SRSF5	Y	Y				Y
SRSF8						
SRSF9	Y		Y			Y
SXL			Y			
SYNCRIP		Y		Y		
TARBP1		Y				Y
THOC1		Y				
TIAL1		Y		Y		Y
U2AF1		Y				
U2AF1L5	Y					
U2AF2	Y		Y			
YBX			Y			Y
ZNF638		Y				

Publications arising from this thesis

Adamo, P., Porazinski, S., Rajatileka, S., Jumbe, S., Hagen, R., Cheung, M., Wilson, I. and Ladomery, M. (2017) The oncogenic transcription factor ERG represses the tumour suppressor gene PTEN in prostate cancer cells. *Oncology Letters.* **14** (5), pp. 5605-5610.

Jumbe, S.L., Porazinski, S.R., Oltean, S., Mansell, J.P., Vahabi, B., Wilson, I.D. and Ladomery, M.R. (2019) The evolutionarily conserved cassette exon 7b drives ERG's oncogenic properties. *Translation Oncology.* **12** (1), pp.134-142