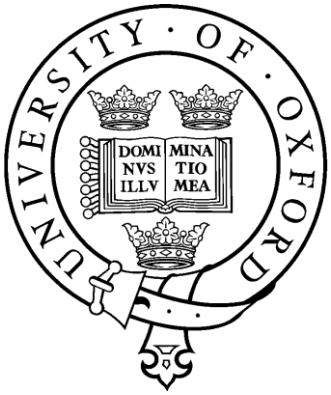
Transcriptome-wide Study of the Role of Arginine Methylation on E2F1



Alice Poppy Roworth
Kellogg College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Michaelmas 2017



Contents

Acknowledgements **4**

Abstract **5**

List of abbreviations **6**

List of tables **9**

List of figures **10**

Chapter 1. Introduction **14**

1.1 The hallmarks of cancer 15

1.2 Biological functions of E2F1 21

1.2.1 Properties of the E2F family 21

1.2.2 E2F1 and cell cycle control 22

1.2.3 E2F1 and apoptosis 24

1.2.4 Post-translational modifications and control of E2F1 26

1.3 Arginine methylation 33

1.3.1 Mechanism of arginine methylation 33

1.3.2 PRMTs and cancer 36

1.4 RNA processing 42

1.4.1 Mechanism of splicing 43

1.4.2 Co-transcriptional splicing 45

1.4.3 Alternative splicing in cancer 49

1.4.4 Splicing factors as oncogenes and tumour suppressors 52

1.4.5 The roles of PRMTs and tudor domain containing proteins in splicing 53

1.5 Research objectives 55

Chapter 2. Materials and methods **56**

2.1 Cell lines and culture 57

2.1.2 Generation and maintenance of Tet-On cell lines 57

2.1.3 Generation of CRISPR/Cas9 knockout cell lines 57

2.2 Drug treatments 59

2.3 Transfection of plasmid DNA 59

2.4 Silencing of gene expression using silencing RNA (siRNA) 61

2.5 DNA transformation into competent cells 61

2.6 Site directed mutagenesis 62

2.7 Immunoblotting 62

2.8 Immunoprecipitation 64
2.9 Immunofluorescence 65
2.9.1 Ectopic proteins 65
2.9.2 Endogenous proteins 65
2.10 Antibodies 65
2.11 Chromatin immunoprecipitation 66
2.12 Polymerase chain reaction (PCR) 68
2.12.1 Semi-quantitative PCR 68
2.12.2 Quantitative PCR 68
2.13 Reverse-transcription 69
2.14 Flow cytometry 72
2.15 Clonogenic (colony formation) assay 72
2.14 Growth curve 73
2.15 Scratch assay 73
2.15 RNA sequencing 73
2.15.2 Library preparation 73
2.15.2 Bioinformatics 74
2.16 RNA-immunoprecipitation (RIP) 75
2.17 RIP sequencing 76
Chapter 3. Generation and characterisation of cell lines 77
3.1 Introduction 78
3.2 Generation of inducible cell lines 80
3.3 Methylation of E2F1 by PRMT1 and PRMT5 and influence on interaction partners 83
3.4 Effect of E2F1 mutants on target gene specificity 86
3.5 The effect of arginine methylation on E2F1 90
3.4 Generation of CRISPR/Cas9 knockout cell lines 96
3.4.1 Overview 96
3.4.2 Generation of stable, knockout cell lines 98
3.4.3 Characterisation of E2F1 CRISPR/Cas9 knockout cell lines 100
3.5 Chapter summary 104
Chapter 4. Identification of novel E2F1 target genes by RNA-sequencing 107
4.1 Introduction 108
4.2 Sample preparation 110
4.3 Bioinformatics analysis of differential expression 111

4.4 Validation of differential expression analysis 123
4.5 Bioinformatics analysis of exon usage 126
4.6 Validation of alternative splicing 131
4.7 Chapter Summary 139
Chapter 5. E2F1 and p100-TSN can bind to E2F1 target gene RNA as well as components of the spliceosome 141
5.1 Introduction 142
5.2 p100-TSN and E2F1 can bind to E2F1 target gene RNA 144
5.3 The E2F1-RNA interaction is dependent on p100-TSN 145
5.4 E2F1 requires an intact DNA binding domain to bind to RNA 149
5.5 E2F1 is able to associate with the splicing machinery 153
5.6 E2F1 presence alters p100-TSN specificity 157
5.7 Enhancing the p100-TSN interaction with E2F1 increases RNA binding in DuCaP cells 159
5.8 Chapter summary 163
Chapter 6. Sequencing of RNA bound in complexes with E2F1 and p100-TSN 164
6.1 Introduction 165
6.2 Both E2F1 and p100-TSN can bind multiple RNA species 167
6.3 Identification of isoform specific binding 171
6.4 Chapter summary 179
Chapter 7. Discussion 184
7.1 Summary of findings 185
7.2 The arginine methylation status of E2F1 impacts on transcriptional control 188
7.4 E2F1 can bind to target gene RNA and influence alternative splicing 196
7.5 Conclusion 204
References 206
Appendices 226
Appendix 1. Sequence alignments for CRISPR cell lines 226
Appendix 2. Hierarchical clustering of RNA-seq data 234

Acknowledgements

Firstly I would like to thank my supervisor, Professor Nick La Thangue for his time, support and guidance through this project. In addition thanks go to my mentors, Jutta Mohlenbrink and Shonagh Munro who always made time for me, answered many questions and taught me so much.

I would also like to thank many of my colleagues, in particular Amanda Coutts and Simon Carr for their help and unfailing ability to answer every question I posed and willingness to discuss various aspects of the project as it unfolded. I would also like to thank Alexander Kanapin and Anastasia Samsonova of the Bioinformatics Core at Oxford for all their hard work, help and advice with bioinformatics analysis of the RNA-seq and RIP-seq. Additionally I am very grateful to Ed Hookway who taught me how to perform library preparation for sequencing and gave up time to help. Thanks also to Fatemeh, who inspires me, as well as Mimi who gave no end of support through tough times. Many other friends have also been a great support network and have put up with me cancelling plans, or being absent, in order to finish experiments or look after my cells – thank you for understanding and sticking by me.

Eternal thanks and gratitude go to my parents for their unfailing love and support through the past four years, and the many years before that. I would not have been able to do it without you.

Finally, I would like to thank Sia, whose music is a source of inspiration and reminded me to keep going.

Abstract

(283 words)

The transcription factor E2F1 can have diverse phenotypic effects on the cell, from stimulating cell cycle progression and proliferation, to inducing apoptosis. These seemingly opposite effects are regulated by a multiplicity of post-translational modifications. Three sites of arginine methylation have been shown to cause a switch in E2F1 function, with PRMT5 mediated symmetric dimethylation at R111/113 associated with a proliferative outcome and asymmetric dimethylation at R109 by PRMT1 resulting in apoptosis. Importantly, these sites of arginine methylation by PRMT1 or PRMT5 were shown to be mutually exclusive, eliciting an either / or response. Furthermore, a reader for the R111/113Me2s mark was found to be the tudor-domain containing transcription co-activator and RNA binding protein p100-TSN. In this study we have used genome-wide sequencing techniques to explore the role of arginine methylation on E2F1 and identify how it is exerting these diverse effects. We found that, although there were few transcription level changes upon expression of an R111/113K mutant derivative compared to the wild-type expressing cells, these cells exhibited dramatic changes at the RNA isoform level across many E2F1 target genes, including increased levels of proapoptotic isoforms of p73 and BCL-X. Interestingly, the vast majority of these differentially spliced genes did not overlap with the genes showing transcription level changes, suggesting a new role for E2F1 to modulate target gene expression via RNA processing. Furthermore, we discovered that E2F1 can bind to nascent RNA via its interaction with p100-TSN and that this complex is spliceosome associated. Additionally, E2F1 can influence the RNA isoforms of p73 and BCL-X found interacting with p100-TSN. Thus, we present a model suggesting that E2F1 may be directly impacting on RNA processing by aiding assembly of splicing components as transcription occurs.

List of abbreviations

Abbreviation	Definition
ADMA	Asymmetric dimethyl arginine
AML	Acute myeloid leukaemia
CADOR	Chromatin associated domain array
cGAMP	Cyclic GMP-AMP
cGAS	Cyclic GMP-AMP synthase
ChIP	Chromatin immunoprecipitation
CLIP	Cross-linking immunoprecipitation
CRISPR	Clustered regularly interspaced short palindromic repeats
DBD	DNA binding domain
DP	Dimerisation partner
DSB	Double strand break
EMT	Epithelial-mesenchymal transition
ESE	Exonic splicing enhancers
ESS	Exonic splicing silencers
HDAC	Histone deacetylase
HLH	Helix-loop-helix
HR	Homologous recombination
IP	Immunoprecipitation
ISE	Intronic splicing enhancers
ISS	Intronic splicing silencers
Me2a	Asymmetrically dimethylated
Me2s	Symmetrically dimethylated
MMA	Monomethyl arginine

MTA	5'-deoxy-5'-methylthioadenosine
NHEJ	Non-homologous end joining
OSC	Ovarian serous adenocarcinoma
PAD	Peptidyl arginine deiminase
PAM	Protospacer adjacent motif
PCA	principal component analysis
PI	Propidium iodide
PRC2	Polycomb repressor complex 2
PRMT	Protein arginine methyl transferase
RIN	RNA integrity
RIP	RNA immunoprecipitation
RIP-seq	RNA immunoprecipitation sequencing
RNAPII	RNA polymerase II
RNA-seq	RNA sequencing
RT-PCR	Reverse transcription polymerase chain reaction
rtTA	Reverse tetracycline controlled trans-activator
SDMA	Symmetric dimethyl arginine
sgRNA	Single-guide RNA
siRNA	Small interfering RNA
SND	Staphylococcal nuclease domain
SRE	Splicing regulatory element
STING	Stimulator of interferon genes
TA	Transactivation domain
Tet-On	Tetracycline inducible
TRE	Tetracycline responsive elements

U2AF	U2 auxiliary factor

List of tables

Table number	Title	Page
1.1	Properties of human protein arginine methyl transferases	34
2.1	List of potential sgRNA targets and primers	58
2.2	List of plasmids	60
2.3	List of siRNA sequences	61
2.4	List of primers used for point mutagenesis	64
2.5	List of primers used for ChIP qPCR	67
2.6	List of RT-PCR primers	70
4.1	GSEA Gene ontology analysis results	121
4.2	List of novel E2F1 target genes identified in the RNAseq	124
4.3	List of genes showing altered exon use and their properties	133
5.1	Spliceosome related proteins identified in the E2F1 mass-spectrometry analysis	155

List of figures

Figure number	Title	Page
1.1	The hallmarks of cancer	16
1.2	Properties of E2F1	20
1.3	Post-translational modifications on E2F1	27
1.4	Arginine methylation	37
1.5	RNA processing and alternative splicing	46
3.1	Diagrams of the tTet-On system	79
3.2	Generation of Tet-On cell lines	82
3.3	Arginine methylation of E2F1 is abrogated in the mutant derivative cell lines	84
3.4	Interaction of E2F1 mutant derivatives with binding partners	85
3.5	The E2F1 mutant derivatives can bind to target gene promoters	87
3.6	Effect of E2F1 overexpression on target gene expression	89
3.7	The effect of overexpression of E2F1 and mutant derivatives on the cell cycle	91
3.8	The effect of overexpression of E2F1 and mutant derivatives on short term cell growth	93
3.9	The effect of overexpression of E2F1 and mutant derivatives on long term cell growth and clonogenic potential	95
3.10	The effect of overexpression of E2F1 and mutant derivatives on response to DNA damage	97
3.11	CRISPR/Cas9 mediated knockout of E2F1	99
3.12	Validation of E2F1 knockout by CRISPR/Cas9	101
3.13	Proliferation rate of E2F1 knockout cells	103

3.14	DNA damage response in E2F1 knockout cells	106
4.1	Diagram of library preparation protocol for RNA sequencing	109
4.2	RNA-seq sample quality control	112
4.3	RNA-seq sequencing quality control	114
4.4	Heatmap of differential expression compared to pTRE	116
4.5	E2F1 target genes identified form RNA-seq results	118
4.6	Overlap of differential expression analysis	120
4.7	Target validation of novel E2F1 target genes	122
4.8	Target validation of novel E2F1 target genes II	125
4.9	Analysis of exon use across the cell lines	128
4.10	Gene ontology analysis of splicing changes	130
4.11	Gene structure and major splice isoforms of alternatively spliced genes identified in the RNA-seq analysis	132
4.12	Alternative splicing in E2F1 mutant derivative expressing cell lines	136
4.13	Isoform expression is altered after depletion of E2F1 or p100-TSN	138
5.1	Diagram of RNA-immunoprecipitation protocol	143
5.2	E2F1 and p100-TSN can bind to E2F1 target gene RNA	146
5.3	Silencing of p100-TSN inhibits the ability of E2F1 to bind to target gene RNA	148
5.4	Silencing of p100-TSN alters E2F1 binding to chromatin of target genes	150
5.5	PRMT5 activity is required for E2F1 to bind to target gene RNA	152

5.6	Generation and validation of DNA binding domain mutants of E2F1.	154
5.7	DNA binding domain mutants of E2F1 are unable to bind to RNA	156
5.8	E2F1 interacts with snRNU4 and snRNU6	158
5.9	E2F1 alters p100-TSN binding to splice isoforms	160
5.10	Hormone depletion increases E2F1 binding to target genes in DuCAP cells.	162
6.1	Sample preparation for RIP sequencing	166
6.2	E2F1 can bind to different pools of RNA in the presence and absence of p100-TSN	168
6.3	E2F1 can impact the RNA species bound to by p100-TSN	170
6.4	E2F1 and p100-TSN can bind shared RNA species as well as separate pools of RNA species which have different cellular functions	172
6.5	Identifying exon junctions in the RIP-seq analysis enables inference of isoform binding	174
6.6	Candidate genes which show specific isoform binding to E2F1	176
6.7	Validation of the isoform specific binding to E2F1 by MECOM and SENP7	178
6.8	SENP7 is a putative E2F1 target gene	180
6.9	SENP7 expression in cancer	182
7.1	E2F1 links transcription with the splicing machinery through arginine methylation-dependent interaction with p100-TSN	187

7.2 Possible mechanism for p100-TSN recruitment of snRNP components to nascent RNA of E2F1 target genes 199

Chapter 1. Introduction

1.1 The hallmarks of cancer

Hanahan and Weinberg's seminal article from 2000 (Hanahan and Weinberg, 2000), updated in 2011 (Hanahan and Weinberg, 2011), shows that all cancers have a group of similar properties which they require to become viable tumours [Figure 1.1]. Initially these were described as: self-sufficiency in growth signals, insensitivity to anti-growth signals, evasion of apoptosis, tissue invasion and metastasis, limitless replicative potential and sustained angiogenesis (Hanahan and Weinberg, 2000). More recently, four more properties have been included: avoiding immune destruction, tumour promoting inflammation, genome instability and mutation, deregulating cellular energetics (Hanahan and Weinberg, 2011). Whilst not every tumour has all of these properties, a number of them are required for pathogenesis.

Self-sufficiency in growth signals

In order for normal cells to re-enter the cell cycle from quiescence, paracrine or endocrine growth signals must be sent and received. Cancer cells avoid this requirement by having constitutively active pro-growth signalling pathways (for example, PI3K, RAS, MEK, ERK), through either mutation of cell surface receptors (HER2, EGFR) or intracellular signalling components (RAS) resulting in autocrine stimulation. Additionally, neighbouring cancer cells or other cells in the tumour micro-environment can also produce growth factors to increase paracrine stimulation of the tumour (Hanahan and Weinberg, 2000, Hanahan and Weinberg, 2011).

Insensitivity to anti-growth signals

As cell growth is highly regulated, as well as requiring pro-growth signals for positive regulation, there is also inhibitory control which must be overcome in order for the cell to divide. These inhibitory mechanisms are frequently lost in tumours through mutation of tumour suppressor genes such as PTEN, pRB and INK4B, which usually inhibit the pro-growth pathways mentioned above. Furthermore, normal cells will stop dividing once they can contact



Figure 1.1 The hallmarks of cancer

The 10 hallmarks of cancer which contribute to tumorigenesis and progression. Adapted from (Hanahan and Weinberg, 2011).

other cells (contact inhibition) whereas cancer cells can grow into very dense tumours with no contact inhibition (Hanahan and Weinberg, 2000, Hanahan and Weinberg, 2011).

Evasion of apoptosis

Constitutive pro-growth signalling usually triggers cell death as there are multiple mechanisms in the cell which recognise inappropriate signalling – for example deregulation of pRB leads to high E2F1 which then triggers apoptosis. Additionally, in solid tumours the centre of the tumour mass is frequently hypoxic and starved of nutrients as blood flow is limited. Thus, for the cancer cell to survive under these conditions it must also be able to evade the proapoptotic signalling resulting from this. Frequently, tumours overexpress antiapoptotic signalling molecules such as BCL2, BCLX, MCL2 and display defects in proapoptotic signalling such as the p53 pathway (Hanahan and Weinberg, 2000, Hanahan and Weinberg, 2011).

Tissue invasion and metastasis

Many cancer cells gain the ability to spread from the initial tumour, into the surrounding tissue and can migrate into the blood stream. As the cells migrate back out of the blood they can form new tumours in different tissues. It is thought this is stimulated by lack of resources in the area of the primary tumour leading to the cells searching for areas which have higher nutrients and lower competition for them. A process during development and wound healing known as epithelial-mesenchymal transition (EMT) is hijacked by cancer cells to enable invasion and metastasis. Inactivation of the anchoring protein E-cadherin and upregulation of N-cadherin which helps cell migration are just two ways cancer cells can start to migrate from the primary site. Other necessary changes are release of proteins and enzymes to breakdown the extra-cellular matrix and induction of changes in surrounding, non-cancerous cells which may form positive feedback loops to drive EMT (Hanahan and Weinberg, 2000, Hanahan and Weinberg, 2011).

Limitless replication potential

Normal cells are only able to replicate a set number of times (usually quoted as 60-70) due to telomere shortening of the chromosomes on each round of DNA replication. Once the telomeres become too short, the cells senesce or undergo apoptosis. In order to overcome this and be able to divide indefinitely, cancers cells must maintain their telomere length – this is via upregulation / reactivation of telomerase which, although active in stem cells, is not expressed in differentiated cells (Hanahan and Weinberg, 2000, Hanahan and Weinberg, 2011).

Sustained angiogenesis

In solid tumours, the cells must be able to induce new capillary growth to provide oxygen and nutrients to cells in the centre of the tumour, which are distant from the existing vasculature. Thus, many tumours are able to release angiogenic signalling factors such as vascular endothelial growth factor, and reduce release of anti-angiogenic factors such as thrombospondin-1 (Hanahan and Weinberg, 2000, Hanahan and Weinberg, 2011).

Avoiding immune destruction

Immune surveillance can, and should, recognise cancer cells and trigger apoptosis in these cells, particularly during the early stages of tumour formation. Thus cancer cells which survive to become tumours have to evade this response, by either becoming less immunogenic to evade detection, for example via upregulation of PDL1, or through suppression of the immune response through secretion of inhibitory signalling molecules such as TGF-β (Hanahan and Weinberg, 2011).

Tumour promoting inflammation

Although inflammation surrounding a tumour was originally believed to aid an immune response to the cancerous cells, further studies have shown that this inflammation can actually promote tumour growth and inhibit apoptosis though availability of growth factors and

survival signalling, increase angiogenesis and increase metastatic potential as the extracellular matrix is being remodelled. Additionally, production of reactive oxygen species in the tumour microenvironment from inflammation can increase mutagenesis and genetic instability of the cells (Hanahan and Weinberg, 2011).

Genome instability and mutation

To gain many of the other cancer hallmarks, genomic mutations must keep occurring in the cancer cells. However, this is limited normally by DNA damage repair mechanisms, DNA polymerase proofreading and proapoptotic signals if the lesions cannot be repaired. Loss of repair mechanisms, or recognition and reduced apoptotic potential drive high genomic instability in cancer cells resulting in vast mutation rates which enables tumour progression (Hanahan and Weinberg, 2011).

Deregulating cellular energetics

In healthy cells, energy production is through glycolysis and the citric acid cycle, but cancer cells largely rely on glycolysis alone which requires less oxygen consumption, although this is observed even in non-hypoxic conditions. Relying on glycolysis is very inefficient as it yields 18 fold less molecules of ATP than oxidative phosphorylation, thus the cells must increase glucose uptake to increase ATP production. The reliance on glycolysis is believed to enable the products to go into other biosynthetic pathways to increase nucleoside and amino acid production to enable faster growth of the cell (Hanahan and Weinberg, 2011).

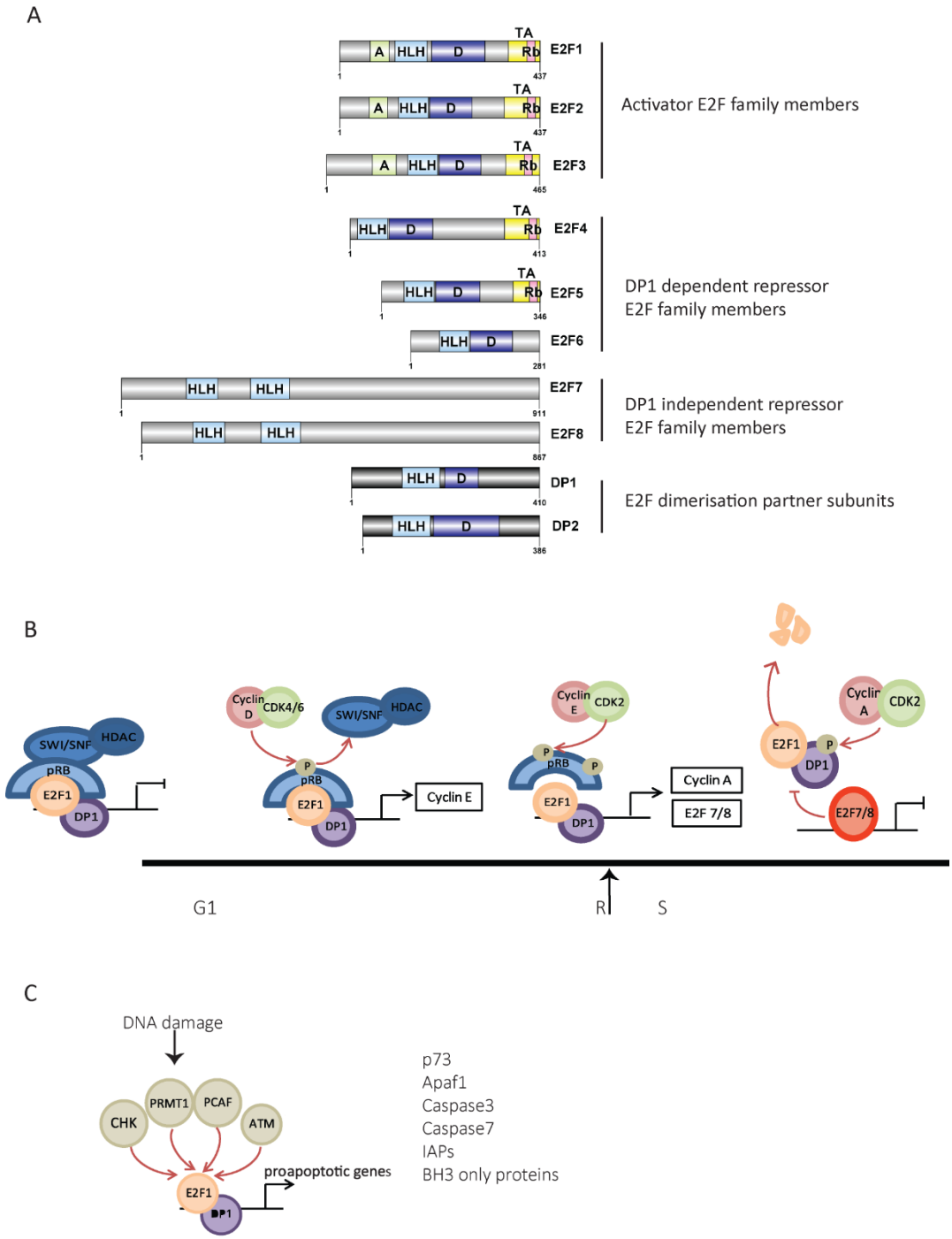


Figure 1.2 Properties of E2F1.

A. Diagram of E2F family members and DP subunits. A is cyclin A binding site, HLH is the helix-loop-helix DNA binding domain, D is the dimerisation site (DP1 binding site), TA is the transactivation domain and Rb is the pRb binding site.

B. In G1 cyclin D-CDK4/6 phosphorylate pRb, reducing transcriptional repression of E2F1 target genes. This is reinforced by cyclin E-CDK2 phosphorylation on pRb. E2F1 dependent transcription is inhibited during S phase by E2F7/8 as well as degradation following phosphorylation by cyclin A-CDK2.

C. Following DNA damage, several proteins modify E2F1 resulting in transcriptional activation of proapoptotic target genes. Adapted from (Roworth et al., 2015).

1.2 Biological functions of E2F1

1.2.1 Properties of the E2F family

The E2F family of proteins are classed as activators or repressors of transcription [Figure 1A], which regulate transcription of genes that are involved in multiple cellular processes, including cell growth and cell cycle progression (Bieda et al., 2006, Wu et al., 2001), autophagy (Polager et al., 2008), DNA repair (Ogara et al., 2013, Black et al., 2005), and apoptosis (Müller et al., 2001). E2F1, 2 and 3 are transcriptional activators whereas E2Fs 4-8 are the repressor proteins [Figure 1.2A]. There is a large degree of functional redundancy and target gene overlap observed between E2F family members, although some specificity does exist as E2F1 is arguably more involved in the checkpoint response to DNA damage and apoptotic signalling through a subset of specific target genes (DeGregori et al., 1997, Black et al., 2005, Ingram et al., 2011). E2Fs bind to specific DNA motifs in the proximal promoters of target genes with E2F1 having selectivity for TTTSSCGC, although the recent availability of ChIP-seq data suggests that these sites are perhaps not as strict as initial studies have shown, and there are multiple other potential binding sequences. However, the vast majority (80%) of E2F1 peaks occur within 1kb of the transcriptional start site (Bieda et al., 2006, Cao et al., 2011).

All E2F proteins apart from E2F7 and 8 (which have two DNA binding domains) require heterodimerisation with the Dimerisation Partner (DP) protein family to form the double helix-loop-helix capable of associating with the promoters of target genes (Wu et al., 1995). There are three DP proteins: DP1 and DP2 are activators and when dimerised with activator E2Fs result in positive regulation of transcription via recruitment of transcriptional co-activators to the transactivation domain. DP3 (also called DP4) is a repressor DP subunit and inhibits activity of the activator E2Fs to negatively regulate gene transcription (Ingram et al., 2011, Milton et al., 2006). Although many of the E2F/DP target genes overlap, there is some evidence to

suggest that that the combination of E2F and DP subunit in the heterodimer can alter the DNA binding consensus. For E2F1, the change in consensus sequence is minimal when changing between a DP1 or DP2 heterodimer; however, for E2F4 this appears to be more dramatic (Tao et al., 1997). There is little indication of DP1/2 selectivity in binding to E2F proteins (Tao et al., 1997, Wu et al., 1995).

The activator E2F/DP dimer is inhibited by pocket proteins pRB, p107 and p130. The pocket domains of the retinoblastoma family proteins bind within the transactivation domain of the E2Fs, thereby inhibiting association with transcriptional co-activators resulting in repression of gene expression. pRB can also recruit chromatin remodelers, such as histone deacetylases (HDACs), to alter local DNA structure and further inhibit transcription (Yao et al., 2008, Burke et al., 2014). The pocket proteins appear to have selectivity for various E2F members with pRB having the least specificity and binding E2Fs 1-5 and the other two RB family members being more selective for E2F4 and E2F5 (Liban et al., 2016), although the relevance of this selectivity *in vivo* remains unclear (Toki et al., 2014). Regulation of the pocket proteins is largely through phosphorylation as only the hypophosphorylated forms can bind to E2F/DP to repress transcription. Cyclin-CDK mediated phosphorylation of this family result in release of E2F/DP and transcriptional activation (Burke et al., 2014, Harbour et al., 1999).

Although it is important to realise the overlapping and complex functions of the E2F/DP and RB families, the rest of this literature review will focus on E2F1, and this should be assumed to be in the heterodimer with DP1/2 which has selectivity for pRB.

1.2.2 E2F1 and cell cycle control

Multiple feedback loops link E2F1 and the cell cycle. E2F1 is activated during G1 and is required for subsequent progression into S phase: during G1, mitogenic stimuli acting through the PI3K and/or MAPK pathways result in MYC activation, which upregulates transcription of E2F1 as well as cyclins D and E (Liu et al., 2015a, Nikolai et al., 2016). Cyclin D-CDK4/6 and cyclin E-

CDK2 successively phosphorylate pRB at several threonine (T) and serine (S) residues, resulting in loss of affinity for E2F1 (Burke et al., 2014). E2F1 is then able to activate transcription of G1/S phase cell cycle related target genes including other cyclins, CDKs and a host of DNA replication machinery [Figure 1.2B] (Roworth et al., 2015). Positive feedback loops amplify these effects via E2F1 dependent upregulation of MYC, cyclin E, CDK2 and E2F1 (Araki et al., 2003, Nikolai et al., 2016).

As cells progress into S phase, another set of E2F1 target genes is upregulated in a well-coordinated fashion. The delay between transcription of these two sets of genes is crucial for maintaining the direction of the cell cycle and to ensure the DNA is replicated only once (Roworth et al., 2015). The mechanism of the delay is not well understood, but may be due to non-canonical binding sites in the promoters, meaning that E2F1 will preferentially bind to those genes which have the classical binding site first. Other sites will only start to be read as levels of E2F1 peak (Schulze et al., 1995). Another theory is that pRB-SWI/SNF repressor complexes may form on the promoters of these so-called late genes. These have been shown to be less responsive to pRB phosphorylation status and require BRG1 phosphorylation at S phase before repression is abrogated (Zhang et al., 2000).

As cells progress into S phase E2F1 is switched off: E2F1 upregulation of cyclin A results in a negative feedback loop as the heterodimer is phosphorylated by cyclin A-CDK2, which both reduces the affinity of the dimer for DNA as well as marking it for degradation via the proteasome through SKP2 mediated ubiquitination (Schulze et al., 1995, Xu et al., 1994, Zhang and Wang, 2006). The repressor E2F7 and 8 are also upregulated during S phase resulting in further inhibition of E2F dependent transcription thus ensuring the cell cycle progresses to the next phase and all previous signalling is abolished (Logan et al., 2005). Interestingly, phosphorylated pRB has also been shown to act in a negative feedback loop to inhibit mTORC2 via a direct interaction in the cytoplasm and therefore inhibit AKT signalling which may also

serve to ensure E2F1 activity is regulated and switched off in a timely manner after G1/S transition. As this feedback loop is destroyed in $pRB^{-/-}$ cells, the pro-survival signalling of AKT may contribute to resistance to E2F1 mediated apoptosis under these conditions (Zhang et al., 2016a).

1.2.3 E2F1 and apoptosis

E2F1 also upregulates transcription of a range of proapoptotic genes after DNA damage. These include p73, caspases 3 and 7, APAF1, NOXA, BIM and SMAC/DIABLO, Staufen2 and $p14^{ARF}$ which acts to induce the p53 response [Figure 1.2C] (Phillips et al., 1999, La Sala et al., 2003, Zhao et al., 2005, Zhang et al., 2016b, Biswas and Johnson, 2012, Gogada et al., 2013, Hershko and Ginsberg, 2004, Lazzerini Denchi and Helin, 2005, Moroni et al., 2001). p73 itself is a p53 related transcription factor which shares many target genes with p53 (Irwin et al., 2000, Tophkhane et al., 2012). Indeed, in the absence of functional p53 in cancer settings, the E2F1-p73 axis becomes the main activator of the apoptotic signalling pathway through their combined upregulation of proapoptotic target genes (Laine and Westermarck, 2014, Roworth et al., 2015). A recent study has shown that in colorectal cancer inhibition of this function of E2F1 is crucial to tumour survival after DNA damage, where PLD1 expression results in inhibition of miRNA mediated downregulation of pRB after DNA damage to reduce E2F1 driven apoptotic signalling by inactivating transcription of proapoptotic target genes (Kang et al., 2017).

In brief, the intrinsic apoptotic pathway is focussed at the mitochondria. BCL2 family proteins BCL2, BCLXL and MCL1 sequester the proapoptotic proteins BAX and BAK at the mitochondrial outer membrane. BH3-only proteins (which contain only a BH3 domain), including BIM, NOXA, PUMA and BAD, are expressed after DNA damage and through this BH3 domain they compete with BAK and BAX for the hydrophobic pocket on BCL2 proteins. This results in release of BAX and BAK which oligomerise, forming pores in the outer mitochondrial membrane to facilitate

cytochrome C release. APAF1 binds to the now available cytochrome C in the cytoplasm and homo-oligomerises driving apoptosome formation and auto-activation of pro-caspase 9. Cleaved caspase 9 then initiates the caspase cascade which results in apoptosis (Roworth et al., 2015, Certo et al., 2006). Both E2F1, p73 and p53 transcriptionally regulate many of the proteins involved in this pathway, therefore after DNA damage there is a significant upregulation of BH3-only proteins which is able to overcome the balance between pro and antiapoptotic proteins at the mitochondrial membrane to ensure an "all-or-none" response (Bertin-Ciftci et al., 2013, Hershko and Ginsberg, 2004). Additionally, SMAC/DIABLO upregulation results in proteasomal degradation of the Inhibitors of Apoptosis (IAPs) which inhibit caspase activation (Cartier et al., 2011, Xie et al., 2006). Due to E2F1 and p73 (and p53) mediated upregulation of other proteins in this pathway such as APAF1 and various caspases, the signal is rapidly amplified to result in a rapid response to damage. E2F1 can also activate transcription of CHK1 which appears to give some resistance to apoptosis as the cells arrest and undergo DNA repair (Bargiela-Iparraguirre et al., 2016).

These seemingly opposing roles of E2F1 (proliferation vs. apoptosis) are highlighted in mouse models where E2F1 loss can induce tissue specific tumour formation (lung and liver amongst others), presumably due to a loss of E2F1 dependent apoptosis (Yamasaki et al., 1998, Yamasaki et al., 1996). However, overexpression of E2F1 can also induce spontaneous tumour formation (most commonly of the vagina, fore-stomach and skin epithelia), suggesting that in these tissues, this may be due to E2F1's role in cell cycle progression (Pierce et al., 1998, Pierce et al., 1999). It is therefore of great interest to elucidate the switch which turns E2F1 from a growth promoting to a death promoting protein.

It is also important to note that E2F1 has been ascribed many other functions – playing roles in autophagy, DNA repair and resistance to chemotherapeutic agents which underline the importance and complexity of the E2F1 biology (Carr et al., 2015, Roworth et al., 2015).

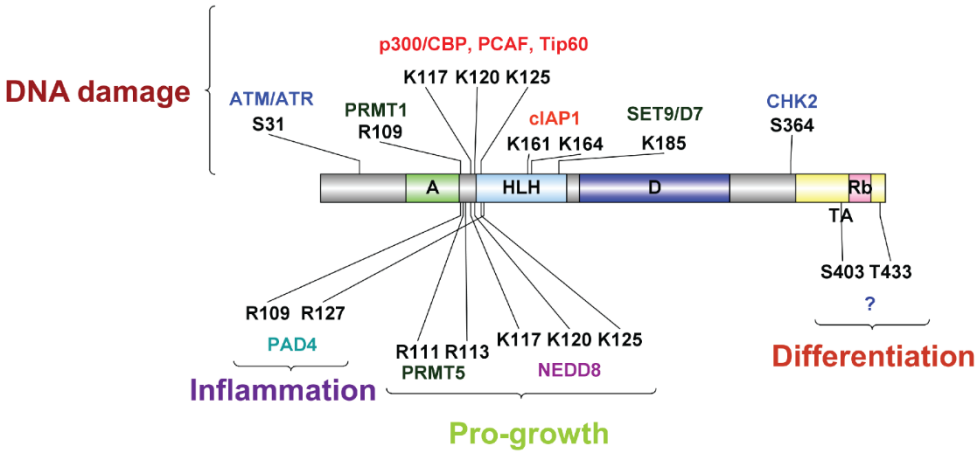
1.2.4 Post-translational modifications and control of E2F1

Post-translational modifications are commonly used as switches of protein function throughout the cell and E2F1 is no exception. An array of modifications have been described, which often interplay with one another, to result in a fine-tuned cellular response to specific stimuli [Figure 1.3A] (Carr et al., 2015, Roworth et al., 2015). Additionally, the level of E2F1 protein also plays a role in determining cell fate – with high levels triggering apoptosis whereas lower expression favours cell cycle progression. Additionally, many modifications affect E2F1 stability to influence the cell fate in response to either mitogenic, or proapoptotic signalling (Shats et al., 2017, Roworth et al., 2015).

Phosphorylation

After DNA damage, E2F1 is rapidly phosphorylated by ATM/ATR at S31 and S364 by CHK2, although these modifications result in an increased stability of the E2F1/pRB complex, under conditions of DNA damage this usually repressive complex actually activates transcription of proapoptotic target genes such as p73 and Noxa, particularly after caspase-mediated cleavage of pRB when a p68 fragment remains bound to E2F1 and stimulates transcription of these target genes [Figure 1.3A] (Stevens et al., 2003, Carnevale et al., 2012, Bertin-Ciftci et al., 2013). This adds another layer of complexity to target gene regulation by E2F1/pRB. Thus, it is likely that a synergy of cross-talk between signalling pathways combines to regulate cell fate via E2F1. AKT can also phosphorylate TopBP1 a chromatin remodelling protein which recruits SWI/SNF to repress transcription. When AKT signalling is active, phosphorylated TopBP1 interacts with E2F1 leading to repression of E2F1 driven transcription at proapoptotic target genes after DNA damage (Liu et al., 2003, Liu et al., 2004).

A



B

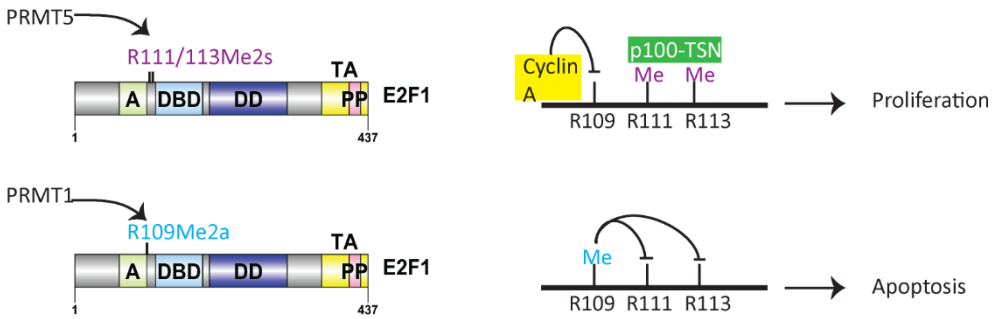


Figure 1.3 Post-translational modifications on E2F1.

A. Diagram of E2F1 showing published post-translational modifications and assigned function. Kinases are in dark blue, Acetyl transferases in red, methyl transferases in green, protein arginine deiminases in light blue and neddylases in purple

B. PRMT5 symmetrically dimethylates E2F1 at residues R111/113 (purple). These modifications are read by p100-TSN and results in cell cycle progression. Conversely, PRMT1 asymmetrically dimethylates E2F1 at R109 (blue) which inhibits SDMA on R111/113, blocking the interaction with p100-TSN, and leading to apoptosis. ADMA is inhibited by cyclin A binding which thereby increases SDMA. Adapted from (Carr et al., 2015).

In differentiated keratinocytes there has been recent evidence for E2F1 phosphorylation at S403 and T433 to drive K11 and K48 linked ubiquitination, subsequent nuclear export and E2F1 degradation during differentiation and in response to DNA damage in these cells [Figure 1.3A] (Singh and Dagnino, 2017, Ivanova et al., 2005). However, the role of phosphorylation at these sites in undifferentiated cells or indeed, other cell types, remains unexplored.

Acetylation

Following phosphorylation of E2F1 after DNA damage, subsequent acetylation at K117/120/125 by p300/CBP, PCAF and Tip60 also serve to increase E2F1 stability in addition to increasing promoter occupancy and transcription by E2F1 at proapoptotic target genes [Figure 1.3A] (Taubert et al., 2004, Nagy and Tora, 2007, Galbiati et al., 2005, Budhavarapu et al., 2012). Accordingly, deacetylation of E2F1 by HDAC1 serves to protect cells from E2F1 dependent apoptosis, this has been observed in MLL where overexpression of RNA polymerase II elongation factor, ELL, enhances recruitment of HDAC1 to E2F1 (Wu et al., 2015, Martinez-Balbas et al., 2000, Zhang et al., 2014b).

NEDDylation

E2F1 has also been shown to be NEDDylated across its lysine residues – particularly at K117,120 and 125 – but in a separate study only loss of all lysines resulted in complete loss of NEDDylation suggesting this modification may be more widespread [Figure 1.3A] (Aoki et al., 2013, Loftus et al., 2012). NEDDylation of E2F1 reduces stability of the protein and also reduces transcriptional activity. NEDDylation is reduced after DNA damage, seemingly via upregulation of the deNEDDylase SENP8 (NEDP1), and results in an increase in transcription at proapoptotic target genes and subsequent apoptosis (Aoki et al., 2013). Thus NEDDylation of E2F1 appears to be consistent with a proliferative cell state via regulation of protein level and deNEDDylated E2F1 connected with a role in DNA damage response and apoptosis. It is important to note that NEDDylation at K117, 120, 125 also inhibits acetylation at these residues, another means

of controlling cellular response, thus deNEDDylation frees these lysine residues for acetylation during the apoptotic response. Interestingly, K185 methylation has been shown to be required for subsequent NEDDylation (Loftus et al., 2012).

Ubiquitination

Interestingly, although E2F1 is stabilised after DNA damage, a few studies have shown that the protein is also more greatly ubiquitinated under these conditions and appears to be enhanced after lysine acetylation by p300/CBP (Budhavarapu et al., 2012). Recently the E3 ligase responsible for the K63 linked ubiquitin chain at K161 and K164 addition was identified as cIAP1 which appears to be critical for E2F1 binding to the promoters of target genes both before and after DNA damage, although the ubiquitination was enhanced after damage. This modification was also abrogated by silencing, or inhibition of the protein arginine methyltransferase (PRMT) 1 suggesting further cross-talk between modifications during the DNA damage response. Silencing of PRMT5 also reduced ubiquitination of E2F1, suggesting a complex role for E2F1 regulation by ubiquitination during normal cell cycle progression as well as during the DNA damage response (Glorian et al., 2017). Further ubiquitination (K11 linked) has been reported at the C-terminus of E2F1 by the E3 ligase APC/C and under normal conditions this appears to target E2F1 for proteasome mediated degradation, but again accumulation of ubiquitin chains after damage was shown to be associated with enhanced protein stability, although the function of this remains unexplored and may be a reflection of many other stabilising factors which occur at this time (Budhavarapu et al., 2012).

The E3 ligase SKP2 involved in S-phase mediated degradation of E2F1 may also play a role in protecting pRB deficient cells from apoptosis as SKP2 mediated degradation of p27 results in enhanced Cyclin A-E2F1 interaction. Cyclin A bound E2F1 has lower transcriptional ability, and therefore inhibits E2F1 dependent apoptosis in $pRB^{-/-}$ cells. It is important to note that the

authors did not look at the impact of SKP2 loss on stabilising E2F1 thereby increasing E2F1 protein levels which can also induce apoptosis (Lu et al., 2014).

Two E2F1 ubiquitin ligases have been discovered; POH1 and UCH37. POH1 acts to stabilise E2F1 protein levels by primarily removing K63-linked ubiquitin chains, although there was some evidence of K11- and K48-linked ubiquitin chain loss but as POH1 is selective for K63 the latter are likely to be due to further cross-talk rather than direct targets for deubiquitination (Wang et al., 2015a). Whilst UCH37 also removes K63-linked ubiquitin chains, in contrast to POH1 this does not affect E2F1 protein stability, but appears to activate transcription of pro-growth target genes during the cell cycle and also of proapoptotic target genes after DNA damage (Mahanic et al., 2015).

Citrullination

E2F1 has been reported to be citrullinated by PAD4, predominantly at R109 and R127 although the authors could not exclude an involvement for R111 and R113 [Figure 1.3A]. Citrullination was shown to increase the affinity of E2F1 for target genes, in particular targeting E2F1 to promoters of genes involved in the inflammatory response (a setting in which PAD4 is over-expressed) including TNFα and Il-1β. Interestingly, citrullination of E2F1 in this setting was also associated with recruitment of BRD4 to E2F1 at the promoters of inflammatory target genes, thus suggesting an interplay between citrullination and lysine acetylation as BRD4 is a bromo-domain containing protein which recognises acetylated lysine residues (Ghari et al., 2016).

Poly(ADP-ribosyl)ation

More recently, E2F1 has been shown to be poly(ADP-ribosyl)ated by PARP1. This reduces transcriptional activity by stabilising the interaction between E2F1 and the transcriptional repressor, BIN1. Loss of poly(ADP-ribosyl)ation results in checkpoint activation and G2/M arrest, thus suggesting that PARP1 may prevent inappropriate re-activation of E2F1 during the cell cycle. It also implies a new role for BIN1 in being able to repress E2F1 independently of

pRB. Additionally, the authors suggest the lack of poly(ADP-ribosyl)ation after serum starvation is a trigger for apoptosis under these conditions (Kumari et al., 2015).

Lysine Methylation

Lysine methylation at K185 is generated by SET9/D7 and is removed by LSD1. However there are conflicting reports as to the functional outcome of lysine methylation at this site [Figure 1.3A]. Initially, Kontaki and Talianidis demonstrated that K185Me was DNA damage dependent, reduced protein stability and inhibited other modifications to protect cells from E2F1 driven apoptosis. Demethylation of E2F1 reverses these effects and drives a proapoptotic transcription profile after DNA damage (Kontaki and Talianidis, 2010). Conversely, a separate study by Xie *et al* showed that after DNA damage K185Me stabilised E2F1 by inhibiting degradation, reduced DNA binding but increased recruitment to proapoptotic target gene promoters and increased apoptosis (Xie et al., 2011). In support of the first study, more recent work has shown that SET9/D7 regulates E2F1 dependent transcription after DNA damage and reduces p73 expression whilst increasing Cyclin E transcription and also gave further evidence to suggest that K185Me destabilises the protein (Lezina et al., 2014). Given these conflicting reports, and the potential interplay with NEDDylation (Loftus et al., 2012) this area requires further study.

Arginine Methylation

E2F1 can be methylated at three arginine residues – R109, R111 and R113 by two protein arginine methyltransferases (PRMTs). PRMT1 asymmetrically dimethylates R109, and PRMT5 symmetrically dimethylates R111 and R113 [Figure 1.3A]. The type and site of methylation can switch E2F1 function in the cell from pro-growth to proapoptotic. Asymmetric arginine dimethylation (Me2a) by PRMT1 at R109 is enhanced after DNA damage and increases E2F1 stability as well as enhancing protein levels of proapoptotic E2F1 target genes p73 and APAF1. Accordingly, depletion of PRMT1, or expression of a methylation deficient mutant (R109K) was

shown to protect cells from apoptosis following DNA damage, compared to wild-type expressing cells. Additionally, loss of PRMT1 resulted in reduced DNA binding activity and transcriptional activity across a panel of different target genes. No reader was identified for asymmetrically methylated R109 [Figure 1.3B] (Zheng et al., 2013).

In contrast, symmetric dimethylation (Me2s) by PRMT5 at both R111 and R113 was shown to support cell cycle progression and also increase E2F1 turnover [Figure 1.3A] (Zheng et al., 2013, Cho et al., 2012). Loss of PRMT5, or expression of methylation deficient E2F1 R111/113K reduced cell growth and sensitised the cells to apoptosis after DNA damage as well as enhancing DNA binding and transcriptional activity (Cho et al., 2012, Zheng et al., 2013).

Given the opposite functional effects these modifications have on E2F1 control of the cell, it is unsurprising that these sites of arginine methylation were shown to be mutually exclusive with either R109Me2a or R111/113Me2s observed, but never both as methylation at one site inhibited binding of the other enzyme to E2F1. Additionally, symmetric methylation by PRMT5 is enhanced after Cyclin A binding to E2F1 and accordingly, this interaction also inhibits asymmetric methylation (Zheng et al., 2013).

Furthermore, the transcriptional co-activator p100-TSN (also called SND1) was identified as a reader of the R111/113Me2s site in a chromatin associated domain (CADOR) array [Figure 1.3B] (Zheng et al., 2013, Su et al., 2015). p100-TSN is a multifunctional protein containing a single tudor domain and five staphylococcal nuclease domains (SND) which enable it to bind to multiple protein and RNA partners (Shaw et al., 2007). It acts as a transcriptional co-activator for multiple transcription factors including STAT6 (Wang et al., 2010), c-Myb (Leverson et al., 1998), Smad2/3/4 (Yu et al., 2017) and TFIIE (Tong et al., 1995) and thus plays roles in multiple cellular processes including growth and inflammation as well as regulating general transcription. In the context of E2F1, p100-TSN can bind to promoters of pro-growth E2F1 target genes in an E2F1 dependent manner, but does not bind to the promoters of

proapoptotic target genes, in line with the R111/113Me2s dependency and the role of this methylation mark on cell growth [Figure 1.3B]. Thus p100-TSN acts as a transcriptional co-activator on pro-growth E2F1 target genes via its recognition of the R111/113Me2s site generated by PRMT5 (Zheng et al., 2013).

1.3 Arginine methylation

1.3.1 Mechanism of arginine methylation

Arginine methylation is a common post-translational modification which has been found to effect many diverse cellular functions, as well as being implicated in cancer progression (Bedford and Clarke, 2009). Protein arginine methyltransferases (PRMTs) are the enzymes responsible for arginine methylation by catalysing the sequential transfer of methyl groups from S-adenosyl-l-methionine (AdoMet) to the guanidino nitrogen of arginine residues – thereby forming monomethyl arginine (MMA). Type I PRMTs (1, 2, 3, 4, 6, 8) then catalyse asymmetric dimethylarginine (ADMA) formation by adding a second methyl group to the same nitrogen atom, whereas type II PRMTs (5 and 9) form symmetric dimethylarginine (SDMA) as the other methyl group is added to the second nitrogen atom [Figure 1.4] (Bedford and Clarke, 2009, Yang and Bedford, 2013). PRMT7 is a type III enzyme capable of only creating monomethyl arginine *in vivo* (Zurita-Lopez et al., 2012). All PRMTs, except PRMT7, appear to require oligomerisation for activity, at least as dimers. However, their substrate specificity is hard to predict as many target proteins have little sequence similarity beyond an RXR / RGG / RG preference for the arginine which will be methylated. Additionally, two crystal structures identified thus far show that PRMT5 and 7 appear to interact with the peptide backbone of the substrate peptide (H4) which is a potential reason for the limited proximal sequence similarity observed between substrates. Therefore it is likely that distal sites may also play a role in substrate selectivity (Morales et al., 2016, Jahan and Davie, 2015). Regulation of PRMT activity

and substrate specificity is directed by protein-protein interactions, post- translational modifications on the PRMT or through masking modifications on the substrate which inhibit PRMT binding / activity (Yang and Bedford, 2013, Jahan and Davie, 2015). One masking modification is citrullination of arginine residues, as once citrullinated these residues are no longer substrates for PRMTs and vice-versa. Peptidyl-arginine deiminase (PAD) enzymes can target mono-methylated or un-methylated arginine residues and catalyse hydrolysis of the guanidino nitrogen atom of arginine to citrulline. This does not represent a true demethylation event as citrulline is neutrally charged and so alters the properties of the arginine residue [Figure 1.4] (Witalison et al., 2015).

Table 1.1 Properties of human protein arginine methyl transferases

PRMT	Type	Sub-cellular Localisation	Example substrates	Expression in cancer (protein atlas)
PRMT1	I	Nuclear	H3R3, H2AR11, E2F1, 53BP1, H4R3, SAM68	Over-expressed in lung, bladder, breast and colon cancers
PRMT2	I	Nuclear and cytoplasmic	H3R8	Overexpressed in breast cancer
PRMT3	I	Cytoplasmic	ribosomal protein S2, p53	Unconfirmed
PRMT4 (CARM1)	I	Nuclear	H3R17, H3R26, H3R42, AIB1	Overexpressed in breast cancer

PRMT5	II	Nuclear and cytoplasmic	H3R2, H3R8, H4R3, H2AR3, E2F1, p53	Overexpressed in B-CLL, DLBCL, lung adenocarcinomas, glioblastomas and prostate cancer
PRMT6	I	Nuclear	H3R2, H3R42, H4R3, H2AR3, H2AR29, DNA pol β	Overexpressed in bladder and lung cancers
PRMT7	III	Nuclear	H3R2, H3R17, H3R19	Overexpressed in breast cancer
PRMT8	I	Plasma membrane of brain tissue		Overexpressed in breast, ovarian and cervical cancers (anti-proliferative effect). Overexpressed in gastric cancer (tumorigenic effect).
PRMT9 (PRMT10)	II	Nuclear and cytoplasmic	SAP145	Overexpressed in non-Hodgkin's lymphoma, melanoma, pancreatic and testicular cancers
PRMT11 (putative)	?	?	?	?

Although JMJD6, was initially identified to be a potential arginine demethylase, this has failed to be reproduced in other work and is likely to be a lysine demethylase. However, more recently other JMJC domain containing proteins have been found to have arginine demethylation abilities, however, these have not yet been tested *in vivo* and the specificity of these demethylases for particular substrates is, of course, unknown (Bottger et al., 2015, Chang et al., 2007, Walport et al., 2016).

1.3.2 PRMTs and cancer

PRMT1

PRMT1, the most abundant arginine methyltransferase in cells, is associated with gene activation: asymmetric methylation of H3R4 by PRMT1 is required for subsequent acetylation at H3K9 and H3K14 which is considered an 'active' chromatin state. Inhibition of methylation represses these marks, and increases heterochromatin formation (Huang et al., 2005). PRMT1 is present both in the cytoplasm and nucleus and appears to be dynamic; for example after DNA damage by γ-irradiation or UV-A, PRMT1 has been observed to leave areas of high staining in the cytoplasm. Additionally, during apoptosis PRMT1 was seen to translocate from the cytoplasm to the nucleus (Herrmann et al., 2005, Suchankova et al., 2014, Goulet et al., 2007). PRMT1 is implicated in DNA damage response and can be recruited by p53 to target promoters to stimulate H4R3 methylation and therefore aid gene expression (An et al., 2004) as well as enhancing recruitment of methylation-dependent recruitment of MRE11 and 53BP1 to sites of DNA damage which activates γH2AX phosphorylation (Boisvert et al., 2005). Furthermore, PRMT1 can methylate BH3-only protein BAD to promote its localisation to the mitochondria, sequestration of BCL-XL, and activation of BAX and BAK (Sakamaki et al., 2011).

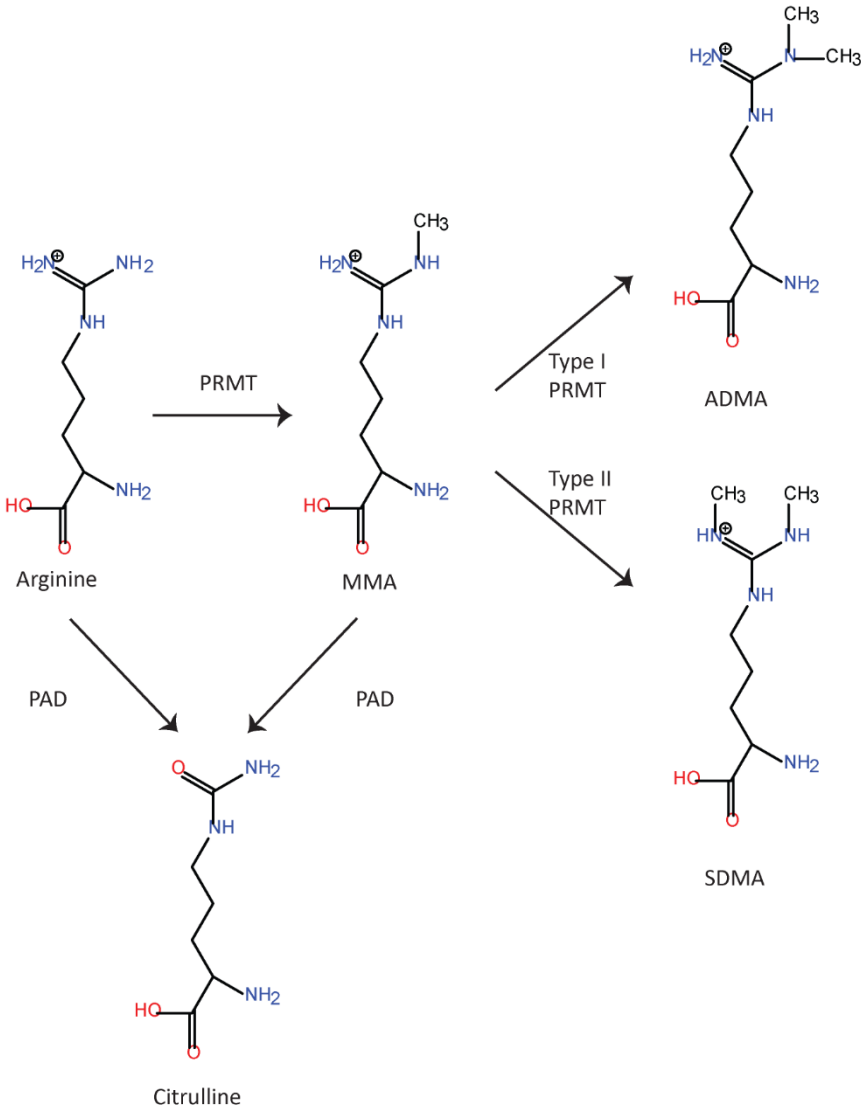


Figure 1.4 Arginine methylation.

Catalysis of arginine modifications, PRMTs catalyse conversion of arginine to dimethylarginine either processively or semi-processively via a mono-methyl intermediate. Type I PRMTs add both methyl groups to the same nitrogen atom creating asymmetric dimethylarginine,type II PRMTs add a methyl group to each nitrogen atom to create symmetric dimethylarginine. PAD enzymes citrinullate arginine or monomethylarginine by deimination to form citrulline . Adapted from (Carr et al., 2015).

Several studies have focussed on identifying the clinical significance of PRMT1 expression in tumourigenesis. ADMA levels are increased in cancer patients and PRMT1 has been found to be upregulated in lung, bladder, breast and colon cancers and is required for leukemic transformation, where it modulates the epigenetic landscape via H4R3 methylation to promoter tumour growth (Mathioudaki et al., 2008, Yoshimatsu et al., 2011, Mathioudaki et al., 2011, Cheung et al., 2016). There are 7 splice variants of PRMT1 (v1-v7) which show different substrate specificity and tissue expression (Goulet et al., 2007). Two of these, PRMT1v1 and PRMT1v2, have been found to be upregulated in breast cancer and colon cancer, where levels correlated with tumour grade (Goulet et al., 2007, Mathioudaki et al., 2008, Mathioudaki et al., 2011). Additional work has shown that PRMT1 can induce EMT in breast cancer via methylation of H4R3 at the ZEB1 promoter. ZEB1 inhibits the tumour suppressor-mediated senescence mechanism after sustained EGFR signalling (Gao et al., 2016). PRMT1 has also been found to methylate the inner centromere protein (INCENP) and is required for cytokinesis during mitosis. This methylation event increases the interaction with Aurora Kinase B required for correct chromosomal alignment and subsequent separation and has been suggested to enhance proliferation of cancer cells (Deng et al., 2015).

PRMT2

PRMT2 is overexpressed in oestrogen receptor positive breast cancer. It interacts with and acts as a transcriptional co-activator for oestrogen, progesterone, androgen and peroxisome proliferator-activated receptor-γ dependent gene transcription although so far only H3R8 has been described as a substrate (Oh et al., 2014, Zhong et al., 2014). It has also been described to bind to pRB-E2F1 to represses E2F1-dependent transcription (Yoshimoto et al., 2006). In line with this, it has been shown that loss of nuclear PRMT2 correlates with enhanced cyclin D expression, grade and progression of the tumour which may reflect the role of PRMT2 on E2F1 (Zhong et al., 2014). PRMT2 can also interact with SAM68 to modulate splicing of BCLX, increasing the antiapoptotic XL isoform, during inflammation (Vhuiyan et al., 2017). Clearly

more work needs to focus on identifying the mechanisms of PRMT2 arginine methylation and the role of this enzyme in cancers.

PRMT3

PRMT3 is negatively regulated by the tumour suppressor DAL1, which inhibits its catalytic activity. Loss of DAL1 in breast cancers would suggest PRMT3 is more active in this setting, although this has yet to be confirmed and the outcome of this has not been explored. Additionally, the function of PRMT3 mediated p53 arginine methylation is also unknown (Lai et al., 2011, Morettin et al., 2015).

PRMT4

PRMT4/CARM1 is overexpressed in aggressive breast cancers and correlates with poor prognosis. Amongst its substrates is the co-activator AIB1, which is stabilised after arginine methylation, and increased levels of AIB1 are linked with oncogenesis. PRMT4 can also act as a co-activator of c-MYB to increase proliferation (Streubel et al., 2013). PRMT4 is also implicated in enhancing WNT and FOS signalling through H3R17 methylation, and can increase stability of target gene mRNA through interaction with AU elements to further increase target gene expression (Fauquier et al., 2008, Ou et al., 2011). PRMT4 also plays a role in E2F1 regulation via methylation of pRB and by direct transcriptional upregulation of E2F1. Methylation of pRB increases subsequent phosphorylation, resulting in E2F1 release and transcriptional activation of E2F target genes (Kim et al., 2015). Methylation of H3R17 by PRMT4 has been seen at the E2F1 promoter, where it upregulates transcription of E2F1 and thereby increases cell growth in ER positive breast cancer (Frietze et al., 2008). Interestingly, both CBP and p300 are methylated by PRMT4 after DNA damage to induce cell cycle arrest (Chevillard-Briet et al., 2002) and increased PRMT4 expression also increases sensitivity to chemotherapeutic agents through methylation of MED12 and suppression of p21 mediated

senescence (Wang et al., 2015b). Thus, PRMT4 has a multiplicity of roles in the cell with a diverse set of outcomes.

PRMT5

PRMT5, the major type II PRMT in cells, has largely been associated with tumour growth promotion and survival and is overexpressed in a variety of cancers, including B-CLL and DLBCL, as well as lung adenocarcinomas, glioblastomas and prostate tumours (Gu et al., 2012a, Wang et al., 2008, Gu et al., 2012b, Yan et al., 2014, Sheng and Wang, 2016). This has been reported to be due to deregulation of miRNA suppression of PRMT5 in some cases, and may reflect a wider control mechanism which is lost in cancer cells (Wang et al., 2008). PRMT5 is both cytoplasmic and nuclear and thus it may play different roles depending on its location. Indeed, several studies have looked at location of PRMT5 and tumour grade: high levels of cytoplasmic PRMT5 has been linked with high tumour grade, and with epithelial-mesenchymal transition in lung cancers as well as being related to increased growth of prostate cancer cell lines (Ibrahim et al., 2014). Conversely, nuclear PRMT5 is observed in well differentiated (carcinoid) lung tumours, and is associated with growth inhibition of prostate cancer where it methylates the AR to increase ERG signalling (Shilo et al., 2013, Gu et al., 2012b, Mounir et al., 2016). However, expression of PRMT5 in glioblastoma is mainly nuclear where expression level was shown to correlate with grade (Yan et al., 2014).

PRMT5 requires MEP50 for maximum catalytic activity and is recruited to chromatin by BRG/BRM in SWI/SNF complexes resulting in methylation of histones H3R8 and H4R3 which can result in silencing of several tumour suppressors including ST7 and RBL2. Silencing of the latter subsequently upregulates E2F target genes. (Pal et al., 2004, Wang et al., 2008, Chung et al., 2013). Most notably, this includes EZH2, the catalytic component of polycomb repressor complex 2 (PRC2), among whose proapoptotic target genes are HRK1, CASP10 and HOXA5. Therefore, upon depletion of PRMT5 these genes were expressed resulting in apoptosis, and

increased survival of mice injected with Eμ-BRD2 lymphoma cells (Chung et al., 2013). PRMT5 has also been shown to methylate p53 at three arginine residues which results in upregulation of cell cycle arrest genes (p21) and reduced expression of apoptotic target genes (Jansson et al., 2008). Thus there was increased survival following UV-irradiation in $p53^{+/+}$ cells overexpressing PRMT5.

Due to the role of PRMT5 in cancer, there have been several PRMT5 inhibitors generated which have anti-proliferative and apoptotic effects (Prabhu et al., 2017, Jin et al., 2016, Ye et al., 2017). Moreover, loss of MTAP appears to be a marker for sensitivity to PRMT5 inhibition. MTAP is a kinase required for phosphorylation of 5'-deoxy-5'-methylthioadenosine (MTA - a by-product of polyamine synthesis) and is the rate limiting step in the methionine salvage pathway (Backlund and Smith, 1981). MTA accumulation, due to MTAP loss, can result in inhibition of several enzymes including PRMT5. Thus cells which have reduced MTAP expression also have lower basal PRMT5 activity so they are more sensitive to further inhibition of PRMT5 (Kryukov et al., 2016, Marjon et al., 2016).

PRMT6

PRMT6 is overexpressed in bladder and lung cancers and mainly acts to repress thrombospondin, p53 and p21 transcription through H3R2 methylation. PRMT6 can also directly methylate p21 at R156 which results in cytoplasmic retention of p21 (Nakakido et al., 2015). This negative regulation of p21 is likely to result in further inactivation of the p53 pathway in PRMT6 overexpressing cells (Neault et al., 2012, Stein et al., 2012).

PRMT7

PRMT7 expression is correlated with to Epithelial-to-mesenchymal transition (EMT), metastasis aggressiveness and poor survival in breast cancer. So far this has been shown to be via upregulation of MMP9 (required for extracellular matrix breakdown) (Yao et al., 2014) and transcriptional repression of E-cadherin through H4R3 methylation (Baldwin et al., 2015).

PRMT8

PRMT8 is expressed normally only in the brain but has been shown to be mutated in skin, ovarian and colon cancers although there no functional consequence for these mutations has been described so far (Forbes et al., 2008, Forbes et al., 2011). More recently, PRMT8 has been shown to be overexpressed in breast, ovarian and cervical cancers where expression was correlated with increased survival. Conversely, high expression in gastric cancers has been linked to poor prognosis (Hernandez et al., 2017).

PRMT9

The role of PRMT9 overall has not yet been well elucidated and its function, if any, in cancers remains elusive. However, according to the protein atlas it is either not expressed, or has low expression in most cancers, the exception being non-Hodgkin's lymphoma, and occasionally in melanoma, pancreatic and testicular cancers (Uhlen et al., 2015, Morales et al., 2016). PRMT9 has been reported to methylate SAP145 R508 which can then bind SMN and facilitate the splicing complex snRNPU2 maturation. Thus, loss of PRMT9 results in reduced and aberrant splicing which is common in cancers (Yang et al., 2015, Hadjikyriacou et al., 2015).

Overall, despite the role of PRMTs and cancer being well studied, there remains a lack of knowledge about many of the molecular roles these enzymes perform in carcinogenesis.

1.4 RNA processing

Many features of RNA processing become deregulated during tumourigenesis – upregulation of miRNAs which target tumour suppressors, downregulation of miRNAs which target oncogenes, changes in lncRNA expression and changes in splicing and isoform expression of mRNAs. Other methods of RNA processing are via alternative start site / use of alternative

promoters within a gene, alternative poly-adenylation sites and alternative splicing of a gene [Figure 1.5A].

Splicing is the process whereby introns are removed from pre-mRNA to leave the exon-only mature mRNA for subsequent translation. Many genes also contain alternative splicing cassettes whereby different exons (and introns) are removed or included to create more than one possible transcript from a single gene [Figure 1.5A]. This can impact on mRNA stability, translation to regulate expression, or incorporation or loss of other exons which can alter protein function via the inclusion or exclusion of domains. This provides cells with a way to regulate function of proteins with a limited number of genes and splicing plays a key role in many cellular pathways including differentiation, immune cell activation, ageing and apoptosis (Fiszbein and Kornblihtt, 2017, Gallego-Paez et al., 2017, Latorre and Harries, 2017, Lin et al., 2016).

1.4.1 Mechanism of splicing

Splicing follows a well mapped mechanism whereby the ends of the introns are defined through snRNA and RNA/protein interactions, the spliceosome is then assembled on the intron to facilitate lariat formation, the 5' splice site is cleaved from the preceding exon before 3' splice site cleavage of the intron. Finally the two exons are ligated before release of the intronic lariat and spliceosome complex which is then recycled (Chen and Manley, 2009, Bertram et al., 2017, Will and Lührmann, 2011).

In the major splicing pathway GU-AG splice sites are ligated [Figure 1.5B]. snRNAU1 binds to the 5' splice site and splicing factor 1 (SF1) to the branch point within the intron – this is called the E' complex (Wu and Maniatis, 1993). Subsequent recruitment of the U2 auxiliary factor (U2AF – a heterodimer of U2AF65 and U2AF35) to the 3' splice site and an intronic poly-pyrimidine tract forms the E complex and define the exon as well as recruiting splicing enhancers (Zuo and Maniatis, 1996, Graveley et al., 2001, Li and Blencowe, 1999, Lacadie and

Rosbash, 2005). The E complex is converted to the A complex in an ATP-dependent mechanism whereby SF1 recruits snRNP2 – an snRNA/protein complex – to the pre-mRNA via protein-protein interactions. SF1 is subsequently detached from the pre-mRNA, leaving snRNP2 in its place – this is stabilised by U2AF interactions (Crisci et al., 2015, Rutz and Seraphin, 1999). Interactions between U1 and snRNPU2 enables the intron to be excised to be defined. Further recruitment of the tri-snRNP (U4, U5, U6 containing snRNA-protein complex) marks the B complex (Nguyen et al., 2015, Liu et al., 2015b). Finally, loss of U1 and U4 from the pre-mRNA results in a catalytically active spliceosome, and lariat formation in the intron, and is referred to as the C complex [Figure 1.5B] (Galej et al., 2016, Bertram et al., 2017).

Another spliceosome complex is also used in cells – this is the U11/U12 minor spliceosome which splices AU-AC or GU-AG junctions. Both major and minor spliceosomes share common subunits, but both also have unique protein components too suggesting a slightly different method of splicing control and mechanism (Turunen et al., 2013, Will et al., 2004). In the minor splicing complex U11 takes the place of U1, U12 in place of U2 and Urp in place of SF1 (Turunen et al., 2013). Additionally, different versions of U4 and U6 are used in the minor spliceosome – U4atac and U6atac (Schneider et al., 2002). Furthermore, the assembly is also altered – U11 and U12 snRNPs are recruited as a dimer, U11 is then lost as the U4atac-U6atac-U5 is recruited. U4atac is then lost, forming the catalytically active complex (Will et al., 1999, Will et al., 2004, Turunen et al., 2013). Although an incorrect snRNU partnership is inhibitory (i.e. U1/U12), there is also evidence of interplay between major and minor spliceosomesas in the absence of a U12 site, U11 can act as an enhancer for the major spliceosome (Verbeeren et al., 2010).

Many other factors regulate this process carefully, resulting in exon preference and alternative splicing – splicing enhancers and silencers which bind to sequences in either the exon or intron. Exonic splicing enhancers (ESEs) are bound to by the serine-arginine rich (SR) protein family of splicing enhancers whereas intronic splicing enhancers (ISEs) appear to have less specificity for

a protein family and have been found to bind both hnRNPs (heterogeneous nuclear RNP) and RNA binding motif (RBM) proteins (Wang et al., 2012, Wang et al., 2007). Both exonic splicing silencers (ESSs), intronic splicing silencers (ISSs) are bound to by hnRNP family members. The combination of enhancers and silencers work in harmony to define which exons are included and which are spliced out with the intron, for example – direct competition between RBM4 and DAZAP1 for the same splicing regulatory element (SRE) – AG(G/T)T(A/T)A – in the intron results in either inhibition of splicing (exon or intron inclusion), or increased splicing (exon and intron exclusion) respectively (Wang et al., 2012). As sequencing technologies have advanced it has become possible to identify many SREs, their position relative to the exon/intron borders and splice sites as well as inferring secondary structures which may help, or hinder splicing factor binding and function (Ke et al., 2011, Goren et al., 2006, Zhang and Chasin, 2004). Additionally, this will help identify mutations which may impact on splice function and isoform expression.

These splicing factors are regulated themselves at transcriptional and post-transcriptional levels, with some splicing factors being tissue-specific, resulting in tissue-specific isoform expression – for example NOVA1 in the brain and FOX1 and FOX2 in heart, muscle and neurons. These SFs bind to specific motifs around the tissue-specific exons – UCAY (where Y is C or U) is an ISE sequence recognised by NOVA1 and UGCAUG in the downstream intron is recognised by FOX1 and 2 (Underwood et al., 2005, Brudno et al., 2001, Ule et al., 2003). Similarly, many post-translational modifications also regulate splicing factor function to control isoform selection and alternative splicing.

1.4.2 Co-transcriptional splicing

Co-transcriptional splicing was first described in (Bauren and Wieslander, 1994) for the BR1 gene where loss of intron 3 was shown to occur during transcription. Since then much research has been carried out to understand this process better. In yeast, it has been shown that splicing

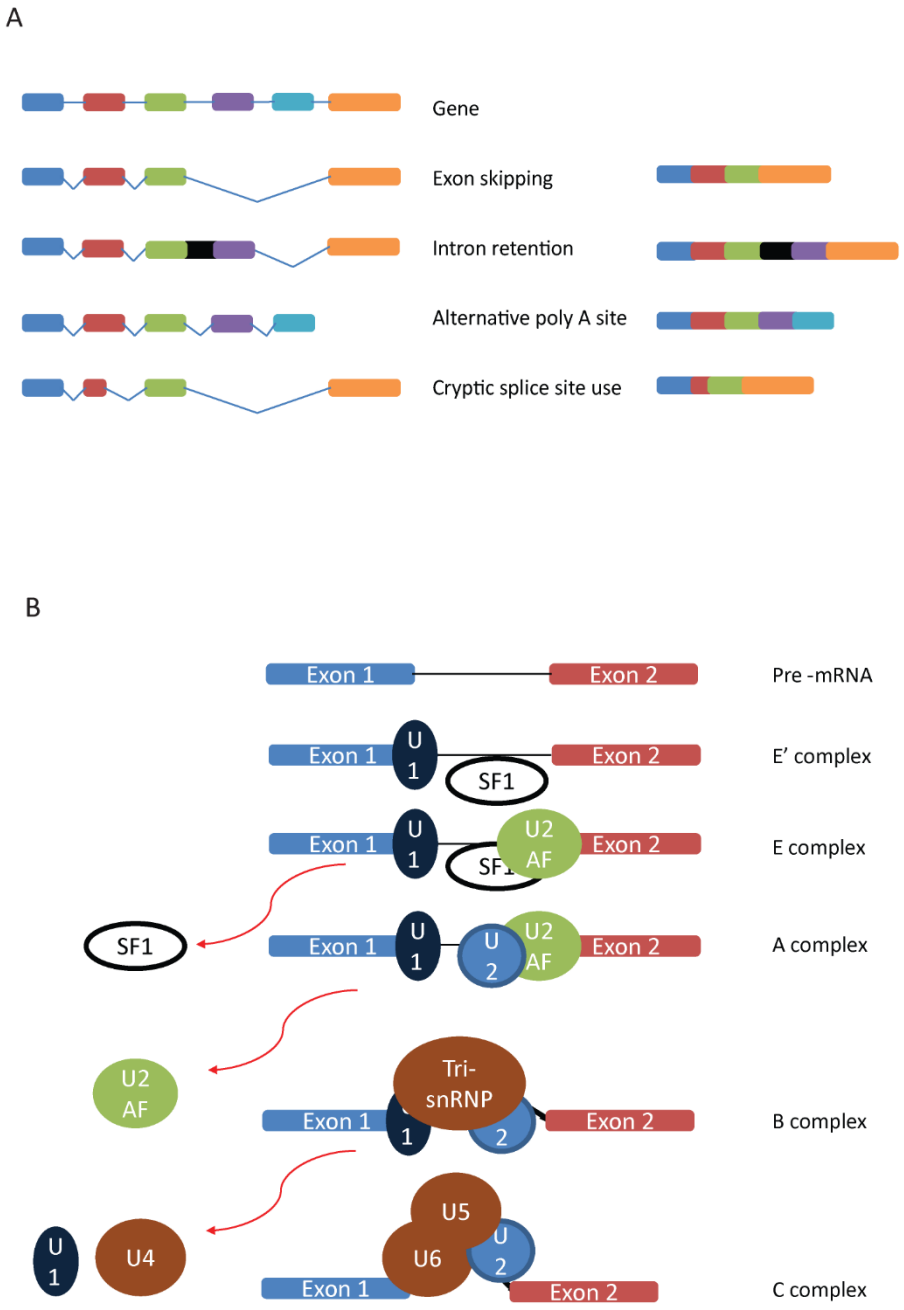


Figure 1.5 RNA processing and alternative splicing.

A. Diagram showing exon-intron structure of a gene (top) and potential RNA processing of the pre-mRNA (left) to form distinct mRNA isoforms (right).

B. Mechanism of alternative splicing via the major spliceosome. U1 and SF1 bind to introns of pre-mRNA to form the E' complex, recruitment of U2AF2 marks the E complex. SF1 is displaced by U2 snRNP in the A complex which enables recruitment of the tri-snRNP and loss of U2AF in the B complex. The mature, catalytically active spliceosome, or C complex, is formed after loss of U4 and U1 from the B complex.

is completed as close as 45bp upstream of the transcribing RNA polymerase II (RNAPII) (Carrillo Oesterreich et al., 2016). Not only is splicing occurring co-transcriptionally, but transcription can impact on splicing control as another way of controlling splicing is through speed of transcription via RNAPII elongation rates. As the majority of splicing is believed to occur co-transcriptionally – where the nascent pre-mRNA is spliced as the rest of the RNA is still being transcribed - the speed of transcription can impact on exon use. Slower transcription speeds generally enable more splicing to occur, whereas faster transcription results in increased intron inclusion, although this can also be gene specific as slower transcription also allows more silencers to accumulate on the nascent RNA. Slower transcription rates can also favour weak splice site use to create alternate isoforms, thus RNAPII mutants which have altered elongation rates or transcription factors which can increase transcription speed can impact on isoform selection, (Braberg et al., 2013, Alpert et al., 2017, Fong et al., 2014, Ardehali et al., 2009). RNAPII pausing at the 3' end of an intron as well as the final exon has also been shown to facilitate efficient splicing in yeast suggesting that co-transcriptional splicing must occur before the end of transcription for maximal efficiency (Carrillo Oesterreich et al., 2010, Alexander et al., 2010).

A major regulator of transcription speed is chromatin structure, and multiple epigenetic marks and chromatin remodelers have been associated with altering splicing. Interestingly, nucleosomes themselves have been indicated to mark spice sites, tending to be positioned centrally in the exon – and weak splice sites, and alternative exons, having less nucleosome occupancy than constitutive exons with strong splice sites (Tilgner et al., 2009, Huang et al., 2012). In addition to nucleosome position within the exon structure, several histone modifications have been identified. These nucleosomes within exons are also enriched for tri-methylation at H3K36, mono-, di- and tri-methylated H3K27, mono-methylated H4K20, H3K79me1 and H2BK5me1 (Schwartz et al., 2009, Tilgner et al., 2009, Andersson et al., 2009). H3K36me3 has been found to recruit splicing regulator PTB, which inhibits exon definition by

U2AF, and PSIP1 which can recruit other splicing factors (Luco et al., 2010, Pradeepa et al., 2012). Histone H4 acetylation has also been linked with enhanced RNAPII elongation rates, splicing factor SRp40 interaction with nascent RNA and through these effects regulating alternative splicing (Hnilicová et al., 2011). Furthermore, the histone marks H3K9me2, H3K27me3 and H3 acetylation (H3K9me2 and H3K27me3) reduce RNAPII processivity and can thus increase alternative exon inclusion (Zhou et al., 2014, Allo et al., 2009). Other chromatin associated factors can also alter RNAPII processivity, the histone chaperone FACT (facilitates chromatin transcription) enables efficient RNAPII transcription through nucleosomes by displacing one H2A-H2B dimer, presumably impacting on splicing of nascent RNA (Hsieh et al., 2013). Furthermore, several splicing factors have been shown to interact with H2A, H2A.X and H2A.Y in mass spectrometry analysis, some of which appear to be specific interactors for a particular H2A variant, although many still require further verification (Fujimoto et al., 2012). In addition to the role of epigenetic marks on splicing control, the chromatin remodelers SWI/SNF have been shown to directly impact alternative splicing as the component BRM has been found to interact with splicing machinery containing snRNAs U1 and U5, and proteins PRP8 (U5 snRNP) and SAM68 (a splicing factor involved in alternative exon inclusion) as well as nascent pre-mRNAs (Tyagi et al., 2009, Batsche et al., 2006). In an interesting possible feedback mechanism between transcription initiation, elongation and splicing decisions, nascent RNA itself has been described to impact local chromatin structure by recruiting accessory proteins (reviewed in (Skalska et al., 2017)).

Coupling of transcription and splicing is via a direct interaction between the C-terminal domain (CTD) of RNAPII and components of the splicing machinery (de la Mata and Kornblihtt, 2006). The pausing of RNAPII at introns has been linked with enhanced CTD phosphorylation which can recruit splicing factors and machinery to the nascent RNA (Alexander et al., 2010). Phosphorylation of the CTD is required for co-transcriptional splicing in Xenopus oocytes (Bird et al., 2004). Furthermore, in yeast, phosphorylated CTD has been found to have over 100

interaction partners, many of which are involved in RNA processing, including CBP2 (splicing factor), CBF5 (required for snRNA U2 processing) and snu56 (U1snRNP protein) (Phatnani et al., 2004). Additional studies in HeLa cells have shown that U2AF65 can bind to the phosphorylated CTD and recruits the essential splicing factor PRP19 complex and loss of this interaction inhibits CTD-dependent splicing activity (David et al., 2011). Furthermore, in human cells the CTD has been shown to recruit splicing factors SRp20 (de la Mata and Kornblihtt, 2006), SC35, SF2/ASF, SRp40, U2 snRNP-associated protein B (Misteli and Spector, 1999) and PSF (Rosonina et al., 2005) amongst others. This firmly links transcription with the splicing machinery.

1.4.3 Alternative splicing in cancer

Given the widespread role of splicing across many cellular pathways, and the role it plays in regulating cellular pathways, it is unsurprising that splicing aberrations are prevalent in many cancers. Splice isoforms of various oncogenes and tumour suppressors can inhibit, or enhance their respective function to drive tumourigenesis.

Splicing in oncogenic pathways

Alternative splicing in oncogenic pathways is common in cancers, MYC upregulates splicing factor SRSF1 (SF2/ASF) which is a key driver of MYC oncogenic potential, and is itself considered an oncogene as loss of SRSF1 reduces proliferation and increases apoptosis whereas overexpression can transform immortal cells. Several targets of SRSF1 have been found: MNK2 (promotes MNK2b expression), S6K1 (promotes expression of isoform 2), TEF1 (exon 5 inclusion), MGEA6, (promotes exclusion of exon 19), BIN1 (exon 12B-12A switch) and may also impact BRCA1 and MYB splicing (Das et al., 2012, Karni et al., 2007). Functional roles for some of these isoforms include: S6K isoform 2 is more oncogenic than isoform 1 and can transform cells and increase colony formation (Karni et al., 2007). SRSF1 can also activate

mTORC1 independently of AKT signalling to drive proliferation and survival (Karni et al., 2008) although the pathway driving this remains unclear.

MYC can also upregulate hnRNPs A1, A2 and H as well as PTB (Rauch et al., 2011, David et al., 2010). Upregulation of hnRNP A1, hnRNP A2 and PTB have been shown to regulate pyruvate kinase (PKM) isoform expression, favouring PKM2, normally expressed only as an embryonic form, but also in tumours where it increases cell proliferation (David et al., 2010). MYC driven upregulation of hnRNP H induces full length RAF expression, inhibiting splicing of a short isoform which lacks the ability to activate pro-growth ERK signalling (Rauch et al., 2011). MYC can also upregulate PRMT5 which plays multiple role in RNA processing, and increases recognition of weak 5' sites in DVL1 (regulates proliferation), ATR (checkpoint protein) and EP400 (chromatin associated transcriptional activator) to enhance cancer cell growth (Koh et al., 2015). Indeed, such is the importance of splicing changes in MYC overexpressing, or MYC driven cancers that inhibition of the spliceosome results in cell death (Hsu et al., 2015).

Alternative splicing can also impact cancer cell metabolism to reduce input to the TCA cycle, driving dependency on glycolysis (Kozlovski et al., 2017) and regulate angiogenesis via VEGF which has multiple splice isoforms, but use of a distal splice site in exon 8 creates B isoforms which are anti-angiogenic and shows weaker signalling than the A isoforms (Harper and Bates, 2008).

Splicing regulates apoptotic and tumour suppressor pathways

Several tumour suppressors are impacted by alternative splicing in cancers, resulting in loss of function. Many of these are part of the apoptotic pathway and are have dramatic alterations of function between isoforms (Schwerk and Schulze-Osthoff, 2005). One of the most well studied is BCLX which has two isoforms – BCL-XL, the full length mRNA, or BCL-XS which loses part of exon 2 using an internal 5' splice site creating a truncated version. Whilst BCL-XL is a BCL-2 antiapoptotic family member, BCL-XS lacks the BH1 and part of the BH2 domain and

becomes proapoptotic and thus alters the balance of splicing in BCL-XL dependent cells causes cell death (Mercatante et al., 2002). Another tumour suppressor, p73, is also alternatively spliced into a pro- and antiapoptotic form of the protein – the full length TAp73 is the proapoptotic version, but several splice isoforms have been described which alter the N-terminus of the protein resulting in truncated versions which lack the transactivation domain and therefore act in a dominant negative fashion to inhibit proapoptotic transcription at target genes (Stiewe et al., 2002). Further members of the intrinsic apoptotic pathway which undergo splicing changes are the BH3-only proteins BID and BIM. There are three isoforms of BIM – extra-long (EL), long (L) and short (S) which show different apoptotic potential – the S isoform being more apoptotic than the EL or L version as it can directly bind to BAX/BAK to activate them (Marani et al., 2002). Similarly, BID is spliced into three isoforms: EL, S, or extra-small (ES) (there is also an L isoform which is derived from an alternate start site). The EL and L isoforms are the most apoptotic, followed by ES whereas the S form is antiapoptotic as it lacks the BH3 domain (Renshaw et al., 2004). BIN1 is a protein which shows a 12B – 12A exon switch upon upregulation of SRSF1, inclusion of 12A inhibits the tumour suppressor function of BIN1 as it no longer binds to, and inhibits MYC (Karni et al., 2007).

Additionally, several caspases have been shown to be alternatively spliced: exon 9 containing caspase-2 yields a truncated protein from an early stop codon, this is antiapoptotic and antagonises cell death (Fushimi et al., 2008). Similarly, the antiapoptotic caspase-9b isoform lacks the catalytic domain through loss of exons 3-6, but can still bind to APAF1 to antagonise the proapoptotic caspase-9a activity (Vu et al., 2013). Similar mechanisms have been described for splice isoforms of caspases -8 and -10 (Himeji et al., 2002, Mühlethaler-Mottet et al., 2011). APAF1 itself is also regulated by splicing and loss of WD40 repeats in the shorter isoforms result in inability to bind to cytochrome c and the apoptosome is not formed, resulting in lack of pro-caspase-9 oligomerisation and inability to cause apoptosis (Benedict et al., 2000).

1.4.4 Splicing factors as oncogenes and tumour suppressors

Oncogenes

Upregulation of certain splicing factors has also been shown to induce tumour progression: hnRNP M induces EMT in breast cancers via altering the major expressed isoform of CD44 to a shorter version missing exon 3 and 10 which is required for mesenchymal transition and metastasis (Xu et al., 2014). Similarly, PRPF6 is often overexpressed in cancers and regulates splicing of pro-growth isoforms including the long, pro-growth isoform of ZAK, (Adler et al., 2014). In many cases, splicing factors are mutated in cancers resulting in changes in splicing regulation and global isoform changes, for example U2AF1 is frequently mutated (S34F heterozygous) in cancers resulting in altered 3' splice site recognition and altered splicing profiles across many biological pathways (Fei et al., 2016).

Tumour suppressors

Some splicing factors have been shown to act like tumour suppressors and reduce cancer cell growth, PRPF8 loss correlates with poor prognosis and splicing defects resulting from poor 3' splice site recognition – with PRPF8 mutants preferring weaker splice sites (Kurtovic-Kozaric et al., 2015). Interestingly, E2F1 has been linked with enhancing splicing of tumour-suppressive isoforms after DNA damage via upregulation of SRSF2 (SC35) resulting in enhancing proapoptotic forms of BCL-X, c-FLIP, caspase-8 and -9 (Merdzhanova et al., 2008), it is unsurprising therefore that in several cancers SRSF2 is mutated and this is correlated with poor prognosis. Mutations in SRSF2 alter the RNA binding selectivity of the splicing factor resulting in altered exon use and isoform (Zhang et al., 2015).

RBM5 is another tumour suppressor – one of its targets is Caspase-2, where it promotes splicing of the full-length, proapoptotic form by inhibiting exon 9 addition (Fushimi et al., 2008). hnRNPU binds to pre-mRNA of caspase-9 to promote splicing of the full-length, proapoptotic isoform, this is inhibited by AKT mediated phosphorylation, often seen in cancers

(Vu et al., 2013), but is stimulated in a protein phosphatase-1 dependent manner after ceramide treatment (Chalfant et al., 2002). TIA-1/TIAR is a splicing factor which is frequently down-regulated in tumours and acts as a tumour suppressor as suppression is associated with enhanced proliferation. TIA-1 recognises the 5' splice site and acts as an enhancer, one of its target RNAs is the death ligand FAS, as well as caspase-7, TNFα and also oncogenes c-MYC, WT1 and FGFR (Sanchez-Jimenez et al., 2015, Izquierdo et al., 2005, Lopez de Silanes et al., 2005).

1.4.5 The roles of PRMTs and tudor domain containing proteins in splicing

Multiple roles for arginine methylation have been observed in regulating splicing through methylation of splicing factors. PRMT5 mediated methylation of multiple Sm proteins, SmB, SmB', SmD1 and SmD3 enables complex formation with the Tudor-domain containing protein, survival of motor neuron (SMN), and subsequently U snRNA to aid assembly of the core snRNP particles (Meister et al., 2001). Loss of PRMT5 causes splicing defects in multiple RNAs including MDM4 resulting in p53 activation and apoptosis, therefore PRMT5 is required for regulating normal cellular splicing (Bezzi et al., 2013). Indeed, MYC can upregulate, and increase nuclear localisation of PRMT5, in cancers to enhance constitutive splicing, and this upregulation is required for cell viability (Koh et al., 2015, Park et al., 2015). PRMT1 methylates multiple hnRNP proteins including Q, A2, R and K and in the case of the first two, this regulates the subcellular localisation to the nucleus, presumably to enable it to play a role in splicing (Wada et al., 2002, Passos et al., 2006, Nichols et al., 2000).

The readers of arginine methylation are Tudor domain containing proteins (Gayatri and Bedford, 2014) and p100-TSN has been found to bind to E2F1Me2s (Zheng et al., 2013). As well as being a transcriptional co-activator, p100-TSN has multiple roles in splicing, it interacts with SmB, SmD1 and SmD3 to act in a similar manner to SMN and assemble the core snRNP (Gao et al., 2012). It also interacts with U1, U2, U4, U5, U6 snRNPs, in particular PRP8 – a component

of the U5 snRNP/ U4, U5, U6 tri-snRNP complex – to increase the efficiency of spliceosome assembly to the pre-mRNA (Yang et al., 2007). Additionally, p100-TSN has been shown to interact with splicing factor SAM68 to regulate exon inclusion in CD44 and increase proliferation via recruitment of the tri-snRNP complex to exon 5 (Cappellari et al., 2014). A separate study showed that SAM68 is regulated by ERK signalling to regulate CD44 splicing (Matter et al., 2002) and can interact with the BRM subunit of SWI/SNF to facilitate co-transcriptional splicing, favouring exon inclusion (Batsche et al., 2006). Furthermore, SAM68 has been shown to regulate BCL-X splicing for the XS isoform and therefore trigger apoptosis by recruiting the repressor hnRNPA1 to block SRSF1 recognition of the alternative 5'SS. This can be inhibited by FYN mediated phosphorylation which abrogates the RNA binding ability of SAM68 (Paronetto et al., 2007).

1.5 Research objectives

Previous work in our lab discovered a switch mechanism for E2F1 to move from a pro-growth to a proapoptotic phase and vice versa, arginine methylation on R111/113 was associated with E2F1 presence on pro-growth target gene promoters and increased cell growth. Conversely, R109 asymmetric dimethylation was observed after DNA damage and corresponded with E2F1 presence on proapoptotic target gene promoters. This work also identified the co-transcriptional activator p100-TSN as a reader for the R111/113Me2s mark. As this initial work focussed on well-known E2F1 target genes and biological outcomes we wanted to investigate the role of arginine methylation on E2F1 in greater depth – focussing on identifying new pathways which may be involved in the E2F1 response and identifying novel target genes which may be impacted by arginine methylation. Furthermore, given the roles of PRMT5 and p100-TSN in alternative splicing, we also wished to determine if E2F1 may impact on RNA processing through recruitment of these proteins to the transcriptional start site. Although there has been much work done to identify methods of co-transcriptional splicing, in particular with RNAPII, to our knowledge no work thus far has looked at the impact of transcription factors on RNA processing decisions of their target genes.

By creating cell lines which could be induced to overexpress conservative R to K mutant derivatives of E2F1 which cannot be arginine methylated we were able to explore the role of arginine methylation of E2F1 at a genome-wide level. Not only did we find potential target genes of E2F1 which showed differences in expression between the methylation-dead mutant derivatives, suggesting they are transcriptionally regulated by arginine methylation of E2F1, but we also discovered that arginine methylation of E2F1 may play a role in regulating RNA processing of its target genes. Furthermore, we have identified a direct link between E2F1, target gene RNA and components of the spliceosome which is dependent on PRMT5 mediated arginine methylation and p100-TSN.

Chapter 2. Materials and methods

2.1 Cell lines and culture

HeLa (cervical), HCT116 (colon) and U2OS (osteosarcoma) were grown in Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplemented with 5% (v/v) foetal calf serum (FCS, Sigma) and 1% (v/v) Penicillin-Streptomycin (Pen-Strep, Gibco). SAOS2 cells (osteosarcoma) and DuCAP cells (prostate) were maintained in DMEM supplemented with 10% (v/v) FCS and 1% (v/v) Pen-Strep. Tet-On inducible cells were maintained in DMEM supplemented with 5% tetracycline negative FCS (Sigma), 1% Pen-Strep, 100µg/ml G418 (Clontech) and 150µg/ml hygromycin (Clontech).

All cells were grown at 37°C in 5% CO_2 and passaged regularly at 80-90% confluency. Sterile phosphate buffered saline (PBS, Oxoid) was used to wash cells before detachment with Trypsin-EDTA (Lonza).

2.1.2 Generation and maintenance of Tet-On cell lines

Constructs containing HA-E2F1 WT, HA-E2F1 R109K, HA-E2F1 R111/113K (KK) and HA-E2F1 R109/111/113K (KKK) were subcloned into pTRE2-hyg vectors and confirmed by sequencing. These constructs, and a vector-only control (pTRE) were transfected into G418-resistant Tet-On U2OS, 48 hours post transfection the selection process was started and cells were from this point cultured as described above. Selection was considered complete when all non-transfected control G418-resistant Tet-On U2OS had undergone apoptosis and the transfected cells were growing well. All inducible cells used were maintained in the same medium as described above and expression of the protein was induced by addition of 1µg/ml doxycycline for 24 hours (unless stated otherwise in figures). Cell passage followed the same principles as described in the section above.

2.1.3 Generation of CRISPR/Cas9 knockout cell lines

The protocol outlined in (Ran et al., 2013) was used. Briefly, guide RNA (sgRNA) to direct Cas9 to the target gene using the design tool available at http://tools.genome-engineering.org.

Forward sgRNA: 5'CACCGGAAACTGACCATCAGTACC, reverse sgRNA: 5'AAACGGTACTGATGGTCAGTTTCC. These sgRNA strands were annealed and phosphorylated using T4 PNK (New England Biolabs), then cloned into the pSpCas9(BB)-2A-Puro plasmid (Addgene) using *BbsI* (Thermo Fisher) and T4 Ligase (New England Biolabs). The plasmid was then transformed into competent cells (XL-Blue, Agilent) as described below and transformed cells underwent selection on ampicillin. Colonies were grown up and the plasmid was sequenced to confirm successful incorporation of the sgRNA.

Table 2.1 List of potential sgRNA targets and primers for sequencing these

Gene name	**Potential cut site (hg19 assembly)**	**Primers (5'-3') for genomic sequencing**
E2F1	Chr20:-32268185	F: GACTCTCAGGCTATGTCCCC R: GCATTCTTCTTCTGGCTGGG
DNML1	chr12:-32897088	F: TGTGAGTTGGCAAGTTCCTCA R: CAGCTCCTCCAGTACTTTCATACT
NROB1	chrX:-30326348	F: AAGTACTTGCCCTGCTTCCAG R: TCCCATGAAGGCACAAACTTTTC
MACROD2	chr20:-16021906	F: CAAGCAACTCAACCCTCTGT R: TCTTAGCCCACACCAAAGGT

Cells were transfected with the plasmid as described below and after 24h 2µg puromycin was added. This was maintained for 48h, when the un-transfected control cells had undergone apoptotsis suggesting selection was complete. Cells were then maintained in DMEM

supplemented with 5% (v/v) FCS and 1% PenStrep. Clones were isolated and grown up before identifying those which had successful E2F1 knockout via SDS-PAGE and immunofluorescence. Those clones then had their genomic DNA extracted (QIAamp DNA Mini Kit, Qiagen) and the top 3 genes identified by the genome engineering tool as most likely to have off-target nicks were sequenced to confirm these were not affected (see table 2.1). Primers were designed to flank the cut site by 250bp in either direction and sequences are in table 2.1.

2.2 Drug treatments

1µg/ml doxycycline was added for 24h to induce expression of tet-responsive E2F1 in the Tet-On cell system.

Etoposide (Etp), JS579-44, JS579-68 and 5-florouridine (5FU) stocks were made in DMSO, doxorubicin was dissolved in water and cisplatin in 150mM NaCl. All solvent-only controls were done with the appropriate solute. Final concentrations of drugs were as follows Etoposide was used at 20µM for 48h, doxorubicin at 2µM for 24h, JS579-44 and JS579-68 at 5µM for 72h.

For hormone depletion experiments cells were seeded and left to settle for 48h, washed twice in PBS and the media replenished – either normal media containing 10% FBS (control) and 1% (v/v) Pen-Strep, or phenol-red free DMEM supplemented with 10% (v/v) charcoal-stripped FBS (Gibco) and 1% (v/v) Pen-Strep (hormone depletion).

2.3 Transfection of plasmid DNA

Cells were transfected with 1µg plasmid DNA using the GeneJuice system (Millipore). 2µl GeneJuice per microgram DNA and Optimem (Gibco) were incubated at room temperature for five minutes. Separately, Optimem and plasmid DNA were incubated for five minutes at room

temperature. The two mixes are combined 1:1 (vol : vol) by gentle pipetting and incubated for a further 5 minutes before dropwise addition of the transfection mix onto dishes pre-seeded with cells. For transient transfections, cells were harvested 48h post-transfection. See table 2.2 for a list of plasmids used.

Table 2.2 List of plasmids used in this study

Plasmid	Tag
pTRE WT-E2F1	HA
pTRE R109K-E2F1	HA
pTRE KK-E2F1	HA
pTRE KKK-E2F1	HA
pTRE empty vector	N/A
pcDNA3.1 WT E2F1	HA
pcDNA3.1 L132E E2F1	HA
pcDNA3.1 R166H E2F1	HA
pcDNA 3.1 empty vector	N/A
pSpCas9(BB)-2A-Puro E2F1 sgRNA	N/A
pSpCas9(BB)-2A-Puro vector	N/A

2.4 Silencing of gene expression using silencing RNA (siRNA)

Pre-warmed Optimem (Gibco) was incubated at room temperature with Oligofectamine (Invitrogen) for five minutes, during which time the siRNA was diluted in a separate volume of Optimem. After five minutes, the siRNA and the Oligofectamine mixes were combined before further 15 minute incubation. The transfection mix was added dropwise to dishes pre-seeded with cells, with serum-free Optimem. The media was topped up with serum (to 5% v/v) the following day and plates were harvested 72h post-transfection. Unless stated otherwise, the total concentration of siRNA was 25nM per dish, for double knockdown 12.5nM of each individual siRNA (or topped up with non-targeting, if appropriate) was used so as not to exceed a final concentration of 25nM.

Table 2.3 List of siRNA sequences used

siRNA	Sequence (sense strand) 5'-3'
siGFP (NT)	AGCUGACCCUGAAGUUCUU
siE2F1	CUCCUCGCAGAUCGUCAUCUU
siTSN	AAGGAGCGAUCUGCUAGCUACdTdT
siPRMT1	GGACAUGACAUCCAAAGAUU
siPRMT5	CCGCUAUUGCACCUUGGAAUU

2.5 DNA transformation into competent cells

XL Blue cells (Agilent) were thawed on ice and mixed with 1µg plasmid DNA. This was incubated on ice for 5 minutes. Heatshock was carried out at 42°C for 45 seconds before returning the mixture to ice for a 5 minute recovery. Luria Broth (Sigma-Aldrich) was added to cells were

transferred to 37°C for 1 hour before plating on LB agar plates containing appropriate antibiotic for selection.

Linear DNA was purified using PCR clean up or gel extraction kit as required (Qiagen) as per manufacturer's instructions. Plasmid DNA was purified using Miniprep kit, (Qiagen) or Maxiprep kit (Invitrogen) as required.

2.6 Site directed mutagenesis

Primers were designed using QuikChange Primer design tool available at http://www.genomics.agilent.com/primerDesignProgram.jsp and the QuikChange site-directed mutagenesis kit (Agilent) used as per manufacturer's instructions. Plasmid DNA, primers, dNTPs, buffer and *Pfu turbo* (Agilent) were mixed and underwent PCR in a thermo cycler. *DpnI* was then added to digest the template plasmid to retain only the new plasmid containing the mutation. The plasmid was transformed into super-competent cells (XL Blue, Agilent) for subsequent DNA isolation via mini-or maxi-prep. Incorporation of the mutation was confirmed by sequencing.

2.7 Immunoblotting

Cells were harvested and pelleted at 1000 revolutions per minute (rpm) for 5 minutes at 4°C, the pellet was washed once in ice-cold PBS before lysis in RIPA buffer (50mM Tris-HCl pH8, 150mM NaCl, 1% IGEPAL, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulphate (SDS), 0.2mM sodium orthovanadate and protease inhibitor cocktails (PI [1M PMSF, 1µg/ml leupeptin, 1µg/ml aprotinin, 1µg/ml pepstatin A]) and incubated on ice for 30minutes. The lysate was then centrifuged at 13000rpm (table top centrifuge) for 10 minutes at 4°C to remove cellular debris. Protein concentration was determined via Bradford assay (Bio-Rad).

Normalised samples were mixed with SDS sample buffer (62.5mM Tris-HCl pH6.8, 2% SDS, 8% glycerol, 0.1% bromophenol blue, 300mM β-mercaptoethanol) and boiled at 95°C for 5 minutes to denature the proteins.

Samples were separated by size using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Acrylamide gels were constituted acrylamide (variable percentage), 375mM Tris-HCl pH8.8, 0.1% SDS, 0.15% ammonium persulphate (APS), 0.07% tetramethylethylenediamine (TEMED); stacking gel: 5% acrylamide,125mM Tris-HCl pH6.8, 0.1% SDS, 0.15% APS, 0.28% TEMED, and carried out in Laemmli running buffer (25mM Tris, 192mM glycine, 0.1% SDS) at 200V until fully resolved. ColorPlus Prestained Protein ladder (New England Biolabs) run alongside samples to estimate molecular weight. Proteins were transferred onto nitrocellulose or PVDF membranes (latter pre-rinsed in methanol for 2 minutes) (Whatman) using the Bio-Rad wet blot system: towbin transfer buffer (25mM Tris, 192mM glycine, 10% methanol) was used at 400mA for 2 hours or 150mA overnight. Membranes were stained with Ponceau S (Sigma) to confirm successful transfer before removal of the stain and blocking in 5% milk in PBS-Tween 0.1%, or 5% BSA in TBS-Tween 0.1% for one hour at room temperature. Membranes were incubated with antibody on a rocker either overnight at 4°C or for 2 hours at room temperature. The membranes were washed three times in PBS-Tween 0.1%, or TBS-Tween 0.1%, as appropriate, before incubation with secondary horse radish peroxidise (HRP)-conjugated antibody for 1 hour at room temperature. A further 5 washes were carried out before addition of chemiluminescent substrate (2.5mM luminol, 396µl coumaric acid, 100mM Tris-HCl pH8, 0.02% (v/v) hydrogen peroxide). Signals were captured by exposing X-ray films (Kodak) to the membranes before developing on the AGFA CP1000 automatic developer.

Table 2.4 List of primers used for point mutagenesis used in this study

Primer name	Forward primer sequence 5'-3'	Reverse primer sequence 5'-3'
E2F1 L132E	CGCTATGAGACCTCAGAGAATCTGACCAC CAAG	CTTGGTGGTCAGATTCTCTGAGGTCTCATAG CG
E2F1 R166H	AGGGTGCAGAAGCGGCACATCTATGACA TGACC	GGTCATGTCATAGATGTGCCGCTTCTGCACC CT

2.8 Immunoprecipitation

Cells were harvested as above and lysed in modified RIPA buffer (50mM Tris-HCl pH7.4, 150mM NaCl, 1% IGEPAL, 1mM EDTA, 1mM NaF, 0.2mM sodium orthovanadate and protease inhibitor cocktails [1M PMSF, 1µg/ml leupeptin, 1µg/ml aprotinin, 1µg/ml pepstatin A]) and incubated on ice for 50 minutes. The lysate was then centrifuged at 13000rpm (table top centrifuge) for 10 minutes at 4°C to remove cellular debris. Protein concentration was determined via Bradford assay and normalised samples were adjusted to 600µl before addition of 1µg specific antibody or IgG control and 30µl TrueBlot beads (Rockland Inc.) for 1-2h. Bead-bound proteins were washed 5 times in modified RIPA buffer then resuspended in SDS sample buffer and boiled for 95°C for 5 minutes before resolution via SDS-PAGE.

2.9 Immunofluorescence

2.9.1 Ectopic proteins

Cells were seeded onto glass coverslips and treated as described in the figures. They were fixed in 4% paraformaldehyde for 5 minutes before being washed. Cells were permeablised with 0.5% Triton X-100 in PBS for 5 minutes and washed again before incubation with primary antibody for 1 hour in a humid chamber. Coverslips were washed 5 times before incubation with Alexa Fluor fluorescence conjugated secondary antibody (Invitrogen) for 1 hour in the dark in a humid chamber. Coverslips were washed again before fixation to glass slides using Vectashield mounting medium with DAPI (4',6-diamidino-2-phenylindole). Proteins were visualised by BX60 transmitted-reflective light fluorescence microscope (Olympus) and analysed with Openlab software (Improvision).

2.9.2 Endogenous proteins

Cells were seeded as above and fixed with 2% paraformaldehyde for 5 minutes, washed once in PBS before a further methanol fixation at -20°C for 20 minutes. Coverslips were rinsed three times in PBS then incubated for five minutes in PBS-Tween (0.05%). A 20 minute blocking step was carried out in 1% BSA in PBS-Triton X-100 (0.5%). Primary antibodies were incubated as before, secondary antibodies were incubated for 30 minutes to reduce background staining; all wash steps and slide fixation were carried out as described above.

2.10 Antibodies

E2F1 (KH95 and C-20), DP-1(K20), p53 (DO1), BCLX (H5) all from Santa Cruz; β-Actin, FLAG peptide monoclonal antibody M2, anti-HA coupled agarose beads, anti-Flag peptide monoclonal antibody M2 coupled agarose (Sigma). Antibodies against PRMT1 and H4R3me2s were purchased from Abcam. Anti-PRMT5, anti-phospho-Histone H2A.X (Ser139) (Millipore),

anti-HA1.1 (Covance). anti-cleaved-caspase 3, anti-cleaved-PARP, anti-asymmetric dimethyl-arginine (adma-R), anti-symmetric dimethyl-arginine (sdma-R) were from Cell Signalling Technologies. Anti-Rb from BD Pharmingen, mouse and rabbit secondary antibodies were from GE Healthcare and anti-p100TSN (SND1) from Bethyl Laboratories.

2.11 Chromatin immunoprecipitation

Proteins were cross-linked to DNA using 1% formaldehyde for 10 minutes. The reaction was stopped by adding glycine to a final concentration of 125mM for a further 10 minutes. The media was removed and cells washed before harvesting in lysis buffer I (5mM Tris-HCl pH8.0, 85mM KCl, 0.5% Igepal) for 20 mintues, then the nuclei lysed for a further 10 minutes in lysis buffer II (50mM Tris-HCl pH8.0, 10mM EDTA, 1% SDS). A Bioruptor Pico (Diagenode) was used to sonicate samples for 10 cycles of 30 seconds on / 30 seconds off at 4°C. After sonication samples were diluted in ChIP dilution buffer (20mM Tris-HCl pH8.0, 1mM EDTA, 0.01% SDS, 150mM NaCl, 1% Triton X-100) and spun at 13000 rpm for 10 minutes to remove debris.

Protein concentration was measured and equal amounts (1-2mg) were made up to 1ml and pre-cleared for 2h at 4°C with IgG and pre-blocked (1mg/ml BSA and 400µg/ml salmon sperm DNA) protein A/G agarose beads. The lysates were spun down at 1000rpm for 5 minutes and the supernatant was then transferred to new microcentrifuge tubes and incubated with either IgG or specific antibody overnight at 4°C. Pre-blocked agarose A/G beads were added for 2 hours before washing. Endogenous ChIPs were washed twice in wash buffer I (20mM Tris-HCl pH8.0, 150mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton-X100), four time in wash buffer III (10mM Tris-HCl pH8.0, 250mM LiCl, 1mM EDTA, 1% Igepal, 1% DOC) and twice in TE; ectopic pull-downs were washed twice in wash buffer I, twice in wash buffer II (20mM Tris-HCl pH8.0, 500mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton-X100) then twice in wash buffer III and finally twice in TE buffer.

Protein-DNA complexes were eluted from the beads beads using 142µl elution buffer (1%SDS, 0.1M $NaHCO_3$) at 65°C for 15 minutes and the eluates transferred to a fresh tube and a second elution step carried out on the remaining beads. 250mM NaCl (final concentration) was added and reverse cross-linking carried out at 65°C overnight. 20µg Proteinase K (Fisher) was added and incubated for 1 hour at 55°C before the DNA was purified using the QIAquick PCR Purification Kit (Qiagen) as per manufacturer's instructions.

ChIP primers are listed in table 2.5

Table 2.5 List of primers used for ChIP qPCR in this study

ChIP Primer	Forward primer sequence 5'-3'	Reverse primer sequence 5'-3'
E2F1	TAAAGCCAATAGGAACCGC	AGGGCTCGATCCCGCTCCG
Albumin	TGGGGTTGACAGAAGAGAAAAGC	TACATTGACAAGGTCTTGTGGAG
APAF1	CTTGGCCAGGCTGGTCTTGAAT	GCAGCCATTCAAAATTATGACACAT
CDC6	GGCCTCACAGCGACTCTAAGA	CTCGGACTCACCACAAGC
Cyclin A	CTGCTCAGTTTCCTTTGGTTTACC	CAAAGACGCCCAGAGATGCAG
p73	TGAGCCATGAAGATGTGCGAG	GCTGCTTATGGTCTGATGCTTATG
BCLX	ACAAGGCGATGCTTCTTCTAT	TTCATTCAGTCAGTCAGTCACTC
DEF6	GGGAGCCACGTGGATACAG	GAGCTGGGACTTGGAGAC
ETV1	CGTCACTTTAAGGAGCCAG	GGGATACTTGGGCCTTTG
FGF4	GGTGCAGGACCAGGCCTC	GAAGCTCCTAGTGGGCGCAG

LRRC4	CTCCGGCTTTGTGCGGAG	CTCACTCCGGCTCGCTC
KCNIP	GATTGAGCAGCTACCCTGTG	GCATCCACATGCATCTGAGGC
NTSR4	CCCTTCCTAGCCGGTGCG	GAGCTGTTGAGGCGCATG
SAGE1	CGGCAGCACCTCACAAAATG	GCCTGTCAGCTCCAAATTCAC

2.12 Polymerase chain reaction (PCR)

2.12.1 Semi-quantitative PCR

cDNA/ chromatin was amplified using Paq5000 (Agilent) according to manufacturer's instructions. The reaction was run as follows: denaturation at 95°C for 5 minutes, 95°C for 3 seconds, 61°C annealing temperature for 1 min and 72°C for 1 min to a total of 35 cycles, and a final 72°C elongation for 5 minutes. PCR products were visualised on 2% agarose gel with 0.5µg/ml ethidium bromide. Primers are listed in table 3.

2.12.2 Quantitative PCR

cDNA/ chromatin was amplified and measured using the Brilliant III Ultr-fast Sybr Green master mix (Agilent) according to manufacturer's protocol on a MX3005P QPCR system (Stratagene). Analysis was performed using the ΔΔCT method where raw CT values were normalised to the input (for ChIP) or the control treatment (RT-PCR) followed by normalisation to IgG (for ChIP) or the control gene (RT-PCR). These were converted into fold enrichment. The entire equation is as follows:

$\Delta CT = CT(\text{treatment or specific IP}) - CT(\text{control or input})$

$\Delta\Delta CT = \Delta CT\ (\text{target gene or specific IP}) - \Delta CT\ (\text{control gene or IgG IP})$

$\text{Fold enrichment} = 2^{-\Delta\Delta CT}$

Results shown are means of three independent experiments, run in triplicate with standard error shown. Significance was determined using the student's t-test.

In cases where a single representative experiment is shown (indicated in figure legend) the standard deviation is shown. This is calculated as follows:

Standard deviation (s) of ΔCT:

$s = (s1^2 + s2^2)^{1/2}$; where s1= standard deviation of CT for sample, s2= standard deviation of CT for actin control

Conversion to fold change:

$2^{-\Delta\Delta Ct}$ with ΔΔCT + s and ΔΔCT – s; Where the minimum and maximum fold change (according to the standard deviation range) is calculated. The error bars reflect these values.

Minimum fold change = $2^{-(\Delta\Delta Ct + s)}$

Maximum fold change =$2^{-(\Delta\Delta Ct - s)}$

2.13 Reverse-transcription

Cells were harvested and homogenised in TRIzol Reagent (Thermo Fisher), RNA extraction was continued as per manufacturer's protocol using RNA-grade glycogen as a carrier. RNA levels were measured on a Nanodrop and equal amounts were used for subsequent cDNA synthesis. *DNAseI* (Roche) was added to extracted RNA and and RNase Inhibitor (RNasIn, Promega) added before incubation at 37°C for twenty minutes. EDTA was added to stop the reaction, and the DNAse was inactivated by heating to 75°C for ten minutes. The RNA was then incubated with oligoDT (Invitrogen) or random hexamers (Invitrogen) for 5 minutes at 70°C before addition of dNTPs, MMLV reverse transcriptase and buffer (Promega). Thermo cycling was carried out as per manufacturer's instructions for MMLV-RT.

Table 2.6 List of RT-PCR primers used in this study

RT-PCR Primer	Forward primer sequence 5'-3'	Reverse primer sequence 5'-3'
E2F1	ATGAGACCTCACTGAATCTGACCACC	AGTCACAGTCGAAGAGGTCTCTG
GAPDH	CCATCAATGACCCCTTCATTGACC	GAAGGCCATGCCAGTGAGCTTCC
Actin	GGGATGTTTGCTCCAACCAA	TGGCGCTTTTGACTCAGGAT
CDC6	GCCACAGCTGTTGAACTTCC	ATAGCTCTCCTGCAAACATCCA
Cyclin A	CCAGTGCACTTGCCAGTTGTT	GCTGCTGCTGGAAGACGAAATC
Total p73	CTTCAACGAAGGACAGTCTGCTC	CAGGATGGTGGTGAATTCCGT
TAp73	CCTCTGGAGCTCTCTGGAACC	CAGATTGAACTGGGCCATGACAG
DNp73	CTCGCCACGGCCCAGTTC	CGAAGGTGGAGCTGGGTTG
DN'p73	CATCTGTCATGGGATTCAGCC	GAGGCTCCGCAGCTAGTGAG
DNex2p73	CGGCTGCAGGGAACCAGAC	GATGTAGTCATGCCCTCCAGG
DNex23p73	CGGCTGCAGGCCCAGTTC	CGAAGGTGGAGCTGGGTTG
Cyclin D1	CCAGAGTGATCAAGTGTGAC	GGGACATCACCCTCACTTAC
APAF1	TGAGCCCACTCAACAGCAAA	TGTCCTTACACTGGAAGAAGAGA
Total BCLX	CACAGCAGCAGTTTGGATGC	ATATGCTGTCCCTGGGGTGA
BCLXL	TAAACTGGGGTCGCATTGTG	AGGTAAGTGGCCATCCAAGC
BCLXS	GCAGTAAAGCAAGCGGTGAG	GTTCCACAAAAGTATCCTGTTCAAAG

DEF6	GAGCCTGGAGATAAGCGTCC	CCCCACCATTGAGGGACTTC
ETV1	CATACCAACGGCGAGGATCA	GCCACCTCTTCAGGCTCAAT
FGF4	TCTACTGCAACGTGGGCATC	TGAAGAAGGGCGAGCCATAG
LRRC4	TCCTTCCGAGATGGAAAGAGG	ATGGCTGCACACAGAATCCA
KCNIP	GATGAGCTGGAGATGACCATGG	CCACTGGGGCACTCATTTTTG
NTSR4	GTCCTACGTGCAGTGGTCAT	AGTTGGCAGAGACGAGGTTG
snRNU1	GGCAGGGGAGATACCATGATC	CCCCCACTACCACAAATTATGC
snRNU2	CTGATACGTCCTCTATCCGAGGAC	CCAGGTCGATGCGTGGAG
snRNU4	CTTTGCGCAGTGGCAGTATC	CAGTGCCGACTATATTGCAAGTC
snRNU5	ATACTCTGGTTTCTCTTCAGATCGC	CTTGCCAAAGCAAGGCCTC
snRNU6	CTCGCTTCGGCAGCACATATACT	ACGCTTCACGAATTTGCGTGTC
snRNU11	GCTTCTGTCGTGAGTGGCAC	CAACGATCACCAGCTGCCC
snRNU12	CGCCCGAATCCTCACTGC	GCAGATCGCAACTCCCAGG
Total SAGE1	CACCATGGCCTGATGCTACC	GATGGAATACCAGCTGCAGCC
SAGE1 201	CACCATGGCCTGATGCTACC	GACCAGGTTGACAAGACGTTATCAG
SAGE1 203	CACCATGGCCTGATGCTACC	CAGTTGACAAGTCATAACTAGGTTGTG
Total RBM38	GCTGACCCCGCACTACATC	GAGGCGGCGTATGGGTACTG
RBM38 V1	CTACGGCTTCGTGACCATGG	GGTCAGCCCGTAAGTCCG

RBM38 V3	GAGAGCCCCAGGTGACCAT	GGTCAGCCCGTAAGTCCG
RBM38 V4	CTACGGCTTCGTGACCATGG	GAGCCTCCAGACGGGCTGAC
Total SENP7	CAGCCTGAAGACCTTAACAG	GGAAATCAGAGTGAGTTCCAC
SENP7 V5	CACCAGAAAGGGGCTCAC	CCACTGTTAAGGTCTTCAGGC
SENP7 V3	CATCACCAGAAAGGATACCCAGAG	GCTCTGATTGCAAGTTGTCTG
SENP7 E5	CACCAGAAAGGCAACTCAAAG	GGCAATGAGTCTGATTGCACC
Total MECOM	GATATCCACGAAGAACGGC	GCTGAGTGAGGAGTACTGC
MECOM V4	CTGGCCACAAATAATGAGTGTG	GCATGGCTCTTGAATATTGAGG
MECOM V7	CTGGCCACAAATCTTAGACG	GCATGCAACAAGGTTGTGC
MECOM 2b	GGCCATAGAAATCTAAAGATCTTAG	GCATGCAACAAGGTTGTGC

2.14 Flow cytometry

Cells were harvested via trypsinisation and the pellet washed once in PBS before fixing in ice-cold 75% ethanol overnight at 4°C. Cells were pelleted and washed once in PBS before resuspension in PBS containing 125U/ml RNaseA (Qiagen) and 40µg/ml propidium iodide (Sigma) and incubated at 4°C in the dark for 20 minutes. Samples were run and analysed on a Flow Cytometer (BD C6 Accuri).

2.15 Clonogenic (colony formation) assay

Cells were seeded in 6 well plates in triplicate (number of cells/ well described in results for each experiment) and left to settle overnight. For inducible cells, doxycycline was added the

following morning to induce protein expression and was topped up every 72h over the 10 day period. After 10 days, cells were washed twice in PBS before methanol fixation for 20 minutes on ice. After removal of methanol, 0.5% crystal violet stain was added to each well for 10 minutes at room temperature. The colonies were washed thoroughly in water and left to dry before counting.

2.16 Growth curve

$1x10^5$ cells were seeded on day 0 in triplicate, if required doxycycline was added 8h later. Cells were harvested by trypsinisation and a counted using a flow cytometer (BD C6 Accuri). Population doubling time was worked out according to the formula: (DT) = 2 logn/log (Tn/Ti), where Ti is initial cell number, Tn is final cell number and n is the time (days) between Ti and Tn. Proliferation rate as a percentage of control were then calculated from the equation: Proliferation rate = doubling time (control) X 100/doubling time.

2.17 Scratch assay

Cells were allowed to come to confluency after induction of E2F1 with doxycycline before a scratch was made across the monolayer. The scratch was measured at time 0 and again after 8h to assess the wound healing potential.

2.18 RNA sequencing

2.18.2 Library preparation

RNA was isolated and DNase treated as described previously. PolyA selection and library preparation of mRNA was carried out using NEBNext Poly(A) mRNA Magnetic Isolation Module (E7490) and NEBNext Ultra Directional RNA Library Prep Kit for Illumina (E7420) as per

manufacturer's instructions. Briefly, 1µg RNA was mixed with magnetic Oligo(d)T beads and non-bound species were removed by washing. The mRNA was eluted from the beads and fragmented at 95°C for 15 minutes followed by first strand synthesis and subsequently second strand synthesis. The cDNA was purified using SPRI beads (Fisher, see below). The ends of the cDNA were repaired before adaptor ligation and amplification of adaptor-ligated DNA. The library was then purified before final quality control on a Bioanalyser. The library was sequenced on an Illumina NextSeq and bioinformatics was completed by Alexander Kanapin of the Bioinfomatics Core at Department of Oncology, University of Oxford.

SPRI beads (Sera-mag SpeedBeads, Fisher) were washed twice in 0.5M EDTA then resuspended in buffer containing 20% PEG (Sigma), 2.5M NaCl, 10mM Tris-HCl pH8.0, 1mM EDTA and stored at 4°C. They were allowed to come to room temperature for at least 30 minutes before use.

2.18.2 Bioinformatics

Quality control checks for FASTQ reads were done with QC FastQC software package. The reads were aligned with the human reference genome hg19 build using STAR short reads aligner version 2.5.1b. The read counts for RefSeq gene models were calculated with htseq-count version 0.6.1p1**.**

Differential expression analysis was performed with DESeq2 Bioconductor package (version 1.12.4) using adjusted p-value of 0.01 as a threshold for significantly differentially expressed genes. Differential exon usage was assessed by DEXSeq Bioconductor package (version 1.1.8.4) and results were filtered for those with an adjusted p value of <0.01. The TRANSFAC database release 2015.3 was used to identify E2F1 binding motif consensus and promoter sequences which carried them. Gene set enrichment analysis was done by GSEA tool by the Broad Institute with an FDR of 25%, those with a q<0.05 were considered significant. Additional GO analysis was done using DAVID 6.7 with BP_GO_FAT terms, KEGG pathway and OMIM disease (Huang et al., 2009a, Huang et al., 2009b).

2.19 RNA-immunoprecipitation (RIP)

Cells were washed once with PBS before UV cross-linking at 900mJ/cm^2 using a Stratalinker (Stratagene). Dishes were places immediately on ice and RIP lysis buffer (50mM Tris-HCl pH8.0, 150mM NaCl, 1mM $MgCl_2$, 10% glycerol, 1% NP-40, 1mM DTT, 0.2mM sodium orthovanadate and PI cocktails) added directly to the plate. The lysate was collected and lysis continued at 4°C on a rotor for ten minutes before centrifugation at 13,000 rpm for ten minutes. Protein concentration was calculated using Bradford Reagent, as described previously. For protein, 5% input was taken, and the samples boiled in SDS-sample buffer. For RNA, 10% input was taken and 1µg proteinase K added for thirty minutes at 37°C before addition of TRIzol.

Protein A/G agarose beads were washed four times in RIP lysis buffer. 30µl of beads, 1µg of non-specific IgG and 0.1mg/ml heparin-sulphate were added to the lysate to pre-clear for one hour at 4°C on a rotor. Following a centrifugation at 0.5g for two minutes, the supernatant was transferred to a fresh tube and 1µg of non-specific IgG, or specific antibody was added. Binding was allowed to occur for one hour at 4°C whilst tumbling, before addition of 30µl of beads for a further hour. The beads were washed four times in RIP lysis buffer and resuspended in 400µl RIP lysis buffer this was divided into two – for protein the beads were dried and resuspended in SDS sample buffer before boiling. For RNA extraction, an equal amount of RIP extraction buffer (350mM NaCl, 10mM Tris-HCl pH7.5, 10mM EDTA, 0.1% SDS, 7M urea) was added, along with 15µg proteinase K and incubated at 37°C for thirty minutes before addition of Trizol. Protocol adapted from (Niranjanakumari et al., 2002).

TRIzol extraction proceeded as described previously. Following resuspension of RNA, RNAseIN and DNaseI was added to each sample and incubated at 37°C for thirty minutes. EDTA was added to stop the reaction, and the DNase was inactivated by heating to 75°C for ten minutes. cDNA synthesis was carried out using random hexamers, as described in section 2.13.

2.20 RIP sequencing

RIP was carried out as described above. Following RNA extraction and DNAse treatment equal volumes of sample were taken and underwent ribodepletion using GeneRead rRNA Depletion Kit (Qiagen). Libraries were prepared as before using NEBNext Ultra Directional RNA Library Prep Kit for Illumina. The library was sequenced on an Illumina NextSeq and bioinformatics was completed by Alexander Kanapin of the Bioinfomatics Core at Department of Oncology, University of Oxford. FASTQ files for two biological replicates in each condition were trimmed as described above. The reads were aligned to the human genome build hg19 by gsnap aligner v. 2017-04-21 with two mismatches allowed. The RIPSeq analysis was performed with RIPSeeeker R package v. 1.18.0 with the parameters as follows: uniqueHit=TRUE, assignMultihits=TRUE, rerunWithDisambiguatedMultihits=TRUE and automatic bin size selection. EnsEMBL bioMart build 75 was used for functional annotation of the RIPSeq results.

Chapter 3. Generation and characterisation of cell lines

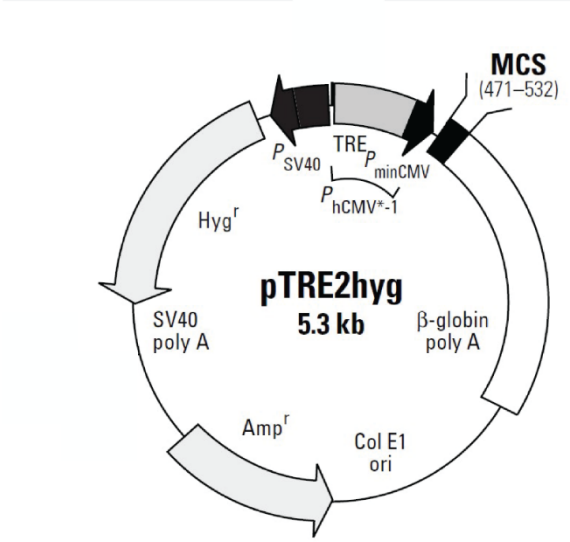
3.1 Introduction

Previous work in our lab has established that arginine methylation of E2F1 by PRMT1 and PRMT5 results in diverse cellular outcomes. Crucially, these methylation sites were mapped to R109Me2a and R111/113Me2s, with PRMT1 mediated asymmetric dimethylation of R109 being observed after DNA damage and resulting in E2F1 recruitment to proapoptotic target gene promoters and subsequent cell death. Conversely, symmetric dimethylation of residues R111 and R113 by PRMT5 was apparent in cells under normal growth conditions and was associated with E2F1 presence at cell cycle regulator target genes, and cell cycle progression (Cho et al., 2012, Zheng et al., 2013).

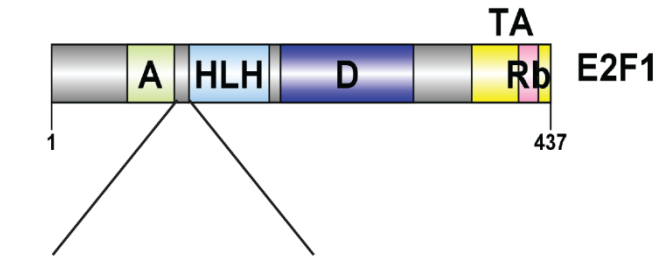
These initial studies focussed on known E2F1 target genes and cellular outcomes, so to further investigate the impact of these modification sites on E2F1 and their role in global cellular function the tetracycline-inducible (Tet-On) system was used to generate stable cell lines expressing E2F1 wild-type (WT), the point mutant R109K, the double mutant R111/113K (KK), and the triple mutant R109/111/113K (KKK). These R-K mutations inhibit PRMT mediated arginine methylation at the indicated amino acids, and although lysine has a smaller side chain, they are both similarly charged, meaning there should be minimal impact on protein folding around these sites.

The parental U2OS Tet-On cell line stably expresses the regulatory plasmid. This plasmid contains both a G418 resistance gene to enable selection of those cells which include this protein, and the reverse tetracycline controlled trans-activator (rtTA). rtTA is a fusion of the E.coli Tet-repressor protein with the Herpes Simplex virus VP16 activation domain and is constitutively expressed. This yields a transcription factor which is sensitive to the presence of tetracycline, and activates transcription at tetracycline responsive elements (TRE) only after exposure to tetracycline.

A



B



107- PARGRGRHPGKG - 115 WT

107- PA**K**GRGRHPGKG - 115 R109K

107- PARG**K**G**K**HPGKG - 115 R111/113K (KK)

107- PA**K**G**K**G**K**HPGKG - 115 R109/111/113K (KKK)

Figure 3.1 Diagrams of Tet-on system.

A. Plasmid map of pTRE2hyg showing tet-response elements, the multiple-cloning site (MCS) and hygromycin resistance gene.

B. Diagram of E2F1 showing domains: A – cyclin A binding, HLH – DNA binding domain, D – dimerisation domain, TA – transactivation domain, Rb – pRB binding site. Below is the position and sequence of the wild type (WT) and relevant mutations shown in green.

The response plasmid used here is the pTRE2hyg vector [Figure 3.1A] which contains the TRE in the $P_{hCMV*-1}$ promoter sequence upstream of the cloning site. This allows expression of the cloned protein only after addition of tetracycline when the rtTA can bind to the TRE and stimulate gene expression. The response plasmid also contains a hygromycin resistance gene enabling efficient selection and maintenance of cells after transfection, by adding hygromycin to the media. It is also worth noting, that although still regarded as a Tet-On system, the tetracycline derivative doxycycline is also able to activate this system, and as it has a longer half-life (48 hours vs 24 hours) and is not light sensitive it is more routinely used.

CRISPR/Cas9 is a relatively new genome editing tool. The use of guide RNA, complementary to the gene of interest, adjoined to Cas9 nuclease results in specific localisation of the enzyme to the gene. Cas9 is a nuclease which creates double-strand breaks (DSB) 3bp upstream of a 5'-NGG (PAM) site. Subsequent repair of the DSB by the cell is through non-homolgous end-joining – an error prone DNA repair mechanism which results in indel and frameshift mutations. Ultimately this makes the gene unviable and the cell can no longer make protein from this gene creating a knockout cell line as the RNA becomes non-functional and subsequently undergoes nonsense mediated decay. It is also possible to create specific gene-level mutations by providing a repair template for the cell to use in homology-directed recombination after the DSB. Therefore, this gives an advantage over short term silencing RNA (siRNA) as the genome knockout is complete, whereas no siRNA is efficient enough to completely silence protein expression. A CRISPR/Cas9 knockout E2F1 model in U2OS cell was also generated to validate some of the observations made from the over-expression system.

3.2 Generation of inducible cell lines

Hemagglutinin (HA)-tagged E2F1 had previously been mutated using site-directed mutagenesis to create the point mutant R109K, the double mutant R111/113K (KK) and the triple mutant

R109/111/113K (KKK) [Figure 3.1B]. By designing primers to these which contained additional BamHI and NotI at the 5' and 3' ends respectively it was possible to subclone these, and the wild-type (WT) into the pTRE2hyg vector. Once the sequences had been confirmed, the HA-E2F1.pTRE E2F1 mutant derivatives along with a pTRE vector-only control (pTRE) were transfected into parental (rtTA containing) U2OS cells, along with an untransfected control. After 48h, hygromycin selection was started (in addition to the G418 selection) and cells were maintained in this medium for all subsequent passages. Selection had been completed after 2 weeks, when untransfected control cells were no longer viable and the transfected cells started to expand. Doxycycline was added and after 24h the cells were harvested to assess expression of the HA-E2F1 proteins via immunoblot [Figure 3.2A]. All cell lines expressed HA-E2F1 after addition of doxycycline, except for the vector only control. Additionally, all cell lines appeared to express relatively equal amounts of HA-E2F1 protein [Figure 3.2A]. To confirm this further, and to ensure the protein was localised in the nucleus, immunofluorescence was carried out. Cells were plated on coverslips and protein expression induced with doxycycline or vehicle control for 24h, before fixation with formaldehyde and methanol, anti-HA was used to recognise ectopic E2F1 and fluorescent-conjugated secondary antibody was used to visualise the protein. Coverslips were mounted onto slides with mounting medium containing DAPI stain to visualise the nucleus [Figure 3.2B]. Individual cells appear to express differing amounts of HA-E2F1, which is expected of a polyclonal pool of cells which may all have different transfection efficiencies, but again, these are dependent on doxycycline addition with little leakage being observed. Additionally, for all cell lines, HA-E2F1 was present in the nucleus as expected.

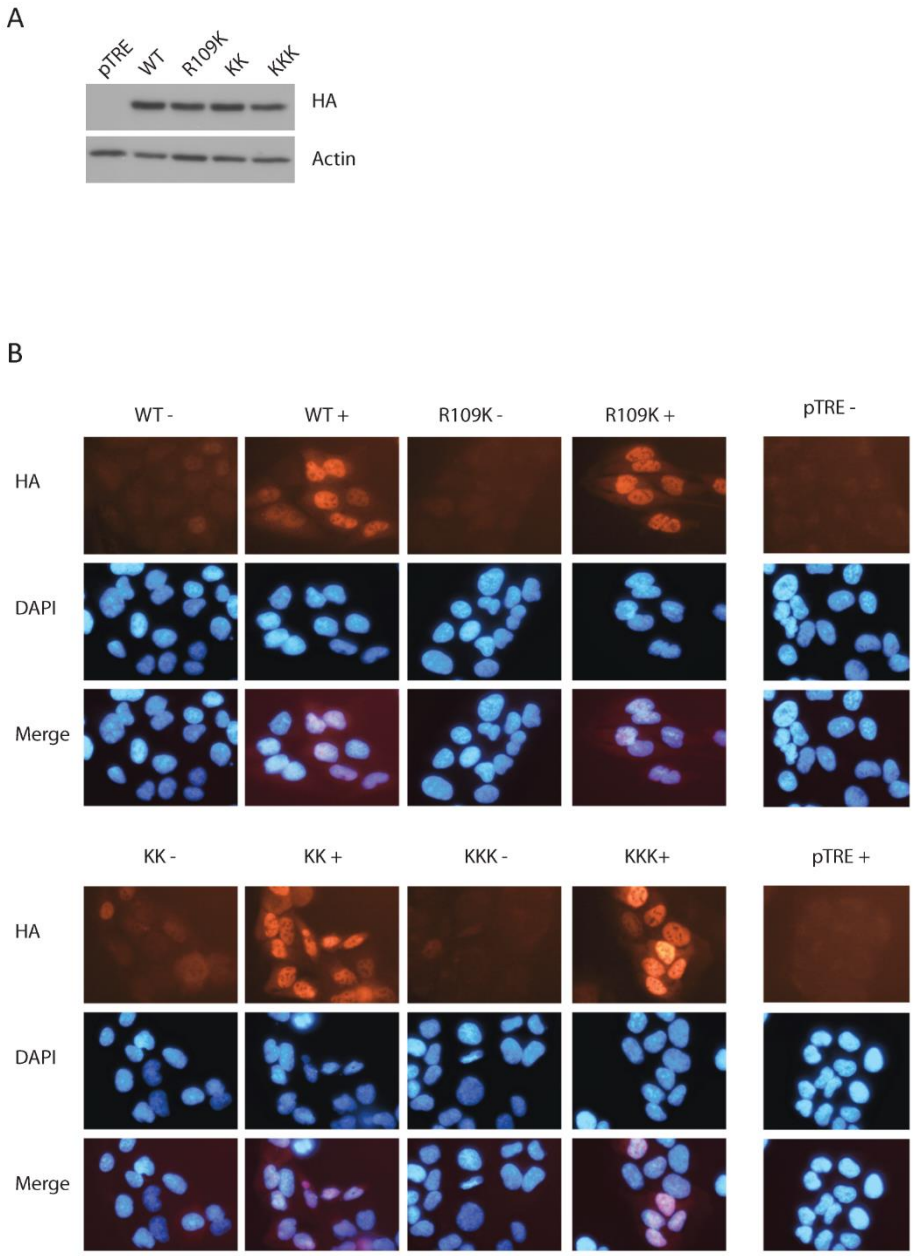


Figure 3.2 Generation of Tet-On cell lines.

Parental U2OS cells containing rtTA were transfected with 5µg HA-tagged pTRE-E2F1 WT, or the point mutants R109K, R111/113K (KK) or R109/111/113K (KKK) , or vector only control (pTRE) and grown under hygromycin selection until non-transfected cells had undergone apoptosis. Expression was induced with 1µg/ml doxycycline for 24 hours before harvest.

A. Immunoblot showing induction of expression of ectopic E2F1 using anti-HA and anti- β-actin as a loading control

B. Cells were plated on coverslips and induced with doxycycline (+), or vehicle control (-), for 24 hours before fixation with PFA and methanol. Coverslips were incubated with anti-HA and fluorescent secondary antibody before mounting with mounting medium containing DAPI stain and visualised on a fluorescent microscope.

3.3 Methylation of E2F1 by PRMT1 and PRMT5 and influence on interaction partners

Having established inducible cell lines expressing the E2F1 WT and mutants it was then important to confirm that they were able to be methylated *in vivo*. As no specific antibody for site specific methyl-E2F1 exists, proteins were isolated via immunoprecipitation using anti-HA agarose before immunoblotting with pan-specific anti-methylarginine antibodies. Anti-RMe2s is specific for symmetric dimethyl arginine, whereas Anti-RMe2A is specific for asymmetric dimethyl arginine. After visualisation of the methyl arginine, blots were stripped and reprobed with anti-HA to ensure the methyl arginine bands overlapped with the anti-HA signal, indicating arginine methylation of the ectopic protein. As expected, HA-E2F1 WT and R109K were symmetrically dimethylated, but crucially symmetric methylation was inhibited by the double R111/113K mutant and the KKK mutant [Figure 3.3A]. Unfortunately, due to high unspecific background staining from the anti-RMe2A antibody, and despite multiple rounds of optimisation, we were unable to conclusively show asymmetric methylation of WT and KK HA-E2F1. However, further phenotypic tests on the cell lines were conclusive with the published data and demonstrated that R109K is protected against PRMT1 mediated methylation and conferred resistance to apoptosis (as described below), suggesting that the mutant derivatives lines are asymmetrically methylated appropriately.

The ability of the mutants to bind to known E2F1 interaction partners was investigated [Figure 3.4]. Anti-HA agarose was used to immunoprecipitate ectopically expressed E2F1 or the mutant derivatives. Immunoblotting was carried out to identify that known binding partners of E2F1 could be isolated with the ectopic protein. As expected there was no change in the ability of the mutants to bind to the pRB, the RGRGR motif is distant to the pRB binding domain so it would be unlikely to impact on this interaction [Figure 3.1]. There was also no difference in the ability of the mutant derivatives to bind E2F1's heterodimerisation partner

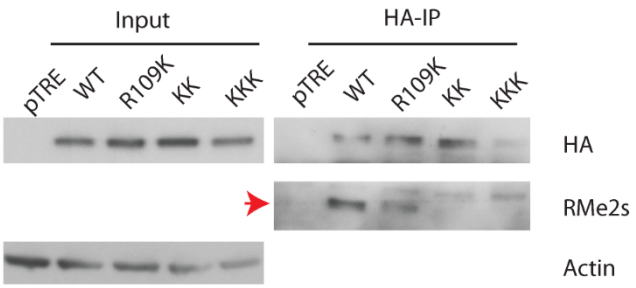


Figure 3.3 Arginine methylation of E2F1 is abrogated in the mutant derivative cell lines.

Tet-On cells containing HA-tagged E2F1 and the indicated point mutants were induced with 1µg/ml doxycycline for 24 hours and immunoprecipitated with anti-HA agarose. Arginine methylation of E2F1 was detected using pan-specific anti-symmetric dimethyl-arginine. Following visualisation of SDMA the blot was stripped and reprobed with anti-HA to ensure the SDMA band overlayed with the E2F1 signal. Anti β-actin was included as a loading control.

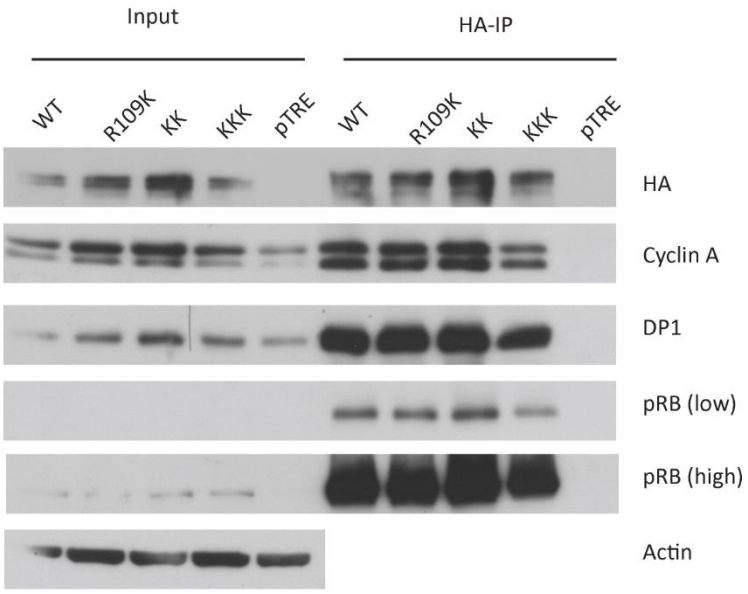


Figure 3.4 Interaction of E2F1 mutant derivatives with binding partners.

Tet-On cells containing HA-tagged E2F1 and the indicated point mutants were induced with 1µg/ml doxycycline for 24 hours A. E2F1 was immunoprecipitated with anti-HA agarose. Interactions with known binding partners, pRB, DP1 and Cyclin A was determined by immunoblotting using specific antibodies as indicated. Anti-β-actin was used as a loading control.

DP1 between the different cell lines. Thus, the effect of the methylation sites was unlikely to be due to dimerisation partner changes, although, due to antibody efficiency, this could not be confirmed with the minor dimerisation partners DP2 and DP3. Cyclin A (in combination with CDK2) binds to its binding site in E2F1 during S phase, which phosphorylates the E2F1-DP dimer to reduce DNA binding activity and increase turnover – previously it was described that methylation at R111/113 enhances the E2F1-Cyclin A interaction and results in increased E2F1 turnover (Zheng et al., 2013). However, in this cell system there were no changes observed in E2F1 binding to Cyclin A between the mutants [Figure 3.4A].

3.4 Effect of E2F1 mutants on target gene specificity

Having confirmed the mutant E2F1 were able to undergo arginine methylation in the inducible cell line system, and could interact with appropriate binding partners, it was then important to establish that these E2F1 mutant derivatives were able to bind to the promoters of known E2F1 target genes and activate transcription. Furthermore, as previous work has described changes in target gene specificity dependent on the methylation marks at R109 and R111/113 it was also crucial to determine whether these were affected appropriately by the mutations. For example, in normal growth conditions it would be expected that symmetric methylation of R111/113 by PRMT5 would correspond with E2F1 dependent cell cycle progression and reduced chromatin occupancy. Conversely, we have previously shown that under conditions where E2F1 is asymmetrically methylated by PRMT1 at R109, this results in E2F1 dependent apoptosis and enhanced presence at target gene promoters (Zheng et al., 2013). Therefore we would expect the mutations to impact on these. Firstly, to confirm the ability of the ectopic proteins to bind to DNA, chromatin immunoprecipitation (ChIP) was carried out. This technique combines immunoprecipitation of a protein of interest, in this case HA-E2F1, whilst bound to DNA. It is

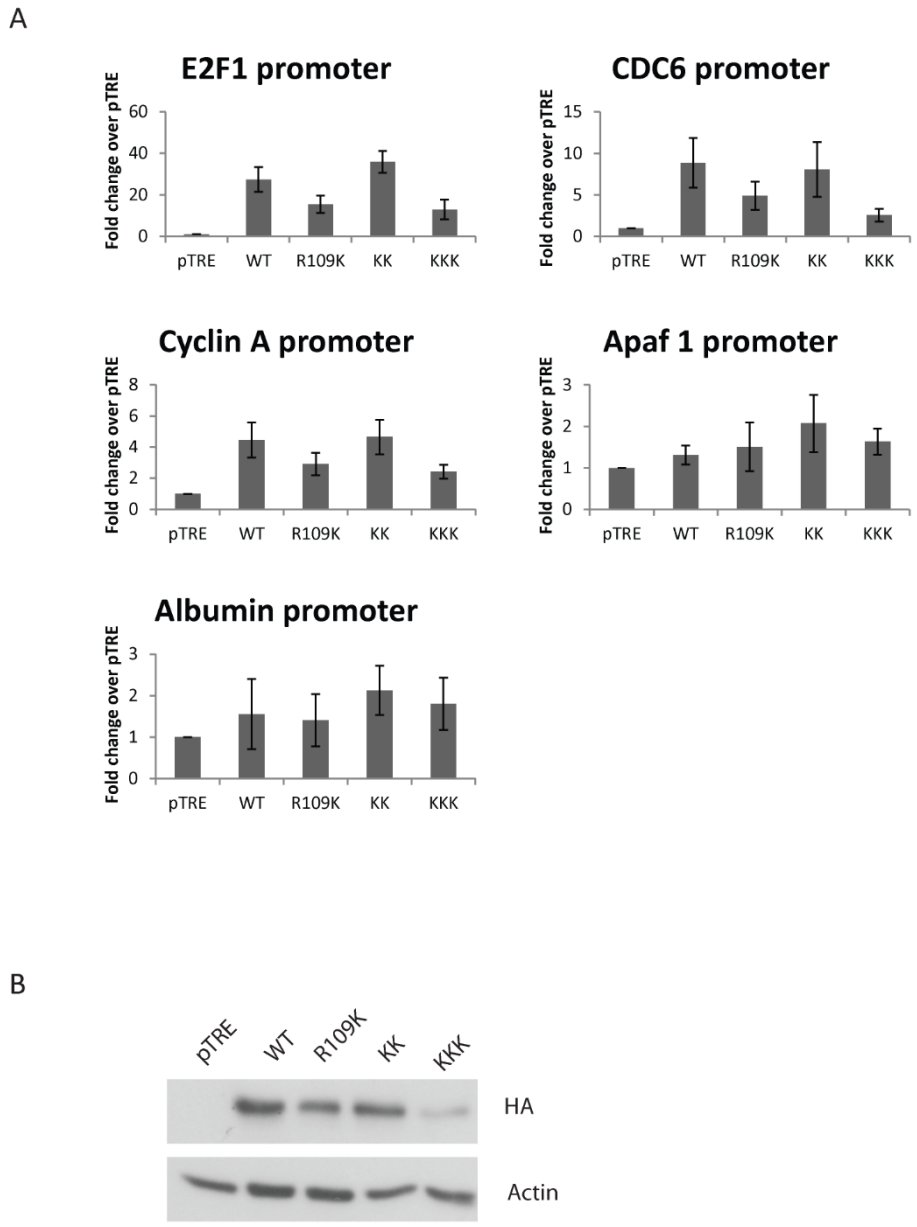


Figure 3.5 The E2F1 mutant derviatives can bind to target gene promoters.

Tet-On cells containing HA-tagged E2F1 and the indicated point mutants were induced with 1µg/ml doxycycline for 24 hours. Protein was crosslinked to DNA using 1% formaldehyde before cell lysis. DNA fractionation was carried out via sonication before immunoprecipitation using anti-HA. The cross-links were reversed and protein removed before DNA isolation

A. qPCR was carried out on the DNA using primers as indicated. Results shown are averages of three independent experiments with standard error shown.

B. Protein expression was measured via immunoblot using anti-HA to determine expression of the E2F1 derivatives, and anti-β-actin as a loading control.

then possible to isolate the DNA fragments which have been isolated with the protein and identify specific regions using semi-quantitative or quantitative PCR using primers to known promoter regions. E2F1 and all the mutant derivatives were able to bind to chromatin of known E2F1 target genes and the binding was enriched compared to the empty vector control, this was not the case for the control promoter Albumin which is not an E2F1 target gene, and was included as a negative control [Figure 3.5]. However, no significant differences between the E2F1 mutant derivatives' ability to bind at different promoters was observed. Additionally, there appeared to be loss of KKK expression as the cell lines were passaged [Figure 3.5B], and this effect was observed consistently, although no other cell line appeared to show this [data not shown].

As the E2F1 mutant derivatives demonstrated an ability to bind chromatin, we next asked whether we could observe appropriate changes in RNA levels of E2F1 target genes, which would suggest that the E2F1 mutant derivatives were transcriptionally active. After induction of expression for 24h, RNA was isolated and subject to qRT-PCR to measure the mRNA levels of a selection of E2F1 target genes. As expected, E2F1 target gene RNA levels were enhanced upon induction of expression of the different mutant derivatives, above the levels observed in the pTRE control [Figure 3.6A]. However, once again there was no significant difference between the different E2F1 mutant derivatives in their abilities to upregulate those genes tested in this experiment [Figure 3.6A]. Although it is well known that relative levels of mRNA do not directly relate to levels of protein made (Liu et al., 2016), the protein levels of a selection of target genes was assessed – [Figure 3.6B] APAF1, TS and DHFR showed no changes upon addition of doxycycline and E2F1 induction. In contrast, Cyclin A did show upregulation at the protein level after E2F1 expression in WT, R109K and KK, but not in KKK. Interestingly the basal level of Cyclin A was also increased in the uninduced cells compared to pTRE suggesting there may be some low-level HA-E2F1 expression in these cells.

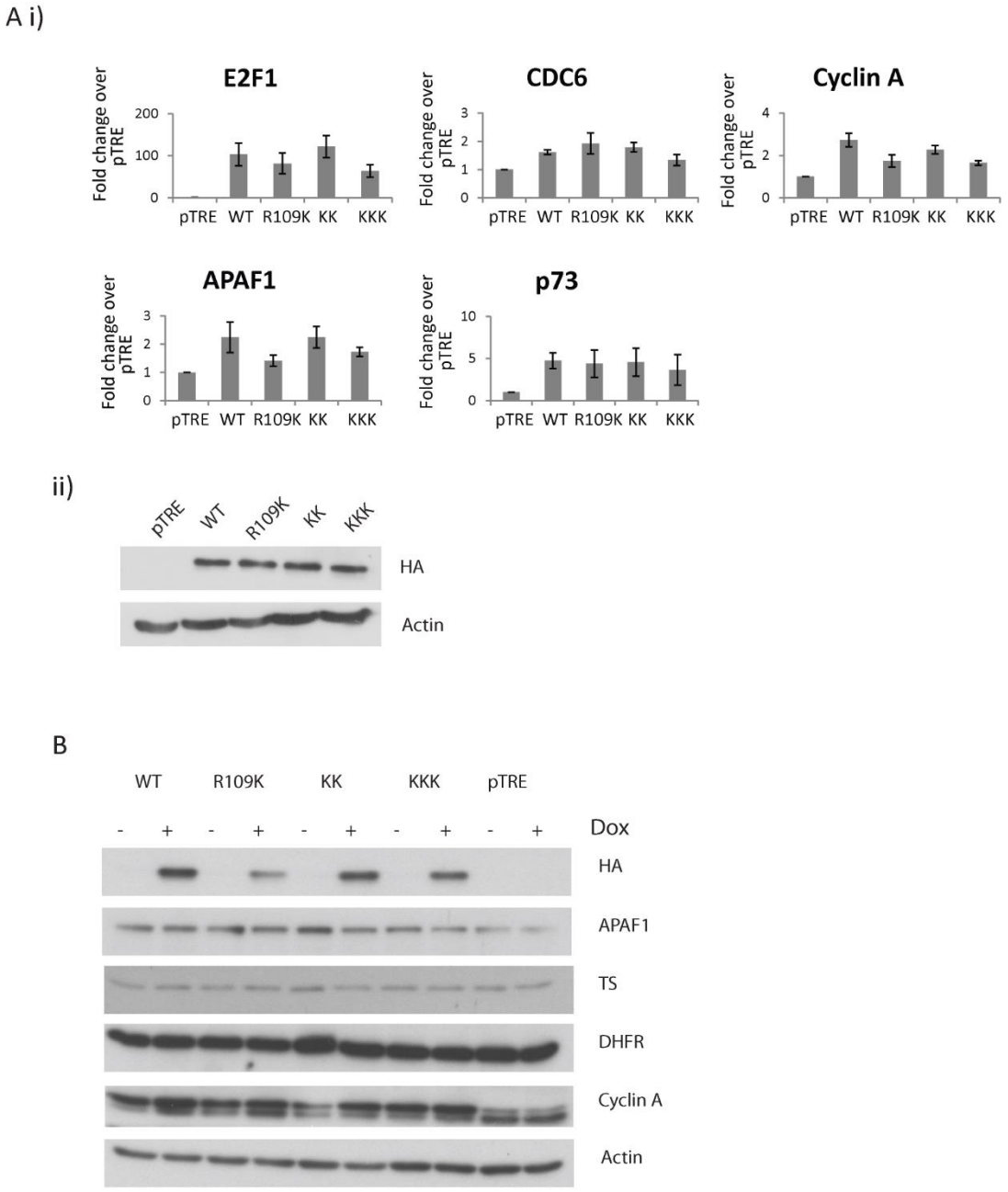


Figure 3.6 Effect of E2F1 over-expression on target gene expression.

A. Tet-On cells containing HA-tagged E2F1 and the indicated point mutants were induced with 1µg/ml doxycycline for 24 hours i) RNA was extracted with TRIzol and underwent reverse transcription before qPCR of the cDNA. Results shown are averages of three independent experiments with standard error shown ii) protein expression of the constructs is shown.

B. Cells were induced with 1µg/ml doxycycline, or vehicle control, for 24 hours and protein expression of E2F1 target genes after Cyclin A, DHFR, APAF1 and TS were measured using specific antibodies as indicated. Anti-HA was used to determine expression of the E2F1 derivatives after doxycycline addition, and anti-β-actin as a loading control.

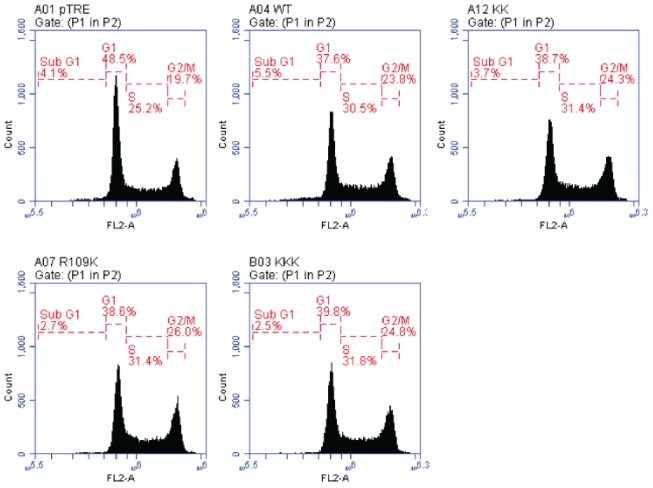
3.5 The effect of arginine methylation on E2F1

Once it had been established that the E2F1 mutants were transcriptionally active, we asked what effect overexpression of these mutants had on the cells and if this was comparable to the effects seen in transiently transfected cells, or the corresponding siRNA depletion effect of PRMT1/5 seen previously (Zheng et al., 2013). In these systems it was observed that depletion of PRMT1, or overexpression of the R109K mutant (which cannot be methylated by PRMT1) resulted in increased cell proliferation and reduced sensitivity to DNA damage. Conversely, depletion of PRMT5, or over-expression of the KK mutant resulted in greater apoptosis following DNA damage, and reduced growth rates (Cho et al., 2012, Zheng et al., 2013).

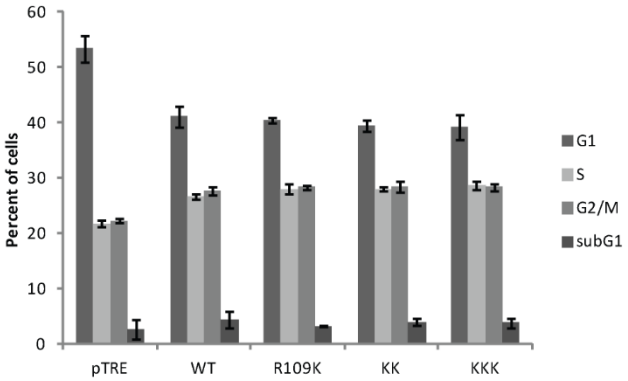
Cell cycle profiles of the cell lines growing asynchronously were analysed by flow cytometry to assess whether there were any changes in cell cycle phase [Figure 3.7]. Cells were seeded in triplicate and doxycycline used to induced protein expression for 24h. The cells were tyrpsinised and fixed in ice-cold ethanol overnight, prior to staining with propidium iodide (PI). 20,000 cells were used for cell cycle analysis on the flow cytometer [Figure 3.7A]. There were minor differences in G1 content between pTRE and the E2F1 overexpressing cells, with a reduction in the number of cells in G1 and a corresponding increase in cells in S phase and G2/M phases, suggestive of an increased G1/S phase transition upon overexpression of E2F1 [Figure 3.7B]. However, there were no significant differences between the different mutant derivative expressing cell lines after 24 hours of protein induction.

Although the cell cycle profiles looked similar over the short term, the rate of growth over a prolonged period of time was affected by the E2F1 mutant derivatives. Equal numbers of cells were plated in triplicate, and doxycycline was added the following day. Cells were then trypsinised and counted over each day for 4 days and, growth curves were drawn [Figure 3.8A] the doubling time and the proliferation rate as a percentage of pTRE were

A



B



C

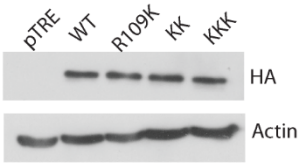


Figure 3.7 The effect of overexpression of E2F1 and mutant derivatives on the cell cycle.

Tet-On cells containing HA-tagged E2F1 and the indicated point mutants were induced with 1µg/ml doxycycline for 24 hours in triplicate. Cells were detached with trypsin and fixed in ethanol overnight. The cell pellets were washed, RNase treated and stained with 40µg/ml propidium iodide before analysis on a flow cytometer.

A. Example cell cycle profiles and phase gating are shown.

B. Percent of cell population in each cell cycle phase was calculated. Results shown are averages of three independent experiments with standard error shown

C. Example protein expression of the constructs is shown. Protein expression was measured via immunoblot using anti-HA to measure induced E2F1 expression and anti-β-actin as a loading control.

calculated [Figure 3.8B and C]. Cells expressing WT and KK E2F1 showed a significantly reduced proliferation rate and increased population doubling time compared to the pTRE control (Day 1 compared to Day 3). In contrast, cells overexpressing R109K and KKK showed no significant difference in growth compared to the control. This is as expected from previous work which suggests that E2F1 overexpression is proapoptotic, and also in line with other work from our lab which showed that R109Me2a by PRMT1 is required for this proapoptotic function of E2F1. Loss of this site in the R109K mutant means E2F1 loses its apoptotic potential and stimulates proliferation through the PRMT5 driven R111/113Me2s modification (Zheng et al., 2013). Interestingly, the KKK mutant appeared to behave similarly to the R109K mutant, which is in contrast to previous results which showed that ectopic expression of KKK resulted in a phenotype more similar to KK in transiently transfected cells (Zheng et al., 2013).

A long term (10 day) clonogenic assay was carried out to assess growth rate and clonogenic potential after overexpression of E2F1 or the mutant derivatives over a longer period of time. In this experiment, an equal number of cells were seeded in triplicate at day 0 and doxycycline added 8h later and again every 72 hours to maintain expression of the E2F1 mutant derivatives. At day 10, the cells were fixed in methanol and stained with crystal violet before the colonies were counted [Figure 3.9A]. The R109K and KKK mutant derivatives showed greater growth potential compared to WT and KK [Figure 3.9B]. pTRE demonstrated an increased clonogenic potential over WT and KK and, to a lesser extent, R109K and KKK. These results imply that E2F1 overexpression is inhibitory to growth, but this effect could be partially abrogated by mutation of the R109 site. Additionally, over the longer term the differences in growth potential between the cell lines, observed in the growth curve, were more pronounced [Figure 3.8, Figure 3.9]. Thus, as expected, loss of the R111/113 site resulted in reduced cell growth and proliferation whereas loss of R109 had the opposite effect.

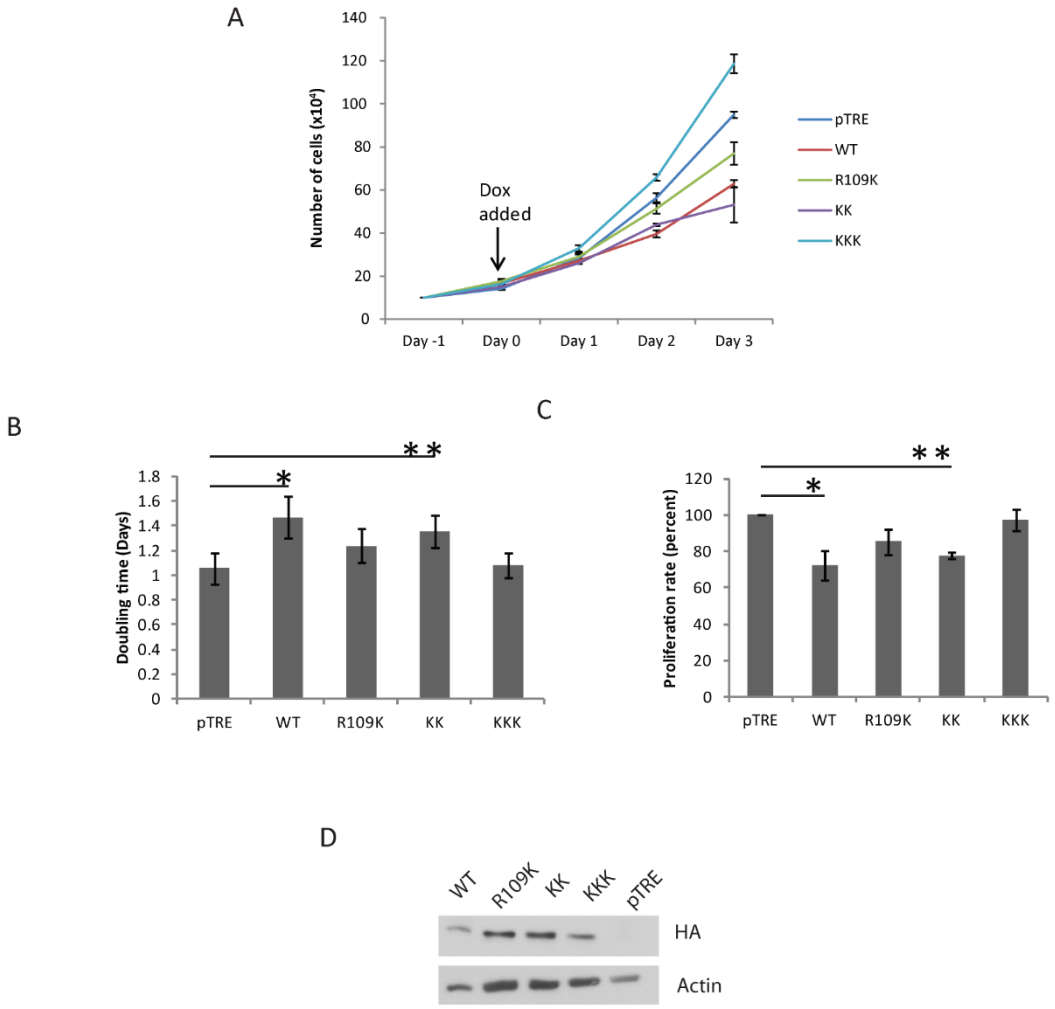


Figure 3.8 The effect of overexpression of E2F1 and mutant derivatives on short term cell growth

Equal numbers of Tet-On cells containing HA-tagged E2F1 and the indicated point mutants were plated in triplicate and 1µg/ml doxycycline was added 8h later. Cells were harvested and counted at the time points shown.

A. Growth curves were drawn for each cell line, example curve from one experiment, plated in triplicate, is shown.

B. The doubling time (in days) for each cell line was calculated. Results shown are averages of three independent experiments with standard error indicated . Significance was determined with a T-Test * $p<0.05$, ** $p<0.005$

C. The proliferation rate as a percentage of the pTRE rate was calculated. Results shown are averages of three independent experiments with standard error indicated. Significance was determined with a T-Test * $p<0.05$, ** $p<0.005$

D. Induction of expression was measured via immunoblot using anti-HA to determine expression of the E2F1 derivatives after doxycycline addition, and anti-β-actin as a loading control.

Subsequently, the effect of DNA damage on cells overexpressing E2F1 or the mutant derivatives was investigated. Both etoposide and doxorubicin (topoisomerase II inhibitors which induce double strand breaks) were used to induce apoptosis. The proapoptotic, active form of caspase-3 (cleaved caspase-3) is an early stage marker for apoptotic signalling and this enzyme is responsible with other caspase family members for subsequent apoptosis. PARP is cleaved by caspase-3 and also caspase-7 during apoptosis, and thus the cleaved form of this protein can also be used as a marker for apoptosis. Cells were seeded and doxycycline added 8h later to ensure expression of ectopic E2F1 derivatives prior to DNA damage. 24h after seeding 2µM doxorubicin (or water control) was added [Figure 3.10A i] or 20µM etoposide (or DMSO control) [Figure 3.10A ii]. Doxorubicin treated cells were harvested 24h after addition of the DNA damaging agent and etoposide treated cells were harvested 48h following addition of the DNA damaging agent. Overexpression of E2F1 resulted in an increase in both cleaved caspase-3 and cleaved-PARP compared to the pTRE control [Figure 3.10A]. This is consistent with previous results, and earlier work which have shown that high levels of E2F1 causes apoptosis (Qin et al., 1994). Interestingly, cells that expressed E2F1-KK showed highest basal expression of both apoptotic markers, which is also in line with previous results that showed KK can sensitise cells to apoptosis (Zheng et al., 2013). Upon treatment with either doxorubicin [Figure 3.10A i] or etoposide [Figure 3.10A ii], levels of the apoptotic markers, cleaved caspase-3 and cleaved PARP increased and this was most apparent in the cells overexpressing E2F1. However, the levels of these apoptotic markers were lower in R109K and KKK cell lines compared to WT and KK where higher expression of both cleaved caspase-3 and cleaved PARP was observed [Figure 3.10A]. Again, this is consistent with previous work which showed that the R109K mutation can protect cells against DNA damaged induced E2F1-dependent apoptosis and that overexpression of E2F1 drives an apoptotic response (Zheng et al., 2013, Shats et al., 2017).

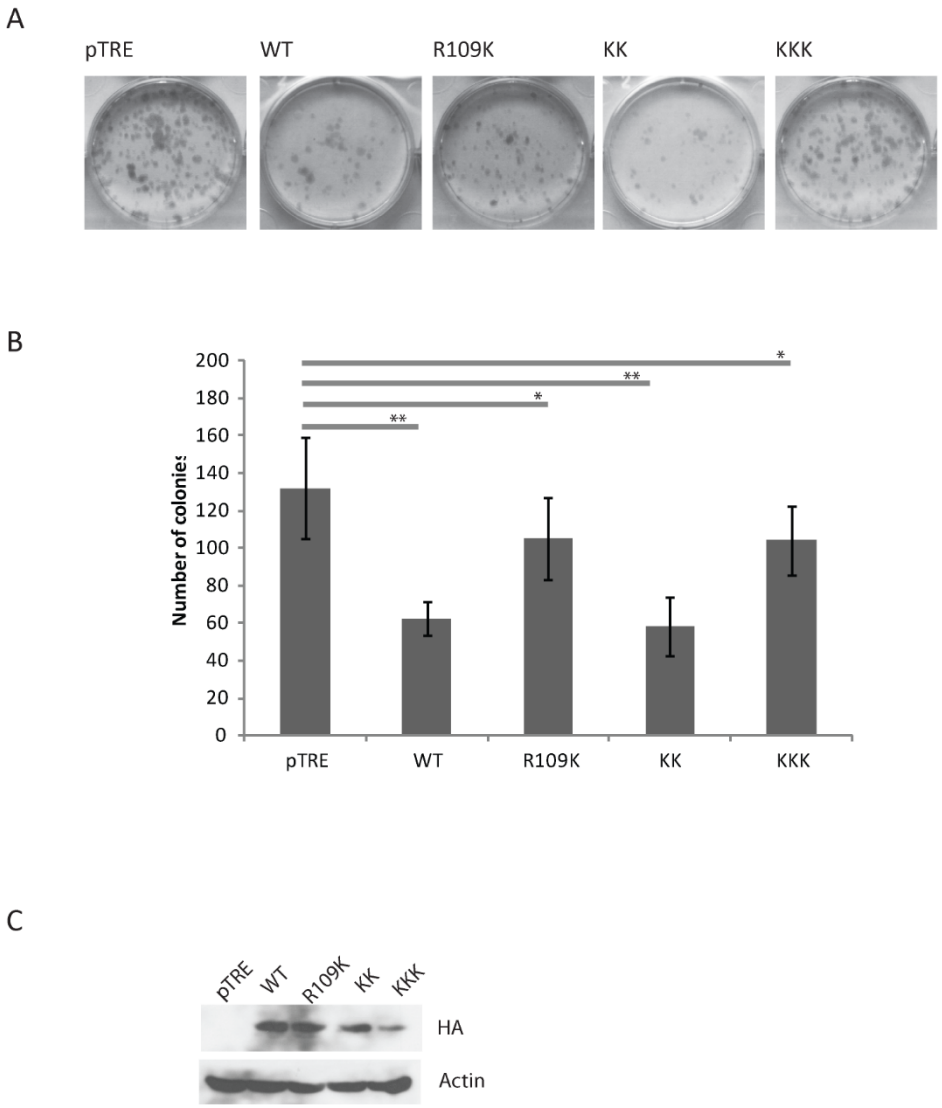


Figure 3.9 The effect of overexpression of E2F1 and mutant derivatives on clonogenic potential.

Equal numbers of Tet-On cells containing HA-tagged E2F1 and the indicated point mutants were plated in triplicate. Protein expression was induced with 1µg/ml doxycycline after 8h and again every 72h. After 10 days cells were fixed and stained and the colonies counted.

A. Photograph showing example colony formation for each cell line.

B. Average number of colonies of three independent experiments for each cell line with standard error shown. Significance was determined with a T-Test * $p<0.05$, ** $p<0.005$.

C. Induction of expression was measured via immunoblot using anti-HA to determine expression of the E2F1 derivatives after doxycycline addition, and anti-β-actin as a loading control.

Cell viability was also assessed after DNA damage by PI staining and FACS cell cycle analysis [Figure 3.10B]. Cells were plated in triplicate and doxycycline added after 8h to induce protein expression. 24h after seeding 20µM etoposide was added, cells were trypsinised 48h later, and fixed in ethanol. PI was used to stain the cells before analysis on a flow cytometer. All the cell lines showed an accumulation of cells in G2/M after etoposide addition [Figure 3.10B i] – as would be expected of a topoisomerase inhibitor which triggers G2/M arrest (Clifford et al., 2003). The sub-G1 population can be used as a marker for apoptosis as these cells have hypo-diploid DNA content which is consistent with DNA fragmentation during cell death [Figure 3.10B ii]. Thus, again R109K expressing cells were less able to induce apoptosis after DNA damage compared to WT expressing cells, but expression of KK was seen to sensitise the cells to apoptosis.

3.4 Generation of CRISPR/Cas9 knockout cell lines

3.4.1 Overview

A new method of introducing mutations at the genomic level is by harnessing cas9 endonuclease activity. Cas9 is a DNA enzyme which creates a double strand break upstream of a photospacer adjacent motif (PAM) which varies depending on the Cas9 enzyme used. The enzyme can be targeted to specific points of the genome (providing there is an appropriate PAM site) by adding a 20 nucleotide guide sequence homologous to the region of interest. This creates a specific double strand break in the DNA at the selected site resulting in non-homologous end joining (NHEJ) as all copies of the gene within the cell will have been targeted by Cas9. The resulting indel mutation from NHEJ should create frameshift mutations which result in non-functional, or truncated protein usually with a short half-life. If targeted to the 5' end of the gene, this has the ability to completely abrogate protein function and is essentially a knock-out model [Figure 3.11A] (Ran et al., 2013).

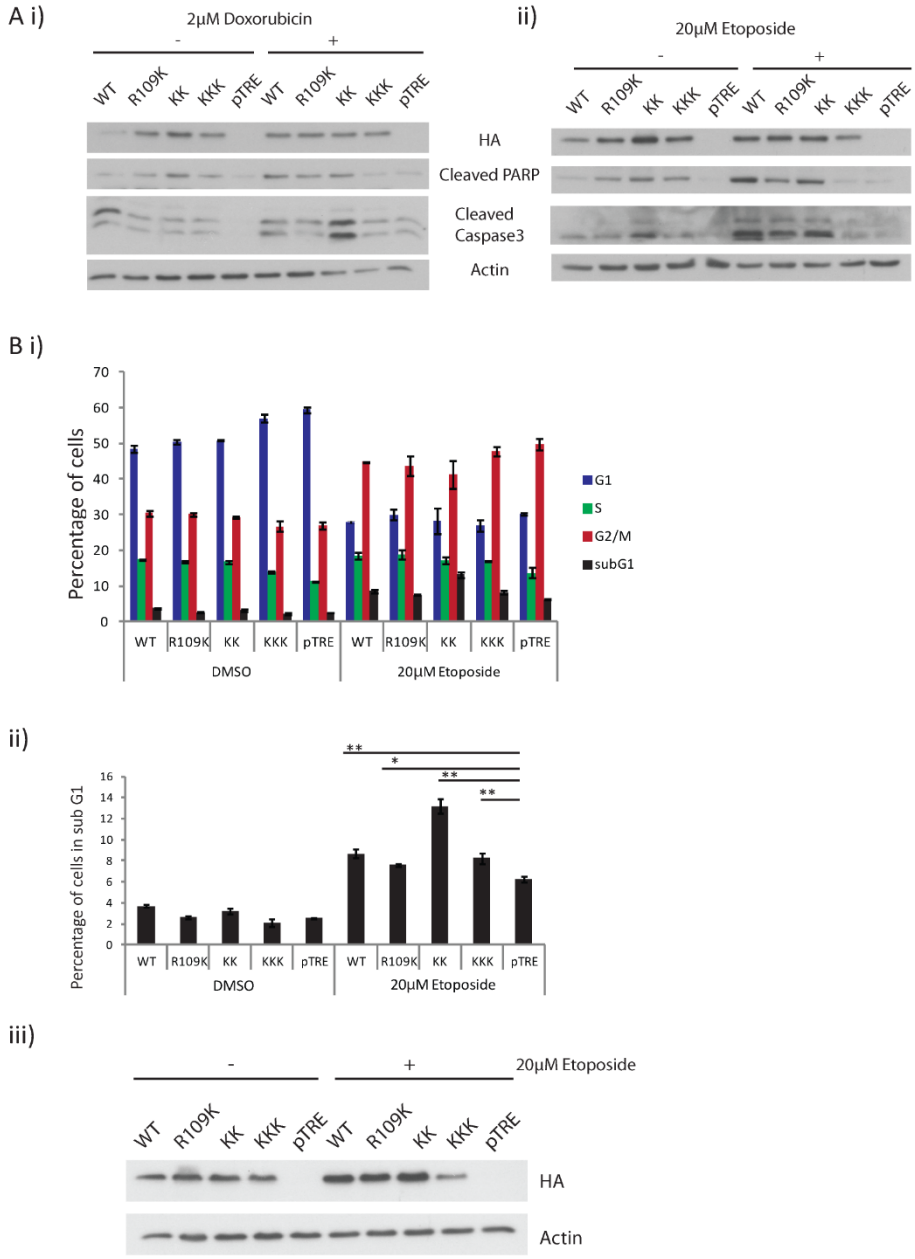


Figure 3.10 The effect of overexpression of E2F1 and mutant derivatives on response to DNA damage.

Tet-On cells containing HA-tagged E2F1 and the indicated point mutants were plated and 1μg/ml doxycycline was added 8h later. 2μM Doxorubicin or 20 μM Etoposide was added as indicated and cells harvested after a further 24h (doxorubicin) or 48h (etoposide).

A. Immunoblots showing protein levels of apoptotic markers, cleaved PARP and cleaved caspase-3, as well as HA to determine expression of the E2F1 derivatives after doxycycline addition, and anti-β-actin as a loading control.

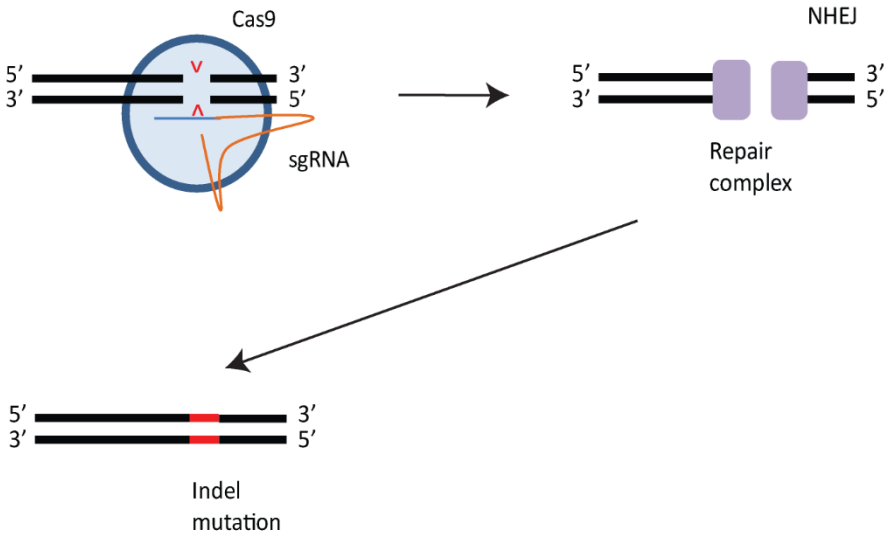
B. Cells were detached with trypsin and fixed in ethanol overnight. The cell pellets were washed, RNase treated and stained with 40μg/ml propidium iodide before analysis on a flow cytometer. i) all cell cycle stages are shown ii) only subG1 phase is shown. Results shown are averages of three independent experiments, run in triplicate with standard error indicated. Significance was determined with a T-Test * $p<0.05$, ** $p<0.005$. iii) Induction of expression was measured via immunoblot using anti-HA to determine expression of the E2F1 derivatives after doxycycline addition, and anti-β-actin as a loading control.

It is also possible to use this system to create specific mutations in the genome by using Cas9 to create a double strand break in the gene, and adding an oligonucleotide template which the cell can use to complete homologous recombination instead of NHEJ. By ensuring there are homology arms to the DNA, specific point mutations can be incorporated into the oligonucleotide which will then be used as the repair template for the genomic DNA. Therefore, these mutations will be included into the genome resulting in endogenous proteins with specific mutations (Ran et al., 2013).

3.4.2 Generation of stable, knockout cell lines

sgRNA towards E2F1 was designed using an online tool (see methods) and cloned into the Cas9-puro vector before transfection into U2OS cells [Figure 3.11B]. Short term puromycin selection was carried out and the surviving cells were plated singly to form colonies ensuring monoclonal, homozygous knockouts as any cells still expressing E2F1 are likely to outgrow $E2F1^{-/-}$ cells in a culture. Cas9-puro containing no sgRNA inserts were used to create the $E2F1^{+/+}$ control cell lines (*WT)* and a monoclonal line was established from these in parallel with the knockout cells. A *WT* polyclonal cell line was also maintained as a separate control. Genomic DNA was isolated from all four cell lines and sequenced across the predicted cut site of E2F1, and the top three potential off target sites identified by the sgRNA design programme – these were NROB1, DNMT3 and MACROD1. Alignments shown in Appendix 1 confirm that there were no off-target cut sites at these genes in either $E2F1^{-/-}$ cell line, although there were some clonal variations between the polyclonal and monoclonal *WT* cell lines. In the E2F1 alignment the Cas9 cut site can be seen at the NGG motif resulting in repair of the break, leading to a frameshift mutation in one cell line and an early stop codon in the second [Figure 3.11C and Appendix 1].

A



B

5′ 281-GGAAACTGACCATCAGTACCTGG-304

sgRNA target	Cut site	PAM sequence

C

WT A	99- QYLAESSGPARGRGRHPGK (end of exon)
WT B	99- QYLAESSGPARGRGRHPGK (end of exon)
E2F1$^{-/-}$ C	99- QLPGREQWASSGQRPPSRKRYLWSLGWADPL-STOP
E2F1$^{-/-}$ D	99- QFSLR-STOP

Cut site

Figure 3.11 CRISPR/Cas9 mediated knockout of E2F1

A. Diagram of CRISPR/Cas9 mechanism: sgRNA mediated targeted cutting of genomic DNA to create double strand breaks followed by non-homologous end joining repairs creating indel mutations in the genomic DNA.

B. Diagram showing sgRNA target sequence, the PAM site and cut site of E2F1 genomic DNA. C. Genomic DNA from clonal cell populations was isolated and the region around the E2F1 cut-site was sequenced. Expasy software was used to translate the coding sequence

Complete knockout of E2F1 at the protein level in both clones C and D was assessed by immunoblot [Figure 3.12A] and also confirmed similar expression level of E2F1 in the *WT* polyclonal ($E2F1^{+/+}$ A) and monoclonal ($E2F1^{+/+}$ B) cell lines. Although, as the antibody was generated against a C-terminal peptide, we cannot rule out expression of an N-terminal truncated version of the protein. However, the genomic sequence indicated that the protein would be truncated before the DNA binding domain, yielding a non-functional product, if any protein is produced [Figure 3.11C]. E2F1 mRNA expression was also measured in these cells [Figure 3.12B] and, as expected, the expression level of E2F1 was significantly reduced to less than half that of the *WT* cells. This suggests that the RNA could be more unstable after the Cas9 mediated mutation of the genomic DNA. The RNA levels of known E2F1 target genes were also measured to further confirm E2F1 deficiency [Figure 3.12C]. Across the three target genes, there were non-significant differences between the two *WT* clones and, separately, between the knockout clones which suggested that the different clonal cell lines behaved similarly. Importantly, there was significant reduction in expression of CDC6 and Cyclin A in the $E2F1^{-/-}$ cell lines compared to the *WT* cells, as well as a reduction in APAF1 mRNA expression after Cas9 mediated deletion of E2F1, although this was not significant.

3.4.3 Characterisation of E2F1 CRISPR/Cas9 knockout cell lines

Once the cell lines had been validated, the effect of stable knockout of E2F1 was investigated. First, their growth was assessed in a short-term growth curve, where equal numbers of cells were plated in triplicate, trypsinised and counted each day for three days [Figure 3.13]. As expected, given E2F1's role in cell cycle control, $E2F1^{-/-}$ cells showed reduced proliferation rates by approximately 20% [Figure 3.13B], corresponding with an increased doubling time of 1.25-1.5 days compared to 1 day for the *WT* cells [Figure 3.13C]. Thus, there were moderate differences in growth between the two $E2F1^{-/-}$ clonal cell lines, but these were not significant.

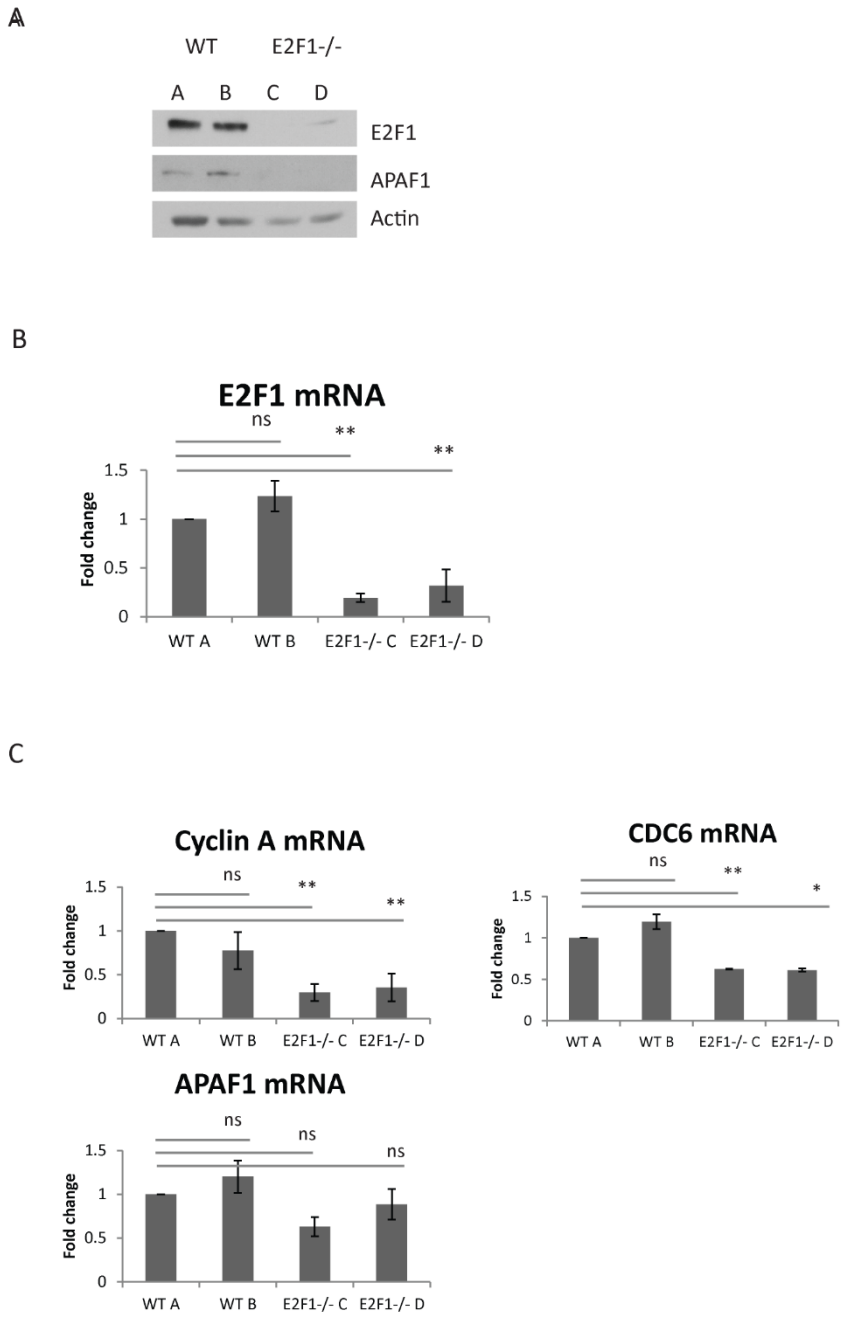


Figure 3.12 Validation of E2F1 knockout by CRISPR/Cas9

U2OS cells were transfected with pSpCas9(BB)-2A-Puro A containing guide RNA to E2F1.

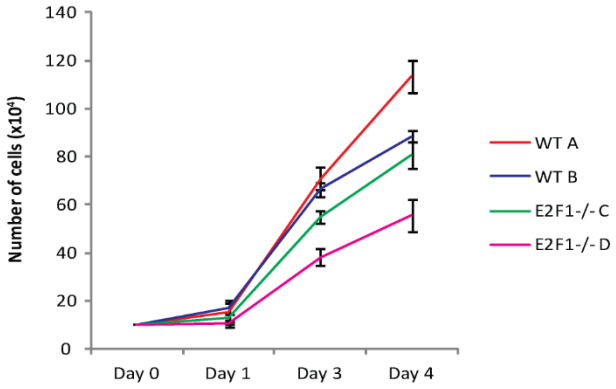
A. Immunoblot showing protein level of endogenous E2F1 in the cell lines was depleted after Cas9 mediated disruption of the E2F1 gene. B-actin is included as a loading control.

B. RNA from the indicated cell lines was extracted, reverse transcribed and subject to qPCR. mRNA levels of E2F1 are shown . Average of three independent experiments. Significance was determined with a T-Test ** $p<0.005$.

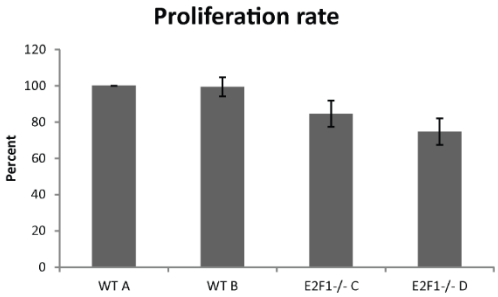
C. mRNA levels of the indicated E2F1 target genes were analysed by qPCR. Average of three independent experiments (CDC6 is average of two independent experiments). Significance was determined with a T-Test * $p<0.05$, ** $p<0.005$.

Subsequently, the effect of DNA damage on these cell lines was investigated. 20μM etoposide was used to induce double stranded-breaks in the DNA and cells were harvested after 48h, immunoblotting was carried out to analyse expression of apoptotic markers [Figure 3.14]. Compared to both *WT* cell lines, the E2F1$^{-/-}$ clones had reduced cleaved PARP and cleaved caspase-3 expression after DNA damage, thus suggesting they were more resistant to apoptosis under these conditions. Interestingly, although stabilisation of p53 occurred in all the cell lines after DNA damage, there was also reduced expression in the E2F1$^{-/-}$ cells suggesting that there was also a reduced p53 response in these cells which is consistent with reports suggesting E2F1 can both increase p53 transcription and phosphorylation induced stabilisation (Lindstrom and Wiman, 2002). *WT* B cells showed high levels of constitutive apoptosis, suggesting clonal differences in this cell line which may have higher basal levels of apoptosis.

A



B



C

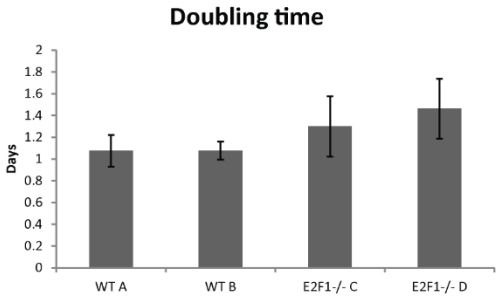


Figure 3.13 Proliferation rate of E2F1 knockout cells.

Equal numbers of *WT* or E2F1$^{-/-}$ U2OS cells were plated in triplicate and allowed to grow. At the time points shown the cells were trypsinised and counted.

A. Growth curve over 4 days of a single representative experiment.

B. The proliferation rate for each cell line was calculated as a percentage of WT A shown are averages of three independent experiments with standard error.

C. The average doubling time for each cell line was calculated, shown are averages of three independent experiments with standard error.

3.5 Chapter summary

In this chapter we have described the generation of cell lines that will be valuable tools for further studies. Firstly, Tet-On inducible cell lines which overexpress E2F1, or the mutant derivatives R109K, R111/113K and R109/111/113K. The R to K mutation is relatively conservative as there is no charge difference and only a small size difference between these amino acids. Therefore there should be minimal impact on protein folding or function other than modifications made on those residues. Overall, the overexpressing cells showed the expected phenotypes with R109K having increased growth potential and reduced sensitivity to apoptotic stimuli due to its inability to be methylated by PRMT1, and the KK mutant having the opposite effects as it is no longer a target for PRMT5. For this reason, these cell lines can be used to examine the function of arginine methylation at these sites on E2F1. However, as the KKK mutant behaved differently than expected, and also showed consistently reduced expression as the cells aged, we removed this cell line from future experiments.

It is worth noting that overexpression models are not the most elegant method of studying protein function due to artefacts which can occur such as altered localisation and interactions as well as increasing cell stress. Additionally, as E2F1 functions as a heterodimer with the DP subunits, the endogenous levels of DP1/2 will act as a rate-limiting step on E2F1 function in this system. There is also the matter of endogenous E2F1 still being present in these cells which may impact on results as overexpression of E2F1 will result in endogenous levels of E2F1 also increasing due to E2F1 being its own target gene. Regardless of these flaws in the method, overexpression of mutant proteins has distinct benefits over the other methods available, such as silencing the enzymes which create these post-translational modifications. By inhibiting, or silencing PRMTs we would expect many off-target effects unrelated to their function on E2F1, as they have multiple targets across different pathways in the cell. The creation of point-mutants will allow a more direct measure of the function of arginine methylation on E2F1.

Ideally, endogenous point mutants could now be made using CRISPR/Cas9 technology to induce a double-strand DNA break whilst providing a repair template containing the mutations to induce the cell to undergo homologous recombination and stably introduce the mutations into the genome which would therefore be expressed at endogenous levels. Whilst this was attempted during making the CRISPR/Cas9 E2F1$^{-/-}$ cells, we found that transfecting a long (150bp) DNA oligonucleotide repair template alongside the CRISPR/Cas9 plasmid was highly inefficient with no cells surviving selection despite multiple attempts at optimisation.

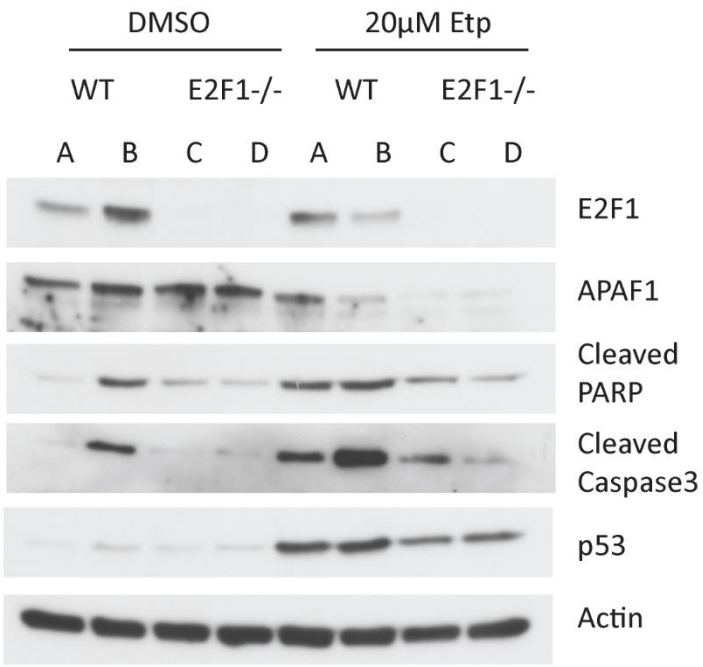


Figure 3.14 DNA damage response in E2F1 knockout cells

E2F1 Crispr U2OS cells were treated with DMSO or 20µM Etoposide. After 48h cells were harvested and immunoblotting was carried out. Protein expression of cleaved PARP, cleaved caspase-3, and p53 were used as indicators of apoptosis. Constitutive knockout of E2F1 was confirmed by probing for endogenous E2F1 and anti-β-actin was used as a loading control. Example experiment of 3 independent experiments is shown.

Chapter 4. Identification of novel E2F1 target genes by RNA-sequencing

4.1 Introduction

Once the Tet-On inducible cell lines had been validated, they could be used as a tool to further study the impact of arginine methylation on E2F1 function. As previous work in our lab had looked at known E2F1 target genes, we decided to expand this to a global study with the view to finding potential new E2F1 target genes and functions of arginine methylation.

RNA sequencing is a powerful tool which enables a genome-wide analysis of genes up, or downregulated under different conditions. The information this gives, can identify not only potential new target genes of E2F1, but will also identify how these genes are affected by the methylation status and may also identify new pathways which can be impacted by methylation of E2F1. Additionally, it is possible to identify isoform use and look at potential differences in splicing between the cell lines.

There are multiple parts to an RNA sequencing experiment: sample acquisition, library preparation, sequencing and analysis [Figure 4.1]. Each step requires understanding of the limitations of the system as there are many biases which can affect the output. Samples need to be in triplicate in order for statistical analysis to be effective and it is important that they are highly similar as otherwise you may not see an effect as one sample may cancel out the effect of another. There are two methods of library preparation – poly-A selection, or ribo-depletion. In the first, beads ligated to dT(n) are used to enrich for poly-adenylated RNAs (usually mRNA, although some non-coding RNA is also poly-adenylated), this removes the problem of rRNA contamination, but any non-poly-A RNA species such as non-coding RNA and miRNAs are lost from the analysis. The ribo-depletion method removes rRNA from your RNA pool, leaving all other RNA (mRNA and non-coding RNA) to be sequenced. However, this remains fairly inefficient and many of the sequencing reads can be taken up with rRNA contamination, reducing the depth of sequencing. As coding RNA remains the better studied of these, downstream analysis and validation was focussed on protein coding RNA over

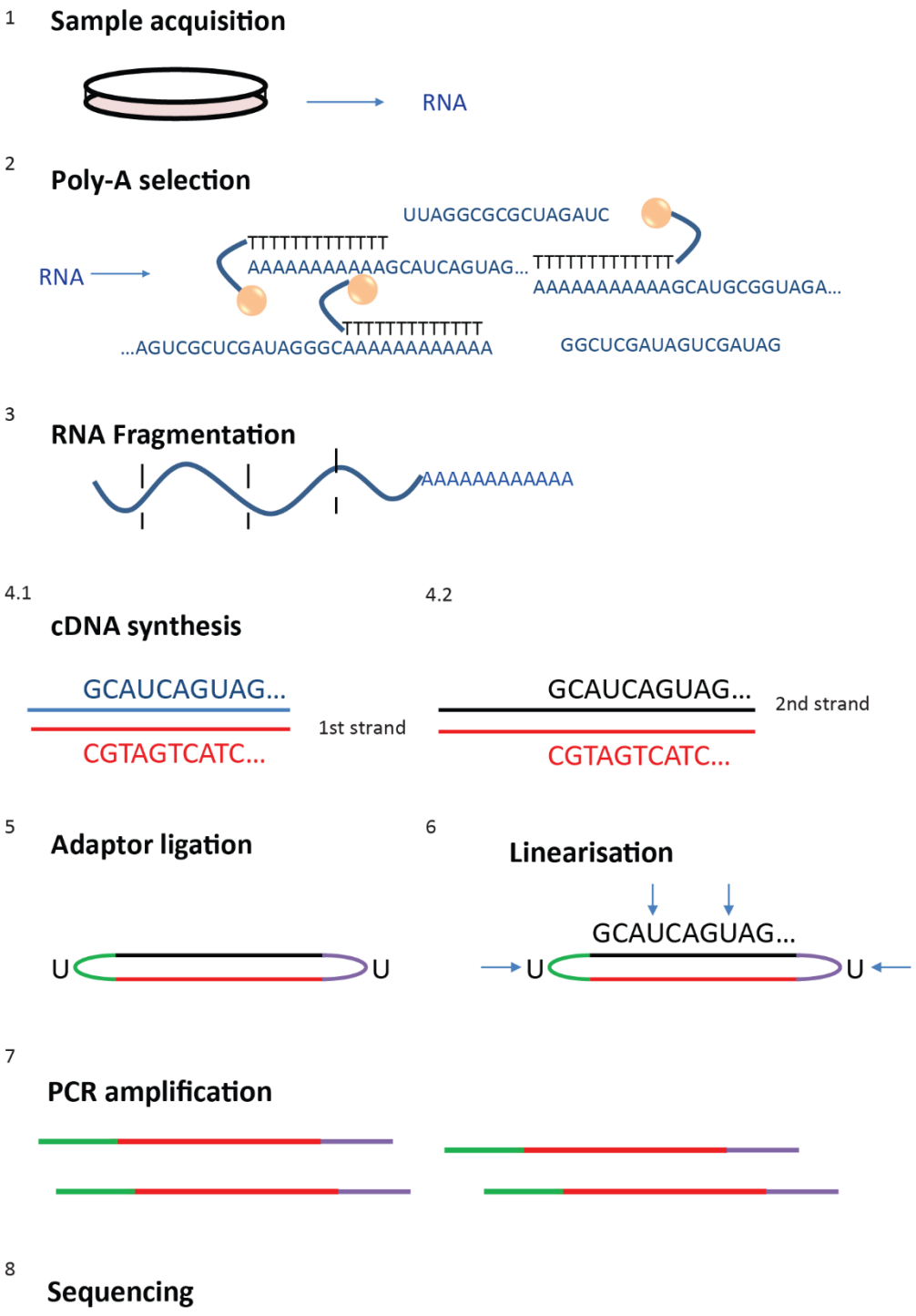


Figure 4.1 Diagram of library preparation protocol for RNA sequencing.

Steps are numbered in order. RNA is shown in blue .1. Samples are obtained and RNA is extracted. 2. poly-A selection purifies mRNA using n(dT) ligated beads. 3. The mRNA is fragmented. 4. Random primers are used for first strand cDNA synthesis (first strand is shown in red). Second strand (black) synthesis uses dUTP. 5. Adaptors are ligated to the cDNA. 6. Excision of uridine bases linearises the cDNA and removes the second strand to create single stranded cDNA with adaptors at both ends. 7. adaptor-ligated cDNA is amplified via PCR. 8. Sequencing is carried out.

the non-coding RNA. Therefore, for the purpose of this study, poly-A selection was chosen in order to get better sequencing depth, even if this meant losing most non-coding RNA from the experiment.

The RNA is then fragmented and first strand cDNA synthesis using random primers is carried out. Second strand synthesis uses dUTPs which ensures the two strands of cDNA are distinct. Subsequently, the adaptors are ligated onto the cDNA – which forms a circular cDNA species. This is linearised, and the second strand is removed by USER enzyme which excises uridine bases causing strand fragmentation. The single stranded cDNA is then amplified by PCR before sequencing [Figure 4.1].

4.2 Sample preparation

Cells were plated at the same density, doxycycline was added 24h later. After a further 24h cells were harvested, half of the cell pellet was lysed in TRIzol for RNA extraction and the other half was used for immunoblotting to analyse protein expression.

To ensure the samples from each replicate were good enough quality to go onto library preparation and sequencing, an aliquot of RNA was taken for cDNA synthesis and qRT-PCR analysis of target genes was performed to ensure the cells were behaving as expected and to confirm there was high reproducibility between the samples, with small standard error [Figure 4.2A]. Additionally, the RNA quality for each sample was analysed by nanodrop (ensuring a 260/230 ratio of >2.2, and a 260/280 ratio >2.0). These ratios are considered 'pure' for RNA and show that it is free from contamination by protein or organic solvents used during RNA extraction. The quality of the RNA was subsequently analysed on the Bioanalyser which uses a RNA integrity (RIN) score to ensure the size of the RNA species is acceptable – i.e. it ensures that the RNA has not degraded [Figure 4.2B]. Typically, RIN scores over 8 are considered ideal

for sequencing but lower scored RNA is also acceptable as RIN values between 7-10 are classified as little to no degradation of the RNA. 5-7 is partial degradation and lower than 5 is degraded. Given that all samples, except WT2, were over 8 and WT_2 had a RIN score of 7.4 we decided that these samples were of a high enough quality to proceed with library preparation.

Once 3 replicates had shown to be reproducible under qRT-PCR for a small selection of target genes, and had good quality RNA they were processed for library preparation using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina with the NEBNext Poly(A) mRNA Magnetic Isolation Module as per manufacturer's instructions (detailed in Chapter 2). A final quality control to determine concentration of each cDNA sample to ensure equal amounts of sample were included into the library, these were pooled and the combined libraries were run on the Tapestation to determine the average size (base pair) of the libraries – just over 300bp [Figure 4.2B ii].

4.3 Bioinformatics analysis of differential expression

All bioinformatics analysis was done in collaboration with Alexander Kanapin of the Bioinformatics Core at Old Road Campus Research Building, Oxford. Initial bioinformatics analysis of the FASTQ reads using FastQC software served as a quality control for the experiment [Figure 4.3]. The reads were aligned using STAR short reads aligner version 2.5.1b to the hg19 reference genome build. All samples except WT_2 showed aver 20 million mapped reads suggesting good sequencing coverage and low contamination [Figure 4.3A].

Principal Component Analysis (PCA) is a measure of similarity between samples and identifies the top expressed genes in each sample before comparing them. Thus, replicates should cluster closely on the graph and there should be distinction between the clusters of different

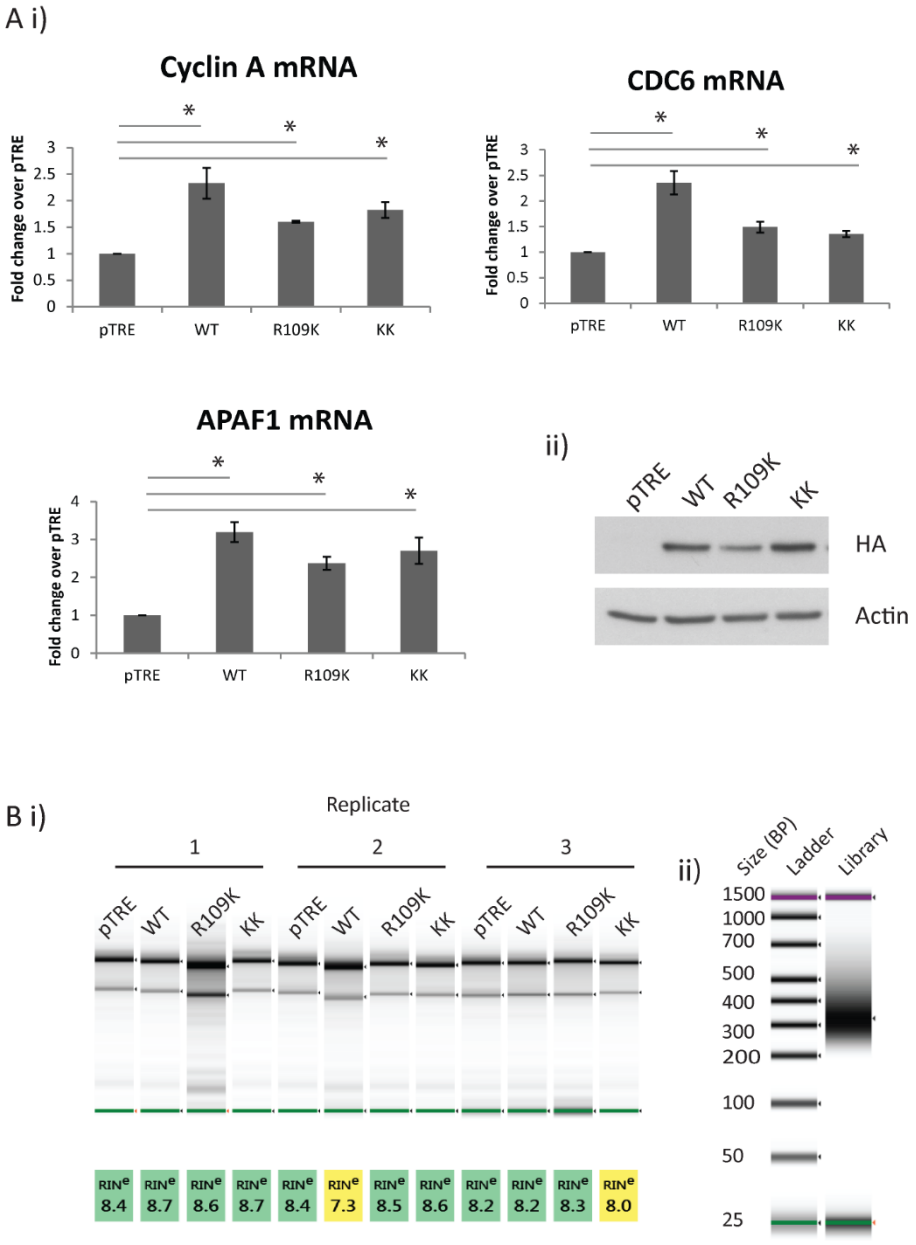


Figure 4.2 RNA-seq sample quality control.

A. Equal numbers of cells were seeded and 1µg/ml doxycycline added for 24h. The RNA was extracted and i) 1µg was taken for cDNA synthesis and qPCR to assess reproducibility between three independent experiments. 1µg of RNA was subsequently processed for sequencing. qPCR of cDNA across three target genes as an average of three independent experiments. Significance was determined with a t-test * $p<0.05$, ** $p<0.005$ ii) Protein expression was measured via immunoblot using anti-HA and β-actin as a loading control.

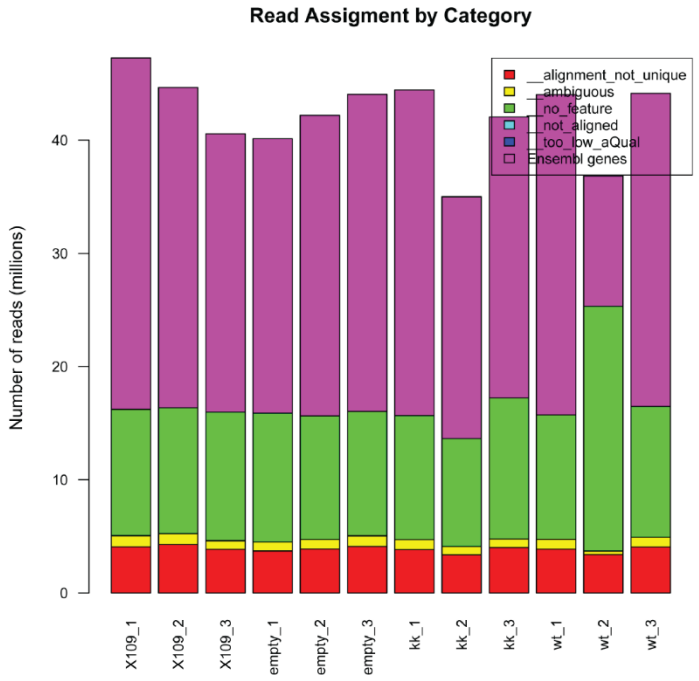
B. i) Tapestation analysis was carried out before library preparation to assess RNA quality and size. ii) Tapestation analysis of the pooled libraries was also carried out to confirm average fragment size.

samples if there are differences in gene expression in the different treatments. Clustering on the PCA plot and in the hierarchical clustering [Figure 4.3B and Supplementary Figure 2] showed that the empty vector (pTRE) clustered distinctly from the rest of the samples, and R109K was also distinctly separated from WT and KK. However, WT and KK appeared to cluster together suggesting that these showed very similar gene expression. Again, WT_2 was very different to the remainder of the samples, again signifying that this sample was poor quality. From this point, WT_2 was removed from further analysis.

Differential expression of genes was analysed using DESeq2 package (Bioconductor) which allows comparisons between gene expression in each treatment to be analysed. Figure 4.4 is a heatmap of the differential expression between pTRE and each other cell line as indicated. From this we can see that overall there are a group of genes which are differentially expressed in all three cell lines compared to pTRE, as would be expected from the PCA clustering. Also, there are some differences between cell lines and perhaps the gene expression profile of R109K is more dissimilar from WT and KK, whereas the latter two have fewer obvious differences in gene expression, which was also indicated by the PCA plot and hierarchical clustering.

ChIP-seq experiments have identified the majority of E2F1 binding sites are located in the -900 to +100 region (Bieda et al., 2006). In order to identify potential E2F1 target genes from our dataset, we got -900 to +100 sequences for all the genes identified as differentially expressed and scanned these for E2F1 binding motifs [Figure 4.5a]. It is important to note that this does not confirm that these genes are E2F1 targets, and similarly, we may have excluded direct E2F1 target genes from this analysis if they do not contain a canonical binding motif and/or the motif lies outside the proximal promoter. Nevertheless it helps narrow down the large dataset to those which have a good possibility of being direct targets of E2F1, rather than E2F1 having an indirect effect on

A



B

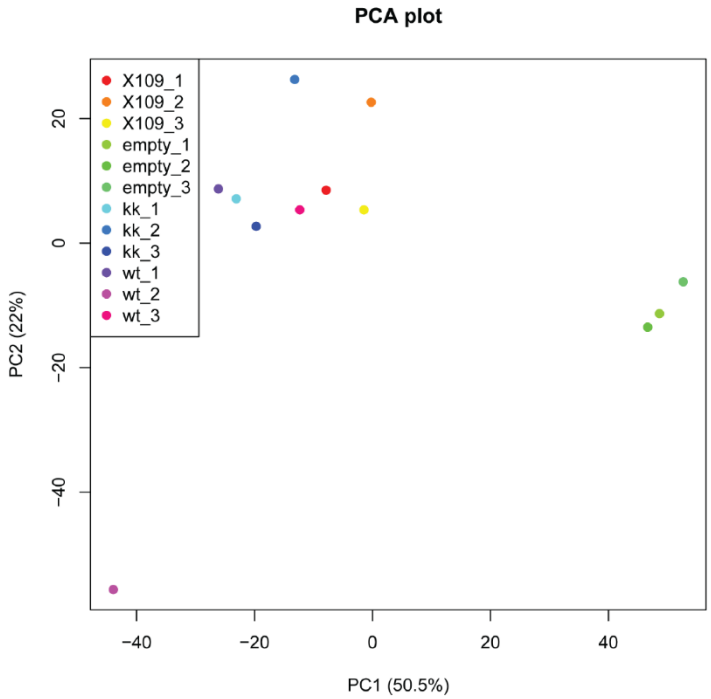


Figure 4.3 RNA-seq sequencing quality control.

A. Read assignment of reads after alignment to hg19 genome build. Pink reads are aligned with Ensembl genes, green reads have no assigned feature, yellow reads are ambiguous, red reads are not unique reads.

B. PCA plot of reads for each sample, pTRE in green, R109K in orange, KK in blue and WT in pink.

transcription through modulation of other factors. Over 50% of the total genes in each comparison (except for WT vs KK which has so few differentially expressed target genes in total) were identified as potential E2F1 target genes. Of these, the majority (55%) were upregulated when compared to pTRE [Figure 4.5B].

When the data was filtered further to identify those genes differentially-regulated with a fold change of 2 or greater [Figure 4.5C], the effect of all vs pTRE was exacerbated, with over 70% of E2F1 target genes being upregulated. As E2F1 is a transcription factor we would expect that the majority of genes would be up-regulated after over expression of E2F1. It is somewhat surprising that there is a large quantity of potential E2F1 target genes which are downregulated when compared to pTRE. This could be due to E2F1-dependent upregulation of inhibitory factors, a negative feedback response, or a role for E2F1 in transcriptional inhibition for select target genes. Additionally, examination of genes with a 2 fold change in expression level revealed no differences in the WT vs KK differential expression comparison, further underlying the similarity between these two cell lines [Figure 4.5C]. Interestingly however, in the WT vs R109K and KK vs R109K comparisons R109K is seen to have a reduced number of transcripts upregulated compared to the other two cell lines. This may indicate that R109K has a reduced ability to activate transcription, or that it has additional repressive functions.

When compared to the pTRE control there were clear differences between gene expression of WT, R109K and KK. In total, 1095 E2F1 target genes were shared between all three as showing differential expression from the control. Of these, 350 were downregulated and 744 were upregulated compared to pTRE [Figure 4.6]. Whilst all cell lines shared a pool of genes which were commonly regulated compared to pTRE, there were also distinct sets of genes which were unique to each comparison. In concordance with the analysis which suggested

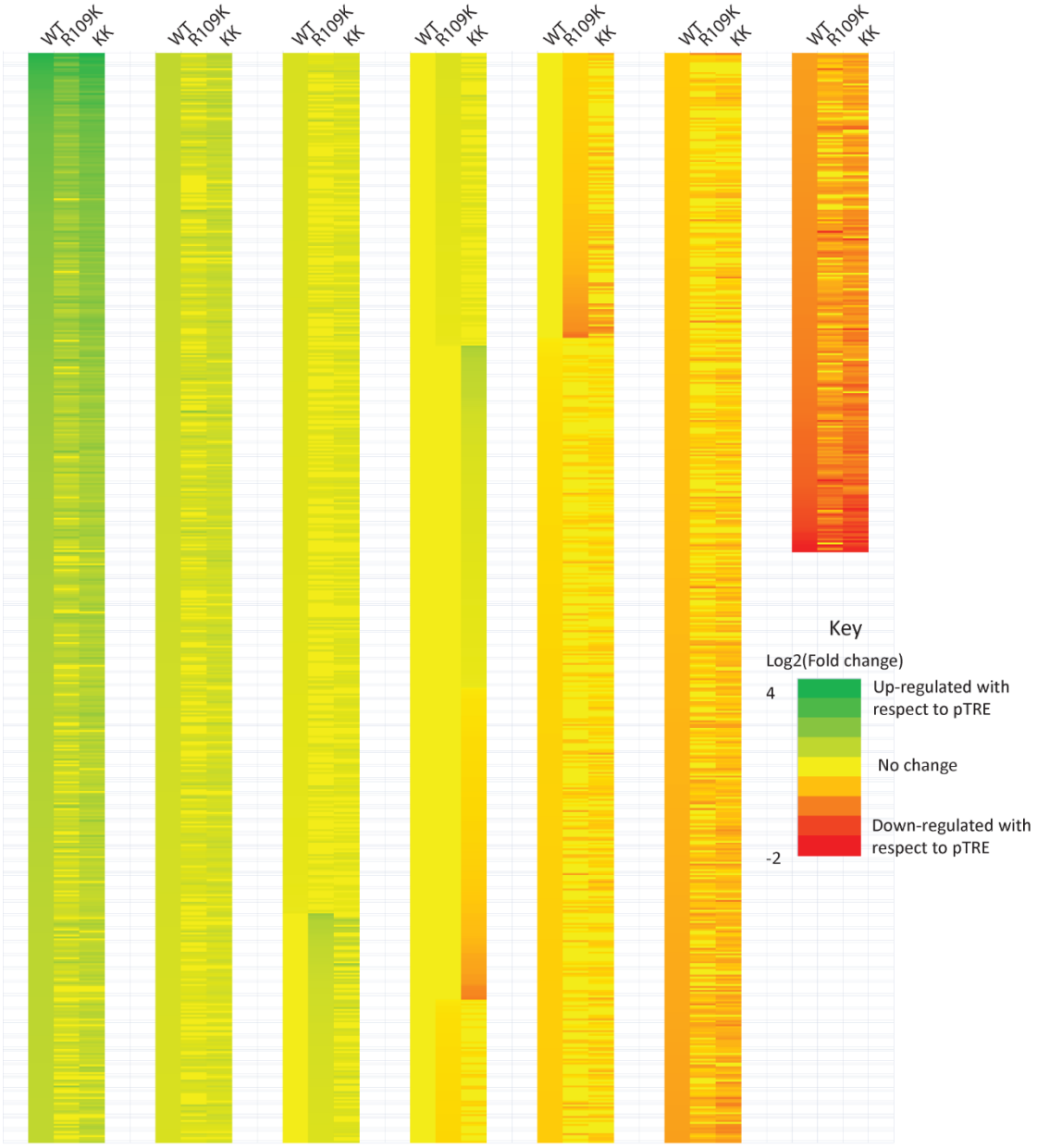


Figure 4.4 Heatmap of differential expression compared to pTRE.

Differential expression of genes compared to pTRE was calculated using DESeq2 on Bioconductor. The log2 fold change compared to pTRE for each RNA species in each cell line was converted into a heatmap to show differences between cell lines when compared to pTRE, with green being the most upregulated (maximum 16 fold change) and red being most downregulated (minimum -4 fold change).

that WT and KK showed similar profiles, WT and KK also showed the most similarity in target gene expression, sharing 84 downregulated genes and 185 upregulated genes compared to pTRE. However, both also had distinct pools of target genes which were not shared which was unexpected from the differential expression analysis of WT vs KK that suggested there were no genes differentially regulated between these two cell lines. In contrast, WT and R109K had the least amount of overlap between gene pools, as only 9 genes were downregulated by both, and 7 upregulated by both.

Gene ontology (GO) analysis of genes both shared and unique to each of these comparisons in the Venn diagram was run using DAVID 6.7, with the BP_FAT GO categories with the most amount of genes enriched ($p<0.05$) indicated on the diagram [Figure 4.6A] (Huang et al., 2009a, Huang et al., 2009b). This shows that DNA replication related genes were enriched in R109K and KK compared to pTRE, G-protein coupled receptor signalling was enriched in WT and R109K, and cell-surface receptor signalling and regulation of transcription related genes were enriched in WT and KK with respect to pTRE. Genes uniquely enriched in R109K compared to pTRE had GO terms 'ion transport' and 'positive regulation of proliferation', whereas genes uniquely upregulated in KK versus pTRE were enriched for GO terms 'regulation of transcription' and 'cell-cell signalling'. Finally, GO terms enriched in genes uniquely upregulated in WT compared to pTRE were 'ion transport' and 'phosphate metabolic process'.

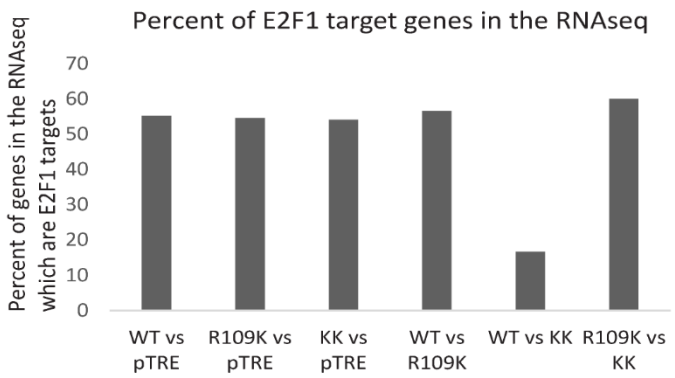
Using the data from the differential expression analysis we used GSEA tool (Broad Institute) to identify biological pathways which may be impacted in the different conditions. GSEA uses gene ontology terms to categorise gene function and weights pathways based on both number of genes affected in a certain pathway as well as using the fold change of up- or down-regulation to assess more fully how a particular treatment is affecting cell biology.

A

E2F1 motifs where S: Guanine / Cytosine, W: Adenine / Thymine, N: Any, K: Guanine / Thymine, R: Purine, V: Guanine / Cytosine / Adenine

NNNSSCGCSAANN
WWTGGCGGCAAA
NTTTCGGGCS
NKTSSCGC
NTTSGCGG
TTTGGCGCCAAA
NTTSGCGG
TTTSGCGS
NGGGCGGGARV

B



C

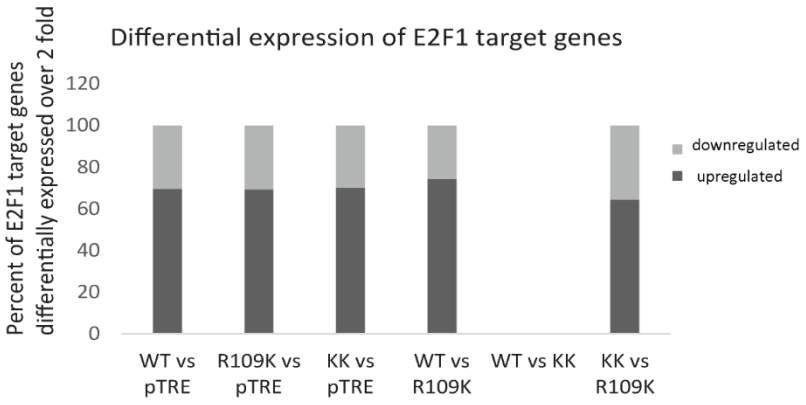


Figure 4.5 E2F1 target genes identified from RNAseq results

A. E2F1 binding motifs scanned for in the promoters (-900 - +100) of genes.

B. Percent of genes identified in each differential expression analysis which contained one or more E2F1 binding motifs in the promoter.

C. Up-or down-regulation of the potential E2F1 target genes which showed differential expression over 2 fold.

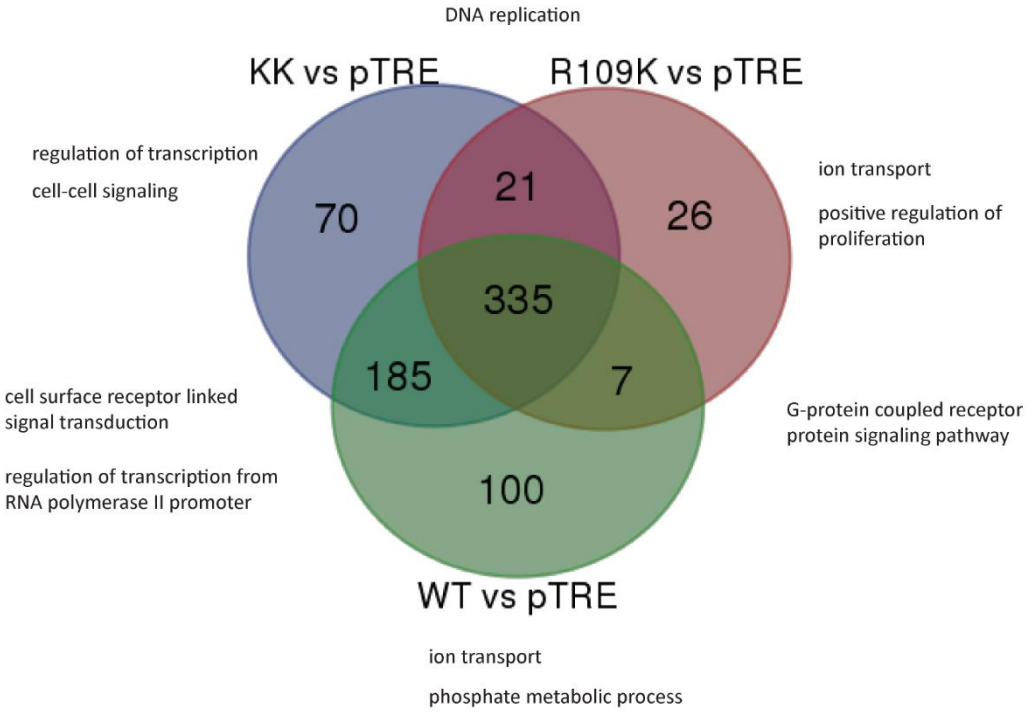
Table 4.1 shows those pathways which are significantly (FDR adjusted q<0.05, FDR<25%) enriched in each comparison. Across all three E2F1 expressing cell lines there is an enrichment of E2F1 responsive genes compared to the control as was anticipated. Interestingly, there are also spermatogenesis and WNT pathway enrichments in all three cell lines compared to the pTRE cells which suggests a global role for E2F1 in these pathways which may not be influenced by arginine methylation. G2/M checkpoint and pancreatic beta cell signalling events are upregulated in WT and R109K compared to pTRE, but these are not in the KK vs pTRE comparison which may reflect some role of arginine methylation on regulation of E2F1 target genes in these pathways. The G2/M checkpoint is particularly interesting given the role of E2F1 in apoptotic signalling.

Upregulated in pTRE vs the HA-E2F1 inducible cells were EMT and angiogenesis pathways [Table 4.1] This was unexpected as E2F1 has been linked to EMT and metastatic potential (Engelmann and Pützer, 2012). Here, it is likely that as E2F1 is being overexpressed, the cells are in a more stressed state than in the control treatment due to the high levels of E2F1, and this is likely to impact the EMT potential of these cells.

WT and KK cells displayed no significant differences in the pathways which they upregulated, or downregulated compared to each other and comparisons with R109K showed similar enrichments: TNFα signalling through NFκB and increased KRAS signalling. WT vs R109K also showed an increase in androgen response signalling, but as the cell line we used was osteosarcoma this is less obvious although there have been few reports of osteosarcoma responding to androgens (Wiren et al., 1997), it is also likely that similar growth factor pathways are induced in these cells as have been identified in androgen responsive settings.

Oddly, R109K appears to have more enrichment of E2F1 and MYC dependent target genes compared to WT or KK and there was an enrichment of genes involved in oxidative phosphorylation in R109K compared to WT.

A



B

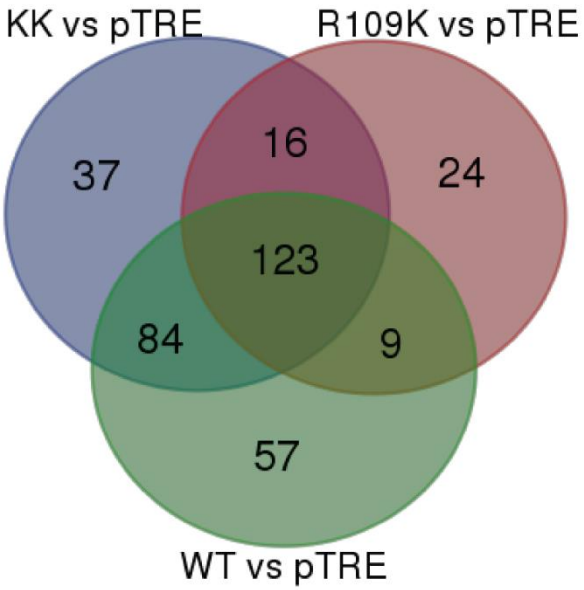


Figure 4.6 Overlap of differential expression analysis

A. Venn diagram showing numbers of genes 2-fold upregulated with respect to pTRE, which have E2F1 binding motifs in their proximal promoter, shared or unique between cell lines. Top DAVID GO 6.7 (BP_FAT) terms enriched ($p<0.05$) for each pool of genes is indicated around the diagram.

B. Venn diagram showing numbers of genes 2-fold downregulated with respect to pTRE, which have E2F1 binding motifs in their proximal promoter.

Table 4.1 Gene ontology analysis

Cell line	Up-regulated with respect to	Pathway	Score	FDR q value
WT	**pTRE**	Pancreas beta cells	1.66	0.020
		Spermatogenesis	1.60	0.023
		G2M checkpoint	1.58	0.021
		TNFα signalling via NFκB	1.56	0.021
		WNT Beta Catenin signalling	1.56	0.018
		E2F targets	1.55	0.016
		UV response up	1.46	0.038
R109K	**pTRE**	E2F targets	1.94	0.001
		G2M checkpoint	1.76	0.003
		WNT Beta Catenin signalling	1.65	0.012
		Spermatogenesis	1.57	0.021
		Pancreas beta cells	1.52	0.030
KK	**pTRE**	Spermatogenesis	1.64	0.034
		WNT Beta Catenin signalling	1.52	0.045
		E2F targets	1.51	0.036
pTRE	**WT**	EMT	1.83	0.013
		Angiogenesis	1.80	0.007
pTRE	**R109K**	EMT	2.02	0.000
		Angiogenesis	1.73	0.007
		UV response down	1.58	0.023
		Coagulation	1.52	0.031
pTRE	**KK**	EMT	1.85	0.002
		Angiogenesis	1.78	0.001
WT	**R109K**	TNFα signalling via NFκB	1.90	0.001
		KRAS signalling up	1.78	0.003
		Androgen response	1.55	0.037
KK	**R109K**	KRAS signalling up	1.65	0.020
		TNFα signalling via NFκB	1.57	0.037
R109K	**KK**	E2F targets	1.63	0.022
		Myc targets	1.51	0.044
R109K	**WT**	Myc targets	1.76	0.002
		Oxidative Phosphorylation	1.61	0.016
		E2F targets	1.50	0.029

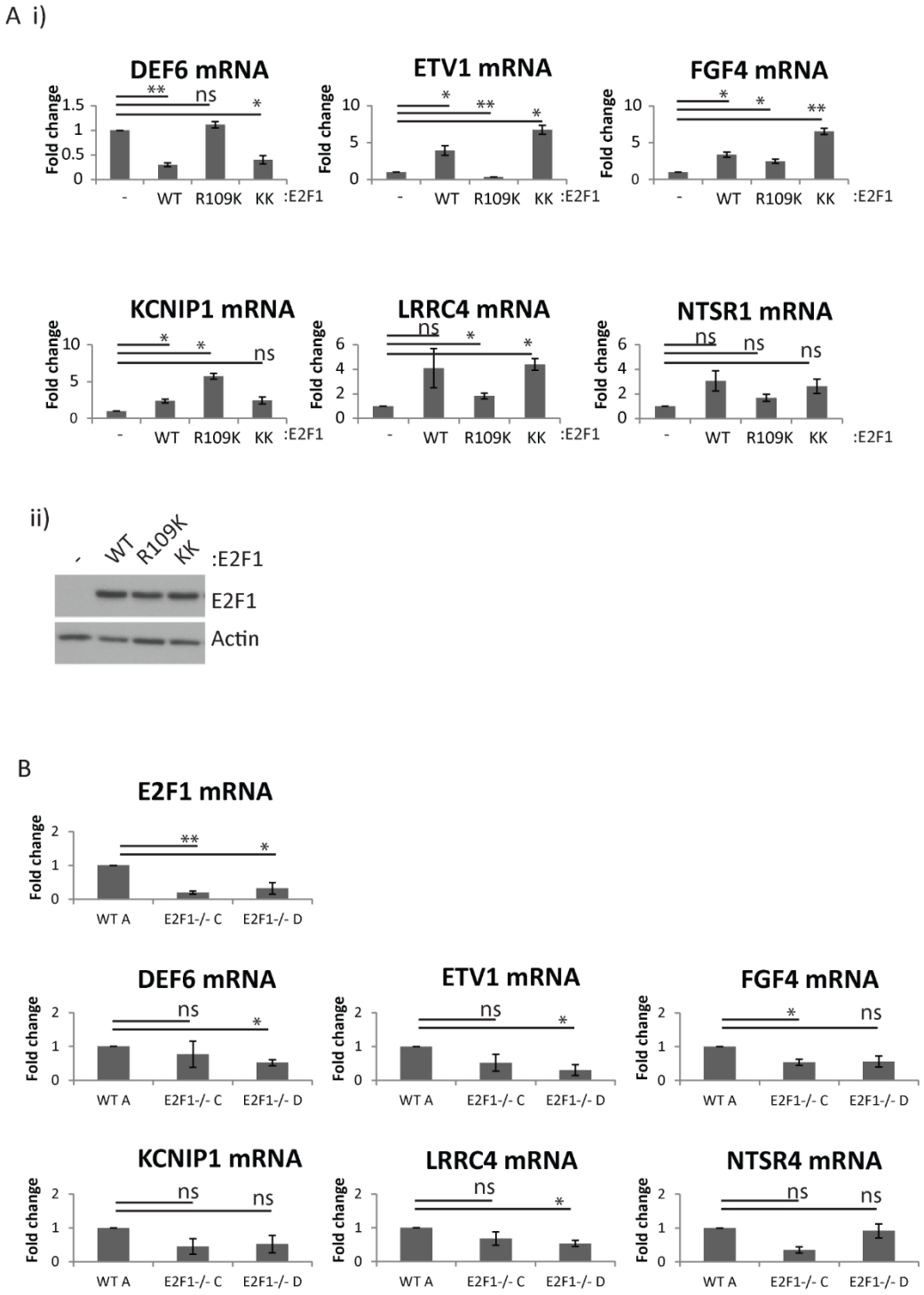


Figure 4.7 Target validation of novel E2F1 target genes

A. Equal numbers of cells were seeded and 1µg/ml doxycycline added for 24h. i) RNA was extracted with TRIzol, cDNA was synthesised and subject to qPCR with the indicated primers. Average of three independent experiments * $p<0.05$, ** $p<0.005$ ii) Protein expression was measured via immunoblot using anti-HA for ectopic E2F1 expression and anti-β-actin as a loading control

B. i) Crispr *WT* or E2F1$^{-/-}$ U2OS cells were harvested and the RNA was extracted using TRIzol. cDNA synthesis was carried out before analysis by qPCR for the primers indicated. Average of three independent experiments * $p<0.05$, ** $p<0.005$.

It is important to note that GO analysis is a fairly loose term especially as RNA levels don't necessarily correlate with protein levels. However, it is a useful tool to give an overview of how a treatment can affect different cell pathways and can provide some functional overviews and an idea of how the signalling may be occurring.

4.4 Validation of differential expression analysis

As discussed previously, there are many biases and potential for false positives during a sequencing experiment, therefore it is vital that we validate the experiment on a smaller scale by looking for a selection of potential target genes identified by the RNA-seq. We identified 6 potential novel target genes of E2F1 which showed upregulation with respect to pTRE in at least one of the cell lines and also showed differences in expression between the mutants. Additionally, we confirmed manually, and by using publically available E2F1 ChIPseq data, their potential E2F1 binding motifs in the promoter. These genes are outlined in Table 4.2 below.

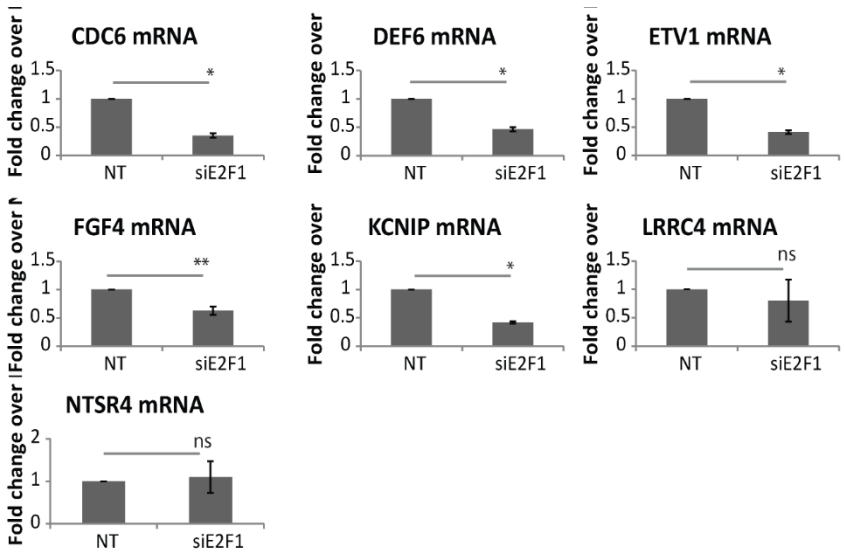
Primers were designed and qRT-PCR carried out on RNA generated under the same conditions as used for the RNA-seq [Figure 4.7A]. Similarly to results from the RNA-seq, there was greater DEF6 expression in R109K cells compared to WT, and KK. However, this was not due to enhanced expression, but it appears that DEF6 expression is reduced in the WT and KK expressing cells, compared to pTRE, and this inhibition of expression is not seen in R109K. This is in contrast to KCNIP, which showed enhanced expression across all cell lines compared to pTRE (although this is not significant for KK), with highest expression levels observed in R109K cells. Again, this confirmed results from the RNA-seq [Table 4.2]. For all other RNA species, the trend is similar – with WT and KK showing increased expression over pTRE and R109K [Figure 4.7A].

Table 4.2 List of novel E2F1 target genes identified in the RNA-seq

Gene name	Function	Differential expression in RNA-seq analysis
DEF6	DEF6 guanine nucleotide exchange factor (for RAC and CDC42) involved in migration and cell polarity	Increased in R109K compared to KK or WT
KCNIP	Potassium Voltage-Gated Channel Interacting Protein 1, calcium sensor.	Upregulated in R109K compared to KK or WT
ETV1	ETS variant 1, transcription factor involved in proliferative signalling	Upregulated in WT and KK compared to R109K
FGF4	Fibroblast growth factor 4, cell growth and survival pathways	Upregulated in WT and KK compared to R109K
LRRC4	Leucine rich repeat containing 4. Function unknown	Increased in WT compared to R109K
NTSR1	Neurotensin receptor 1, GPCR family member	Increased in WT compared to R109K

This trend is most pronounced in ETV1 where R109K significantly reduces mRNA expression relative to pTRE, but WT and KK increase it significantly. In both FGF4 and LRRC4 all cell lines increase mRNA expression above pTRE, and the effect is greatest for WT and KK expressing

A



B

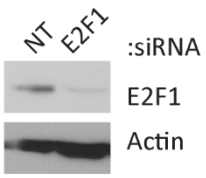


Figure 4.8 Target validation of novel E2F1 target genes II

A. U2OS cells were treated with siNT or siE2F1 for 72h. The RNA was extracted and subject to qRT-PCR for the mRNA species shown. Average of two independent experiments with standard error is shown * $p<0.05$, ** $p<0.005$.

B. Protein expression was measured with an immunoblot to ensure depletion of E2F1 after siRNA treatment, using anti-E2F1 and anti-β-actin as a loading control.

cells, with R109K showing a reduced ability to enhance expression. NTSR4 also follows this trend, but the upregulation is not significant. In order to confirm these genes as E2F1 responsive, firstly we used the E2F1$^{-/-}$ cripsr cells to determine if there was an impact of E2F1 loss on RNA expression of these genes [Figure 4.7B]. The *WT* polyclonal cell line A was used as a control alongside both monoclonal E2F1$^{-/-}$ cell lines. For all genes there was a reduction in expression of RNA in both the E2F1$^{-/-}$ cell lines compared to the *WT* cells, which is consistent with these genes being positively regulated by E2F1 and in at least one of the CRISPR cell lines this reduction was significant except for KCNIP and NTSR4. Additionally, for all genes (except NTSR4) both the E2F1$^{-/-}$ cell lines have a similar degree of reduction in expression. The differences between both CRISPR cell lines, could be due to clonal differences. Additionally, these cells may utilise rescue mechanisms through other pathways after long-term silencing of E2F1, in order to re-establish expression of genes required for viability and growth. Therefore, to confirm the results from the CRISPR cell lines, short-term siRNA depletion of E2F1 expression was carried out. RNA levels of the potential target genes in E2F1 depleted cells was compared to those treated with non-targeting (NT) siRNA [Figure 4.8]. After short-term E2F1 silencing DEF6, ETV1, FGF4 and KCNIP all show reduced RNA levels to half suggesting their expression is controlled by E2F1. Both LRRC4 and NTSR4 show no significant difference between the NT and the siE2F1 treated cells suggesting that they may not be regulated by E2F1.

4.5 Bioinformatics analysis of exon usage

As p100-TSN is known to have a role in splicing via its interaction with SAM68, we wanted to know whether E2F1 may impact splicing through recruitment of p100-TSN via the R111/113Me2s recognition. It is possible to detect isoform use in next-generation sequencing analysis by identifying exon-exon boundaries and determining the number of reads at each

junction. In this way, exon and isoform expression can be inferred. To compare between different conditions it is also important to take the gene expression into account and normalise for any differences in general expression before looking at differences in exon use. The DEXSeq Bioconductor package was used to estimate exon use across each transcript. It uses a formula: $\frac{E}{T}$ where E is number of transcripts from the gene that contain this exon and T is the number of all transcripts from the gene. When the ratio across different conditions are compared it is possible to infer differences in exon use and therefore alternative splicing, although changes in 5' and 3' exons may also indicate alternative transcriptional start sites or poly-adenylation sites respectively. By building up a picture of different exon use across a transcript it is possible to surmise potential changes in isoform under different conditions.

Relative use of each exon is shown in the heatmap [Figure 4.9A], where each horizontal line represents a single exon with red having the highest use of the exon and blue showing the lowest use between the cell lines. This also infers that KK shows increased exon gain and shows the strongest difference from pTRE and WT. R109K shows less of a trend across the board when compared to WT and pTRE with fewer strong differences.

Interestingly, 83% of the genes identified as having alternative exon use were identified as potential E2F1 target genes in ChIPseq data publically available (ENCODE) [Figure 4.9B]. Additionally, those genes which showed alternative exon use were largely a different set of genes than those identified in the differential expression analysis [Figure 4.9C]. The Venn diagrams show there were very few genes which showed both differential expression compared to pTRE and also had a different splice pattern – only 9 genes of those upregulated across all the cell lines compared to pTRE were also subject to alternative splicing, and only 6 genes from those down-regulated compared to pTRE were also seen in the alternative splicing analysis [Figure 4.9C]. This suggests that E2F1 may be impacting on splicing of a subset of its

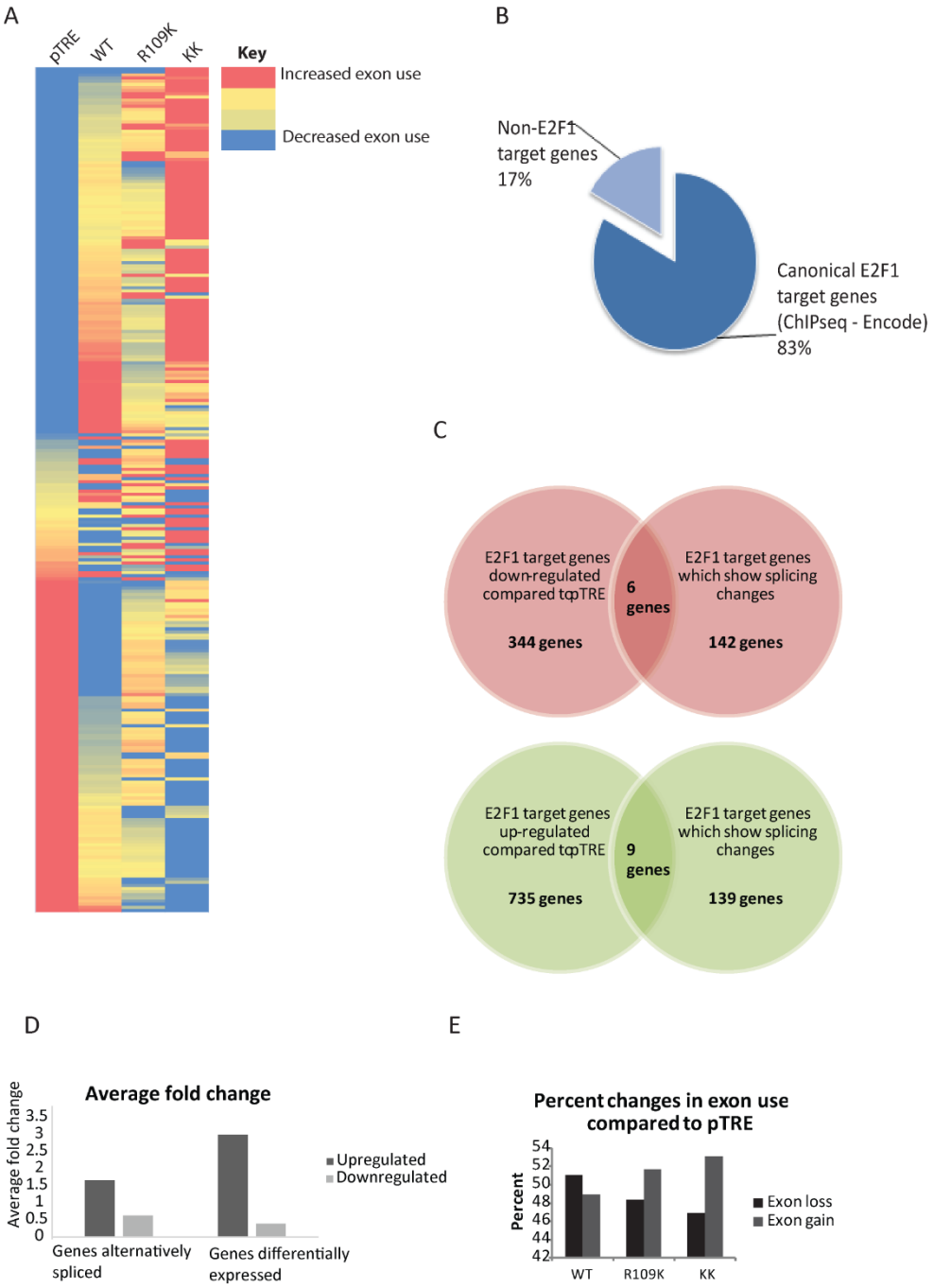


Figure 4.9 Analysis of exon use across the cell lines.

A. Heatmap showing use of each exon that showed statistically significant changes across the cell lines, with blue being lowest use and red being highest use.

B. Pie chart showing the percentage of genes identified in the splicing analysis which have E2F1 ChIPseq peaks in their promoter.

C. Venn diagram showing overlap between genes identified in the differential expression analysis and those identified as being alternatively spliced.

D. Average fold change of differentially expressed E2F1 target genes, compared with the differential expression of those genes identified in the exon use analysis.

E. Fold change in exon use compared to pTRE was calculated and expressed as a percent exon gain / loss.

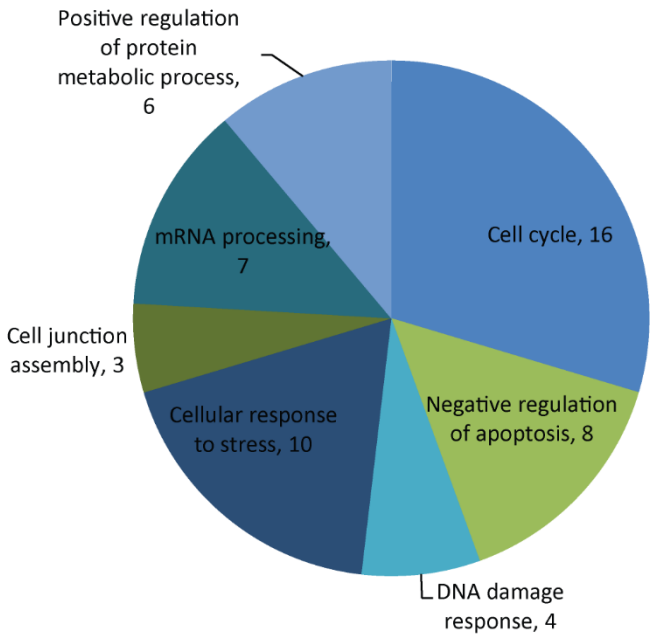
target genes as an additional, and separate method of gene expression control. When the average fold change of genes which showed differences in exon use was identified from the differential expression data, 78 genes failed to meet the cut-off value of 1 fold change, and the average fold change for the rest of the genes was 1.7 for the upregulated genes and 0.6 fold change for the downregulated genes [Figure 4.9D]. These changes are markedly reduced from those genes considered to be significantly differentially expressed in the differential expression analysis [Figure 4.6] which showed an average positive fold change of 3 fold, and an average downregulation of 0.3 fold [Figure 4.9D]. This finding further supports the hypothesis that E2F1 is acting on alternative RNA processing of a subset of genes which do not show strong, classical transcriptional regulation by E2F1.

Fold change of the use of each exon across the different cell lines compared to pTRE was calculated and represented as percent use [Figure 4.9E]. This revealed that more genes had increased exon loss in the WT cell line, whereas inhibition of methylation in the KK and R109K mutants results in increased exon gain. Again, the strongest difference was in the KK cell line which suggests there may be aberrant splicing in the absence of R111/113 methylation and loss of p100-TSN recruitment.

The overall number of genes identified as alternatively spliced was low but several pathways were enriched in DAVID 6.7 (BP_FAT) GO analysis (Huang et al., 2009a, Huang et al., 2009b) [Figure 4.10A]. These included cell cycle and negative regulation of apoptosis, which would be expected of E2F1 targets as these are two major functions of E2F1. Although, by the very nature of alternative splicing, the function of the different splice forms will impact the function of the protein and many genes still do not have annotated functions for different isoforms so the impact these have on the designated pathways is unclear.

As E2F1 has previously been shown to impact on splicing through transcriptional regulation of splicing factors (Merdzhanova et al., 2008), we searched for genes involved in splicing in

A



B

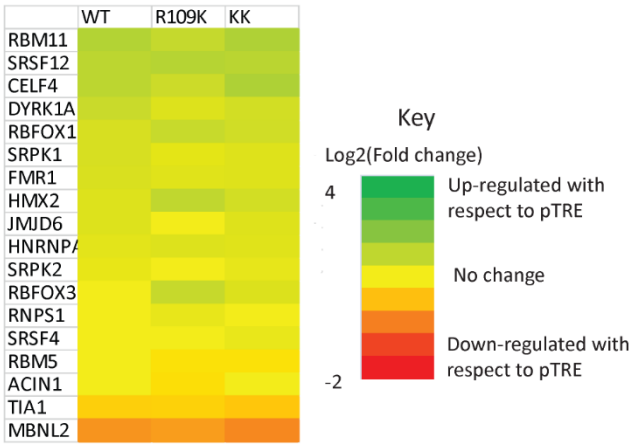


Figure 4.10 Gene ontology analysis of splicing changes.

A. Genes identified in the DEXseq analysis were analysed in DAVID 6.7 functional annotation (BP_FAT) to determine gene ontology and pathway enrichment $p<0.05$.

B. Heatmap showing differential expression compared to pTRE of splice related genes (identified from GO terms) identified in the RNA-seq (extracted from Figure 4.4).

the differential expression analysis data to identify any potential target genes which were up, or downregulated with respect to pTRE and how expression of the different E2F1 mutant derivatives impacted their expression [Figure 4.10B]. The heatmap colouring and key is taken from the heatmap showing all differential expression comparisons in Figure 4.4 and showed that there were few splicing factors differentially expressed compared to pTRE, and those that were, show only minor changes in expression (from 1.5 log2fold change to -1.4 log2fold change). Additionally, there are no clear differences in expression between the different cell lines, particularly WT and KK, suggesting that transcriptional changes in splicing factor expression is not the reason for the changes we observed in exon use changes.

4.6 Validation of alternative splicing

To confirm the exon use changes identified in the RNA-seq under the different cell lines we identified two genes, RBM38 (a known E2F1 target gene (Feldstein et al., 2012)) and SAGE1 (a novel E2F1 target gene) to follow up. Additionally, we included two known E2F1 target genes which have well studied splice isoforms with assigned functions – TP73 and BCL-X (also known as BCL2L1). Details about these are outlined in Table 4.3 below and in Figure 4.11. Although the latter two genes were not found in the RNA-seq analysis, we felt that given their dual functions in apoptosis, with TAp73 and BCL-XS being proapoptotic and the other isoforms, BCL-XL and multiple versions of DNp73, being antiapoptotic these may well be targets of methylation dependent E2F1 induced splicing changes. It is important to note that although we focussed on the well characterised isoforms, both BCL-X and TP73 have additional splice isoforms. RBM38 and SAGE1 also have additional isoforms, but these are non-protein coding and we focussed on those isoforms most likely to be changing according to the RNA-seq analysis.

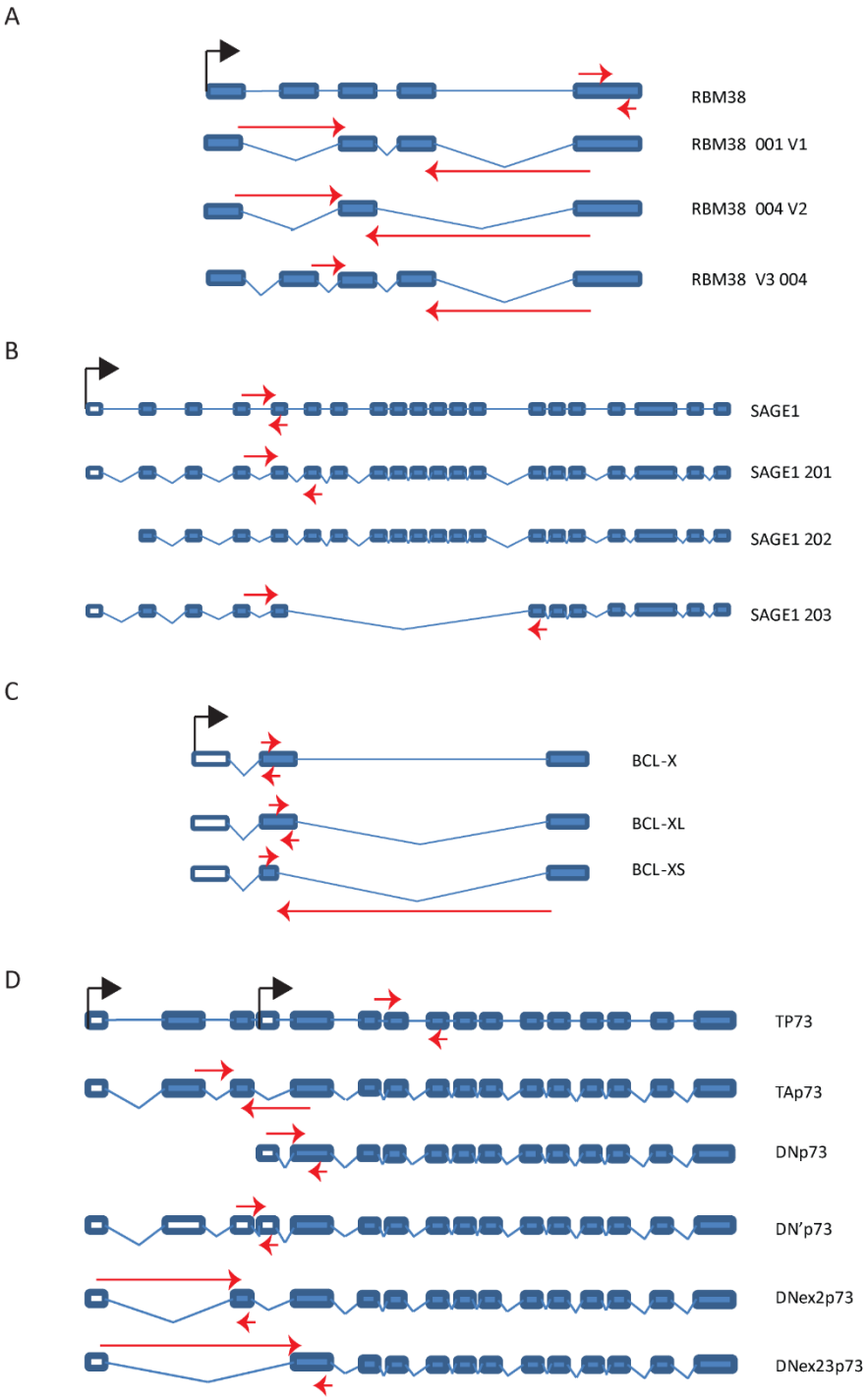


Figure 4.11 Gene structure and major splice isoforms of alternatively spliced genes identified in the RNA-seq analysis.

Unshaded boxes represent untranslated exons. Primers are shown in red.

A. RBM38 gene and three protein coding isoforms.

B. SAGE1 gene and two protein coding isoforms.

C. BCL-X gene and two splice isoforms – use of a cryptic splice site in BCL-XS truncates exon 2.

D. TP73 gene and a selection of protein coding isoforms.

Table 4.3 List of genes showing altered exon use and their properties

Gene name	Function	Isoforms (protein coding)	Isoform function	DEXseq predictions
RBM38	RNA binding motif protein 38, stabilises CDKN1A mRNA to induce cell cycle arrest in a p53 dependent manner	**V1** **V3** **V2**	V1 is full length. V3 contains a truncated RNA binding motif so may be non-functional. V2 is missing exon 3 but this has not been annotated with a domain so function is unknown	KK and WT more likely to express V3 than R109K.
SAGE1	Sarcoma Antigen 1, function unknown but highly expressed in tumours	**201** **202** **203**	201 is full length and differs from 202 only in an un-translated exon. 203 has deletion of 8 WD40 repeats	R109K more likely to drive expression of full length (201/202).

				KK more likely to express 203 and least likely to express 201/202.
TP73	P73 is a proapoptotic transcription factor	**TAp73** **DNp73** **DN'p73** **DNex2p73** **DNex23p73**	TAp73 is the full length isoform with proapoptotic activity. DNp73 (all versions) are truncated, with a loss of the trans-activation domain, making it anti-apoptotic.	N/A
BCL-X (BCL2L1)	BCL2 like protein, involved in apoptotic signalling at the mitochondria	**BCL-XL** **BCL-XS**	BCL-XL is an anti-apoptotic protein which sequesters BH3-only proteins	N/A

	BCL-XS is a proapoptotic BH3 domain containing protein.

Primers were designed to be specific for each splice isoform as well as including primers which would amplify all splice isoforms to enable normalisation to the total levels of gene expression to account for any alterations in expression under the different E2F1 mutant cell lines, or treatments. The total levels of mRNA for each gene were normalised to actin, to determine gene expression changes upon expression of E2F1 and the mutant derivatives whereas the levels of isoform expression were normalised to the total amount of each RNA species, to adjust for any changes in overall gene expression [Figure 4.12]. For RBM38, total levels of RNA were similar across all cell lines, albeit slightly lower in KK. The major isoform across all three cell lines was V3, but expression of this was reduced in R109K compared to WT or KK. Additionally, KK showed a minor increase in V1 and V2. Similarly to RBM38, SAGE1 total RNA levels were equal across the cell lines. SAGE1 201 was enhanced in R109K compared to both WT and KK and this corresponded with a slight reduction in SAGE1 203. For BCL-X, again total levels of mRNA were similar across all three cell lines expressing E2F1, and the differences in isoform expression were small for BCL-XL. However the proapoptotic version, BCL-XS was increased in KK, where it was expressed at equal levels to BCL-XL. Conversely, BCL-XS showed the lowest expression in R109K expressing cells and this expression was lower than the expression levels of BCL-XL. This change in ratio between XL and XS may be one reason for the higher basal levels of apoptosis seen in KK and it could be anticipated that it sensitises cells to apoptosis. The trend of KK expressing higher levels of proapoptotic isoforms was additionally observed for p73 as well – TAp73 is upregulated in KK compared to WT or R109K, whereas

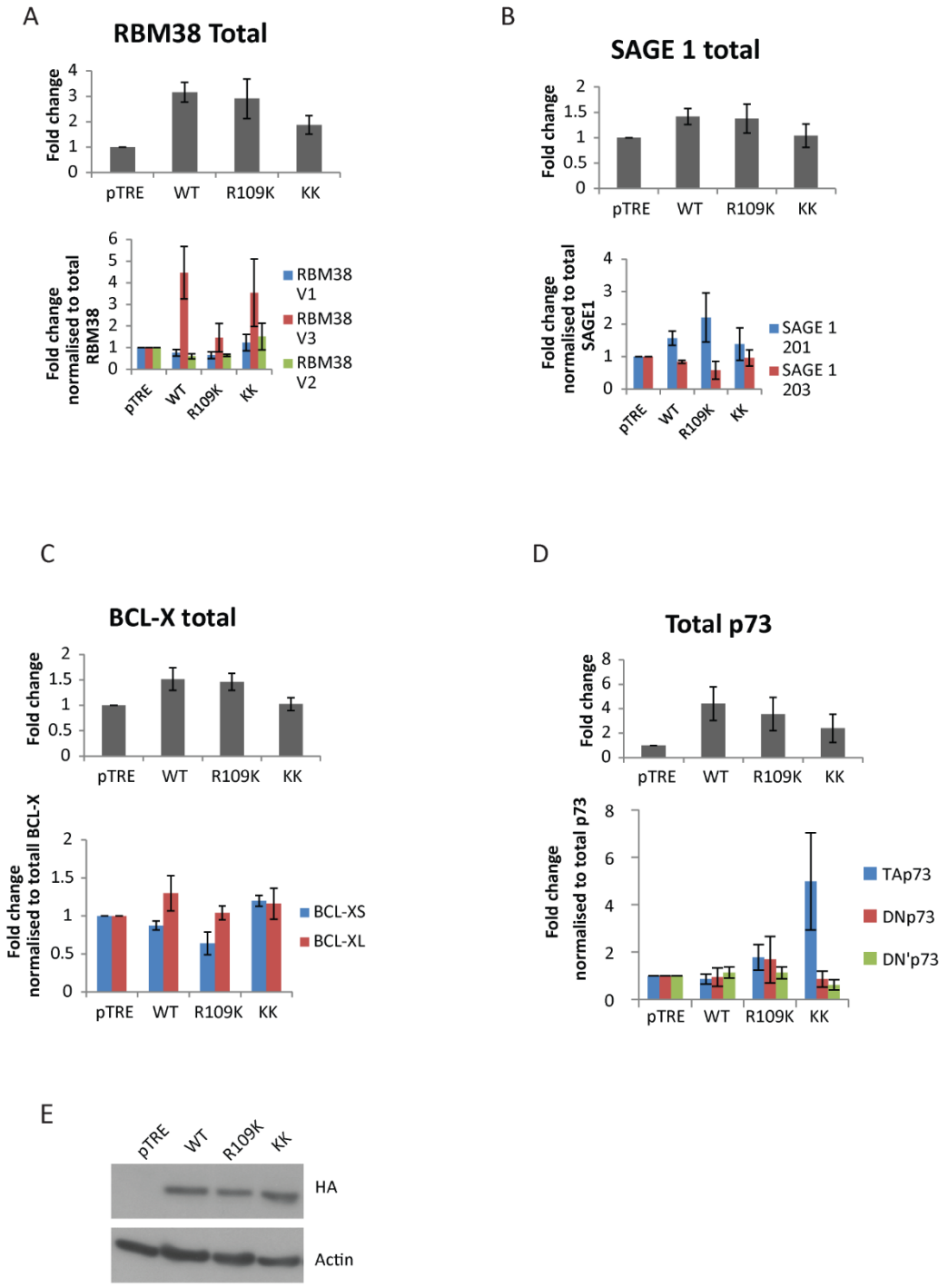


Figure 4.12 Alternative splicing in E2F1 mutant derivative expressing cell lines.

E2F1 expression was induced with 1µg/ml doxycycline for 24h before harvest. RNA was extracted with TRIzol and cDNA synthesised before qPCR with primers as indicated. Shown is average of three independent experiments. Total RNA levels were normalised to Actin, splice isoforms were normalised to total RNA of that species. A. RBM38. B. SAGE1. C. BCL-X. D. TP73. E. Protein expression was measured via immunoblot using anti-HA to measure induced E2F1 expression and anti-β-actin as a loading control.

DNp73 and DN'p73 remain fairly stable across all cell lines. Primers for DNex2p73 and DNex23p73 were unsuccessful in producing a single PCR product so were not included in this analysis.

To understand the role of E2F1 in isoform selection further, E2F1 and p100-TSN (the reader of the R111/113Me2s mark) were depleted using siRNA and levels of isoform expression were analysed [Figure 4.13]. Oddly, despite overexpression of E2F1 inducing expression of RBM38 V1, silencing E2F1 expression did not have the contrasting effect, as all three isoforms were slightly elevated after both E2F1 and p100-TSN depletion [Figure 4.13A]. However, for SAGE1 there was a reduction in 201 isoform levels after treatment with siE2F1, and to a lesser extent a decrease in 203, which is what we would expect as the overexpression increased 201 and had a minimal effect on levels of 203 [Figure 4.13B]. Interestingly, loss of p100-TSN increased levels of 201, but not 203, suggesting that expression of these isoforms may be differentially regulated by the tudor domain containing protein [Figure 4.13B]. For BCL-X, depletion of E2F1 and, to a lesser extent, p100-TSN resulted in an increase in BCL-XL expression, and a slight reduction in BCL-XS although this was not significant [Figure 4.13C]. Depletion of E2F1 resulted in a significant reduction in TAp73 expression, and a minor increase in DNp73 and DN'p73. DNp73 also showed a slight increase in expression after p100-TSN depletion, but in this case DN'p73 showed lower expression [Figure 4.13D]. This follows the trend observed in the inducible cell lines where TAp73 appeared to be the most differentially regulated isoform [Figure 4.12D], however we might have anticipated that the siTSN treatment acted more like the KK cell line – resulting in an increase in TAp73, rather than a reduction.

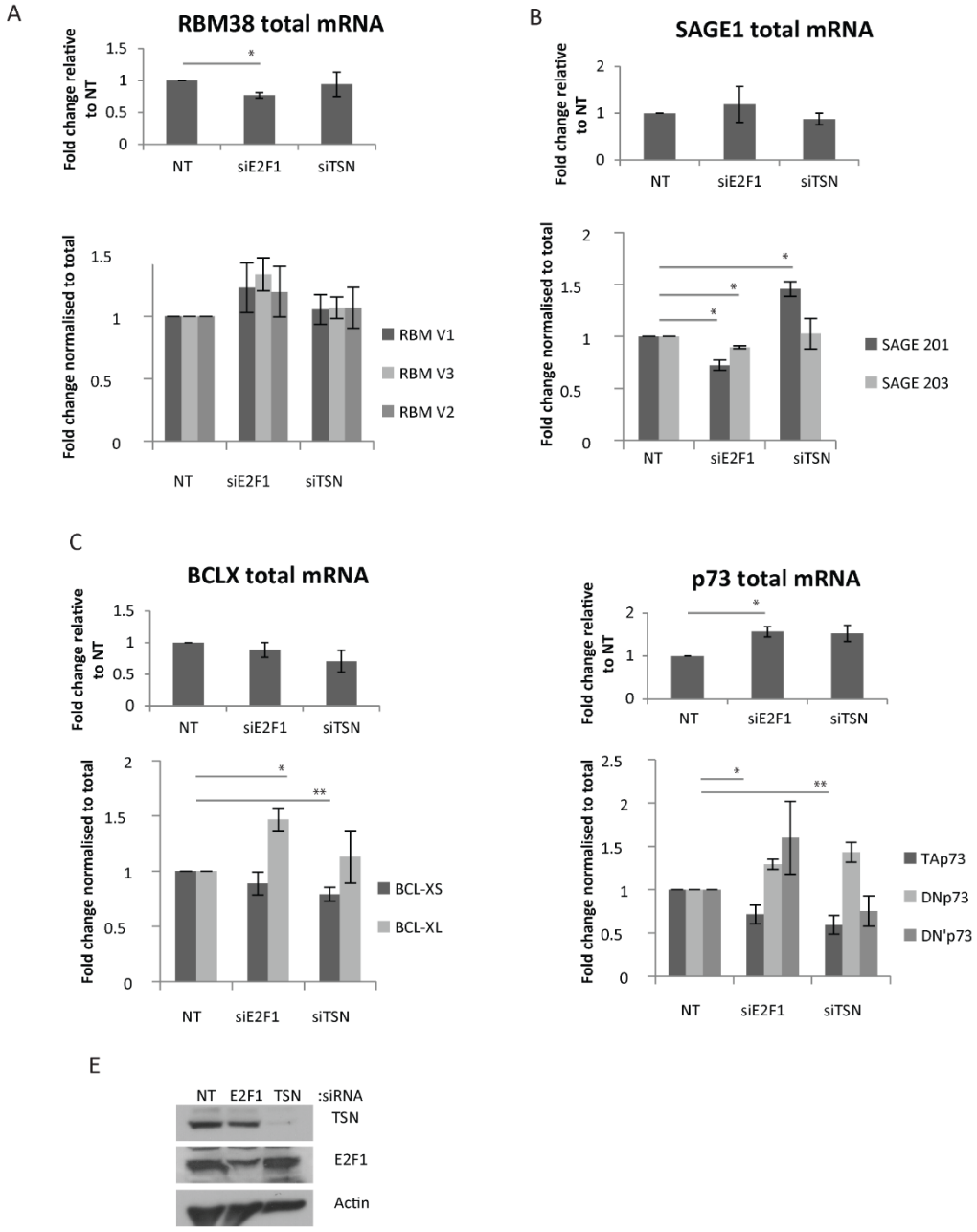


Figure 4.13 Isoform expression is altered after depletion of E2F1 or p100-TSN.

U2OS cells were treated with siRNA against E2F1, p100-TSN or NT control for 72h. RNA was extracted with TRIzol and cDNA synthesised before qPCR with primers as indicated. Total RNA levels were normalised to Actin, splice isoforms were normalised to total RNA of that species. Average of three independent experiments * $p<0.05$, ** $p<0.005$. A. RBM38. B. SAGE1. C. BCL-X. D. TP73. E. siRNA efficiency was measured via immunoblot using anti-E2F1 or anti-p100-TSN specific antibodies and anti-β-actin as a loading control.

4.7 Chapter Summary

After carefully ensuring we had good quality biological replicates and RNA quality we successfully carried out an RNA sequencing experiment to identify further changes in gene expression between wild-type E2F1 and the mutants R109K, and KK. R109K cannot be methylated by PRMT1, and therefore is likely to be methylated at R111/113 by PRMT5; conversely, KK cannot be methylated by PRMT5 and is likely to be methylated at R109 by PRMT1 (Zheng et al., 2013).

Overall we observed that there were greater differences between the gene expression profiles of WT and R109K or KK and R109K than between WT and KK which had a very similar subset of target genes. This was observed in the validation experiments where expression levels of potential E2F1 target genes identified in the RNA-seq were often similar between WT and KK, with R109K showing, usually a reduction in gene expression compared to these. In two cases we observed that R109K was able to enhance expression of two target genes above levels observed for WT and KK. This is consistent with the observation that silencing of PRMT1 generally reduces E2F1 target gene expression (Zheng et al., 2013). Our results here expand this to include novel target genes and also pose a role for methylation at R109 by PRMT1 to have inhibitory effects in specific cases, such as that seen for KCNIP and DEF6. Whilst we identified many potential new target genes for E2F1 which showed changes in expression upon overexpression of E2F1 or the mutant derivatives, and for many of these depletion of E2F1 resulted in a reduction of the target gene expression, we were unable to unequivocally prove these are direct target genes. ChIP would need to be carried out to determine if E2F1 could bind to the promoter regions of these target genes in order to confirm these are direct targets of E2F1, and not indirectly regulated. However, due to time constraints we were unable to confirm this and further studies are required to determine if the genes identified here are legitimate targets of E2F1.

Additionally, we identified many changes in exon use between the different cell lines, with KK showing the greatest difference in exon use to WT. This suggests that whilst R109K (or inhibition of PRMT1 mediated methylation of E2F1) impacts gene expression, KK (or inhibition of PRMT5 mediated methylation of E2F1) is having a greater impact on isoform use of a shared subset of target genes which show little differential expression changes. We were able to confirm that, for a selection of E2F1 target genes, the methylation status of E2F1 has an impact on isoform selection and that E2F1 may be playing a role in alternative splicing decisions via PRMT5 mediated symmetric dimethylation of R111/113.

Chapter 5. E2F1 and p100-TSN can bind to E2F1 target gene RNA as well as components of the spliceosome

5.1 Introduction

The effects we see on isoform expression of E2F1 target genes, such as increased expression of BCL-XL, and TAP73 in KK overexpressing cells, as well as the differences in exon use observed in the RNA-seq experiment between WT and KK suggested there may be a role for arginine methylation at R111/113 by PRMT5 in mediating isoform selection and our subsequent work focussed on this, rather than PRMT1 mediated R109 methylation. Given that we did not see differential expression of various splicing factors, we hypothesised that the change in splicing may be due to the presence of p100-TSN at the promoters of E2F1 target genes when E2F1 is symmetrically dimethylated. P100-TSN could then act as a scaffold to recruit RNA processing factors to the nascent RNA as transcription starts. To this end, we wished to investigate whether E2F1 itself, through p100-TSN, could associate with RNA of E2F1 target genes and the dependency of this on PRMT5 mediated methylation. Furthermore we wished to examine the impact of E2F1 on the ability of p100-TSN to bind to target gene RNA.

In order to study protein-RNA interactions we designed and optimised an RNA-immunoprecipitation (RIP) protocol [Figure 5.1]. Once cells are ready to harvest, they must be cross-linked. UV crosslinking is the most widely used method used for RNA cross-linking with formaldehyde being a second option. Although formaldehyde is more efficient, it is also a potent DNA-protein cross-linker. In contrast, UV cross-linking is highly selective for RNA-protein complexes over DNA-protein complexes despite having much lower efficiency (Brimacombe et al., 1988). As both E2F1 and p100-TSN are DNA associated it was important to limit the DNA contamination of the samples to ensure no false positives were observed, so UV cross-linking was chosen. Following cross-linking, immunoprecipitation using specific antibodies or non-specific IgG as a control is then carried out. So far CLIP (cross-linking immunoprecipitation) techniques have been used to identify direct RNA-protein

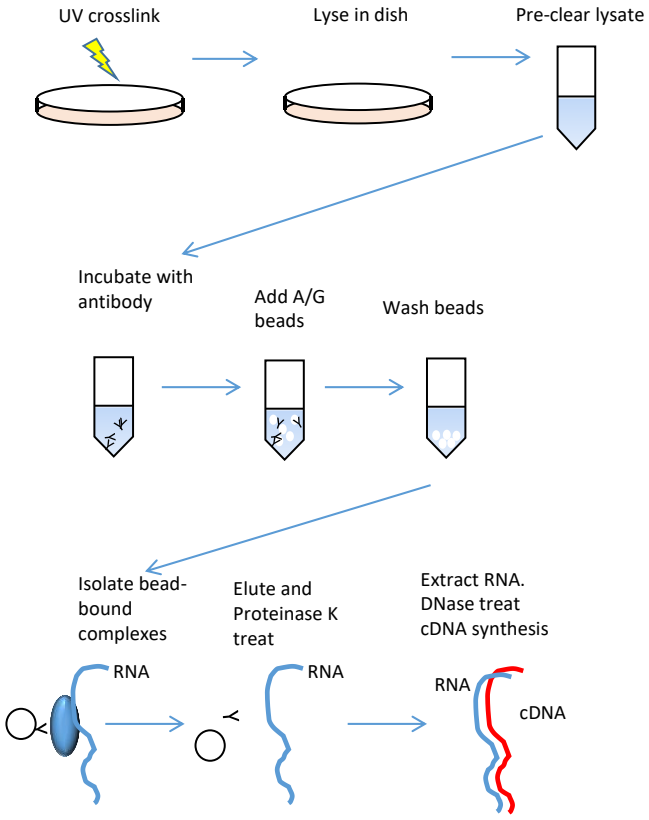


Figure 5.1 Diagram of RNA-immunoprecipitation protocol.

Diagram showing RNA-IP (RIP) experiment set up. Prior to harvest, cells are subject to UV cross-linking. The lysates are pre-cleared before addition of primary antibody. Protein A and G beads are added for immunoprecipitation. The beads are washed, the RNA is eluted and proteinase K treated. RNA is extracted using TRIzol before DNase treatment, cDNA synthesis and qPCR analysis.

interactions so the stringency of the lysis buffer and immunoprecipitation buffer is high to eliminate any protein-protein interactions to identify only directly bound RNA species (CLIP protocol reviewed in (Darnell, 2010)). As we did not hypothesise that E2F1 would bind directly to RNA, but rather in a complex with p100-TSN, lower stringency was required, so gentler buffers were used and the immunoprecipitated complex was not run on an agarose gel before RNA isolation. Instead, the complexes were eluted directly from the beads and proteinase K treated before the RNA was isolated using TRIzol, similar to a protocol adapted from (Niranjanakumari et al., 2002). As both E2F1 and p100-TSN are chromatin associated proteins, the RNA was extensively DNase I treated before cDNA synthesis to ensure only RNA was measured.

5.2 p100-TSN and E2F1 can bind to E2F1 target gene RNA

Firstly, to determine if p100-TSN can bind to E2F1 target gene RNA and to ensure the RIP protocol was stringent, after cross-linking, the samples were divided in two and RNAse A was added to the lysis buffer of one sample during cell lysis and immunoprecipitation with the other being treated with lysis buffer alone. p100-TSN was immunoprecipitated and the RNA extracted, and subsequently the RNA was subjected to extensive DNase treatment before qRT-PCR of the cDNA of E2F1 target genes [Figure 5.2A]. We found that p100-TSN could specifically associate with a selection of E2F1 target gene RNA relative to the IgG control, including TAp73, DNp73 and cyclin D1. Importantly, this was abolished in the RNase A treated samples suggesting this was a specific interaction, and was not due to chromatin contamination.

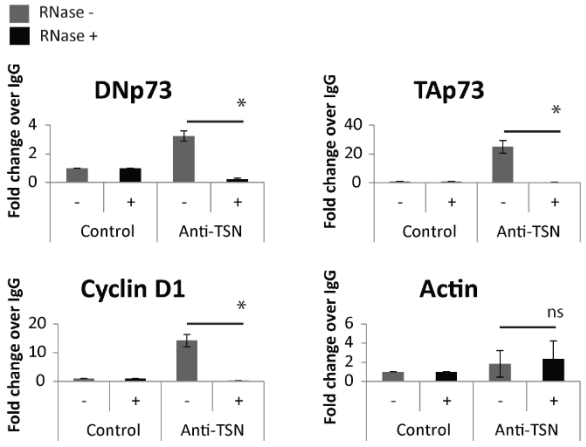
To determine if E2F1 could also associate with the nascent RNA, an E2F1 RIP was performed, using the same target gene RNA as we had successfully seen interacting with p100-TSN. Again, RNase A treatment was carried out on half of the samples during lysis and immunoprecipitation as a negative control [Figure 5.2B]. It was found that E2F1 could indeed

be used to isolate RNA of these target genes (DNp73, TAp73 and cyclin D1) relative to the IgG control, and once again this was significantly impaired in the RNase A treated lysates, confirming isolation of RNA rather than chromatin. Importantly, the E2F1 could not recruit actin RNA (not an E2F1 target gene) suggesting this is observed only for E2F1 target genes, and is not a general function of E2F1.

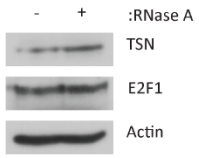
5.3 The E2F1-RNA interaction is dependent on p100-TSN

Given the role of p100-TSN in splicing and the changes in exon use observed upon mutation of the p100-TSN binding site on E2F1, we wished to know if this E2F1-RNA interaction was dependent its interaction with p100-TSN, or if the role of p100-TSN as a co-transcriptional activator was separate to its role in splicing. To this end, U2OS cells were treated with siRNA against E2F1 and p100-TSN, including a non-targeting control, before RIP was carried out [Figure 5.3]. The siE2F1 treatment was included as a second negative control to confirm no unspecific binding of RNA to the E2F1 antibody. Immunoblots confirmed silencing of both proteins, and also indicated successful isolation of both E2F1 and p100-TSN under RIP conditions, which was abolished after siRNA mediated depletion [Figure 5.3A]. E2F1 could be observed to bind to RNA species of its target genes: TAp73, DNp73 and Cyclin D1, but not to actin, which is not an E2F1 target gene [Figure 5.3B]. Additionally, this was abrogated after silencing of E2F1, and the reduction of this interaction was significant across the three target genes which confirmed a specific interaction with E2F1 protein and not an off target effect of the antibody. Furthermore, this interaction was also inhibited after silencing of p100-TSN which suggested that p100-TSN was required for E2F1 to be to be in an RNA bound complex. It also confirmed that the interaction of E2F1 with RNA was unlikely to be a direct interaction

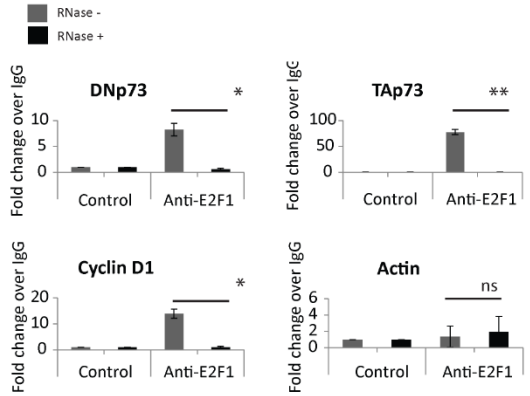
A i)



ii)



B i)



ii)

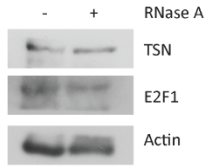


Figure 5.2 E2F1 and p100-TSN can bind to E2F1 target gene RNA.

U2OS cells were grown to 80% confluency before crosslinking. After crosslinking, half the dishes were lysed in normal RIP lysis buffer, the other half were lysed in RIP lysis buffer containing 20µg/ml RNaseA

A. anti-TSN was used to isolate p100-TSN RNA complexes. i) cDNA was analysed by qPCR for the RNA species indicated Results shown are averages of three independent experiments with standard error shown. Significance was determined with a t-test * $p<0.05$, ** $p<0.005$. ii) protein expression of inputs was measured by immunoblotting and probing with anti-E2F1 and anti-p100-TSN with β-actin as a loading control.

B. anti-E2F1 was used to isolate E2F1-RNA complexes. i) cDNA was analysed by qPCR for the RNA species indicated .Results shown are averages of three independent experiments with standard error shown. Significance was determined with a t-test * $p<0.05$, ** $p<0.005$. ii) protein expression of inputs was measured by immunoblotting and probing with anti-E2F1 and anti-p100-TSN with β-actin as a loading control.

and more relies on a tertiary complex with p100-TSN forming the scaffold between transcription factor and nascent RNA.

However, it was possible that loss of p100-TSN could have the above effect if silencing of the co-transcriptional activator was causing E2F1 to be lost from the promoters of these target genes. To investigate this, an E2F1 ChIP was carried out in the presence and absence of p100-TSN, to investigate the role of p100-TSN on E2F1 target gene promoter occupancy [Figure 5.4]. Depletion of p100-TSN by siRNA was seen to impact on E2F1 presence on target genes promoters in a gene specific manner. Whilst loss of p100-TSN reduced E2F1 presence at the CDC6 promoter, it had no effect on E2F1 binding at the BCL-X promoter, and increased E2F1 recruitment to the p73 promoter, although these were not significant changes [Figure 5.4]. There was no recruitment of E2F1 to the Cyclin D promoter under these conditions. Given the results of the RIP suggested that loss of p100-TSN inhibited E2F1 binding to p73 RNA [Figure 5.3], the ChIP data imply that this is not through mere loss of E2F1 at the promoter of this target gene, but was due to a more direct inhibition of the association of E2F1 with RNA. This also suggested that the E2F1 RNA interaction is not merely through the chromatin-RNA hybrid during transcription as increased E2F1 binding to the p73 promoter in the absence of p100-TSN would be expected to also increase the interaction between E2F1 and p73 RNA if the RNA interaction was occurring through the DNA-RNA hybrid.

p100-TSN recognises and binds to the symmetrically dimethylated R111/113 (PRMT5 mediated) marks on E2F1, therefore it would be expected that inhibition of PRMT5, the enzyme responsible for these methylation events, would mimic the effect of silencing p100-TSN. Several PRMT5 inhibitors have been designed and published – the commercially available Epizyme compound EPZ015666 has been shown to be both potent and selective for PRMT5 *in vitro* and *in vivo* (Chan-Penebre et al., 2015). A colleague investigated the IC50 for this compound and a closely related analogue, J579-44, designed in house. Details of the

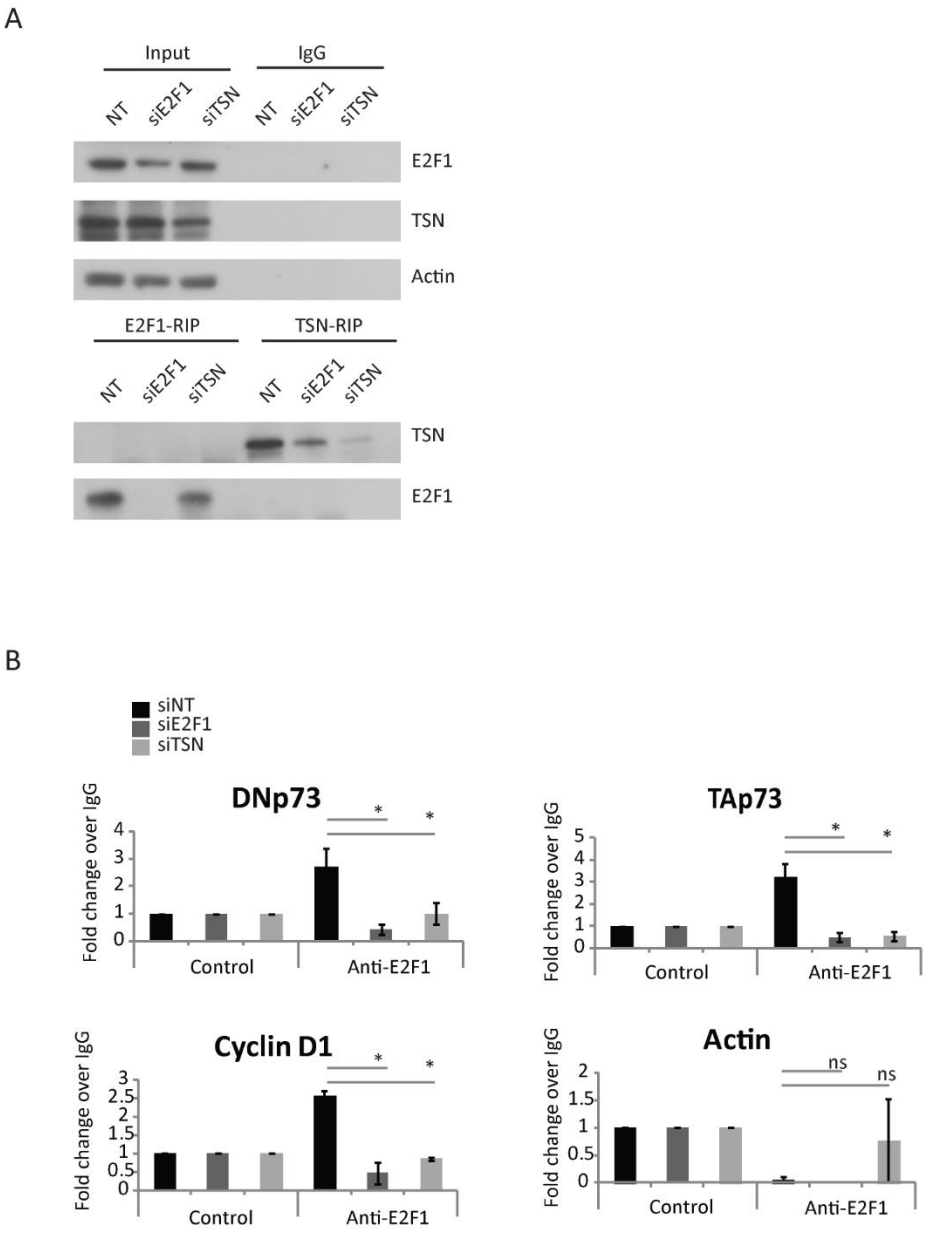


Figure 5.3 Silencing of p100-TSN inhibits the ability of E2F1 to bind to target gene RNA.

U2OS cells were transfected with siRNA against E2F1, p100-TSN or NT control for 72h. Cells were subject to UV crosslinking before lysis and RIP using anti-E2F1 was carried out.

A. Input samples, along with aliquots of the RIPs underwent immunoblotting and probed with with anti-E2F1 and anti-p100-TSN to confirm successful silencing of the proteins and protein pull-down under RIP conditions, with anti-β-actin used as a loading control.

B. Anti-E2F1, RIP cDNA was analysed by qPCR for the RNA species indicated. Results shown are averages of three independent experiments with standard error shown * $p<0.05$, ** $p<0.005$.

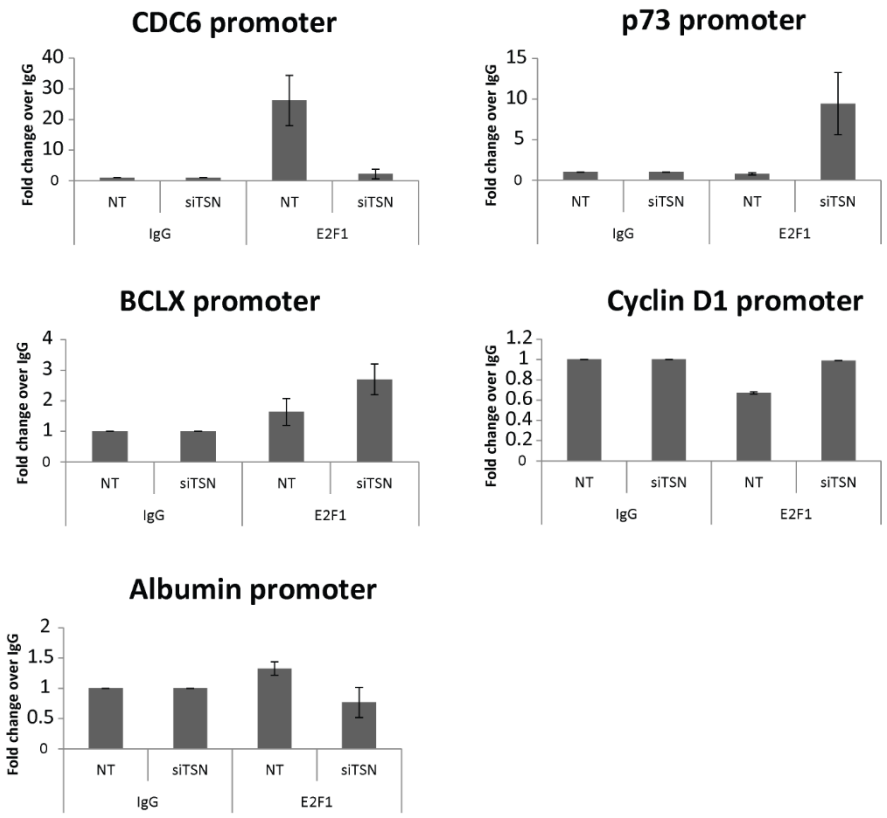
compounds are included in Figure 5.5A for reference.

Symmetric methylation of H4R3 was used as a marker for PRMT5 inhibition as PRMT5 is the major enzyme responsible for this post-translational modification and specific antibodies against this methyl-mark are available. Optimisation had shown that for U2OS this concentration of drug over this time period resulted in inhibition of PRMT5 – confirmed by immunoblot showing a reduction in for H4R3 methylation – but limited toxicity (data not shown). U2OS cells were treated with 5µM J579-44, or the equivalent volume of DMSO for 72h before an E2F1 RIP was carried out. There was a reduction in H4R3me2s signal after PRMT5 inhibition, whereas the total histone H4 levels remain stable [Figure 5.5B i]. In cells treated with DMSO, E2F1 was able to interact with target gene RNA, including TAp73, DNp73 and Cyclin D1 but there was no specific interaction with actin RNA, once again confirming specific interaction of E2F1 with target gene RNA [Figure 5.5B ii]. Inhibition of PRMT5, abrogated the interaction of E2F1 with all three target gene RNA species, which resulted in significant reductions in the amount of RNA present in the RIP. Thus, methylation at R111/113 was shown to be required for E2F1 association with target gene RNA, which was consistent with the role of p100-TSN in the RNA binding ability of E2F1.

5.4 E2F1 requires an intact DNA binding domain to bind to RNA

As most splicing is believed to occur co-transcriptionally, and E2F1 is a transcription factor, we reasoned that this interaction between E2F1 and RNA would be also happening co-transcriptionally at the chromatin. To confirm this, fractionation of cells subsequent to UV cross-linking and before RNA IP was attempted, to identify the cellular fraction the interaction was occurring in. However, it was discovered that the UV cross-linking may be too

A



B

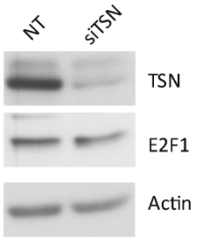


Figure 5.4 Silencing of p100-TSN alters E2F1 binding to chromatin of target genes.

U2OS cells were treated with 25nmol siRNA against p100-TSN or NT control for 72h before crosslinking using 1% formaldehyde. Cells were lysed and DNA fractionation was carried out via sonication before immunoprecipitation using anti-E2F1. Subsequently the DNA was purified.

A. Chromatin was analysed by qPCR for the promoters indicated. Average of three independent experiments with standard error shown * $p<0.05$, ** $p<0.005$.

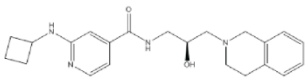
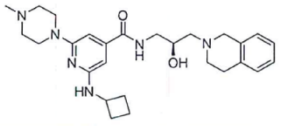
B. Immunoblotting was carried out to confirm successful silencing of p100-TSN and to ensure equal levels of E2F1 in both conditions. B-actin was used as a loading control.

harsh for subsequent fractionation and resulted in poor separation of the fractions [results not shown]. The subsequent approach was to create DNA-binding deficient mutants of E2F1.

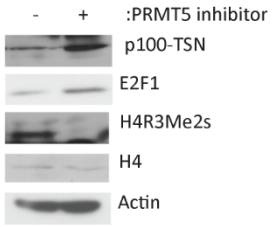
Two point-mutants of E2F1, L132E and R166H, have been described to be unable to bind to DNA (Woods et al., 2007, Yu et al., 2011, Cress et al., 1993, Campanero and Flemington, 1997). Mutagenesis of a HA-tagged E2F1 pcDNA3.1 construct was carried out and immunofluorescence was performed to confirm their expression remained nuclear [Figure 5.6A]. Subsequently, ChIP using anti-HA was carried out to verify these mutants were unable to bind to a selection of E2F1 target gene promoters [Figure 5.6B]. WT-E2F1 was recruited to the promoters of a selection of target genes, CDC6, p73, E2F1, and to a lesser extent Cyclin D1, relative to that of the empty vector control, and this recruitment was greater than that observed for the non-target gene promoter, albumin, which was included as a negative control. In contrast, both R166H and L132E mutants were not recruited to any of these promoters relative to background IgG and showed no difference in recruitment compared to albumin, despite an enhanced protein expression compared to WT [Figure 5.6B ii].

Following confirmation that the DNA binding domain mutants of E2F1 were unable to bind to target gene promoters, equal amounts of pcDNA, WT-E2F1, L132E-E2F1 and R166H-E2F1 were transfected into U2OS cells. Cells were harvested after 48 hours and RIP carried out using anti-HA to isolate ectopically expressed E2F1 or the DNA binding domain point mutants [Figure 5.7A]. Once again the immunobloting showed WT-E2F1 was expressed at lower levels than L132E-E2F1 or R166H-E2F1, despite transfection of equal amounts of plasmid DNA, consistent with previous reports that these mutants are more stable than the WT protein (Campanero and Flemington, 1997, Cress et al., 1993) [Figure 5.7B]. However, only WT-E2F1 was able to interact with three different target gene RNA species (but not the non-target gene actin) investigated compared to the pcDNA control, as the DNA-binding domain mutants failed to isolate RNA from these genes relative to pcDNA. Additionally the difference between the RNA

A

Compound	Structure	Molecular weight	IC50
EPZ015666		380.49	0.0355±0.0567
J579-44		478.63	0.0161±0.0288

B i)



ii)

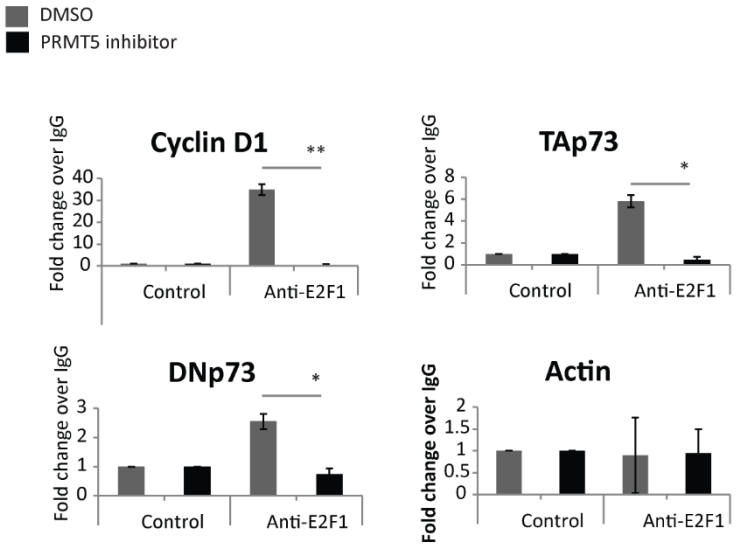


Figure 5.5 PRMT5 activity is required for E2F1 to bind to target gene RNA.

A. Table showing structure, molecular weight and IC50 of PRMT5 inhibitors.

B. U2OS cells were treated with 5μM J579-44 or the equivalent volume of DMSO for 72h. Cells were UV crosslinked before anti-E2F1 RIP was carried out. i) Expression of proteins was determined by immunoblot, methylation of H4R3 (H4R3me2s) was used as an indicator of PRMT5 inhibition, total H4 levels were measured along with E2F1 and p100-TSN, β-actin was used as a loading control. ii) cDNA was analysed by qPCR for the promoters indicated. Average of three independent experiments with standard error shown * $p<0.05$, ** $p<0.005$.

isolated between the WT and the mutants was statistically significant across all RNA species (TAp73, DNp73 and Cyclin D1) measured [Figure 5.7A]. Thus, the ability of E2F1 to bind to, or be bound to DNA was shown to be important for its interaction with specific RNA species.

5.5 E2F1 is able to associate with the splicing machinery

Subsequently, we reasoned that if E2F1 could bind to RNA through a protein with known functions in splicing processes, it may be possible that E2F1 could be found to be associated with the splicing machinery through the same protein – p100-TSN. Previous work by our lab carried out mass spectrometry on E2F1 binding proteins before and after PRMT5 over-expression (Lu, 2014). Briefly, Tet-On cells expressing HA-PRMT5 were generated and E2F1 was immunoprecipitated from doxycycline induced and uninduced cells before isolation of the proteins by gel electrophoresis and silver-staining. In gel-trypsinisation was carried out and the fragments analysed by mass spectrometry. Upon reanalysis of the data, a selection of splicing factors which were present after E2F1 pull-down were discovered (see Table 5.1). Of these, two (highlighted) showed differences between the PRMT5 high-expressing cells and the uninduced cells. PRP31 and SmD1 (also called snRNP D1) were observed to only interact with E2F1 after induction of PRMT5, and were not isolated in the IgG negative control immunoprecipitation. It would be expected that cells expressing high levels of PRMT5 to have increased E2F1-p100-TSN interaction and thus be potentially more RNA bound. The inclusion of proteins involved in splicing in a complex with E2F1 after overexpression of PRMT5 was consistent with this and suggested that the E2F1-p100-TSN-RNA complex may be linked with splicing machinery. SmD1 is an snRNP scaffold protein, which has been shown to be an integral part of the core U1, U2, U4, and U5 snRNP complexes (Chari et al., 2008, Grimm et al., 2013). PRP31 is a key component of the snRNU4/U6 complex and is required for snRNAU5 addition into the tri-snRNP, essential for spliceosome assembly and catalysis (Makarova et al., 2002).

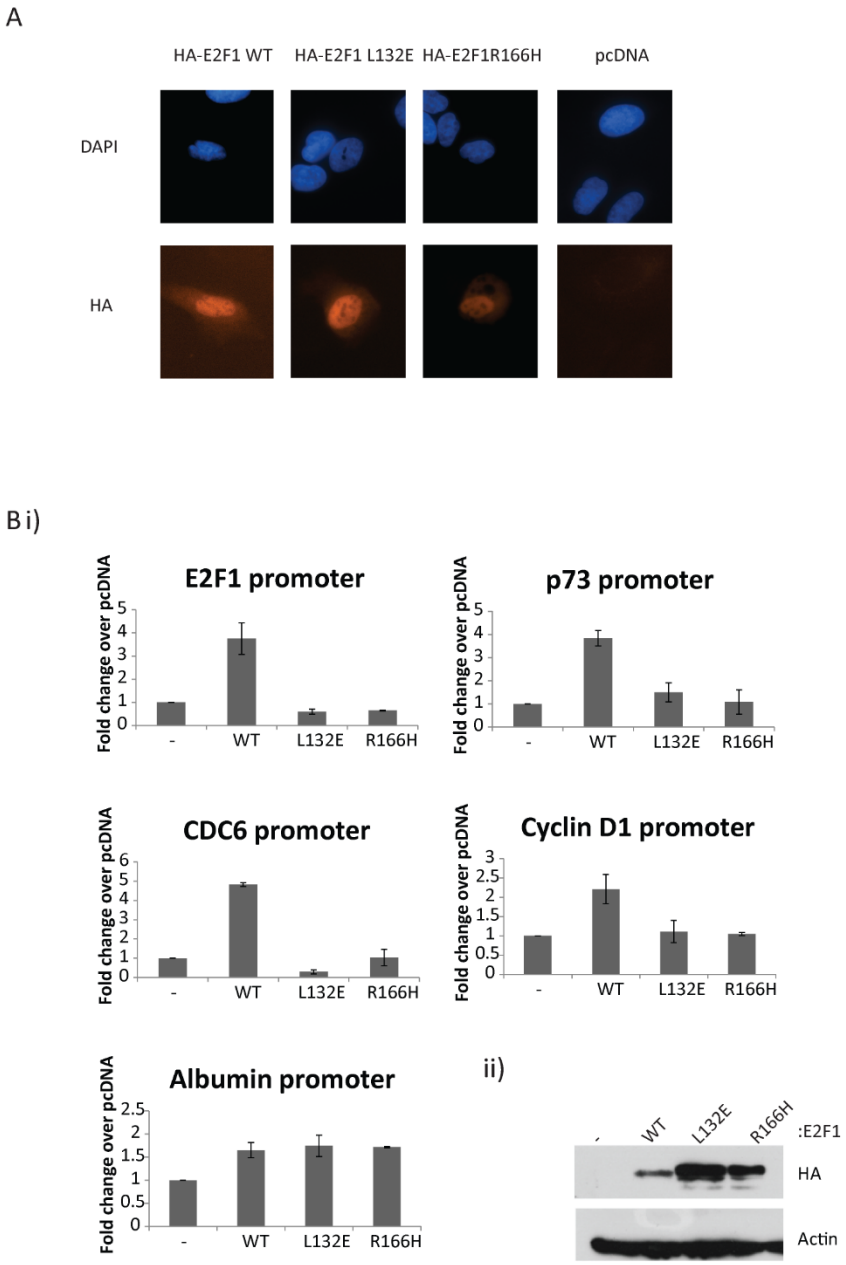


Figure 5.6 Generation and validation of DNA binding domain mutants of E2F1.

Point mutants of E2F1 which are known to abrogate DNA binding, L132E and R166H, were generated.

A. Cells were plated on coverslips and transfected with 1µg of plasmid DNA. Immunoflourescence using anti-HA and fluorescent conjugated secondary antibody was carried out to confirm nuclear localisation of the mutants.

B. 1µg of plasmid DNA was transfected into cells before ChIP was carried out. i) 1% formaldehyde was used to crosslink protein and DNA complexes before cell lysis and immunoprecipitation using anti-HA. DNA was purified and chromatin was analysed by qPCR to determine E2F1 binding to promoters of well known target genes – indicated on the graphs. Average of two independent experiments with standard error shown. ii) Example protein expression of the transfected constructs was determined by immunoblot using anti-HA and β-actin as a loading control.

Unfortunately, a suitable antibody for PRP31 was not found so the protein-protein interaction could not be measured via immunoprecipitation. However, by designing primers to both snRNU4 and snRNU6 it was possible to measure the interaction of E2F1 with the spliceosome via RIP analysis. cDNA produced in the RNase A treated and untreated RIP [Figure 5.2] was amplified with the primers to snRNU4 and U6. RNA for both of these species was identified in the E2F1 RIP and, importantly, this was significantly inhibited after RNase A treatment (n=2) [Figure 5.8A]. Furthermore, when cDNA generated in the RIP after PRMT5 inhibition [Figure 5.5] was used for qRT-PCR analysis, there was also an enrichment in snRNAU4 and U6 binding to E2F1 relative to IgG background which was abrogated after PRMT5 inhibition [Figure 5.8B], although this was not significant for snRNU6. Finally, by using cDNA from the DNA binding domain mutants [Figure 5.7], WT-E2F1 was found to be associated with snRNU6 but this interaction was inhibited in the RIP with the L132E mutant, and reduced in the R166H mutant RIP [Figure 5.8C].

Table 5.1 Spliceosome related proteins identified in the E2F1 mass-spectrometry analysis

Protein identified in Mass-spectrometry analysis	Score	EmPAI
hnRNP E1	177	0.64
hnRNP U	72	0.09
hnRNP H	72	0.16
snRNP D1	71	0.73
snRNP D2	59	0.31
PRP31	57	0.07
hnRNP M	36	0.1
hnRNP K	28	0.08

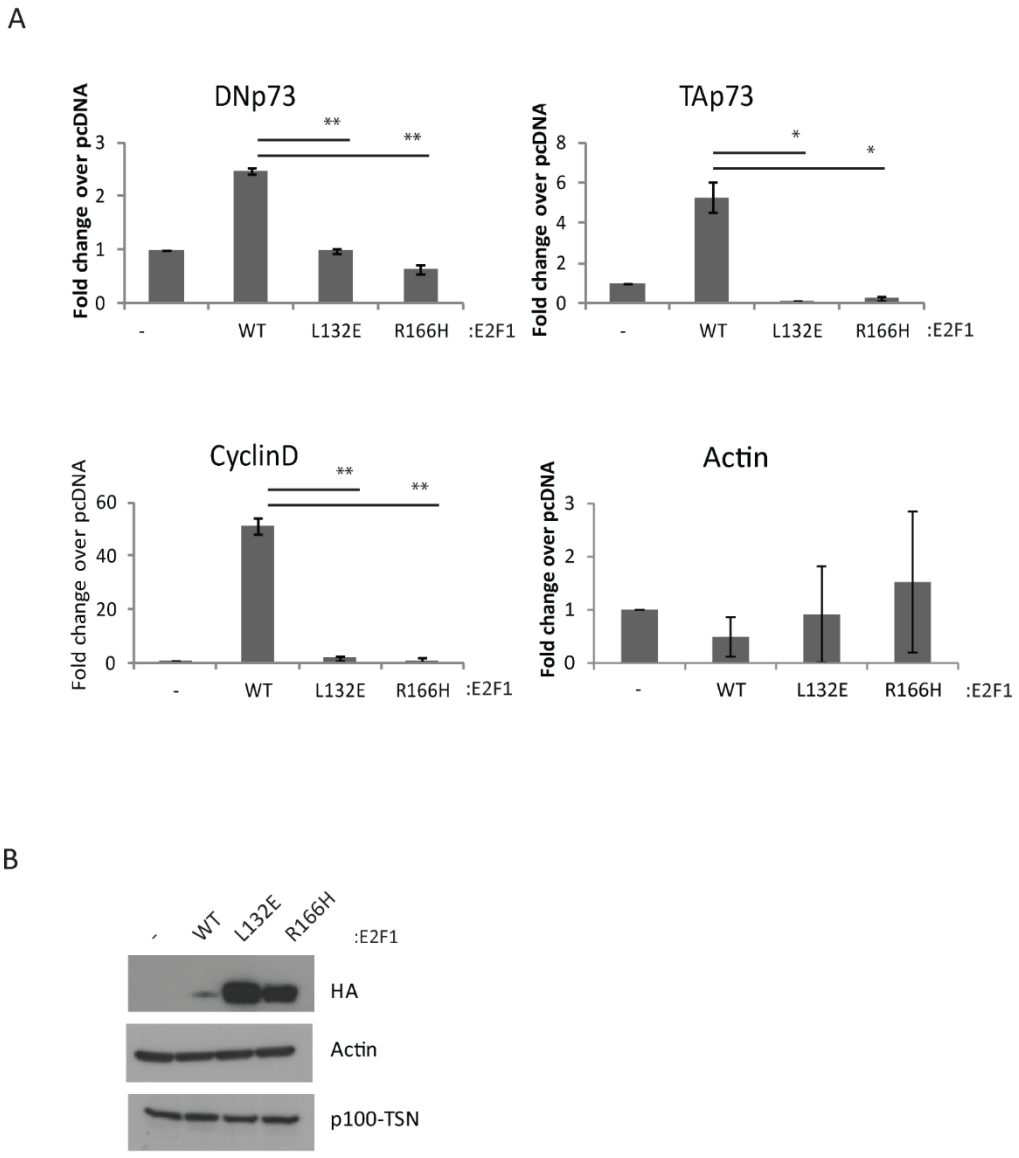


Figure 5.7 DNA binding domain mutants of E2F1 are unable to bind to RNA.

1µg of plasmid DNA, pcDNA, WT-E2F1, L132E-E2F1 or R166GH-E2F1 was transfected into cells for 48h before UV crosslinking and cell lysis. RIP was carried out using anti-E2F1 and RNA extracted.

A. cDNA was analysed by qRT-PCR for the RNA species indicated. Average of three independent experiments with standard error shown. Significance was determined with a t-test * $p<0.05$, ** $p<0.005$.

B. Protein expression was measured via immunoblot using anti-HA and β-actin as a loading control.

5.6 E2F1 presence alters p100-TSN specificity

Having established that E2F1 can associate with RNA in a TSN-dependent manner we next wanted to determine if E2F1 could influence the interaction of RNA species with p100-TSN. A p100-TSN RIP was performed after silencing of E2F1, p100-TSN or a non-targeting control. Once again, depletion of p100-TSN in the pull-down provided an additional negative control to ensure the RNA species identified were being specifically pulled-down by the protein, not by the antibody. Intriguingly, loss of E2F1 appeared to have a profound impact on the species of RNA able to bind to p100-TSN [Figure 5.9A]. The interaction between p100-TSN and DNp73 was significantly reduced after E2F1 silencing, and p100-TSN silencing. However, the interaction between p100-TSN and TAp73, which was not significantly different to siTSN under non-targeting siRNA treatment, increased after E2F1 silencing. This switch in splice isoform interaction after silencing E2F1 was unexpected. We also investigated BCL-XS and BCL-XL to determine if there was a similar trend. Interestingly, BCL-XS only bound to p100-TSN above IgG control in the absence of E2F1, similarly to DNp73 [Figure 5.9A]. Conversely, p100-TSN was able to bind to BCL-XL most strongly in the presence of E2F1, and this was reduced, albeit not significantly, after E2F1 silencing [Figure 5.9A and B]. This suggests an important role of E2F1 and p100-TSN in RNA processing and splicing decisions.

In order to infer a function of the RNA binding properties of E2F1 and p100-TSN in this context, the results seen in the RIP can be compared to the expression data in Figure 4.13. After depletion of p100-TSN, when E2F1 was no longer bound to either DNp73 or TAp73 [Figure 5.9A], under the same conditions there was a reduction in expression of TAp73, despite no significant change in overall p73 expression [Figure 4.13D]. This suggests that p100-TSN may help to stabilise TAp73, or promote splicing of TAp73. Indeed, the increased association between TAp73 and p100-TSN in the presence of E2F1 [Figure 5.9A] further suggests this, as loss of E2F1 was again associated with a reduction in TAp73 expression

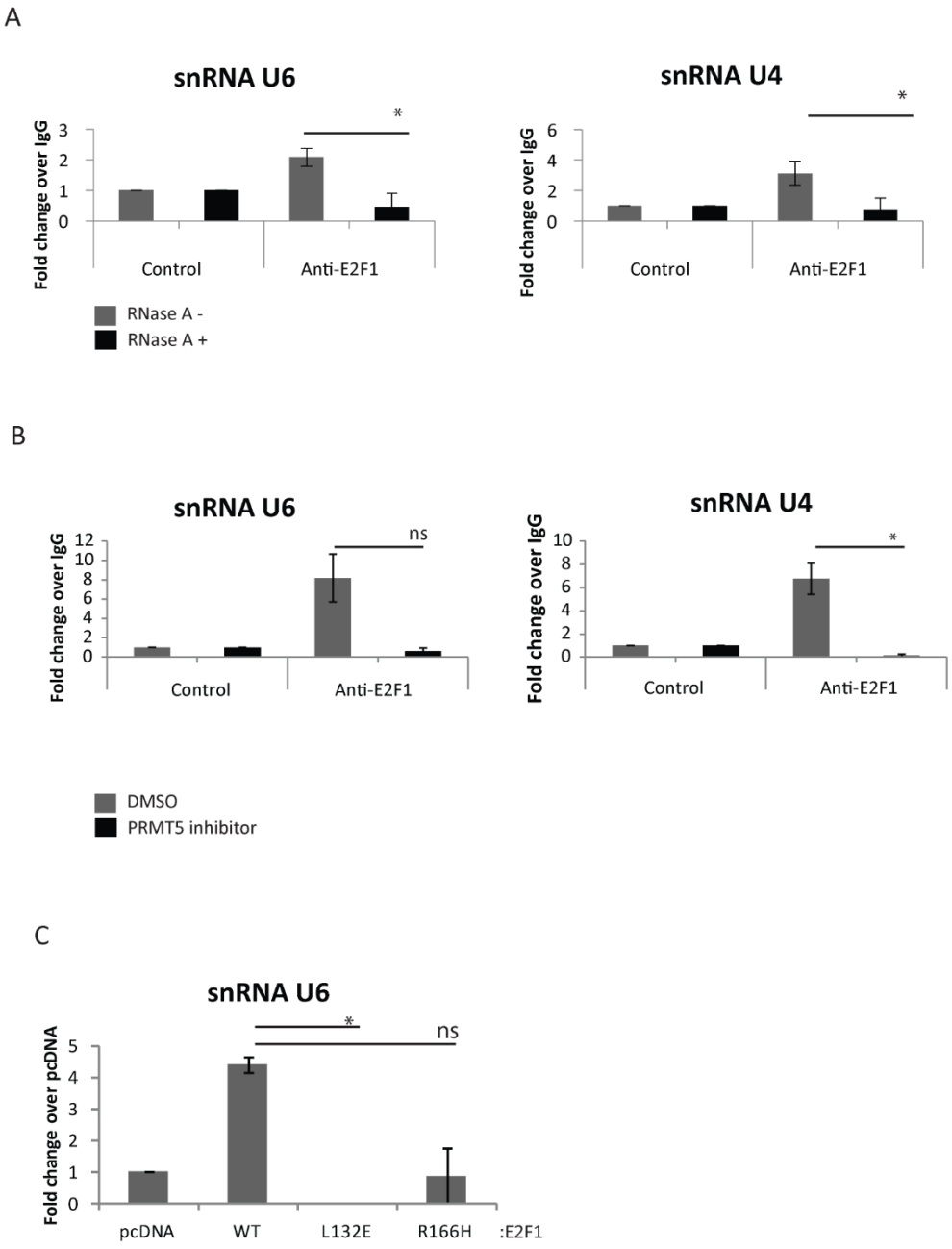


Figure 5.8 E2F1 interacts with snRNU4 and snRNU6.

cDNA from previous RIP experiments was analysed by qPCR to investigate whether E2F1 can interact with RNA elements of the Tri-snNRP complex.

A. cDNA from the RIP treated with RNase A {Figure 5.2] was analysed. Results shown are averages of two independent experiments with standard error shown * p<0.05, ** p<0.005.

B. cDNA from the RIP treated with J579-44 [Figure 5.5] was analysed. Results shown are averages of three independent experiments with standard error shown * p<0.05, ** p<0.005

C. cDNA from the RIP transfected with DNA binding domain E2F1 mutants [Figure 5.7] was analysed. Results shown are averages of three independent experiments with standard error shown * p<0.05, ** p<0.005.

[Figure 4.13D]. Furthermore, in the absence of E2F1, there was a greater interaction of p100-TSN with DNp73 [Figure 5.9A] and this was consistent with a slight increase in DNp73 expression under the same conditions [Figure 4.13D]. Thus there is a strong link to suggest that E2F1 and p100-TSN are promoting expression of TAp73 and, subsequently, loss of either p100-TSN or E2F1 results in an increase in DNp73 expression.

For BCL-X, depletion of E2F1 resulted in an increase in the interaction of BCL-XS with p100-TSN [Figure 5.9A] and under the same conditions we observed a reduction in BCL-XS expression [Figure 4.13C]. Concordantly, we identified a reduction in BCL-XL bound to p100-TSN after E2F1 depletion [Figure 5.9A] and this coincided increased BCL-XL expression [Figure 4.13C]. Therefore, in this case, the interaction of BCL-X RNA with p100-TSN may be inhibitory to expression of a particular isoform depending on the presence of E2F1. Therefore, the role of p100-TSN in regulating isoform expression, dependent on E2F1, is complex and almost certainly is impacted by other factors. Further work is required to fully understand the mechanisms behind this.

5.7 Enhancing the p100-TSN interaction with E2F1 increases RNA binding in DuCaP cells

Previous work in our lab established that the tudor domain containing protein TDRD1 is also able to bind to symmetrically dimethylated E2F1 and competes with p100-TSN for the R111/113me2s mark (unpublished data). TDRD1 is expressed in a hormone-dependent manner only in a subset of breast and prostate cancer cell lines. Therefore, this provides a model to further confirm the E2F1 dependency on p100-TSN for RNA binding. In normal growth media, which contains phenol red (which activates the oestrogen receptor) and normal FBS which contains other hormones, p100-TSN will be competing with TDRD1 for

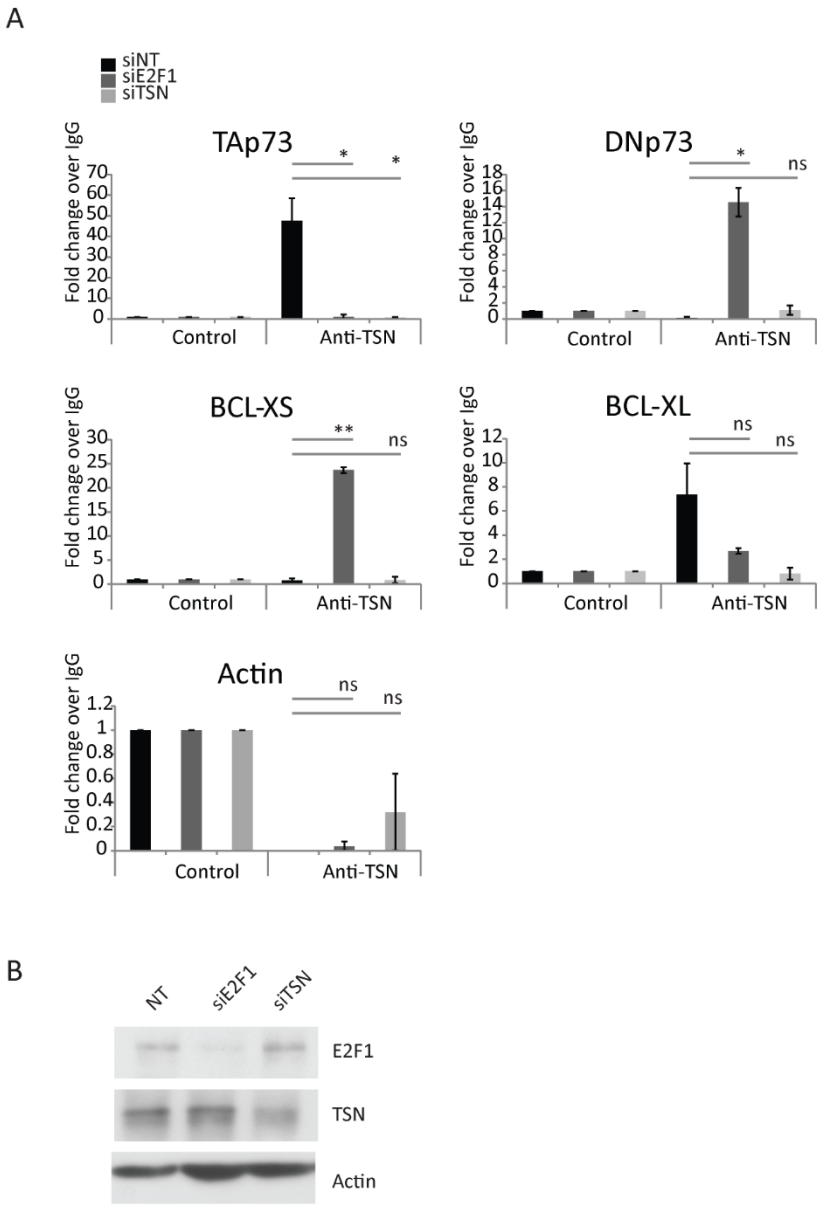


Figure 5.9 E2F1 alters p100-TSN binding to splice isoforms.

U2OS cells were treated with 25nmol siRNA against E2F1, p100-TSN or NT control for 72h. UV was used to crosslink protein-RNA before cell lysis and RIP was carried out using anti-p100-TSN.

A. cDNA was analysed by qPCR for the RNA species indicated. Results shown are averages of three independent experiments with standard error shown * $p<0.05$, ** $p<0.005$.

B. Immunoblotting was carried out to confirm successful silencing of the proteins E2F1 and p100-TSN using β-actin as a loading control.

E2F1, so under these conditions we would expect lower levels of RNA binding (Mohlenbrink J. unpublished data). However, after using phenol red free media and charcoal stripped serum to inhibit hormone signalling, we would expect TDRD1 levels to drop enabling more p100-TSN to interact with E2F1, increasing the RNA binding seen on E2F1. DuCaP cells, a prostate cancer cell line, were used as a model of this system as they express the isoform of TDRD1 which has previously been found to bind to E2F1 (Mohlenbrink J. unpublished data).

DuCAP cells were plated and after 48h the cells were washed three-times with PBS and media was replaced with either normal media (containing FBS and phenol red), or hormone depleted media (phenol red free, charcoal stripped serum). Cells were incubated for a further 48h before RIP was carried out. In the absence of hormone, the interaction of E2F1 with TAp73 or DNp73 was significantly increased compared to in normal growth media [Figure 5.10A], despite a reduction in E2F1 expression after hormone depletion [Figure 5.10C]. This is consistent with our hypothesis, as in the absence of hormone there would be an enhanced interaction of p100-TSN with E2F1 due to loss of TDRD1 competition for the R111/113Me2s site. Thus the increase in p100-TSN interaction with E2F1 would be expected to increase the target gene RNA association with E2F1. Although we were unable to confirm loss of TDRD1 after hormone depletion by immunoblot.

Under the same conditions, but in a p100-TSN RIP [Figure 5.10B], there was a significant reduction in TAp73 isolated after hormone deprivation compared to the normal media control. In contrast, DNp73 is enriched in the RIP under hormone deprivation conditions compared to the normal media control. This is in contrast to the results observed in the U2OS cell line, where loss of the p100-TSN-E2F1 interaction favoured an interaction with DNp73, and inhibited TAp73 binding. Thus, there are clear cell-type differences which can also influence isoform specific binding.

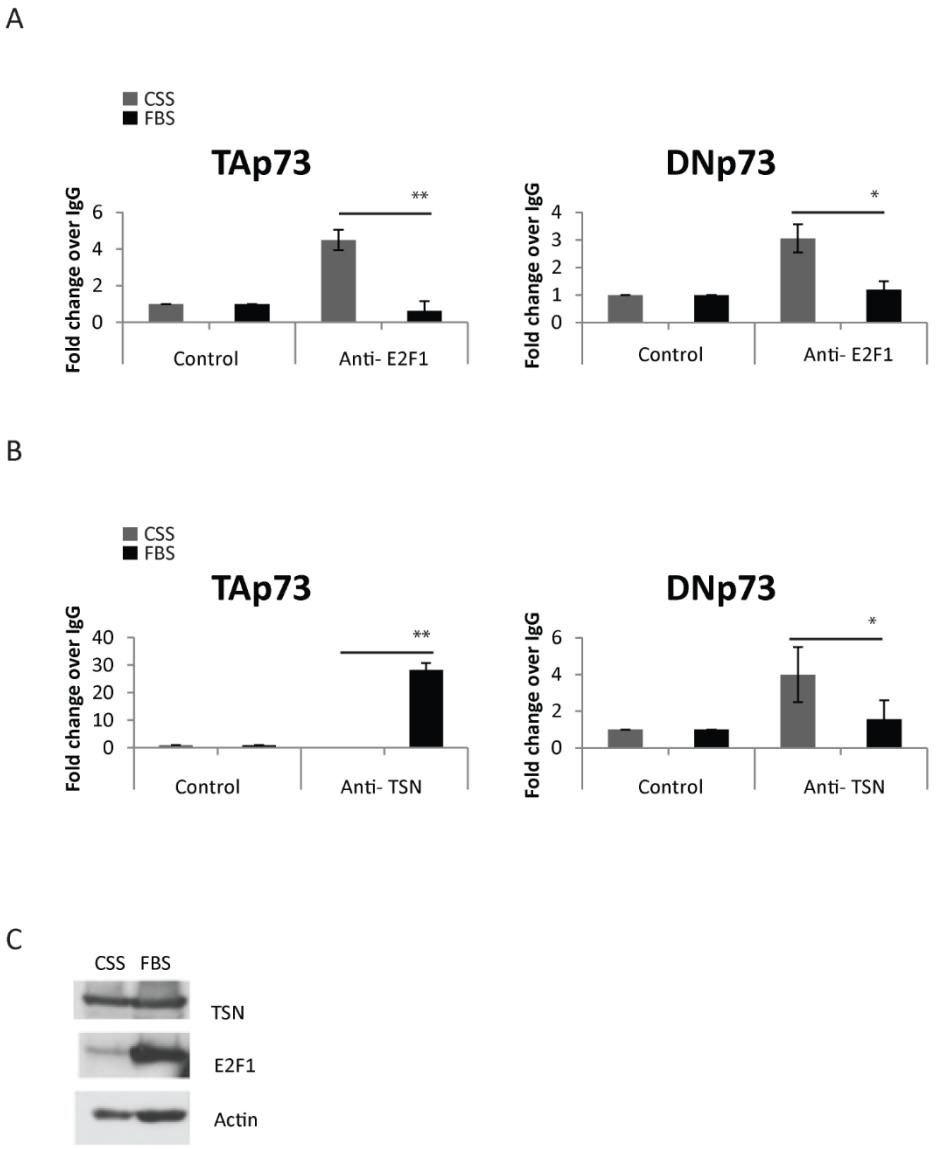


Figure 5.10 Hormone depletion increases E2F1 binding to target genes in DuCAP cells.

DuCAP cells were seeded and grown in normal (FBS) media or homone depleted (CSS) media for 48 hours. Cells were UV crosslinked before lysis.

A. RIP was carried out using anti-E2F1. cDNA was analysed by qPCR for the RNA species shown. Average of two independent experiments with standard error * $p<0.05$, ** $p<0.005$

B. Anti-TSN RIP was carried out and cDNA analysed by qPCR for the species shown. Average of two independent experiments with standard error * $p<0.05$, ** $p<0.005$.

C. Immunoblotting was carried out to determine protein expression of E2F1 and p100-TSN under both conditions using showing example protein expression using β-actin as a loading control.

5.8 Chapter summary

In this chapter we investigated the role of PRMT5 mediated E2F1 R111/113Me2s on RNA processing by looking at direct interactions between target gene RNA and E2F1. It was found that E2F1 can associate with specific target gene RNA and this was dependent on the presence of p100-TSN as well as PRMT5 activity. This is consistent with the previously published data which showed that p100-TSN binds to symmetrically methylated E2F1, generated by PRMT5 (Zheng et al., 2013). Additionally, it was found that E2F1 required an intact DNA binding domain for the RNA interaction. Therefore, we hypothesise that chromatin bound E2F1 is forming a link with the nascent RNA through p100-TSN, a known RNA binding protein. Furthermore, using data from mass spectrometry analysis carried out previously, it was discovered that E2F1 may be in a complex with parts of the spliceosome after PRMT5 upregulation, including PRP38 and SmD1. We were able to confirm this by demonstrating an interaction between E2F1 and snRNAs U4 and U6, which form part of the tri-snRNP complex with PRP38 and SmD1. Once again, this interaction was dependent on PRMT5 activity suggesting that p100-TSN is forming a bridge between transcription and the spliceosome.

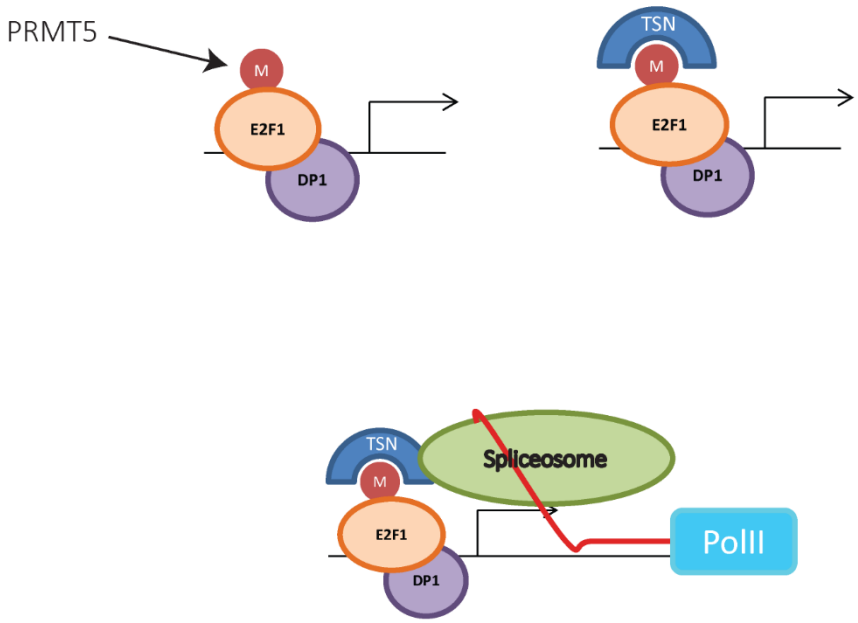
P100-TSN itself was also able to bind to E2F1 target gene RNA, and interacted with different splice isoforms depending on the presence of E2F1. When E2F1 was present, p100-TSN bound to TAp73 and BCL-XL, but in the absence of E2F1 it was associated with DNp73 and BCL-XS. This implies that E2F1 may be having a more direct influence on splicing decisions of its target genes than previously believed, by impacting p100-TSN recruitment to the nascent RNA and spliceosome assembly.

Chapter 6. Sequencing of RNA bound in complexes with E2F1 and p100-TSN

6.1 Introduction

Given that we were able to observe an interaction between E2F1 and various target gene RNA species which was dependent on p100-TSN, and also an interaction between E2F1 and snRNAs associated with the splicing machinery we hypothesised that PRMT5 mediated methylation of E2F1at R111/113 is enabling recruitment of p100-TSN to the promoter of E2F1 target genes. As transcription starts, p100-TSN is then able to recruit the splicing machinery to the nascent RNA as it is transcribed [Figure 6.1A]. In order to more fully explore the role of p100-TSN and E2F1 in RNA processing and to expand the number of RNA species identified as interacting with these proteins, RNA sequencing was carried out on RIP samples. RIP experiments were carried out as before using non-targeting siRNA, siE2F1 or siTSN treated cells and using anti-E2F1 or anti-p100-TSN for the immunoprecipitation. The appropriate siRNA treatment was used as the negative control for the experiment (siE2F1 for the E2F1 RIP and siTSN for the p100-TSN RIP) successful depletion of the proteins was determined by immunoblot [Figure 6.1B]. Following TRIzol extraction, RNA was DNAse treated before library preparation. A trial sequencing run was carried out to determine if the experiment was viable and showed high rRNA contamination, therefore rRNA depletion was carried out after DNase treatment for the subsequent experiment. Samples were submitted in duplicate and sequenced on an Illumina NextSeq. Alexander Kanapin of the Bioinformatics Core (Department of Oncology, University of Oxford) carried out analysis of the sequencing reads.

A



B

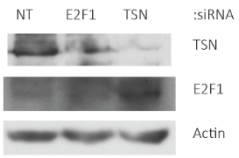


Figure 6.1 Sample preparation for RIP sequencing.

A. Diagram of proposed model linking E2F1 with RNA processing via methylation dependent association of p100-TSN with E2F1.

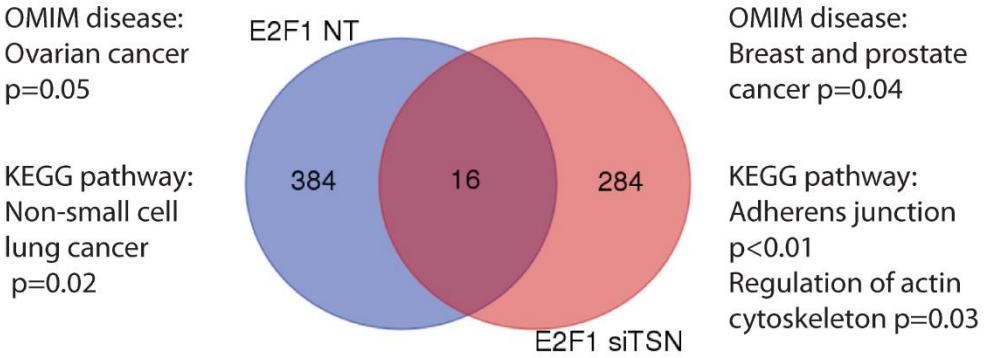
B. U2OS cells were treated with 25nm siRNA against E2F1, p100-TSN and a non-targeting (NT) control for 72h. UV cross linking was carried out before lysis. An aliquot of lysate was taken and immunoblotting carried out to confirm sufficient depletion of the proteins using anti-E2F1, anti-p100-TSN and anti-β-actin as a loading control. The remaining lysates were immunoprecipitated with either anti-E2F1 or anti-p100-TSN before isolation of RNA using TRIzol. The RNA was DNase treated, underwent rRNA depletion prior to library preparation before sequencing on an Illumina sequencer.

6.2 Both E2F1 and p100-TSN can bind multiple RNA species

After peak calling, the gene list was filtered for peaks enriched with an adjusted p-value of less than 0.05. The resulting gene lists were then analysed for overlapping and unique genes. In total, 684 peaks corresponding to unique genes were identified in the E2F1 RIPs. Of these, the majority (400) were in the siNT treated RIP, with 384 unique to this treatment. Somewhat surprisingly, given our previous findings, E2F1 was able to bind to RNA in the absence of p100-TSN – with 300 genes being identified in the siTSN treated RIP [Figure 6.2A]. Only 16 genes were shared between NT and siTSN indicating that E2F1 may bind to a separate pool of RNAs in the absence of p100-TSN. This could reflect another mechanism by which E2F1 can interact with specific RNAs which is not dependent on p100-TSN. GO analysis (DAVID 6.7) (Huang et al., 2009a, Huang et al., 2009b) was run on each separate pool of genes and in genes unique to the siTSN treatment, there was an enrichment of genes seen in breast and prostate cancer (OMIM disease) as well as adherens junctions and cytoskeleton pathway enrichment (KEGG pathway analysis). Gene ontology (BP_FAT) revealed genes involved in GTPase mediated signalling were isolated with E2F1 after p100-TSN depletion [Figure 6.2B]. Interestingly, in the NT treated cells there was a significant enrichment in genes involved in ovarian cancer (OMIM disease) and non-small cell lung cancer (KEGG pathway). Additionally, significant GO terms (BP_FAT) for the RNA enriched in the NT treated cells include protein phosphorylation (signalling pathways), cell adhesion and chromatin modifiers. The only term enriched in the overlapping gene set is ion transport, and there were no enrichments in OMIM disease or KEGG pathways for this set of genes.

In the p100-TSN RIP-seq, 665 unique genes were identified overall and the majority (597) were from the siE2F1 treated cells with 135 genes identified in the NT treated RIP and 67 overlapped between both treatments suggesting these RNAs bind to p100-TSN in an E2F1 independent manner [Figure 6.3A]. This suggested that overall the number of different RNA

A



B

E2F1 RIP	Term (GOTERM_BP_FAT)	Gene Count	P Value
NT	phosphate metabolic process	17	0.015
	phosphorus metabolic process	17	0.015
	protein amino acid phosphorylation	12	0.040
	chromatin modification	7	0.039
	cell-cell adhesion	7	0.040
	positive regulation of developmental process	7	0.041
	histone modification	6	0.005
	covalent chromatin modification	6	0.006
	homophilic cell adhesion	5	0.032
	protein amino acid acylation	4	0.017
	embryonic hindlimb morphogenesis	3	0.011
	hindlimb morphogenesis	3	0.023
	ISG15-protein conjugation	2	0.045
siTSN	regulation of small GTPase mediated signal transduction	5	0.036
Overlap	ion transport	4	0.005

Figure 6.2 E2F1 can bind to different pools of RNA in the presence and absence of p100-TSN.

After peak calling and read alignment, RNA species which were significantly ($p<0.05$) enriched in the E2F1 RIPs compared to the control (siE2F1) were compared.

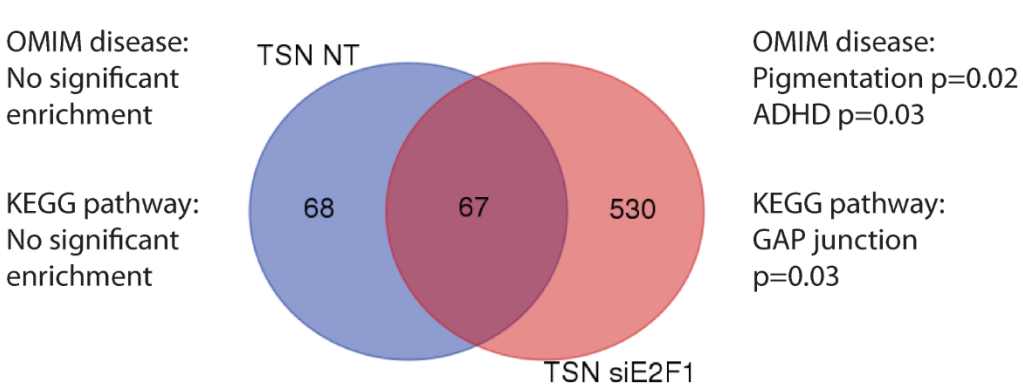
A. Venn diagram showing overlap between RNA species bound to by E2F1 in the presence (NT) and absence (siTSN) of p100-TSN. DAVID 6.7 was used on each separate gene set and OMIM disease and KEGG pathway enrichments are indicated.

B. Significant ($p<0.05$) GO terms (BP_GO_FAT) enriched in each gene set are indicated.

species that p100-TSN can bind to is reduced when E2F1 is present and may reflect E2F1 mediated recruitment of p100-TSN to a more specific set of genes. Indeed, when GO analysis (DAVID 6.7) on these shared genes was completed, three core pathways were enriched significantly – all of which are related to actin cytoskeleton organisation, suggesting there may be an essential regulation by p100-TSN of these core pathways [Figure 6.3B]. Interestingly, GO analysis on genes unique to the NT condition showed no significant enrichment for any analysis, whereas GO analysis on those in the siE2F1 condition were enriched for regulation of the cell cycle, apoptosis, and differentiation – all key pathways regulated by E2F1. This suggested that in the absence of E2F1, the ensuing alterations in these pathways may be also regulated by p100-TSN at the RNA level. Additionally, after depletion of E2F1, RNAs bound to p100-TSN also showed enrichment in GAP junction related genes (KEGG pathway) as well as genes related to pigmentation and ADHD (OMIM disease), although the relevance of the latter two is unclear in a cancer background [Figure 6.3A].

To identify RNAs which can be bound by both E2F1 and p100-TSN we compared all RNAs identified in both E2F1 RIPs and both p100-TSN RIPs [Figure 6.4A]. This shows that there are a core 137 RNA species which were present in both E2F1 and p100-TSN RIPs. However, there was no significant upregulation in pathways (KEGG), disease (OMIM) or GO related gene sets for these genes. Conversely, RNA found in the E2F1 RIPs alone showed enrichment in non-small cell lung cancer related genes (KEGG) and significant GO terms enriched included several signalling pathways, apoptosis, development and cell cycle terms – all E2F1 related pathways. For the RNA unique to the TSN RIPs, there was enrichment in TGFβ signalling pathway (KEGG) as well as cell cycle and immune regulation [Figure 6.4B].

A



B

TSN RIP	Term	Gene Count	P Value
NT	No enrichment		
siE2F	regulation of cell cycle process	5	0.018
	regulation of mitotic cell cycle	5	0.045
	negative regulation of neuron apoptosis	4	0.010
	hindbrain development	4	0.016
	regulation of neuron apoptosis	4	0.045
	regulation of S phase of mitotic cell cycle	3	8.5E-03
	regulation of S phase	3	0.013
	regulation of BMP signaling pathway	3	0.028
	central nervous system neuron development	3	0.032
	central nervous system neuron differentiation	3	0.048
	neurotrophin production	2	0.018
	nerve growth factor production	2	0.018
	nerve growth factor processing	2	0.035
Overlap	cytoskeleton organization	4	0.011
	actin cytoskeleton organization	3	0.025
	actin filament-based process	3	0.028

Figure 6.3 E2F1 can impact the RNA species bound to by p100-TSN.

After peak calling and read alignment, RNA species which were significantly ($p<0.05$) enriched in the p100-TSN RIPs compared to the control (siTSN) were compared.

A. Venn diagram showing overlap between RNA species bound to by p100-TSN in the presence (NT) and absence (siE2F1) of E2F1. DAVID 6.7 was used on each separate gene set and OMIM disease and KEGG pathway enrichments are indicated.

B. Significant ($p<0.05$) GO terms (BP_GO_FAT) enriched in each gene set are indicated.

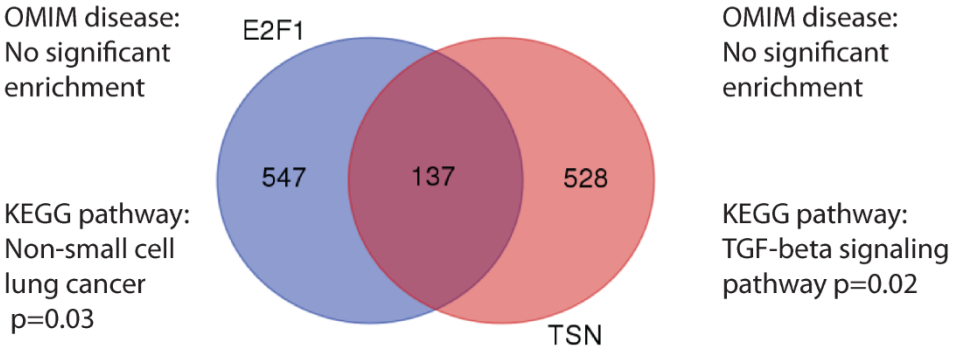
6.3 Identification of isoform specific binding

In order to explore the role of E2F1 and p100-TSN on splicing, peaks from the RIP-seq which crossed exon junctions were identified to infer isoform changes across the different RIPs. In the E2F1 RIP-seq we were able to identify peaks across 27 exon junctions which corresponded to 25 different transcripts over 17 genes in the NT treatment and 8 peaks over 6 genes after depletion of TSN [Figure 6.5A]. In the p100-TSN RIP-seq we identified peaks across 3 exon junctions across three genes after NT treatments, and 21 peaks over 19 genes after depletion of E2F1 [Figure 6.5B].

Two of the genes identified in the E2F1 RIP, but not after depletion of p100-TSN, which also had different peaks across splice junctions were MECOM and SENP7. MECOM (MDS1 and EVI1 Complex Locus) is a complex of two genes – MDS1 and EVI1, which can be found spliced together as EVI1 intron 1 is downstream of the second MDS1 exon [Figure 6.6A]. EVI1 can also be transcribed separately from a downstream promoter in the second intron of MECOM with transcription starting at exon 2b [Figure 6.6A] (Fears et al., 1996). MECOM (and EVI1) is a transcription factor which regulates proliferation, survival and inhibits differentiation. Additionally, MECOM is frequently translocated with AML1 in resulting in overexpression and leukaemic transformation (Buonamici et al., 2003). In the NT treated E2F1 RIP, only the two long (MECOM) isoforms were identified, as splice junctions mapping to Exon 1-2 and 1-3 were sequenced. None of the shorter isoforms, which have the transcription start site at exon 2b, were isolated with E2F1 in the RIP-seq.

SENP7 (SUMO1/Sentrin Specific Peptidase 7) removes SUMO2 and SUMO3 peptides from proteins and is involved in DNA repair and transcription through its role in de-SUMOylation mediated regulation of chromatin remodelers as well as MYC (Shen et al., 2009, González-Prieto et al., 2015, Garvin et al., 2013, Romeo et al., 2015). In the E2F1 RIP-seq, exon junctions 4-6 and 4-7 containing transcripts, which correspond to variant 3 (V3) and variant 5 (V5)

A



B

RIP	Term (GOTERM_BP_FAT)	Gene Count	P Value
E2F1	protein phosphorylation	11	0.032
	nervous system development	8	0.042
	positive regulation of apoptotic process	8	0.051
	ephrin receptor signaling pathway	7	0.000
	Notch signaling pathway	6	0.009
	positive regulation of protein phosphorylation	6	0.014
	homophilic cell adhesion via plasma membrane adhesion molecules	6	0.032
	regulation of Rho protein signal transduction	5	0.013
	transmembrane receptor protein tyrosine kinase signaling pathway	5	0.022
	steroid hormone mediated signaling pathway	4	0.026
	synapse assembly	4	0.031
	vascular endothelial growth factor receptor signaling pathway	4	0.046
	bone resorption	3	0.025
	positive regulation of Rho protein signal transduction	3	0.034
	anion transmembrane transport	3	0.044
	meiotic cell cycle	3	0.055
	negative regulation of negative chemotaxis	2	0.022
	positive regulation of cellular protein localization	2	0.022
	axon midline choice point recognition	2	0.044
	negative regulation of osteoclast development	2	0.044
	osteoclast development	2	0.054
TSN	regulation of cell cycle process	5	0.021
	negative regulation of immune system process	4	0.041
	regulation of S phase of mitotic cell cycle	3	0.009
	regulation of S phase	3	0.014
	regulation of transforming growth factor-beta1 production	2	0.027
	negative regulation of plasma membrane long-chain fatty acid transport	2	0.036
	regulation of plasma membrane long-chain fatty acid transport	2	0.036
	pro-B cell differentiation	2	0.045
Overlap	No enrichment		

Figure 6.4 E2F1 and p100-TSN can bind shared RNA species as well as separate pools of RNA species which have different cellular functions.

Genes significantly enriched in all E2F1 and TSN RIP over their respective controls were compared. A. Venn diagram showing overlap between RNA species bound to by E2F1 and p100-TSN. DAVID 6.7 was used on each separate gene set and OMIM disease and KEGG pathway enrichments are indicated.

B. Significant ($p<0.05$) GO terms (BP_GO_FAT) enriched in each gene set are indicated.

isoforms respectively, were identified. All other isoforms also contain exon 5 [Figure 6.6B], and were not isolated with E2F1 in the RIP-seq. Therefore a common set of primers, designated 'E5', were designed to amplify all other isoforms [Figure 6.6B].

In order to validate the findings of the RIP-seq and also to confirm whether E2F1 was able to bind to specific isoforms for MECOM1 and SENP7, primers were designed (shown as red arrows in Figure 6.6) which would amplify the indicated RNA. Although E2F1 was unable to isolate any MECOM RNA after RIP, SENP7 V5 was isolated with E2F1 [Figure 6.7]. Additionally, this was abrogated by depletion of both E2F1 and p100-TSN, which was consistent with the RIP- seq analysis. However, despite the bioinformatic analysis suggesting SENP7 V3 would also be associated with E2F1, this was not observed when validating the results. Most importantly, as per the RIP-seq, we did not observe any isoforms containing exon 5 as being isolated by E2F1 [Figure 6.7A]. This is the first time we have shown isoform specific binding of an RNA species to E2F1 and suggests strongly that E2F1 has the ability to influence isoform selection of alternatively spliced target genes.

To analyse this further, RNA levels of the SENP7 isoforms were analysed by qRT-PCR on total RNA isolated from U2OS following transfection with siRNA against E2F1 or p100-TSN [Figure 6.8A]. Total levels of SENP7 were slightly increased (although not significantly) upon E2F1 depletion and there was no change after p100-TSN depletion. Additionally, although expression of V3 was marginally reduced after loss of E2F1 expression and increased after p100-TSN depletion, these changes in expression were not significant. However, expression of the V5 isoform was significantly reduced after E2F1 depletion, and to a lesser extent (but not significantly) after p100-TSN depletion. This implies a requirement for E2F1, and possibly p100-TSN, to positively regulate isoform selection of V5. Expression of exon 5 (E5) containing isoforms was not significantly changed after treatment with either siE2F1 or siTSN [Figure 6.8A].

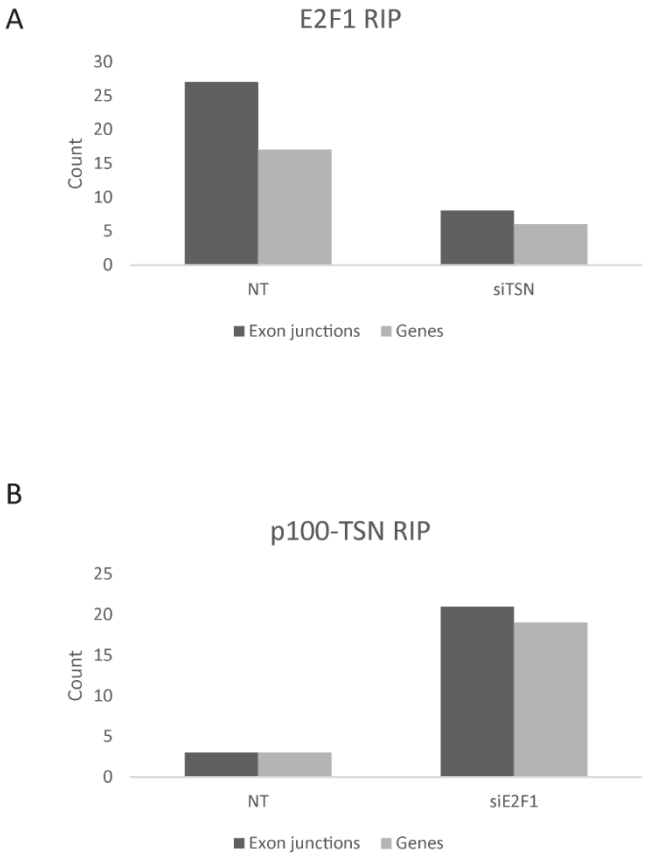


Figure 6.5 Identifying exon junctions in the RIP-seq analysis enables inference of isoform binding.

Peaks were analysed for those crossing exon boundaries. The number of peaks which spanned exon junctions is shown in dark grey and the number of genes these peaks correspond to is shown in light grey for A. E2F1 RIPs and B. p100-TSN RIPs.

The finding that total SENP7 RNA expression was not strongly affected by depletion of E2F1 [Figure 6.8A], would suggest it is not a classic target gene. Furthermore, re-analysis of data from the RNA-seq data revealed that SENP7 was not differentially expressed after overexpression of HA-E2F1. However, this is consistent with the finding from the RNA-seq which showed that those target genes which had altered isoform use were a separate subset of genes from those which were differentially expressed upon induction of E2F1 expression [Figure 4.9]. However, as both the exon use analysis showed that the vast majority of alternatively spliced genes had E2F1 ChIP-seq peaks in their promoters, together with the observation in the RIPs [Figure 5.6C] which showed that E2F1 must be able to bind to chromatin to interact with RNA, would suggest that E2F1 must be able to bind to the SENP7 promoter. Publically available ChIP-seq data has identified E2F4 peaks at an E2F motif within the first intron [Figure 6.8B]. Additionally, when a transcription factor binding site tool Patch 1.0 (Matys et al., 2006) was used to analyse the genomic sequence covering the promoter (-1kb to +1kb), an E2F1 site was predicted within 70bp of the transcription start site [Figure 6.8]. Ongoing work will attempt to determine if either of these sites can be used by E2F1 *in vivo*.

To understand the role SENP7, and in particular V5, may be playing in cancer, we used the publically available isoform expression database ISOexpresso (Yang et al., 2016) which makes use of The Cancer Genome Atlas database (http://cancergenome.nih.gov/). ISOexpresso enables expression of each variant of a gene, as well as total expression across a variety of different tumour types to be analysed. Additionally, for several of the tumour types there are paired normal datasets enabling the user to make comparisons of gene, or isoform, expression between normal and tumour tissues. Overall expression of SENP7 is reduced across most tumour types included in this dataset [Figure 6.9A] compared to the normal tissue, except for in cholangiocarcinoma, pheochromocytoma and paraganglioma, and glioblastoma multiforme. However, of these only the latter expresses V5 [Figure 6.9 B, dark blue bar].

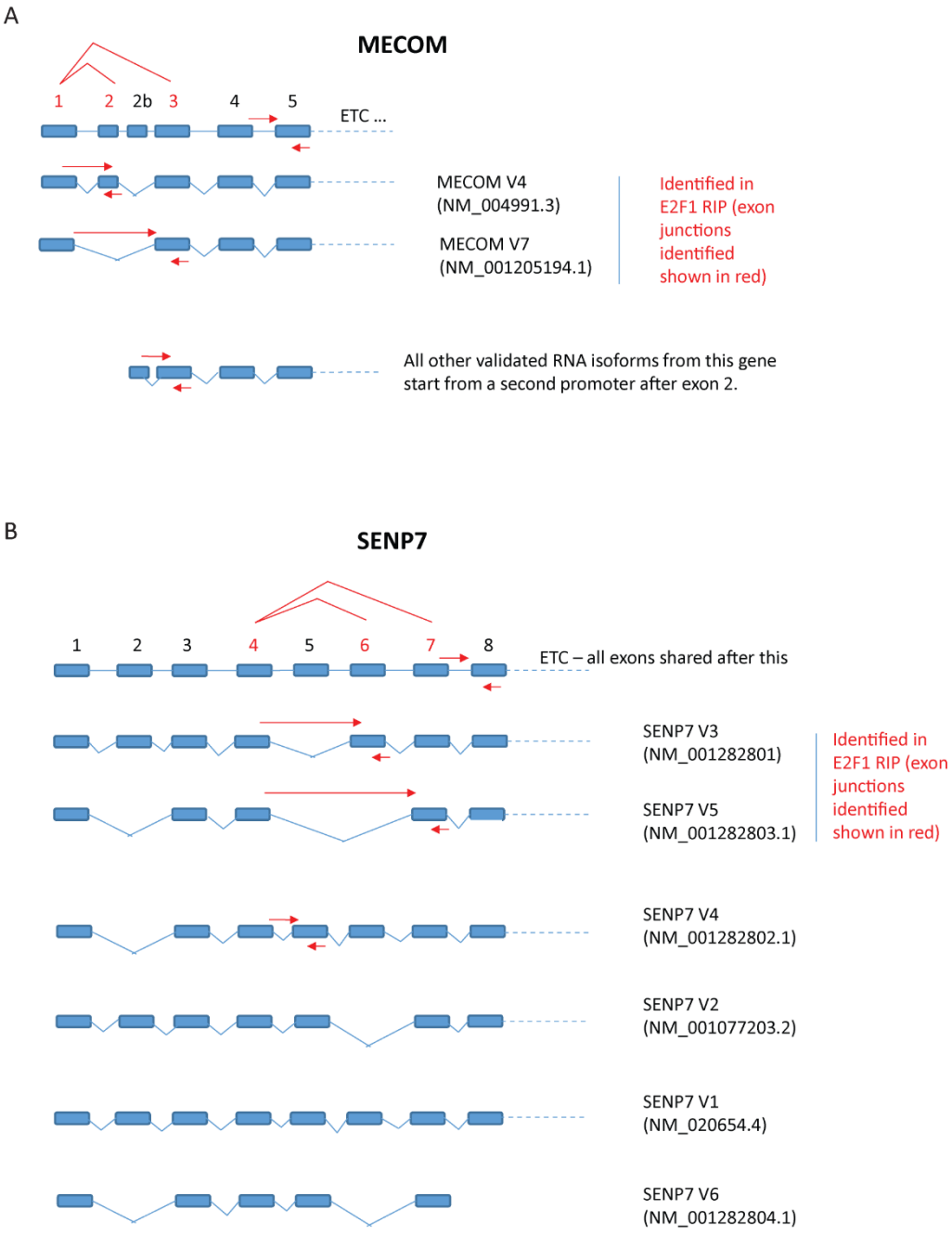


Figure 6.6 Candidate genes which show specific isoform binding to E2F1.

Gene structure of two genes identified in the E2F1 RIP after NT treatment, which were absent in the siTSN treatment and also showed peaks crossing different exon junctions. Exons and splice patterns shown in red represent the exon junctions identified in the RIP-seq analysis, red arrows indicate primer binding sites.

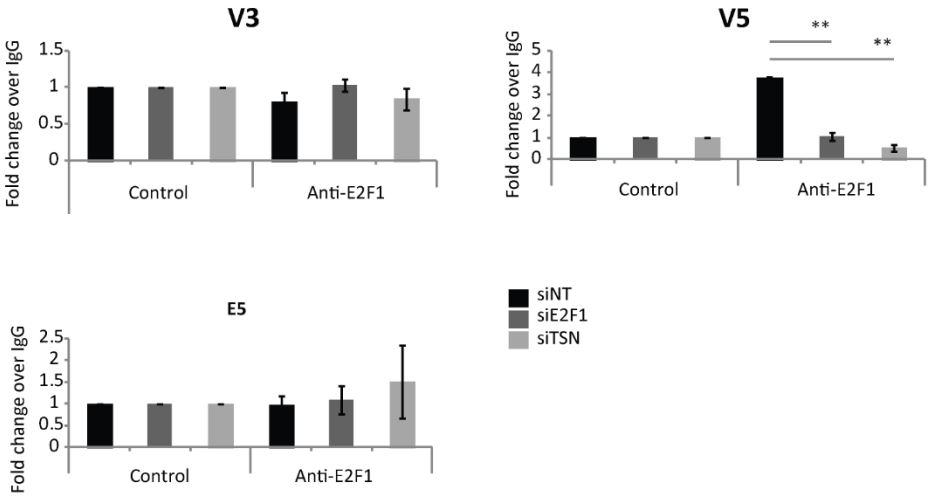
A. MECOM has two validated long isoforms which share all C-terminal exons (not shown) and 5 short isoforms which start from exon 2b, these share exons 3-9 and have some differences in C-terminal exons (not shown).

B. SENP7 has 6 validated isoforms, all of which share a C-terminus structure (not shown) except for V6 which is a short isoform, polyadenylated after exon 7. V3 and V5 are the only isoforms not to contain exon 5 so primers to 'E5' select all other transcripts.

Although for eight tumour types (indicated by asterisks) there is no available normal tissue data available for comparison. Indeed from looking at the isoform expression [Figure6.9B] it is apparent that V5 is the least expressed isoform (at the RNA level) across all settings and it is not expressed in many of the cancers catalogued here [Figure 6.9C]. Additionally, as per the total levels of SENP7 expression, expression of V5 is reduced in tumour compared to normal [Figure 6.9C].

Comparison of V5 expression to the total expression of SENP7 in tumours [Figure 6.9A, Figure 6.9C respectively] showed that the highest expression of both total and V5 is observed in acute myeloid leukaemia. To assess if changes in V5 expression were related to overall changes in total SENP7 expression, relative expression (V5/total) was assessed [Figure 6.9D]. The expression of V5 isoform relative to total SENP7 levels is highest in ovarian serous cystadenocarcinoma. When expression data for E2F1, PRMT5 and p100-TSN was analysed in from the same datasets [Figure 6.9E, F, G] it was found that these were also elevated in ovarian serous cystadenocarcinoma which suggested that there could be a causal link between E2F1, PRMT5 and SENP V5 in this setting. This trend is also observed in oesophageal tumours, but is less apparent across other tumour types. Whilst V5 expression is reduced in oesophageal tumour compared to normal [Figure 6.8C], there is no paired normal for ovarian serous cystadenocarcinoma, so whether this is a tumour related effect, or whether expression is simply high in this tissue is not clear from this dataset. However, this could prove a useful setting in which to conduct future experiments.

A



B

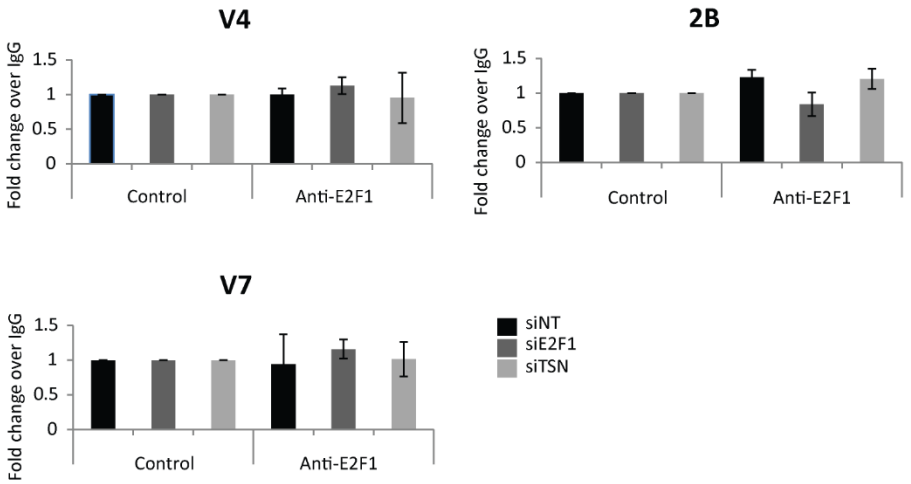


Figure 6.7 Validation of the isoform specific binding to E2F1 by MECOM and SENP7.

cDNA from previous RIP experiments [Figure 5.3] was used to determine if the potential RNA species indicated by the RIP-seq could be isolated by E2F1. U2OS cells were treated with siRNA against E2F1, p100-TSN or NT control for 72h before a RIP was carried out using anti-E2F1.

A. SENP7 isoform specific primers were used to amplify cDNA. Results shown are averages of three independent experiments with standard error shown * $p<0.05$, ** $p<0.005$.

B. MECOM isoform specific primers were used to amplify cDNA. Results shown are averages of three independent experiments with standard error shown * $p<0.05$, ** $p<0.005$.

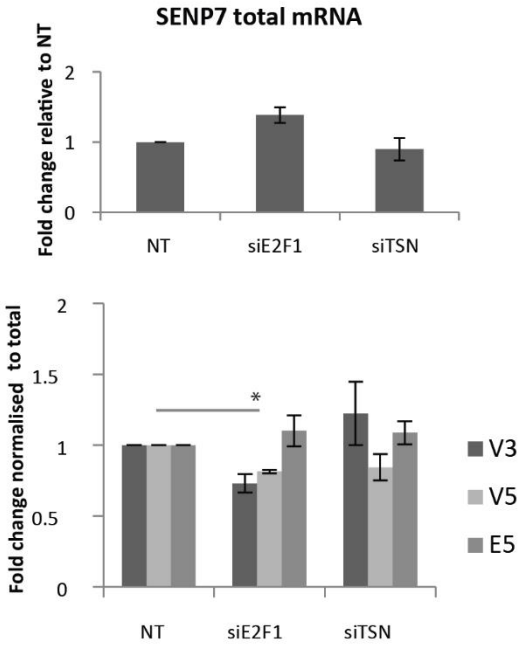
6.4 Chapter summary

In this chapter we have carried out a RIP-sequencing experiment to understand more fully the role of E2F1 in RNA processing. After successful RIP, library preparation and sequencing, we discovered many RNA species associate with E2F1. Unexpectedly, we also found a subset of RNAs which could bind to E2F1 in the absence of p100-TSN, which suggests that in the absence of p100-TSN there may be another scaffold protein forming a bridge between the transcription factor and a separate pool of RNA species. As the pools of RNA species able to associate with E2F1 in the presence of p100-TSN was largely distinct from those bound to E2F1 in the absence of p100-TSN, this may reflect two separate mechanisms utilised by E2F1 to influence RNA processing. Furthermore, GO analysis suggests that different signalling pathways may be influenced by E2F1 dependent RNA processing, and this may be further impacted by the presence or absence of p100-TSN.

We additionally identified RNA species bound to p100-TSN. The number of different RNA species was increased after depletion of E2F1, thus suggesting that E2F1 is sequestering p100-TSN to a specific pool of RNA. Furthermore, when this interaction is inhibited after loss of E2F1, p100-TSN can interact with additional RNA species. Interestingly, GO analysis showed that p100-TSN can bind to cell cycle regulatory RNAs after depletion of E2F1, suggesting a role of p100-TSN in modulating the cell cycle after E2F1 is silenced. This may involve switching typically pro-growth genes into inhibitory splice-isoforms after the loss of E2F1 in order to effect a rapid response to loss of E2F1. Further work in this area will be required to fully understand the role of p100-TSN in this context.

In addition, we identified several peaks which crossed exon boundaries, allowing us to infer isoform specific binding to both p100-TSN and E2F1. Most importantly, we were able to verify that E2F1 is able to interact with SENP7 V5 in a p100-TSN dependent manner. As E2F1 binding to other RNA transcripts of SENP7 was not observed, therefore this interaction

A



B

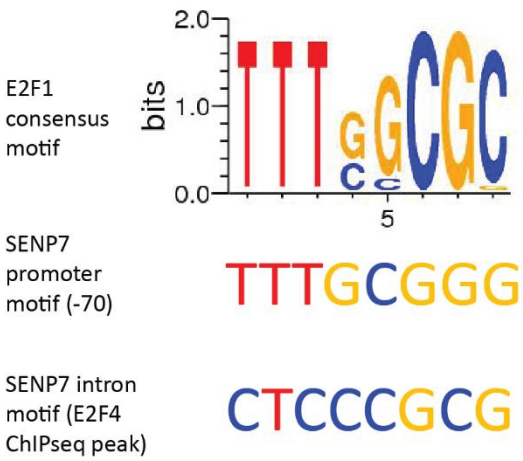


Figure 6.8 SENP7 is a putative E2F1 target gene.

A. U2OS cells were transfected with 25nM siRNA against E2F1, p100-TSN or NT for 72h. RNA was extracted using TRIzol and cDNA synthesised. qPCR was carried out using primers indicated in the figure. Total RNA levels were normalised to actin and isoform specific RNA levels were normalised to the total RNA. Results shown are averages of three independent experiments with standard error shown * $p<0.05$, ** $p<0.005$.

B. Canonical E2F1 DNA binding motif. B. Potential E2F1 motif found in the proximal SENP7 promotor.

C. Motif bound to by E2F4 in publically available ChIP-seq data.

appeared to be isoform specific. By analysing publicly available data we determined that SENP7 is generally reduced in tumour tissue compared to normal, and V5 remains the lowest expressed isoform across all tumour types we were able to get data for. However, V5 expression was relatively increased in OSC compared to total SENP7 levels and we propose that high expression of E2F1, PRMT5 and p100-TSN in this setting may contribute to this. Thus much more work is required to understand the function of SENP7 V5 and importantly on the role E2F1, PRMT5 and p100-TSN have on SENP7 isoform selection.

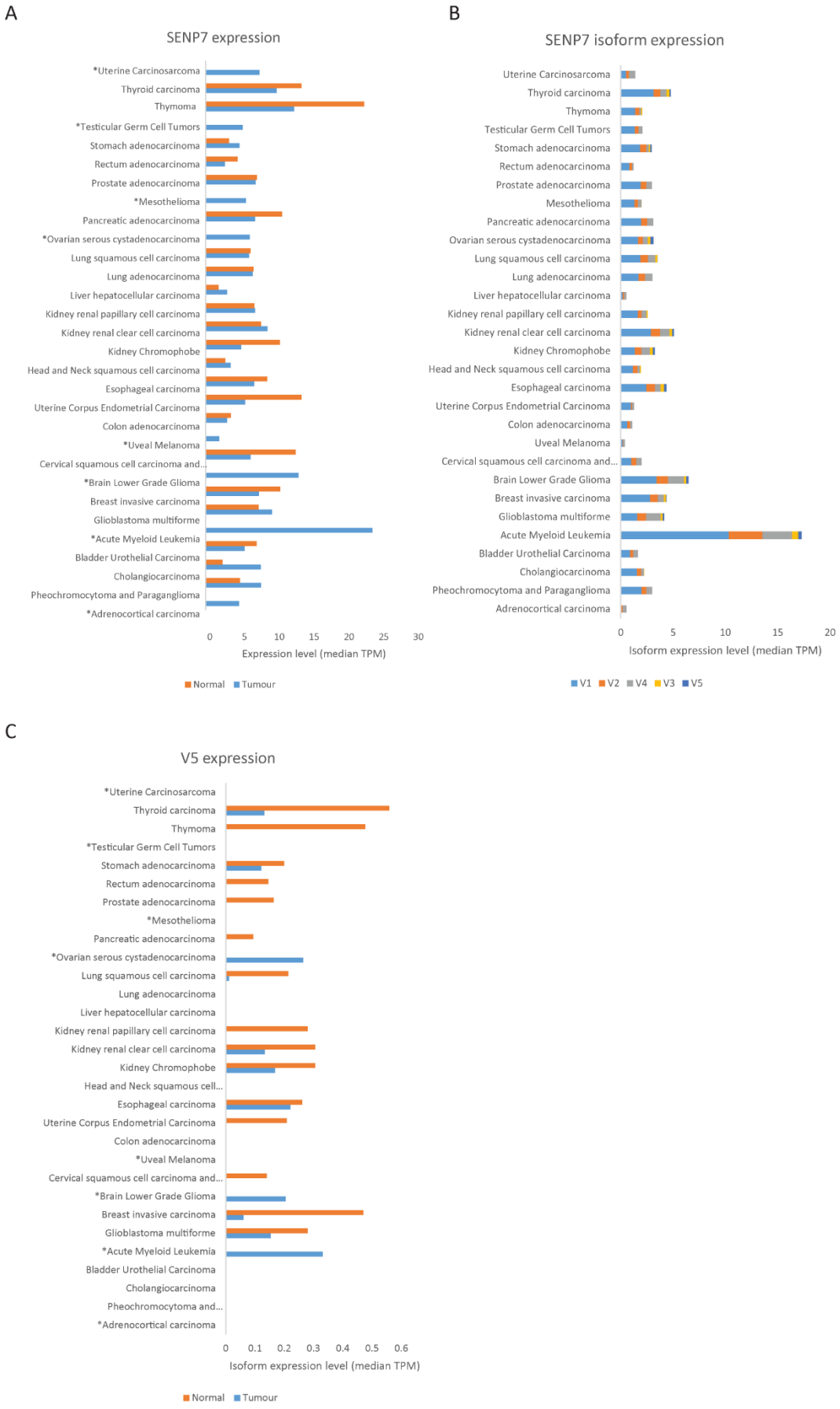


Figure 6.9 (figure legend on following page)

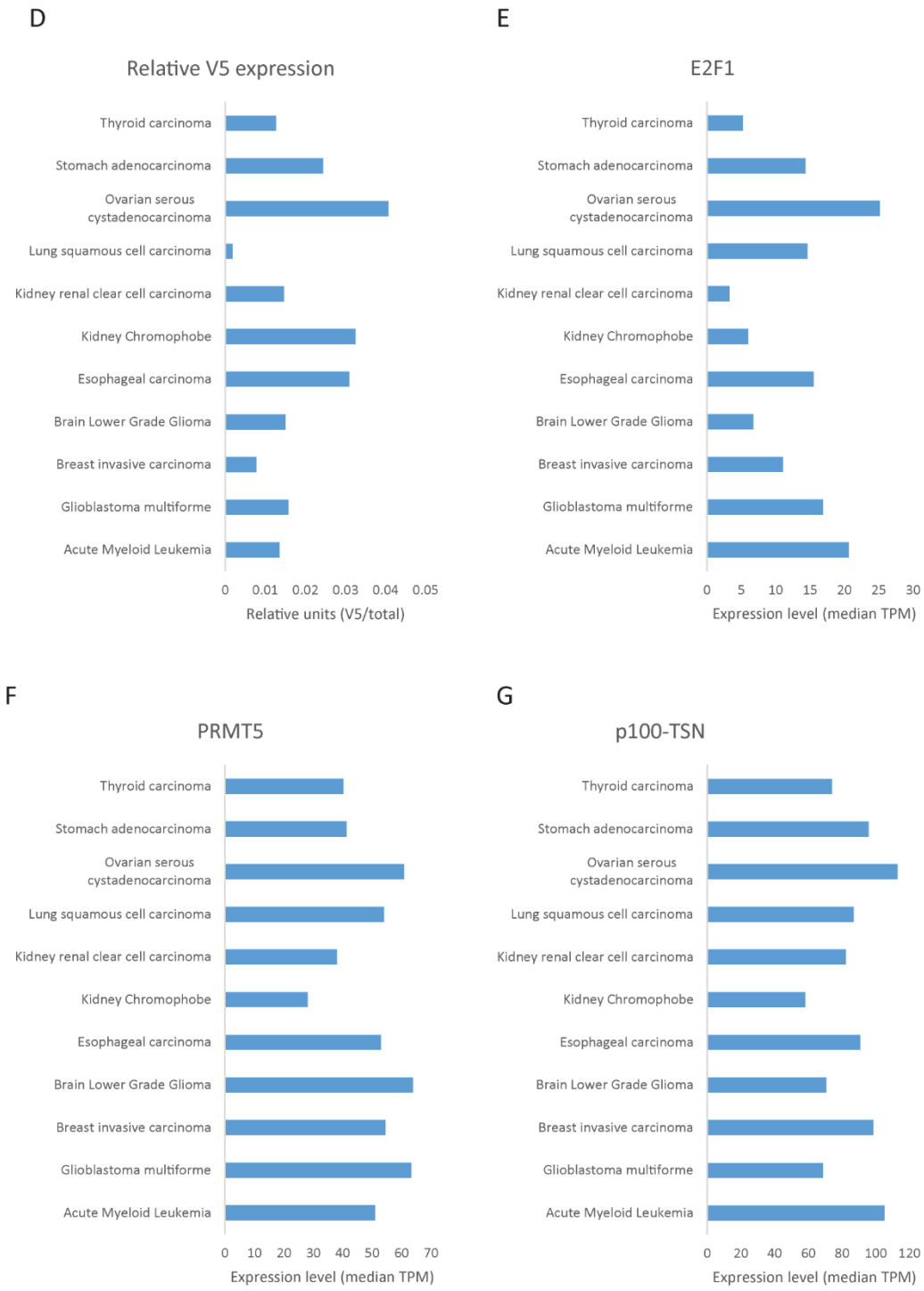


Figure 6.9 SENP7 expression in cancer

RNA expression data was accessed from ISOexpresso using TGCA datasets. A. Total SENP7 expression across tumour and normal samples (asterisk indicates no paired normal data available). B. Isoform expression across tumour samples (V6 data was not available). C. V5 expression across normal and tumour samples (asterisk indicates no paired normal data available). D. SENP7 V5 relative expression (V5/total) in V5 expressing tumour tissue. E. Total E2F1 expression across the V5 positive tumour types. F. Total PRMT5 expression across the V5 positive tumour types. G. Total p100-TSN expression across the V5 positive tumour types

Chapter 7. Discussion

7.1 Summary of findings

In this study we have demonstrated that not only does arginine methylation on E2F1 affect transcription efficiency of a variety of target genes, but also we discovered a novel function of E2F1 in RNA processing which is impacted by methylation status. By utilising both wild-type and methylation dead point mutants of E2F1, R109K and R111/113K, we have been able to carry out a genome-wide analysis to ascertain the impact these methylation sites have on target gene transcription. We discovered that there were few differences in RNA species up, or downregulated compared to the vector control for WT and KK, which cannot be symmetrically dimethylated by PRMT5. In contrast, there was little overlap between target genes when comparing WT or KK to R109K, which cannot be asymmetrically dimethylated by PRMT1. Additionally, R109K overall had a smaller gene pool of upregulated target genes compared to WT or KK, thus suggesting that not only are the specific target genes reduced in number, but that R109K may also be less transcriptionally active than either KK or WT as we have previously identified (Zheng et al., 2013). Crucially, gene ontology analysis revealed that the genes uniquely upregulated in R109K were involved in positive regulation of proliferation – a biological phenotype we have previously observed (Zheng et al., 2013). Interestingly, when we performed the same GO analysis on genes uniquely upregulated in WT we observed enrichment in ion transport and signal transduction, suggesting a link with intracellular signalling and homeostasis. For KK, the enriched GO terms included transcriptional regulation and cell-cell signalling, which included some apoptotic signalling molecules and receptors such as IL-7 and GALR2.

Exon-use analysis of the RNA-seq data revealed that altered methylation status of E2F1 resulted in altered isoform expression which was most apparent in KK compared to WT. Given the fact that most differentially expressed genes were shared between WT and KK, this implied that the phenotypic effects observed in KK may be more likely to be due to alternative splicing

or RNA processing rather than through regulation of transcription of the largely shared set of target genes. Furthermore, the genes identified in the exon-use analysis were either below the threshold for differential expression analysis (lower than 1-fold), or only displayed low differential expression compared to pTRE. Thus, this represents a distinct set of E2F1 target genes which are not regulated transcriptionally by E2F1, but do show exon use differences which are impacted on by arginine methylation status of E2F1. Importantly, we were able to validate these findings for a selection of genes.

As methylation at R111/113 by PRMT5 is required for the interaction between E2F1 and p100-TSN (Zheng et al., 2013), and given the role of this reader protein in RNA processing, we hypothesised that E2F1 may be impacting target gene splicing through p100-TSN. To this end, we demonstrated that both p100-TSN and E2F1 can bind to target gene RNA and, critically, this can be abolished by RNase A treatment. Furthermore, the interaction between E2F1 and the RNA was shown to be dependent on PRMT5 activity, as well as the presence of p100-TSN, but importantly p100-TSN is not required for E2F1 presence at target gene promoters, although it can both positively and negatively influence this depending on the particular target gene. Thus, recognition of methylation at R111 and R113 by p100-TSN is required for E2F1 interaction with target gene RNA. We hypothesised that this interaction is indirect, through p100-TSN itself, or through other protein members of the complex. Additionally, we showed that an intact DNA binding domain is required for the association between E2F1 and target gene RNA, suggesting a co-transcriptional model where nascent RNA is in a complex containing E2F1 and p100-TSN [Figure 7.1A]. Subsequently, we performed a sequencing experiment after RNA immunoprecipitation which showed that although p100-TSN was required for E2F1 binding to a specific set of RNA, E2F1 could bind a separate pool of RNA in the absence of p100-TSN. Therefore, E2F1 may also be able to interact with other RNA species in the absence of p100-TSN possibly through an alternative binding partner, or mechanism [Figure 7.1B].

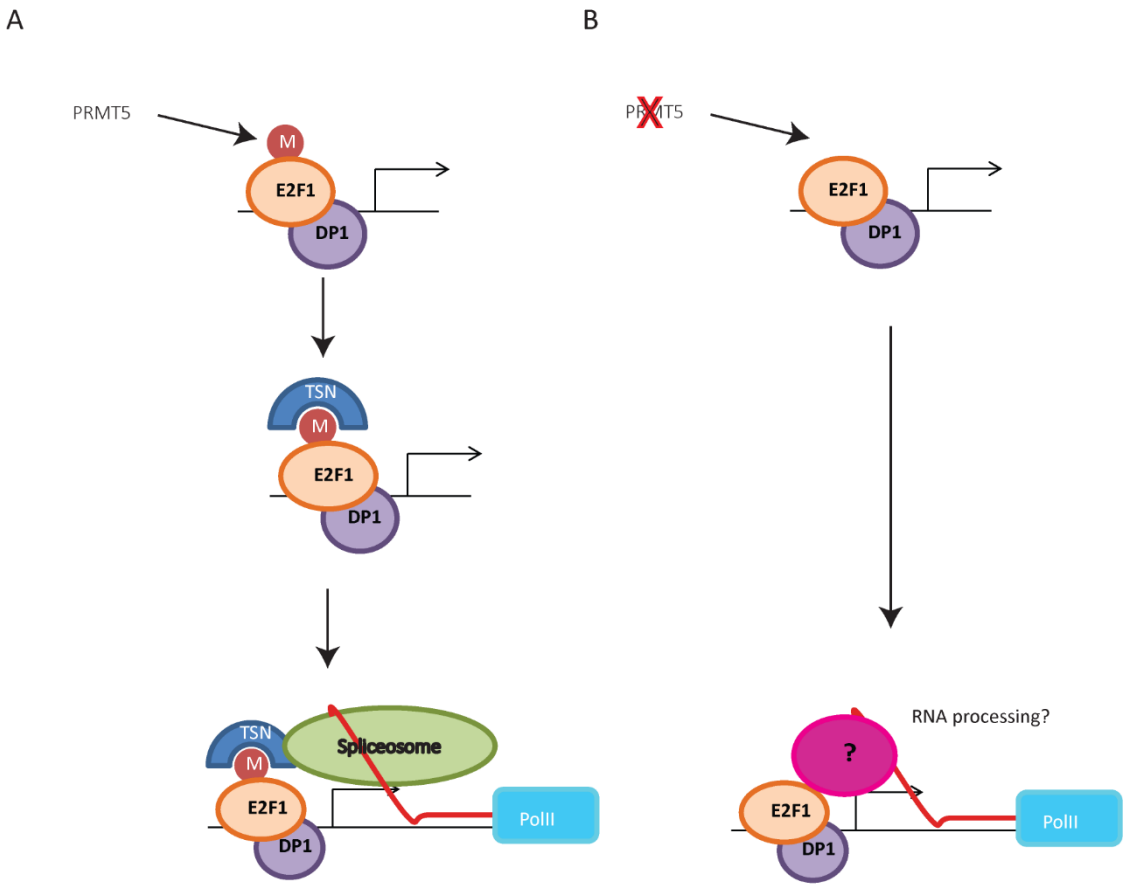


Figure 7.1 E2F1 links transcription with the splicing machinery through arginine methylation-dependent interaction with p100-TSN.

Diagram of the co-transcriptional splicing model showing A. arginine methylation dependent recruitment of the nascent RNA and spliceosome to E2F1, and B. showing the as yet unidentified bridge between E2F1 and the nascent RNA in the absence of PRMT5 or p100-TSN and another potential link with RNA processing

Additionally, we identified a potential link to the spliceosome through analysing mass spectrometry results which showed that two core members of the spliceosome (PRP38 and SmD1) could be co-purified with E2F1 after overexpression of PRMT5. Although we were not able to confirm a protein-protein interaction, we did observe an interaction between E2F1 and snRNA U4 and U6, members of the tri-snRNP complex with PRP31, which was dependent on PRMT5 activity, p100-TSN and an intact DNA binding domain.

We also observed p100-TSN binding to E2F1 target gene RNA (such as p73 and BCLX) and, surprisingly, we found that the presence or absence of E2F1 could influence the isoform of RNA that p100-TSN was bound to. For example, under conditions where p100-TSN was bound to BCL-X RNA this was associated with reduced levels of a particular isoform, whereas for p73, conditions which favoured RNA binding to p100-TSN, resulted in greater cellular levels of that isoform. Finally, when we sequenced the RNA bound to p100-TSN following immunoprecipitation we found that in the absence of E2F1, overall the amount of different RNA species associated with p100-TSN was reduced, and represented a distinct pool of genes compared to the RNAs bound to p100-TSN in the presence of E2F1.

7.2 The arginine methylation status of E2F1 impacts on transcriptional control

As our previous studies have shown, arginine methylation on E2F1 results in an altered transcription profile and subsequent functional outcome, with R109Me2a by PRMT1 causing apoptosis and R111/113Me2s by PRMT5 driving proliferation (Cho et al., 2012, Zheng et al., 2013). Using conservative R to K point mutants, to minimise impact on protein structure, we generated an inducible cell line system expressing wild-type, or methylation dead E2F1 (R109K and R111/113K) to further investigate the impact of arginine methylation on E2F1.

As previously reported, inhibition of PRMT1 mediated methylation on E2F1 at R109 results in reduced chromatin binding and transcription across a variety of target genes suggesting that R109 methylation is required for efficient transcription which can drive apoptotic signalling (Zheng et al., 2013). Concordantly with this, our results showed that the R109K mutant was unable to stimulate transcription of genes to the same level that both WT and KK could – fewer genes were upregulated two-fold above the vector control. Furthermore, when we looked at target genes uniquely enhanced in R109K expressing-cells compared to WT or KK, we observed that for DEF6, the expression was not increased compared to the vector control, but instead expression of DEF6 was repressed in WT and KK compared to pTRE and R109K. For only one gene, KCNIP1, did we observe an increase in expression in R109K that was above expression levels observed in all other cell lines. KCNIP1 is a voltage-gated potassium channel-interacting protein located in intracellular vesicles andit is believed to act as a calcium sensor. A few studies have identified enhanced expression of KCNIP1 in cancer cell lines, but the function of this protein remains largely unexplored (Reinhold et al., 2012, Li et al., 2012). Of course, the reduced transcription rates observed for R109K compared to WT or KK could be consistent with the R109K E2F1 protein having a shorter half-life (Zheng et al., 2013). However, the nature of the Tet-On system should mean there is an excess of protein and the effect of half-life might be expected to be minimal, especially as endogenous expression of the DP subunit would be the limiting factor in transcriptional activation. As previous experiments have shown, loss of asymmetric arginine methylation on E2F1 through silencing PRMT1 or R109 mutation is associated with increased proliferation and a reduction in apoptosis following DNA damage (Zheng et al., 2013). In line with this, we found that genes uniquely upregulated by R109K overexpression were enriched for GO terms involved in positive regulation of proliferation including VASH2 (required for angiogenesis) and SHIP1, a player in the PI3K pathway (Fernandes et al., 2013). VASH2 has recently been identified as a positive regulator of the E2F1 target gene ribonucleotide reductase regulatory subunit M2 (RRM2), via Jun upregulation (Tu

et al., 2017). As well as increasing the conversion of ribonucleotide diphosphate to deoxyribonucleotide diphosphate required for DNA synthesis in proliferating cells, RRM2 can increase resistance to gemcitabine, (Duxbury et al., 2004, Grolmusz et al., 2016). Thus R109K may be able to induce RRM2 directly and indirectly to promote cell cycle progression. Interestingly, another potential target gene unique to R109K is $p27^{Kip1}$, a tumour suppressor which regulates G1 to S phase transition (Agarwal et al., 2013). As R109K has increased proliferation, we would expect that there is deregulation of another part of this pathway to overcome the inhibitory effects of $p27^{Kip1}$, likely through enhanced PI3K pathway signalling (Cunningham et al., 2002, Prasad et al., 2015, Hallstrom et al., 2008).

Both WT and KK showed similar numbers of genes upregulated after overexpression, and many of these upregulated genes were common to both cell lines, thus suggesting similar transcription profiles. Of these genes, several are involved in signalling pathways involved in apoptotic and anti-proliferative signalling, including JMY and HOXA10. JMY is a co-factor of p53 mediated transcription and localises to the nucleus after DNA damage to stimulate the p53 response (Lin et al., 2014, Adams et al., 2012). Similarly, HOXA10 has been shown to increase expression of p53 and therefore influence p53 mediated transcription to elicit anti-tumour effects in breast cancer (Chu et al., 2004, Chen et al., 2012). This fits in with our previous findings that KK and WT are able to elicit a strong proapoptotic response to DNA damage, and from our experiments which showed a higher basal level of apoptosis in these two cell lines (Cho et al., 2012, Zheng et al., 2013).

7.3 PRMT5 mediated arginine methylation of E2F1 at R111/113 can affect RNA processing

The role of PRMT5 in tumorigenesis and progression, as well as in splicing has been well studied (Gu et al., 2012a, Gu et al., 2012b, Sheng and Wang, 2016, Wang et al., 2008, Yan et al., 2014).

PRMT5 mediated methylation of multiple Sm proteins, including SmD1, enables assembly of the core snRNP particles, as well as influencing function of multiple other splicing factors. Thus, PRMT5 activity is crucial for functional splicing (Meister et al., 2001, Koh et al., 2015, Bezzi et al., 2013, Gu et al., 2012a, Gu et al., 2012b). Our data suggests an additional role of PRMT5 in the regulation of splicing through its target E2F1, as loss of the PRMT5 methylation sites on E2F1 resulted in aberrant splicing, despite otherwise functional PRMT5 in cells. Despite the similarities in transcriptional profile between WT and KK, we observed many differences in isoform expression and exon use. The effects we saw on splicing of proapoptotic proteins BCL-XS and TAp73 are consistent with the phenotype we observe in the cells, with KK expressing higher proportions of the proapoptotic isoforms and having enhanced sensitivity to apoptotic stimuli, and R109K expressing lower amounts of the proapoptotic isoforms and showing resistance to apoptosis. This is also consistent with our previous work which identified the R109K mutation as conferring resistance to apoptosis and the KK mutation as sensitising the cells to apoptosis (Zheng et al., 2013).

Several studies have focused on identifying the proteins involved in BCL-X splice site selection with PTBP1 and SRSF2 being key modulators of this (Merdzhanova et al., 2008, Bielli et al., 2014). This is of particular interest as BCL-XL is often overexpressed in cancers and contributes to the antiapoptotic properties observed in many cancer cells. Additionally, overexpression of BCL-X can contribute to resistance to BCL-2 antagonists (Punnoose et al., 2016, Choudhary et al., 2015). It would be interesting, therefore, to investigate the use of PRMT5 inhibitors in combination with BCL-2 antagonists. From the work presented here, we would expect that loss of PRMT5 mediated methylation on E2F1 would increase the amount of BCL-XS splicing and this would help to overcome BCL-XL overexpression mediated resistance in these cells.

TP73 is a complex gene with multiple splice isoforms – many exon changes occur in the C-terminus, but as yet these have remained largely unstudied in terms of function. Instead, much

work has been done on the N-terminal exon changes as loss of part, or all, of the transactivation-domain results in a dominant-negative anti-apoptotic protein (Concin et al., 2004, Cam et al., 2006, Seelan et al., 2002). In addition to splice changes, there is a second promoter within intron 3 that can also be used to produce an N-terminally truncated version of this protein (Concin et al., 2004, Cam et al., 2006). Although we were unable to assess the levels of DNex2p73 and DNex23p73, we saw minimal changes in both DNp73 (second promoter) and DN'p73, which, due to a retained intron with an early stop codon, results in the same protein product as DNp73, despite use of the E2F1 responsive first promoter (Pützer et al., 2003, Cam et al., 2006, Seelan et al., 2002). Interestingly, creation of DN'p73 rather than DNp73 has been shown to be the major driver of truncated p73 isoforms in ovarian cancer, and the role of E2F1 in this setting is believed to contribute to this (Concin et al., 2004). Conversely, we observed increased TAp73 expression in KK expressing cells, which is likely to contribute to the high level of basal apoptosis in addition to increased sensitivity to DNA damage.

We identified RBM38 as a target of alternative exon use in the RNA-seq and confirmed that R109K expression hinders V3 isoform production, although total expression levels of RMB38 did not differ greatly. RBM38 is an RNA binding protein involved in many pathways including p53 response. RBM38 has previously been shown to be a direct transcriptional target of p53 and is upregulated after DNA damage (Shu et al., 2006). RBM38 contains RNA binding motifs in the first and third exon which bind to multiple RNA targets, including $p21^{Cip1}$ which is stabilised by this interaction, resulting in increased $p21^{Cip1}$ expression and cell cycle arrest. Although only the full-length variant 3 RBM38 (sometimes called isoform A) is capable of stabilising $p21^{Cip1}$ and inducing cell cycle arrest (Shu et al., 2006). RBM38 has also been validated as an E2F1 target gene and acts to inhibit E2F1 driven proliferation, in part via the regulation of $p21^{Cip1}$, in a negative feedback loop (Feldstein et al., 2012). Our data suggest that asymmetric arginine methylation at R109 is required for production of full length V3, as the

R109K point mutant was unable to stimulate expression of this isoform to the same levels as WT or KK. This is consistent with the R109K cell line showing greater proliferative potential and reduced response to DNA damage, and may also provide a way of E2F1 escaping this negative feedback loop as increased symmetric methylation by PRMT5 at R111/113 inhibits methylation at R109 by PRMT1 (Zheng et al., 2013). Other RNA targets of RBM38 have also been identified, but for the majority it has not been reported which isoform of RBM38 is capable of binding these targets, or if they have the same effects on stabilisation as the full length version. Generally, these targets are all oncogenes – key to RBM38's role as a tumour suppressor – and include MDM2, an inhibitor of p53, and c-Myc (Li et al., 2017, Xu et al., 2013). Conversely, zonula occludens-1 (ZO-1) mRNA is stabilised by RBM38 which then inhibits TGFβ driven EMT in breast cancer (Wu et al., 2017). Overall, RBM38 is a tumour suppressor with many RNA targets, and in line with this, loss, or reduced expression of RBM38 is associated with tumour formation and oncogenesis (Zhang et al., 2014a). Although the full length isoform has been described as important for destabilising $p21^{Cip1}$ to induce cell cycle arrest, no other reports have identified the function of various isoforms and how these may have alternate impacts on the pathways involved. For example, if these proteins are able to bind to the same targets, but not alter stability, they would act competitively with the full length variant. However, there may be other functions for the smaller isoforms which have not been discovered. Therefore it is of vital importance that researchers begin to assign isoform specific functions to all genes in their research.

Similarly to RBM38, SAGE1 was also identified in the RNA-seq as having altered exon use between the cell lines, with expression of the short isoform 203 reduced in R109K expressing cells. Although the impact of this is unclear as no assigned function has been given to SAGE1 overall, much less the various isoforms which have been described. It is expressed solely in testis in healthy tissue, but is often upregulated in tumours of various origins including patients with head and neck squamous cell carcinoma. Thus it is referred to a cancer/testis antigen

(Atanackovic et al., 2006, Maheswaran et al., 2015). We would expect that the loss of multiple WD40 repeats in SAGE 203 will negatively impact its potential binding partners and this isoform may have reduced function compared to the full length 201, but further research into the function of both the full-length and the short isoform will need to be conducted to assign function.

Speed of transcription has long been linked with splicing efficiency, with increased transcription speeds resulting in reduced splicing and enhanced intron retention, although this increased transcriptional speeds can also reduce the recruitment of silencers onto the pre-mRNA, to result in increased splicing for specific genes. Additionally, faster transcription speeds favour stronger splice site use over weaker alternatives. Thus, E2F1 may be impacting on splicing through increasing transcriptional speed (Braberg et al., 2013, Alpert et al., 2017, Fong et al., 2014, Ardehali et al., 2009). However, our results suggest that for the majority of genes we identified as alternatively spliced depending on E2F1, there was little change in differential expression. Whilst a more thorough analysis of transcriptional speed of the different mutant derivatives would be required to completely rule this effect out, we would expect that greater transcriptional speed would result in increased expression of that RNA species and we did not observe expression changes between the cell lines for those genes which were alternatively spliced. Recent advances in sequencing technologies now enable measurement of the speed of transcription on individual transcripts. Global run-on sequencing (GRO-seq) is performed after a block-release of RNAPII, following release BrUTP is incorporated into the nascent RNA and harvest at specific time points following release results in nucleotide-resolution measurement of elongation speed (Jonkers and Lis, 2015, Core et al., 2008). This would be a useful tool in determining if the elongation rates caused by the various point mutants of E2F1 are globally affecting transcription speed, as well as determining if the specific transcripts we see as being alternatively spliced after over-expression of a particular point mutant have altered elongation rates.

Another possible influence for E2F1 may be having on splicing which we have not addressed in this study is via its roles in chromatin remodelling. Generally this also favours faster transcription – but histone modifications have also been observed to recruit various splicing factors as well as alter nucleosome structure within exons. For example E2F1 can recruit the histone acetyl transferase adaptor TRRAP which can then assemble the TIP60 complex to E2F1 target genes, resulting in acetylation of H3 and H4 (Taubert et al., 2004) and histone H4 acetylation can reduce splicing factor SRp40 association with nascent RNA (Hnilicová et al., 2011). Furthermore, H3K36me3 mark, which is enriched at E2F1 target gene promoters has also been found to recruit splicing factors PTB and PSIP1 to the nascent RNA (Pradeepa et al., 2012, Luco et al., 2010, Roworth et al., 2015). In addition, E2F1 can recruit the chromatin remodelling complex SWI/SNF, through association with ARID1B, to active transcription sites and, as described earlier another component of this complex, BRM, can interact with several splicing components (Tyagi et al., 2009, Batsche et al., 2006, Nagl et al., 2007). Future work on the roles of arginine methylation of E2F1 and the impact this has on chromatin structure will be required to understand if this can affect recruitment of splicing factors to nascent RNA.

Although we identified few splicing factors being differentially expressed after overexpression of the E2F1 mutant derivatives, there is some previous work which shows that E2F1 can regulate expression of splicing factors to affect RNA processing. For example, SRSF2 (SC35) was reported to be an E2F1 target gene required for splicing the proapoptotic isoforms of BCL-X, caspases-8 and -9, as well as c-FLIP in lung adenocarcinoma (Merdzhanova et al., 2008). However, we did not observe an upregulation of SRSF2 after overexpression of WT E2F1 or either of the mutant derivatives compared to pTRE in our U2OS setting. Neither did we observe differential expression of SRSF2 between the different E2F1 derivative cell lines. Thus, the observation we made that BCL-XS was moderately reduced in the R109K expressing cells, and increased in the KK expressing cells was not due to a change in expression of SRSF2. In another study, BCL-XS splicing was shown to be dependent on PTBP1, as PTBP1 bound within exon 2 to

influence splice site selection, and reduced expression of PTBP1 enhanced expression of BCL-XL (Bielli et al., 2014). It would be interesting to determine if the cooperation observed by Merdzhanova and colleagues is due to a more direct complex formation between E2F1 and SRSF2 in the splicing complex, or indeed between PTBP1 and E2F1 or p100-TSN. Additional splicing factors which have been shown to potentially be regulated by E2F1 in response to DNA damage include RNPC-7 *and* SNRPE (Li et al., 2005), but once again these were not identified as differentially expressed upon induction of E2F1 or the mutant derivatives in our study.

Thus, methylation at R111/113 is required to keep the cells in an anti-apoptotic, pro-growth state and this is largely influenced by changes in splice isoform expression, rather than differential expression of different gene sets. Additionally, this change in isoform selectivity is not due to differential expression of splicing factors or altered transcriptional speed but appears to be more directly linked with transcription of the target gene and the spliceosome. This makes PRMT5 a good target for chemical inhibition in cells expressing deregulated E2F1 as high expression levels of the anti-apoptotic variants would be switched to high levels of their proapoptotic counterparts, aiding tumour cell destruction. More work is required to understand to roles of most alternative isoforms described in the literature, as the majority of papers focus on the major, longest isoform (or sometimes do not specify at all) protein product. Thus there is an overall general lack of understanding about how splicing changes can impact individual proteins and this needs to be addressed in future work.

7.4 E2F1 can bind to target gene RNA and influence alternative splicing

Although many links have been found between transcription and splicing, with the majority of splicing believed to happen co-transcriptionally (Carrillo Oesterreich et al., 2016), we have demonstrated for the first time that there is a physical connection between the transcription factor, nascent RNA and the spliceosome. We hypothesise that the transcription factor E2F1

can recruit splicing machinery to the DNA in anticipation of the start of transcription to enable efficient splicing of the nascent RNA. Indeed, we have shown that inhibition of DNA binding by the transcription factor results in loss of the complex formation suggesting this is occurring on the DNA, co-transcriptionally. Further experiments would be required to confirm this – cellular fractionation prior to RIP would determine if the complex is indeed chromatin associated, and mutation of the E2F1 binding motifs in a transcription reporter assay could also be a useful way of measuring this. Additionally, being able to identify protein-protein interactions of E2F1, p100-TSN and the splicing complex would be useful, and optimising antibodies to members of the splicing complexes would be required. This would enable further investigations into the dynamics of the system and additionally allow assessment of any changes which occur after depletion of E2F1 or p100-TSN which would be useful indicators to understand the function of these proteins in splicing, together with other potential factors involved. This is vitally important as our data suggest that E2F1 may be binding to the assembling spliceosome – we saw an interaction between E2F1 and snRNA U4 and U6 and as snRNA U4 is only present in the B complex of the immature spliceosome [Figure 1.5] this suggests that E2F1 is present during spliceosome assembly. Furthermore, as p100-TSN has been shown to facilitate assembly of the spliceosome on pre-mRNA via interactions with PRP8, a core protein in the U4, U5, U6 tri-snRNP complex, and increase the speed for the A to B complex conversion (Yang et al., 2007). Thus, we suggest that E2F1 may be recruiting p100-TSN to the pre-mRNA at the point of the start of transcription, to enable p100-TSN directed spliceosome assembly on the nascent RNA [Figure 7.2A]. In addition to this the RNA species we identified as bound to E2F1 were all amplified using primers across exon-exon junctions (therefore indicating a spliced product). This suggests that the interaction of E2F1 with the spliceosome occurs during assembly and continues once the spliceosome is active through engagement with the RNA. This could indicate protein-RNA interactions downstream of the already spliced exons, as the rest of the nascent RNA may be in the process of being transcribed and spliced [as in Figure 7.2A].

However, the dynamics and organisation of this remain unknown, and other effects of E2F1 on different aspects of RNA processing cannot be ruled out.

For example, our initial hypothesis, that methylation dependent recruitment of p100-TSN was required for the E2F1-RNA link was overturned by the results of the RIP-seq which demonstrated that E2F1 can bind to a different selection of RNA in the absence of p100-TSN. However, the mechanism behind this remains unclear. It would be interesting to identify proteins in a complex with E2F1 in the absence of p100-TSN via affinity purification mass spectrometry in order to shed light onto other potential binding partners which may act as a bridge between E2F1 and RNA, and the role this may have in RNA processing.

7.5 E2F1 can alter p100-TSN's RNA binding specificity

The role of p100-TSN in splicing has been well documented, and our data offer a further level of complexity to this. We have shown that p100-TSN can be recruited by E2F1 to a selection of RNA species, and loss of E2F1 both alters the RNA species p100-TSN is bound to as well as impacting on the RNA isoform p100-TSN is associated with. The function of this is unclear – our investigations revealed that for p73, when p100-TSN was bound to a particular isoform, this was associated with increased levels of this isoform being expressed, suggesting it was enhancing splicing of one isoform over another, or stabilising the RNA of that species. Conversely, for BCL-X, when p100-TSN was bound to a particular isoform this was associated with greater cellular levels of the other isoform, suggesting p100-TSN may be inhibiting expression of the isoform it is bound to. This may be due to regulation by other factors, or through nuclease activity of the SN domains of p100-TSN, more work would be required to determine this.

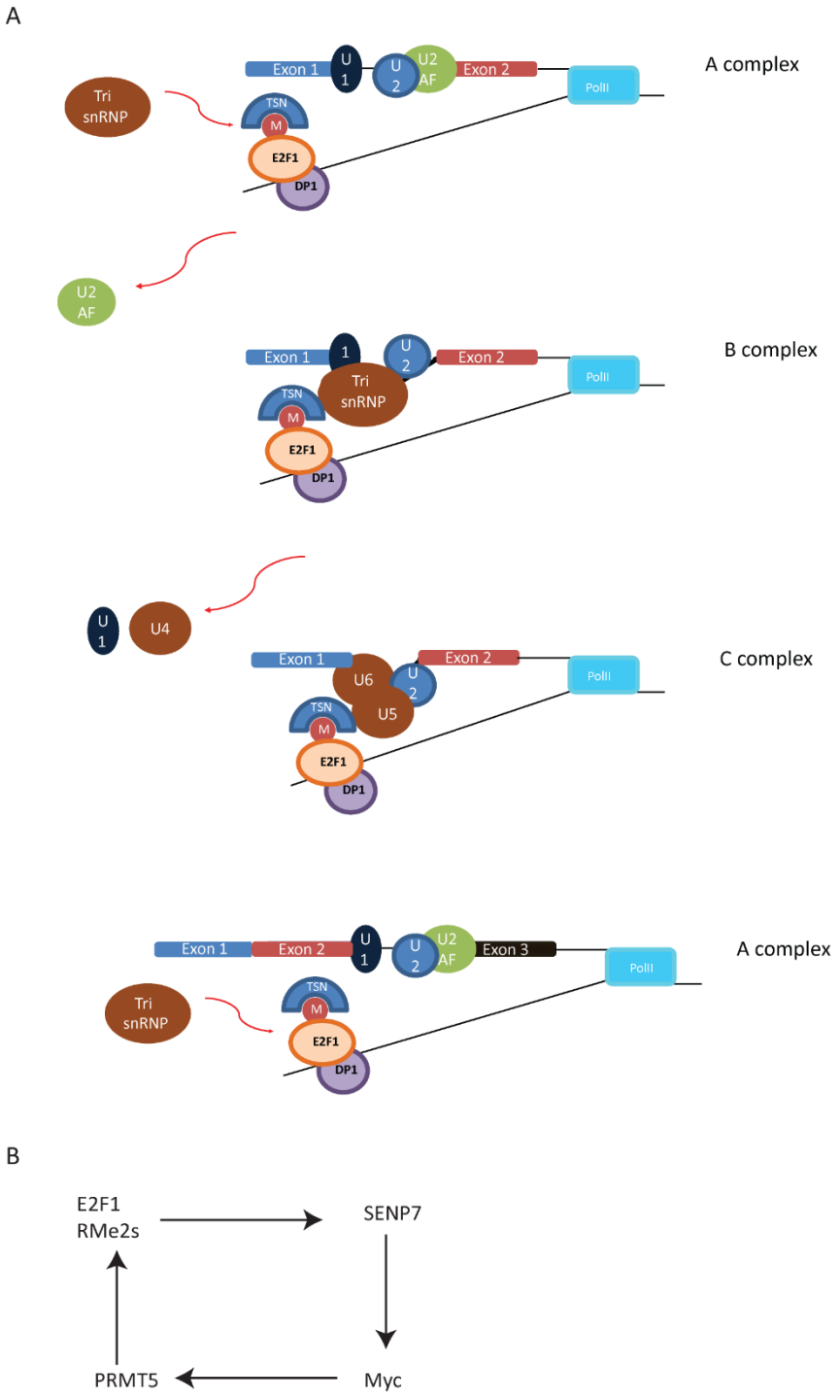


Figure 7.2 Possible mechanism for p100-TSN recruitment of snRNP components to nascent RNA of E2F1 target genes.

A. Diagram showing the E2F1-p100-TSN complex at the promoter, p100-TSN mediated recruitment of the tri-snRNP complex to increase splicing efficiency and continued p100-TSN mediated recruitment of the splicing components to splice sites further along the RNA during transcription.

B. Potential feedback loop between symmetrically methylated E2F1 (R111/113Me2s) which positively regulates SENP7 V5 expression, SENP7 can deSUMOylate Myc, leading to its stabilisation and therefore increased transcription, PRMT5 is a Myc target gene which methylates E2F1 at R111/113.

Our data show that loss of p100-TSN results in reduced levels of BCL-XS, despite enhanced interaction with this isoform after depletion of E2F1. As described earlier, Paronetto *et al* have shown that SAM68 can also regulate BCL-X splicing to favour production of BCL-XS (Paronetto et al., 2007) and PRMT2 mediated re-localisation of SAM68 to the cytoplasm increases BCL-XL (Vhuiyan et al., 2017). This is potentially relevant for our work as SAM68 has also been shown to interact with p100-TSN. Although the Tudor domain of p100-TSN was shown as important for the function of SAM68 to be maintained, the authors did not identify the binding site for SAM68 on p100-TSN (Cappellari et al., 2014). Therefore it would be interesting to identify if E2F1 and SAM68 exist in a complex with p100-TSN or if they compete for the Tudor domain. Additionally, other splicing factors have been shown to be recruited to p100-TSN by the tudor domain including PRP31 and SmD1 (Meister et al., 2001, Gao et al., 2012). This would suggest that they could be competing with E2F1, although according to our mass-spectrometry data we also see a potential E2F1 interaction with both of these proteins. Therefore it would be important to understand the dynamics of this complex further, in order to elucidate a function in the presence and absence of E2F1 or p100-TSN.

7.6 E2F1 can bind to SENP7 V5 RNA and may impact SUMOylation

We were also able to identify an isoform specific interaction between E2F1 and SENP7 V5 RNA which was dependent on the presence of p100-TSN. Loss of E2F1 or p100-TSN resulted in slightly reduced expression of V5 suggesting that splicing of V5 is partially dependent on E2F1 and p100-TSN. SENP7 is a deSUMOylase which has preference for poly-SUMO groups, rather than mono-SUMOylated substrates, and also show preference to SUMO2 or SUMO3 groups (Lima and Reverter, 2008, Alegre and Reverter, 2011, Shen et al., 2009). All reported variants of SENP7 retain their protease C-terminal domains, but have alterations in the N-terminus. The function of these have not been fully explored, but there is evidence to suggest that changes

in these exons can influence SENP7 substrate recognition as several SUMO interaction motifs, as well as other binding motifs have been recognised. For example, a pair of PxVxL HP1 boxes in exons 5 and 6 have been identified as HP1α binding sites (Romeo et al., 2015) And isoform V2, which lacks exon 6, cannot associate with of the chromatin remodeller HP1α. This increases HP1 SUMOylation and favours silencing of transcription at E2F target genes as well as genes required for EMT (Bawa-Khalfe et al., 2012). Thus, SUMOylated HP1α results in senescence driven by epigenetic silencing and specific isoforms of SENP7 can impact this (Zhu et al., 2015). Further investigations on SENP V2 have shown that it is cytoplasmic, and can catalyse deSUMOylation of β-Catenin and Axin1, part of the complex which targets β-Catenin for proteasomal degradation in the absence of WNT signalling. DeSUMOylation of both proteins is required to enhance their interaction and thereby increase β-Catenin degradation. Loss of SENP7 V2 in breast cancer is associated with enhanced WNT signalling through β-Catenin, increased transcription of β-Catenin regulated genes and increased proliferation (Karami et al., 2017). As V5 also lacks exon 6, in addition to exon 5, we can assume that it also lacks the ability to bind to and deSUMOylate HP1α, but this will need to be verified. This may therefore indicate a feedback loop between E2F1 and SENP7 – as enhanced E2F1 and PRMT5 levels would also increase V5 expression, resulting in increased HP1α SUMOylation and silencing of E2F1 target genes. Additionally, the localisation of V5 will also need to be elucidated and the role of V5 on β-Catenin and WNT signalling investigated.

In addition to regulating transcription, deSUMOylation by SENP7 has also been linked with efficient DNA repair by deSUMOylating KAP1. After DNA damage, HP1α recruits SENP7 to chromatin, but is not deSUMOylated itself. Instead KAP1 is deSUMOylated which reduces the affinity of KAP1 for the NuRD complex and methyltransferase SETDB1 which contribute to chromatin condensation. Thus, loss of SUMO-KAP1 relaxes the chromatin structure to allow homologous recombination to take place (Garvin et al., 2013). As this is also dependent on the

HP1α interaction, it would be expected that V5 lacks this function and has a reduced ability to respond to DNA damage.

Cyclic GMP-AMP synthase (cGAS) is a cytosolic DNA sensor which can trigger the immune response. After recognition of foreign DNA it produces cyclic GMP-AMP (cGAMP) which activates the ER-resident stimulator of interferon genes (STING) adaptor. The STING pathway results in the production of interferons and thus, the immune response. SENP7 has been described to potentiate this pathway via deSUMOylation of cGAS and is overexpressed in individuals with the autoimmune disease systemic lupus erythematosus resulting in increased inflammation in this setting (Cui et al., 2017). Although the sites of SUMOylation on cGAS were mapped in this study, the researchers did not identify the motif on SENP7 required for recognition of cGAS so we cannot tell if this is likely to be isoform specific. Nevertheless, it may be an interesting link between E2F1, SENP7 and tumour inflammation and should be further investigated.

SENP7 has also been reported to deSUMOylate Myc, which stabilises the oncogene as SUMOylated Myc is subsequently ubiquitinylated and targeted for proteasome-mediated degradation (González-Prieto et al., 2015). Once again, the site on SENP7 responsible for Myc recognition was not identified so it is unknown if different isoforms of SENP7 will have the same effect on stabilising Myc. This may indicate an interesting feedback loop if SENP7 V5 can deSUMOylate Myc, as Myc has been shown to upregulate PRMT5 and thereby impact on PRMT5 mediated splicing and tumourigenesis (Koh et al., 2015). Thus, symmetrically methylated E2F1 R111/113 would be expected to increase SENP7 V5, which in turn would stabilise Myc, resulting in increased Myc-dependent transcription and upregulation of PRMT5 that would stimulate greater arginine methylation on E2F1 R111/113 [Figure 7.2B]. Of course, to confirm this hypothesis there would need to be much more work to firstly show SENP7 V5 is unequivocally upregulated by E2F1 R111/113Me2s, as well as confirming if Myc is a target

for deSUMOylation by V5, and ongoing work is being done to start understanding this potential positive feedback loop.

Overall, although we have identified a link between SENP7 and E2F1 mediated splicing, we are not able at this time to assign a function to the isoform which E2F1 preferentially selects. Further investigations into the role of V5, and potential binding motifs within the N-terminal regions of SENP7 are likely to aid characterisation of this isoform. Although generally, SENP7 expression is reduced in various cancers compared to normal tissue, we were able to identify a potential link between high expression of PRMT5, p100-TSN, E2F1 and high levels of SENP7 V5, particularly in ovarian serous cystadenocarcinoma therefore this would be a good model for future experiments. It is also worth noting that the expression data from TGCA is based on RNA analysis, so there is a need to assess protein expression of the various isoforms as they may be expressed differently at the protein level compared to the RNA level. In high-grade OSC, over 10% of cases showed homozygous loss of pRB and 80% loss-of-function mutation in p53, suggesting that the E2F1 pathway is not only deregulated, but will be the primary proapoptotic drive in this case (Vang et al., 2009). Furthermore, the median survival rate is only 1.7 years for high-grade OSC and only 9% survival rate over 5 years whereas in low-grade tumours where there is minimal p53 mutation, with the oncogenic drivers more likely to be KRAS, ERBB2 and BRAF and the median survival is 4.2 years with 40% survival after 5 years (Vang et al., 2009). Indeed, the link between deregulated E2F1 and OSC has been previously reported with high expression observed in most OSC cell lines, as well as expression correlating with high grade and reduced survival and deregulated target gene expression (De Meyer et al., 2009, Zhan et al., 2016). In line with this, reduced E2F1 expression as a result of non-steroidal anti-inflammatory treatment, was shown to reduce proliferation of OSC cells in cell lines as well as a xenograft mouse model (Valle et al., 2013). Interestingly, also in this setting there was an increase in the N-terminally truncated isoforms of p73 which served to inhibit apoptotic signalling (Concin et al., 2004, Valle et al., 2013).

As we have shown, PRMT5 mediated methylation of E2F1 is required for E2F1's pro-survival functions and in OSC, which has both high expression of PRMT5 and E2F1, we would expect this to have a key role in ensuring that deregulated E2F1 does not trigger apoptosis. Thus, OSC may prove a good candidate for PRMT5 inhibitors as the loss of E2F1 R111/113Me2s would enable PRMT1 mediated asymmetric methylation at R109 to drive the proapoptotic transcriptional profile of E2F1, as well increasing the proapoptotic splice isoform profiles. This is especially true for those tumours lacking functional p53, as E2F1 will be able to drive p53-independent apoptosis through the p73 and BCL pathways.

7.7 Conclusion

In conclusion, we have demonstrated that arginine methylation of E2F1 can alter the transcriptional profile and functional outcome for the cell. Whilst symmetric dimethylation at R111/113 is associated with a pro-survival phenotype, asymmetric dimethylation at R109 elicits a proapoptotic response. We have shown that this is in part due to changes in target gene expression after arginine methylation of E2F1, but crucially, we have also shown that the effects observed are also mediated by changes in splicing patterns of genes which are not differentially expressed upon methylation status changes. In addition, for the first time, we have shown a physical link between transcription factors, nascent RNA and the spliceosome through the E2F1/p100-TSN axis and identified that E2F1 can direct p100-TSN to specific isoforms of RNA.

Future work in this area should focus on elucidating the mechanism behind E2F1 and spliceosome recruitment and RNA processing particularly in the absence of p100-TSN, in order to understand the dynamics of this complex and the impact it is having on alternative splicing. Additionally, the use of PRMT5 inhibitors should be investigated, with focus on the splicing outcomes for pro- and antiapoptotic target genes and the dependence of this on E2F1. If, as

we expect, BCL-XL expression is reduced after treatment with PRMT5 inhibitors in favour of BCL-XS, it would be interesting to look at combination therapies with BCL-2 antagonists as increased BCL-XL expression is a common mechanism for drug resistance. Another important area of research would be to elucidate functions for various isoforms of genes described in this work, as this will enable greater understanding of the biological effect E2F1 has on cellular outcome by impacting on splicing.

References

ADAMS, C. J., PIKE, A. C., MANIAM, S., SHARPE, T. D., COUTTS, A. S., KNAPP, S., LA THANGUE, N. B. & BULLOCK, A. N. 2012. The p53 cofactor Strap exhibits an unexpected TPR motif and oligonucleotide-binding (OB)-fold structure. *Proc Natl Acad Sci U S A,* 109**,** 3778-83.

ADLER, A. S., MCCLELAND, M. L., YEE, S., YAYLAOGLU, M., HUSSAIN, S., COSINO, E., QUINONES, G., MODRUSAN, Z., SESHAGIRI, S., TORRES, E., CHOPRA, V. S., HALEY, B., ZHANG, Z., BLACKWOOD, E. M., SINGH, M., JUNTTILA, M., STEPHAN, J. P., LIU, J., PAU, G., FEARON, E. R., JIANG, Z. & FIRESTEIN, R. 2014. An integrative analysis of colon cancer identifies an essential function for PRPF6 in tumor growth. *Genes Dev,* 28**,** 1068-84.

AGARWAL, P., SANDEY, M., DEINNOCENTES, P. & BIRD, R. C. 2013. Tumor suppressor gene p16/INK4A/CDKN2A-dependent regulation into and out of the cell cycle in a spontaneous canine model of breast cancer. *Journal of Cellular Biochemistry,* 114**,** 1355-63.

ALEGRE, K. O. & REVERTER, D. 2011. Swapping small ubiquitin-like modifier (SUMO) isoform specificity of SUMO proteases SENP6 and SENP7. *J Biol Chem,* 286**,** 36142-51.

ALEXANDER, R. D., INNOCENTE, S. A., BARRASS, J. D. & BEGGS, J. D. 2010. Splicing-dependent RNA polymerase pausing in yeast. *Mol Cell,* 40**,** 582-93.

ALLO, M., BUGGIANO, V., FEDEDA, J. P., PETRILLO, E., SCHOR, I., DE LA MATA, M., AGIRRE, E., PLASS, M., EYRAS, E., ELELA, S. A., KLINCK, R., CHABOT, B. & KORNBLIHTT, A. R. 2009. Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat Struct Mol Biol,* 16**,** 717-24.

ALPERT, T., HERZEL, L. & NEUGEBAUER, K. M. 2017. Perfect timing: splicing and transcription rates in living cells. *Wiley Interdiscip Rev RNA,* 8.

AN, W., KIM, J. & ROEDER, R. G. 2004. Ordered cooperative functions of PRMT1, p300, and CARM1 in transcriptional activation by p53. *Cell,* 117**,** 735-48.

ANDERSSON, R., ENROTH, S., RADA-IGLESIAS, A., WADELIUS, C. & KOMOROWSKI, J. 2009. Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res,* 19**,** 1732-41.

AOKI, I., HIGUCHI, M. & GOTOH, Y. 2013. NEDDylation controls the target specificity of E2F1 and apoptosis induction. *Oncogene,* 32**,** 3954-64.

ARAKI, K., NAKAJIMA, Y., ETO, K. & IKEDA, M.-A. 2003. Distinct recruitment of E2F family members to specific E2F-binding sites mediates activation and repression of the E2F1 promoter. *Oncogene,* 22**,** 7632-41.

ARDEHALI, M. B., YAO, J., ADELMAN, K., FUDA, N. J., PETESCH, S. J., WEBB, W. W. & LIS, J. T. 2009. Spt6 enhances the elongation rate of RNA polymerase II in vivo. *Embo j,* 28**,** 1067-77.

ATANACKOVIC, D., BLUM, I., CAO, Y., WENZEL, S., BARTELS, K., FALTZ, C., HOSSFELD, D. K., HEGEWISCH-BECKER, S., BOKEMEYER, C. & LEUWER, R. 2006. Expression of cancer-testis antigens as possible targets for antigen-specific immunotherapy in head and neck squamous cell carcinoma. *Cancer Biol Ther,* 5**,** 1218-25.

BACKLUND, P. S., JR. & SMITH, R. A. 1981. Methionine synthesis from 5'-methylthioadenosine in rat liver. *J Biol Chem,* 256**,** 1533-5.

BALDWIN, R. M., HAGHANDISH, N., DANESHMAND, M., AMIN, S., PARIS, G., FALLS, T. J., BELL, J. C., ISLAM, S. & CÔTÉ, J. 2015. Protein arginine methyltransferase 7 promotes breast cancer cell invasion through the induction of MMP9 expression. *Oncotarget,* 6**,** 3013-3032.

BARGIELA-IPARRAGUIRRE, J., PRADO-MARCHAL, L., FERNANDEZ-FUENTE, M., GUTIERREZ-GONZALEZ, A., MORENO-RUBIO, J., MUNOZ-FERNANDEZ, M., SERENO, M., SANCHEZ-PRIETO, R., PERONA, R. & SANCHEZ-PEREZ, I. 2016. CHK1 expression in Gastric Cancer

is modulated by p53 and RB1/E2F1: implications in chemo/radiotherapy response. *Sci Rep,* 6**,** 21519.

BATSCHE, E., YANIV, M. & MUCHARDT, C. 2006. The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat Struct Mol Biol,* 13**,** 22-29.

BAUREN, G. & WIESLANDER, L. 1994. Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell,* 76**,** 183-92.

BAWA-KHALFE, T., LU, L.-S., ZUO, Y., HUANG, C., DERE, R., LIN, F.-M. & YEH, E. T. H. 2012. Differential expression of SUMO-specific protease 7 variants regulates epithelial–mesenchymal transition. *Proceedings of the National Academy of Sciences,* 109**,** 5.

BEDFORD, M. T. & CLARKE, S. G. 2009. Protein arginine methylation in mammals: who, what, and why. *Molecular Cell,* 33**,** 1-13.

BENEDICT, M. A., HU, Y., INOHARA, N. & NÚÑEZ, G. 2000. Expression and functional analysis of Apaf-1 isoforms. Extra Wd-40 repeat is required for cytochrome c binding and regulated activation of procaspase-9. *J Biol Chem,* 275**,** 8461-8.

BERTIN-CIFTCI, J., BARRÉ, B., LE PEN, J., MAILLET, L., COURIAUD, C., JUIN, P. & BRAUN, F. 2013. pRb/E2F-1-mediated caspase-dependent induction of Noxa amplifies the apoptotic effects of the Bcl-2/Bcl-xL inhibitor ABT-737. *Cell Death and Differentiation,* 20**,** 755-64.

BERTRAM, K., AGAFONOV, D. E., LIU, W.-T., DYBKOV, O., WILL, C. L., HARTMUTH, K., URLAUB, H., KASTNER, B., STARK, H. & LÜHRMANN, R. 2017. Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature,* 542**,** 318-323.

BEZZI, M., TEO, S. X., MULLER, J., MOK, W. C., SAHU, S. K., VARDY, L. A., BONDAY, Z. Q. & GUCCIONE, E. 2013. Regulation of constitutive and alternative splicing by PRMT5 reveals a role for Mdm4 pre-mRNA in sensing defects in the spliceosomal machinery. *Genes Dev,* 27**,** 1903-16.

BIEDA, M., XU, X., SINGER, M. A., GREEN, R. & FARNHAM, P. J. 2006. Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. *Genome Research,* 16**,** 595-605.

BIELLI, P., BORDI, M., BIASIO, V. D. & SETTE, C. 2014. Regulation of BCL-X splicing reveals a role for the polypyrimidine tract binding protein (PTBP1/hnRNP I) in alternative 5′ splice site selection. *Nucleic Acids Res,* 42**,** 12070-81.

BIRD, G., ZORIO, D. A. & BENTLEY, D. L. 2004. RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. *Mol Cell Biol,* 24**,** 8963-9.

BISWAS, A. K. & JOHNSON, D. G. 2012. Transcriptional and nontranscriptional functions of E2F1 in response to DNA damage. *Cancer Research,* 72**,** 13-7.

BLACK, E. P., HALLSTROM, T., DRESSMAN, H. K., WEST, M. & NEVINS, J. R. 2005. Distinctions in the specificity of E2F function revealed by gene expression signatures. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 15948-53.

BOISVERT, F. M., DERY, U., MASSON, J. Y. & RICHARD, S. 2005. Arginine methylation of MRE11 by PRMT1 is required for DNA damage checkpoint control. *Genes Dev,* 19**,** 671-6.

BOTTGER, A., ISLAM, M. S., CHOWDHURY, R., SCHOFIELD, C. J. & WOLF, A. 2015. The oxygenase Jmjd6-a case study in conflicting assignments. *Biochemical Journal,* 468**,** 191-202.

BRABERG, H., JIN, H., MOEHLE, E. A., CHAN, Y. A., WANG, S., SHALES, M., BENSCHOP, J. J., MORRIS, J. H., QIU, C., HU, F., TANG, L. K., FRASER, J. S., HOLSTEGE, F. C., HIETER, P., GUTHRIE, C., KAPLAN, C. D. & KROGAN, N. J. 2013. From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. *Cell,* 154**,** 775-88.

BRIMACOMBE, R., STIEGE, W., KYRIATSOULIS, A. & MALY, P. 1988. Intra-RNA and RNA-protein cross-linking techniques in Escherichia coli ribosomes. *Methods Enzymol,* 164**,** 287-309.

BRUDNO, M., GELFAND, M. S., SPENGLER, S., ZORN, M., DUBCHAK, I. & CONBOY, J. G. 2001. Computational analysis of candidate intron regulatory elements for tissue-specific alternative pre-mRNA splicing. *Nucleic Acids Res,* 29**,** 2338-48.

BUDHAVARAPU, V. N., WHITE, E. D., MAHANIC, C. S., CHEN, L., LIN, F. T. & LIN, W. C. 2012. Regulation of E2F1 by APC/C Cdh1 via K11 linkage-specific ubiquitin chain formation. *Cell Cycle,* 11**,** 2030-8.

BUONAMICI, S., CHAKRABORTY, S., SENYUK, V. & NUCIFORA, G. 2003. The role of EVI1 in normal and leukemic cells. *Blood Cells Mol Dis,* 31**,** 206-12.

BURKE, J. R., LIBAN, T. J., RESTREPO, T., LEE, H.-W. & RUBIN, S. M. 2014. Multiple mechanisms for E2F binding inhibition by phosphorylation of the retinoblastoma protein C-terminal domain. *Journal of Molecular Biology,* 426**,** 245-55.

CAM, H., GRIESMANN, H., BEITZINGER, M., HOFMANN, L., BEINORAVICIUTE-KELLNER, R., SAUER, M., HUTTINGER-KIRCHHOF, N., OSWALD, C., FRIEDL, P., GATTENLOHNER, S., BUREK, C., ROSENWALD, A. & STIEWE, T. 2006. p53 family members in myogenic differentiation and rhabdomyosarcoma development. *Cancer Cell,* 10**,** 281-93.

CAMPANERO, M. R. & FLEMINGTON, E. K. 1997. Regulation of E2F through ubiquitin-proteasome-dependent degradation: stabilization by the pRB tumor suppressor protein. *Proc Natl Acad Sci U S A,* 94**,** 2221-6.

CAO, A. R., RABINOVICH, R., XU, M., XU, X., JIN, V. X. & FARNHAM, P. J. 2011. Genome-wide Analysis of Transcription Factor E2F1 Mutant Proteins Reveals That N- and C-terminal Protein Interaction Domains Do Not Participate in Targeting E2F1 to the Human Genome. *The Journal of Biological Chemistry,* 286**,** 11985-11996.

CAPPELLARI, M., BIELLI, P., PARONETTO, M. P., CICCOSANTI, F., FIMIA, G. M., SAARIKETTU, J., SILVENNOINEN, O. & SETTE, C. 2014. The transcriptional co-activator SND1 is a novel regulator of alternative splicing in prostate cancer cells. *Oncogene,* 33**,** 3794-3802.

CARNEVALE, J., PALANDER, O., SEIFRIED, L. A. & DICK, F. A. 2012. DNA damage signals through differentially modified E2F1 molecules to induce apoptosis. *Molecular and Cellular Biology,* 32**,** 900-12.

CARR, S. M., POPPY ROWORTH, A., CHAN, C. & LA THANGUE, N. B. 2015. Post-translational control of transcription factors: methylation ranks highly. *Febs j,* 282**,** 4450-65.

CARRILLO OESTERREICH, F., HERZEL, L., STRAUBE, K., HUJER, K., HOWARD, J. & NEUGEBAUER, K. M. 2016. Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. *Cell,* 165**,** 372-81.

CARRILLO OESTERREICH, F., PREIBISCH, S. & NEUGEBAUER, K. M. 2010. Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol Cell,* 40**,** 571-81.

CARTIER, J., BERTHELET, J., MARIVIN, A., GEMBLE, S., EDMOND, V., PLENCHETTE, S., LAGRANGE, B., HAMMANN, A., DUPOUX, A., DELVA, L., EYMIN, B., SOLARY, E. & DUBREZ, L. 2011. Cellular inhibitor of apoptosis protein-1 (cIAP1) can regulate E2F1 transcription factor-mediated control of cyclin transcription. *Journal of Biological Chemistry,* 286**,** 26406-17.

CERTO, M., DEL GAIZO MOORE, V., NISHINO, M., WEI, G., KORSMEYER, S., ARMSTRONG, S. A. & LETAI, A. 2006. Mitochondria primed by death signals determine cellular addiction to antiapoptotic BCL-2 family members. *Cancer Cell,* 9**,** 351-65.

CHALFANT, C. E., RATHMAN, K., PINKERMAN, R. L., WOOD, R. E., OBEID, L. M., OGRETMEN, B. & HANNUN, Y. A. 2002. De novo ceramide regulates the alternative splicing of caspase 9 and Bcl-x in A549 lung adenocarcinoma cells. Dependence on protein phosphatase-1. *J Biol Chem,* 277**,** 12587-95.

CHAN-PENEBRE, E., KUPLAST, K. G., MAJER, C. R., BORIACK-SJODIN, P. A., WIGLE, T. J., JOHNSTON, L. D., RIOUX, N., MUNCHHOF, M. J., JIN, L., JACQUES, S. L., WEST, K. A., LINGARAJ, T., STICKLAND, K., RIBICH, S. A., RAIMONDI, A., SCOTT, M. P., WATERS, N. J., POLLOCK, R. M., SMITH, J. J., BARBASH, O., PAPPALARDI, M., HO, T. F., NURSE, K., OZA,

K. P., GALLAGHER, K. T., KRUGER, R., MOYER, M. P., COPELAND, R. A., CHESWORTH, R. & DUNCAN, K. W. 2015. A selective inhibitor of PRMT5 with in vivo and in vitro potency in MCL models. *Nat Chem Biol,* 11**,** 432-437.

CHANG, B., CHEN, Y., ZHAO, Y. & BRUICK, R. K. 2007. JMJD6 is a histone arginine demethylase. *Science,* 318**,** 444-7.

CHARI, A., GOLAS, M. M., KLINGENHÄGER, M., NEUENKIRCHEN, N., SANDER, B., ENGLBRECHT, C., SICKMANN, A., STARK, H. & FISCHER, U. 2008. An assembly chaperone collaborates with the SMN complex to generate spliceosomal SnRNPs. *Cell,* 135**,** 497-509.

CHEN, M. & MANLEY, J. L. 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol,* 10**,** 741-754.

CHEN, Y., ZHANG, J., WANG, H., ZHAO, J., XU, C., DU, Y., LUO, X., ZHENG, F., LIU, R., ZHANG, H. & MA, D. 2012. miRNA-135a promotes breast cancer cell migration and invasion by targeting HOXA10. *BMC Cancer,* 12.

CHEUNG, N., FUNG, T. K., ZEISIG, B. B., HOLMES, K., RANE, J. K., MOWEN, K. A., FINN, M. G., LENHARD, B., CHAN, L. C. & SO, C. W. E. 2016. Targeting Aberrant Epigenetic Networks Mediated by PRMT1 and KDM4C in Acute Myeloid Leukemia. *Cancer Cell,* 29**,** 32-48.

CHEVILLARD-BRIET, M., TROUCHE, D. & VANDEL, L. 2002. Control of CBP co-activating activity by arginine methylation. *The EMBO Journal,* 21**,** 5457-5466.

CHO, E.-C., ZHENG, S., MUNRO, S., LIU, G., CARR, S. M., MOEHLENBRINK, J., LU, Y.-C., STIMSON, L., KHAN, O., KONIETZNY, R., MCGOURAN, J., COUTTS, A. S., KESSLER, B., KERR, D. J. & THANGUE, N. B. L. 2012. Arginine methylation controls growth regulation by E2F-1. *EMBO Journal,* 31**,** 1785-97.

CHOUDHARY, G. S., AL-HARBI, S., MAZUMDER, S., HILL, B. T., SMITH, M. R., BODO, J., HSI, E. D. & ALMASAN, A. 2015. MCL-1 and BCL-xL-dependent resistance to the BCL-2 inhibitor ABT-199 can be overcome by preventing PI3K/AKT/mTOR activation in lymphoid malignancies. *Cell Death and Disease,* 6**,** e1593-.

CHU, M. C., SELAM, F. B. & TAYLOR, H. S. 2004. HOXA10 regulates p53 expression and matrigel invasion in human breast cancer cells. *Cancer Biol Ther,* 3**,** 568-72.

CHUNG, J., KARKHANIS, V., TAE, S., YAN, F., SMITH, P., AYERS, L. W., AGOSTINELLI, C., PILERI, S., DENIS, G. V., BAIOCCHI, R. A. & SIF, S. 2013. Protein arginine methyltransferase 5 (PRMT5) inhibition induces lymphoma cell death through reactivation of the retinoblastoma tumor suppressor pathway and polycomb repressor complex 2 (PRC2) silencing. *Journal of Biological Chemistry,* 288**,** 35534-47.

CLIFFORD, B., BELJIN, M., STARK, G. R. & TAYLOR, W. R. 2003. G2 arrest in response to topoisomerase II inhibitors: the role of p53. *Cancer Res,* 63**,** 4074-81.

CONCIN, N., BECKER, K., SLADE, N., ERSTER, S., MUELLER-HOLZNER, E., ULMER, H., DAXENBICHLER, G., ZEIMET, A., ZEILLINGER, R., MARTH, C. & MOLL, U. M. 2004. Transdominant DeltaTAp73 isoforms are frequently up-regulated in ovarian cancer. Evidence for their role as epigenetic p53 inhibitors in vivo. *Cancer Res,* 64**,** 2449-60.

CORE, L. J., WATERFALL, J. J. & LIS, J. T. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science,* 322**,** 1845-8.

CRESS, W. D., JOHNSON, D. G. & NEVINS, J. R. 1993. A genetic analysis of the E2F1 gene distinguishes regulation by Rb, p107, and adenovirus E4. *Mol Cell Biol,* 13**,** 6314-25.

CRISCI, A., RALEFF, F., BAGDIUL, I., RAABE, M., URLAUB, H., RAIN, J. C. & KRAMER, A. 2015. Mammalian splicing factor SF1 interacts with SURP domains of U2 snRNP-associated proteins. *Nucleic Acids Res,* 43**,** 10456-73.

CUI, Y., YU, H., ZHENG, X., PENG, R., WANG, Q., ZHOU, Y., WANG, R., WANG, J., QU, B., SHEN, N., GUO, Q., LIU, X. & WANG, C. 2017. SENP7 Potentiates cGAS Activation by Relieving SUMO-Mediated Inhibition of Cytosolic DNA Sensing. *PLoS Pathog,* 13.

CUNNINGHAM, J. J., LEVINE, E. M., ZINDY, F., GOLOUBEVA, O., ROUSSEL, M. F. & SMEYNE, R. J. 2002. The cyclin-dependent kinase inhibitors p19(Ink4d) and p27(Kip1) are

coexpressed in select retinal cells and act cooperatively to control cell cycle exit. *Mol Cell Neurosci,* 19**,** 359-74.

DARNELL, R. B. 2010. HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley Interdiscip Rev RNA,* 1**,** 266-86.

DAS, S., ANCZUKOW, O., AKERMAN, M. & KRAINER, A. R. 2012. Oncogenic splicing factor SRSF1 is a critical transcriptional target of MYC. *Cell Rep,* 1**,** 110-7.

DAVID, C. J., BOYNE, A. R., MILLHOUSE, S. R. & MANLEY, J. L. 2011. The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes Dev,* 25**,** 972-83.

DAVID, C. J., CHEN, M., ASSANAH, M., CANOLL, P. & MANLEY, J. L. 2010. HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature,* 463**,** 364-8.

DE LA MATA, M. & KORNBLIHTT, A. R. 2006. RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nat Struct Mol Biol,* 13**,** 973-80.

DE MEYER, T., BIJSMANS, I. T., VAN DE VIJVER, K. K., BEKAERT, S., OOSTING, J., VAN CRIEKINGE, W., VAN ENGELAND, M. & SIEBEN, N. L. 2009. E2Fs mediate a fundamental cell-cycle deregulation in high-grade serous ovarian carcinomas. *J Pathol,* 217**,** 14-20.

DEGREGORI, J., LEONE, G., MIRON, A., JAKOI, L. & NEVINS, J. R. 1997. Distinct roles for E2F proteins in cell growth control and apoptosis. *Proceedings of the National Academy of Sciences of the United States of America,* 94**,** 7245-50.

DENG, X., VON KEUDELL, G., SUZUKI, T., DOHMAE, N., NAKAKIDO, M., PIAO, L., YOSHIOKA, Y., NAKAMURA, Y. & HAMAMOTO, R. 2015. PRMT1 promotes mitosis of cancer cells through arginine methylation of INCENP. *Oncotarget,* 6**,** 35173-35182.

DUXBURY, M. S., ITO, H., ZINNER, M. J., ASHLEY, S. W. & WHANG, E. E. 2004. RNA interference targeting the M2 subunit of ribonucleotide reductase enhances pancreatic adenocarcinoma chemosensitivity to gemcitabine. *Oncogene,* 23**,** 1539-48.

ENGELMANN, D. & PÜTZER, B. M. 2012. The Dark Side of E2F1: In Transit beyond Apoptosis. *Cancer Research,* 72**,** 571-575.

FAUQUIER, L., DUBOÉ, C., JORÉ, C., TROUCHE, D. & VANDEL, L. 2008. Dual role of the arginine methyltransferase CARM1 in the regulation of c-Fos target genes. *The FASEB Journal,* 22**,** 3337-3347.

FEARS, S., MATHIEU, C., ZELEZNIK-LE, N., HUANG, S., ROWLEY, J. D. & NUCIFORA, G. 1996. Intergenic splicing of MDS1 and EVI1 occurs in normal tissues as well as in myeloid leukemia and produces a new member of the PR domain family. *Proc Natl Acad Sci U S A,* 93**,** 1642-7.

FEI, D. L., MOTOWSKI, H., CHATRIKHI, R., PRASAD, S., YU, J., GAO, S., KIELKOPF, C. L., BRADLEY, R. K. & VARMUS, H. 2016. Wild-Type U2AF1 Antagonizes the Splicing Program Characteristic of U2AF1-Mutant Tumors and Is Required for Cell Survival. *PLoS Genetics,* 12**,** e1006384.

FELDSTEIN, O., BEN-HAMO, R., BASHARI, D., EFRONI, S. & GINSBERG, D. 2012. RBM38 is a direct transcriptional target of E2F1 that limits E2F1-induced proliferation. *Mol Cancer Res,* 10**,** 1169-77.

FERNANDES, S., IYER, S. & KERR, W. G. 2013. Role of SHIP1 in cancer and mucosal inflammation. *Ann N Y Acad Sci,* 1280**,** 6-10.

FISZBEIN, A. & KORNBLIHTT, A. R. 2017. Alternative splicing switches: Important players in cell differentiation. *Bioessays,* 39.

FONG, N., KIM, H., ZHOU, Y., JI, X., QIU, J., SALDI, T., DIENER, K., JONES, K., FU, X. D. & BENTLEY, D. L. 2014. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev,* 28**,** 2663-76.

FORBES, S. A., BHAMRA, G., BAMFORD, S., DAWSON, E., KOK, C., CLEMENTS, J., MENZIES, A., TEAGUE, J. W., FUTREAL, P. A. & STRATTON, M. R. 2008. The Catalogue of Somatic

Mutations in Cancer (COSMIC). *Current protocols in human genetics,* Chapter 10**,** Unit 10 11.

FORBES, S. A., BINDAL, N., BAMFORD, S., COLE, C., KOK, C. Y., BEARE, D., JIA, M., SHEPHERD, R., LEUNG, K., MENZIES, A., TEAGUE, J. W., CAMPBELL, P. J., STRATTON, M. R. & FUTREAL, P. A. 2011. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res,* 39**,** D945-50.

FRIETZE, S., LUPIEN, M., SILVER, P. A. & BROWN, M. 2008. CARM1 regulates estrogen-stimulated breast cancer growth through up-regulation of E2F1. *Cancer Res,* 68**,** 301-6.

FUJIMOTO, S., SEEBART, C., GUASTAFIERRO, T., PRENNI, J., CAIAFA, P. & ZLATANOVA, J. 2012. Proteome analysis of protein partners to nucleosomes containing canonical H2A or the variant histones H2A.Z or H2A.X. *Biol Chem,* 393**,** 47-61.

FUSHIMI, K., RAY, P., KAR, A., WANG, L., SUTHERLAND, L. C. & WU, J. Y. 2008. Up-regulation of the proapoptotic caspase 2 splicing isoform by a candidate tumor suppressor, RBM5. *Proc Natl Acad Sci U S A,* 105**,** 15708-13.

GALBIATI, L., MENDOZA-MALDONADO, R., GUTIERREZ, M. I. & GIACCA, M. 2005. Regulation of E2F-1 after DNA damage by p300-mediated acetylation and ubiquitination. *Cell Cycle,* 4**,** 930-9.

GALEJ, W. P., WILKINSON, M. E., FICA, S. M., OUBRIDGE, C., NEWMAN, A. J. & NAGAI, K. 2016. Cryo-EM structure of the spliceosome immediately after branching. *Nature,* 537**,** 197-201.

GALLEGO-PAEZ, L. M., BORDONE, M. C., LEOTE, A. C., SARAIVA-AGOSTINHO, N., ASCENSAO-FERREIRA, M. & BARBOSA-MORAIS, N. L. 2017. Alternative splicing: the pledge, the turn, and the prestige : The key role of alternative splicing in human biological systems. *Hum Genet*.

GAO, X., ZHAO, X., ZHU, Y., HE, J., SHAO, J., SU, C., ZHANG, Y., ZHANG, W., SAARIKETTU, J., SILVENNOINEN, O., YAO, Z. & YANG, J. 2012. Tudor Staphylococcal Nuclease (Tudor-SN) Participates in Small Ribonucleoprotein (snRNP) Assembly via Interacting with Symmetrically Dimethylated Sm Proteins. *Journal of Biological Chemistry,* 287**,** 18130-18141.

GAO, Y., ZHAO, Y., ZHANG, J., LU, Y., LIU, X., GENG, P., HUANG, B., ZHANG, Y. & LU, J. 2016. The dual function of PRMT1 in modulating epithelial-mesenchymal transition and cellular senescence in breast cancer cells through regulation of ZEB1. 6**,** 19874.

GARVIN, A. J., DENSHAM, R. M., BLAIR-REID, S. A., PRATT, K. M., STONE, H. R., WEEKES, D., LAWRENCE, K. J. & MORRIS, J. R. 2013. The deSUMOylase SENP7 promotes chromatin relaxation for homologous recombination DNA repair. *EMBO Rep,* 14**,** 975-83.

GAYATRI, S. & BEDFORD, M. T. 2014. Readers of histone methylarginine marks. *Biochim Biophys Acta,* 1839**,** 702-10.

GHARI, F., QUIRKE, A. M., MUNRO, S., KAWALKOWSKA, J., PICAUD, S., MCGOURAN, J., SUBRAMANIAN, V., MUTH, A., WILLIAMS, R., KESSLER, B., THOMPSON, P. R., FILLIPAKOPOULOS, P., KNAPP, S., VENABLES, P. J. & LA THANGUE, N. B. 2016. Citrullination-acetylation interplay guides E2F-1 activity during the inflammatory response. *Sci Adv,* 2**,** e1501257.

GLORIAN, V., ALLEGRE, J., BERTHELET, J., DUMETIER, B., BOUTANQUOI, P. M., DROIN, N., KAYACI, C., CARTIER, J., GEMBLE, S., MARCION, G., GONZALEZ, D., BOIDOT, R., GARRIDO, C., MICHAUD, O., SOLARY, E. & DUBREZ, L. 2017. DNA damage and S phase-dependent E2F1 stabilization requires the cIAP1 E3-ubiquitin ligase and is associated with K63-poly-ubiquitination on lysine 161/164 residues. *Cell Death Dis,* 8**,** e2816.

GOGADA, R., YADAV, N., LIU, J., TANG, S., ZHANG, D., SCHNEIDER, A., SESHADRI, A., SUN, L., ALDAZ, C. M., TANG, D. G. & CHANDRA, D. 2013. Bim, a proapoptotic protein, up-

regulated via transcription factor E2F1-dependent mechanism, functions as a prosurvival molecule in cancer. *Journal of Biological Chemistry,* 288**,** 368-81.

GONZÁLEZ-PRIETO, R., CUIJPERS, S. A., KUMAR, R., HENDRIKS, I. A. & VERTEGAAL, A. C. 2015. c-Myc is targeted to the proteasome for degradation in a SUMOylation-dependent manner, regulated by PIAS1, SENP7 and RNF4. *Cell Cycle,* 14**,** 1859-1872.

GOREN, A., RAM, O., AMIT, M., KEREN, H., LEV-MAOR, G., VIG, I., PUPKO, T. & AST, G. 2006. Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers. *Mol Cell,* 22**,** 769-81.

GOULET, I., GAUVIN, G., BOISVENUE, S. & CÔTÉ, J. 2007. Alternative splicing yields protein arginine methyltransferase 1 isoforms with distinct activity, substrate specificity, and subcellular localization. *Journal of Biological Chemistry,* 282**,** 33009-33021.

GRAVELEY, B. R., HERTEL, K. J. & MANIATIS, T. 2001. The role of U2AF35 and U2AF65 in enhancer-dependent splicing. *Rna,* 7**,** 806-18.

GRIMM, C., CHARI, A., PELZ, J. P., KUPER, J., KISKER, C., DIEDERICHS, K., STARK, H., SCHINDELIN, H. & FISCHER, U. 2013. Structural basis of assembly chaperone- mediated snRNP formation. *Mol Cell,* 49**,** 692-703.

GROLMUSZ, V. K., KARASZI, K., MICSIK, T., TOTH, E. A., MESZAROS, K., KARVALY, G., BARNA, G., SZABO, P. M., BAGHY, K., MATKO, J., KOVALSZKY, I., TOTH, M., RACZ, K., IGAZ, P. & PATOCS, A. 2016. Cell cycle dependent RRM2 may serve as proliferation marker and pharmaceutical target in adrenocortical cancer. *Am J Cancer Res,* 6**,** 2041-2053.

GU, Z., GAO, S., ZHANG, F., WANG, Z., MA, W. & DAVIS, R. E. 2012a. Protein arginine methyltransferase 5 is essential for growth of lung cancer cells. *Biochemical Journal,* 446**,** 235-41.

GU, Z., LI, Y., LEE, P., LIU, T., WAN, C. & WANG, Z. 2012b. Protein arginine methyltransferase 5 functions in opposite ways in the cytoplasm and nucleus of prostate cancer cells. *PLoS ONE,* 7**,** e44033.

HADJIKYRIACOU, A., YANG, Y., ESPEJO, A., BEDFORD, M. T. & CLARKE, S. G. 2015. Unique Features of Human Protein Arginine Methyltransferase 9 (PRMT9) and Its Substrate RNA Splicing Factor SF3B2. *Journal of Biological Chemistry,* 290**,** 16723-16743.

HALLSTROM, T. C., MORI, S. & NEVINS, J. R. 2008. An E2F1-dependent gene expression program that determines the balance between proliferation and cell death. *Cancer Cell,* 13**,** 11-22.

HANAHAN, D. & WEINBERG, R. A. 2000. The Hallmarks of Cancer. *Cell,* 100**,** 57-70.

HANAHAN, D. & WEINBERG, ROBERT A. 2011. Hallmarks of Cancer: The Next Generation. *Cell,* 144**,** 646-674.

HARBOUR, J. W., LUO, R. X., SANTI, A. D., POSTIGO, A. A. & DEAN, D. C. 1999. Cdk Phosphorylation Triggers Sequential Intramolecular Interactions that Progressively Block Rb Functions as Cells Move through G1. *Cell,* 98**,** 859-869.

HARPER, S. J. & BATES, D. O. 2008. VEGF-A splicing: the key to anti-angiogenic therapeutics? *Nat Rev Cancer,* 8**,** 880-887.

HERNANDEZ, S. J., DOLIVO, D. M. & DOMINKO, T. 2017. PRMT8 demonstrates variant-specific expression in cancer cells and correlates with patient survival in breast, ovarian and gastric cancer. *Oncology Letters,* 13**,** 1983-1989.

HERRMANN, F., LEE, J., BEDFORD, M. T. & FACKELMAYER, F. O. 2005. Dynamics of human protein arginine methyltransferase 1(PRMT1) in vivo. *Journal of Biological Chemistry,* 280**,** 38005-10.

HERSHKO, T. & GINSBERG, D. 2004. Up-regulation of Bcl-2 Homology 3 (BH3)-only Proteins by E2F1 Mediates Apoptosis. *Journal of Biological Chemistry,* 279**,** 8627-8634.

HIMEJI, D., HORIUCHI, T., TSUKAMOTO, H., HAYASHI, K., WATANABE, T. & HARADA, M. 2002. Characterization of caspase-8L: a novel isoform of caspase-8 that behaves as an inhibitor of the caspase cascade. *Blood,* 99**,** 4070-8.

HNILICOVÁ, J., HOZEIFI, S., DUŠKOVÁ, E., ICHA, J., TOMÁNKOVÁ, T. & STANĚK, D. 2011. Histone deacetylase activity modulates alternative splicing. *PLoS One,* 6**,** e16727.

HSIEH, F. K., KULAEVA, O. I., PATEL, S. S., DYER, P. N., LUGER, K., REINBERG, D. & STUDITSKY, V. M. 2013. Histone chaperone FACT action during transcription through chromatin by RNA polymerase II. *Proc Natl Acad Sci U S A,* 110**,** 7654-9.

HSU, T. Y., SIMON, L. M., NEILL, N. J., MARCOTTE, R., SAYAD, A., BLAND, C. S., ECHEVERRIA, G. V., SUN, T., KURLEY, S. J., TYAGI, S., KARLIN, K. L., DOMINGUEZ-VIDANA, R., HARTMAN, J. D., RENWICK, A., SCORSONE, K., BERNARDI, R. J., SKINNER, S. O., JAIN, A., ORELLANA, M., LAGISETTI, C., GOLDING, I., JUNG, S. Y., NEILSON, J. R., ZHANG, X. H., COOPER, T. A., WEBB, T. R., NEEL, B. G., SHAW, C. A. & WESTBROOK, T. F. 2015. The spliceosome is a therapeutic vulnerability in MYC-driven cancer. *Nature,* 525**,** 384-8.

HUANG, D. W., SHERMAN, B. T. & LEMPICKI, R. A. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res,* 37**,** 1-13.

HUANG, D. W., SHERMAN, B. T. & LEMPICKI, R. A. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc,* 4**,** 44-57.

HUANG, H., YU, S., LIU, H. & SUN, X. 2012. Nucleosome organization in sequences of alternative events in human genome. *Biosystems,* 109**,** 214-219.

HUANG, S., LITT, M. & FELSENFELD, G. 2005. Methylation of histone H4 by arginine methyltransferase PRMT1 is essential in vivo for many subsequent histone modifications. *Genes Dev,* 19**,** 1885-93.

IBRAHIM, R., MATSUBARA, D., OSMAN, W., MORIKAWA, T., GOTO, A., MORITA, S., ISHIKAWA, S., ABURATANI, H., TAKAI, D., NAKAJIMA, J., FUKAYAMA, M., NIKI, T. & MURAKAMI, Y. 2014. Expression of PRMT5 in lung adenocarcinoma and its significance in epithelial-mesenchymal transition. *Hum Pathol,* 45**,** 1397-405.

INGRAM, L., MUNRO, S., COUTTS, A. S. & LA THANGUE, N. B. 2011. E2F-1 regulation by an unusual DNA damage-responsive DP partner subunit. *Cell Death & Differentiation,* 18**,** 122-132.

IRWIN, M., MARIN, M. C., PHILLIPS, A. C., SEELAN, R. S., SMITH, D. I., LIU, W., FLORES, E. R., TSAI, K. Y., JACKS, T., VOUSDEN, K. H. & KAELIN, W. G. 2000. Role for the p53 homologue p73 in E2F-1-induced apoptosis. *Nature,* 407**,** 645-8.

IVANOVA, I. A., D'SOUZA, S. J. A. & DAGNINO, L. 2005. E2F1 stability is regulated by a novel-PKC//p38[beta] MAP kinase signaling pathway during keratinocyte differentiation. *Oncogene,* 25**,** 430-437.

IZQUIERDO, J. M., MAJOS, N., BONNAL, S., MARTINEZ, C., CASTELO, R., GUIGO, R., BILBAO, D. & VALCARCEL, J. 2005. Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. *Mol Cell,* 19**,** 475-84.

JAHAN, S. & DAVIE, J. R. 2015. Protein arginine methyltransferases (PRMTs): Role in chromatin organization. *Advances in Biological Regulation,* 57**,** 173-184.

JANSSON, M., DURANT, S. T., CHO, E.-C., SHEAHAN, S., EDELMANN, M., KESSLER, B. & LA THANGUE, N. B. 2008. Arginine methylation regulates the p53 response. *Nature Cell Biology,* 10**,** 1431-1439.

JIN, Y., ZHOU, J., XU, F., JIN, B., CUI, L., WANG, Y., DU, X., LI, J., LI, P., REN, R. & PAN, J. 2016. Targeting methyltransferase PRMT5 eliminates leukemia stem cells in chronic myelogenous leukemia. *The Journal of Clinical Investigation,* 126**,** 3961-3980.

JONKERS, I. & LIS, J. T. 2015. Getting up to speed with transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology,* 16**,** 167-177.

KANG, D. W., LEE, S. W., HWANG, W. C., LEE, B. H., CHOI, Y. S., SUH, Y. A., CHOI, K. Y. & MIN, D. S. 2017. Phospholipase D1 Acts through Akt/TopBP1 and RB1 to Regulate the E2F1-Dependent Apoptotic Program in Cancer Cells. *Cancer Res,* 77**,** 142-152.

KARAMI, S., LIN, F. M., KUMAR, S., BAHNASSY, S., THANGAVEL, H., QUTTINA, M., LI, Y., REN, J. & BAWA-KHALFE, T. 2017. Novel SUMO-Protease SENP7S Regulates β-catenin Signaling and Mammary Epithelial Cell Transformation. *Sci Rep,* 7.

KARNI, R., DE STANCHINA, E., LOWE, S. W., SINHA, R., MU, D. & KRAINER, A. R. 2007. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol,* 14**,** 185-93.

KARNI, R., HIPPO, Y., LOWE, S. W. & KRAINER, A. R. 2008. The splicing-factor oncoprotein SF2/ASF activates mTORC1. *Proc Natl Acad Sci U S A,* 105**,** 15323-7.

KE, S., SHANG, S., KALACHIKOV, S. M., MOROZOVA, I., YU, L., RUSSO, J. J., JU, J. & CHASIN, L. A. 2011. Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Res,* 21**,** 1360-74.

KIM, K. Y., WANG, D.-H., CAMPBELL, M., HUERTA, S. B., SHEVCHENKO, B., IZUMIYA, C. & IZUMIYA, Y. 2015. PRMT4-Mediated Arginine Methylation Negatively Regulates Retinoblastoma Tumor Suppressor Protein and Promotes E2F-1 Dissociation. *Molecular and Cellular Biology,* 35**,** 238-248.

KOH, C. M., BEZZI, M., LOW, D. H., ANG, W. X., TEO, S. X., GAY, F. P., AL-HADDAWI, M., TAN, S. Y., OSATO, M., SABO, A., AMATI, B., WEE, K. B. & GUCCIONE, E. 2015. MYC regulates the core pre-mRNA splicing machinery as an essential step in lymphomagenesis. *Nature,* 523**,** 96-100.

KONTAKI, H. & TALIANIDIS, I. 2010. Lysine methylation regulates E2F1-induced cell death. *Molecular Cell,* 39**,** 152-60.

KOZLOVSKI, I., SIEGFRIED, Z., AMAR-SCHWARTZ, A. & KARNI, R. 2017. The role of RNA alternative splicing in regulating cancer metabolism. *Hum Genet*.

KRYUKOV, G. V., WILSON, F. H., RUTH, J. R., PAULK, J., TSHERNIAK, A., MARLOW, S. E., VAZQUEZ, F., WEIR, B. A., FITZGERALD, M. E., TANAKA, M., BIELSKI, C. M., SCOTT, J. M., DENNIS, C., COWLEY, G. S., BOEHM, J. S., ROOT, D. E., GOLUB, T. R., CLISH, C. B., BRADNER, J. E., HAHN, W. C. & GARRAWAY, L. A. 2016. MTAP deletion confers enhanced dependency on the PRMT5 arginine methyltransferase in cancer cells. *Science*.

KUMARI, A., IWASAKI, T., PYNDIAH, S., CASSIMERE, E. K., PALANI, C. D. & SAKAMURO, D. 2015. Regulation of E2F1-induced apoptosis by poly(ADP-ribosyl)ation. *Cell Death and Differentiation,* 22**,** 311-322.

KURTOVIC-KOZARIC, A., PRZYCHODZEN, B., SINGH, J., KONARSKA, M. M., CLEMENTE, M. J., OTROCK, Z. K., NAKASHIMA, M., HSI, E. D., YOSHIDA, K., SHIRAISHI, Y., CHIBA, K., TANAKA, H., MIYANO, S., OGAWA, S., BOULTWOOD, J., MAKISHIMA, H., MACIEJEWSKI, J. P. & PADGETT, R. A. 2015. PRPF8 defects cause missplicing in myeloid malignancies. *Leukemia,* 29**,** 126-36.

LA SALA, D., MACALUSO, M., TRIMARCHI, C., GIORDANO, A. & CINTI, C. 2003. Triggering of p73-dependent apoptosis in osteosarcoma is under the control of E2Fs-pRb2/p130 complexes. *Oncogene,* 22**,** 3518-29.

LACADIE, S. A. & ROSBASH, M. 2005. Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast. *Mol Cell,* 19**,** 65-75.

LAI, Y., SONG, M., HAKALA, K., WEINTRAUB, S. T. & SHIIO, Y. 2011. Proteomic dissection of the VHL interactome. *Journal of proteome research,* 10**,** 5175-5182.

LAINE, A. & WESTERMARCK, J. 2014. Molecular Pathways: Harnessing E2F1 Regulation for Prosenescence Therapy in p53-Defective Cancer Cells. *Clinical Cancer Research,* 20**,** 3644.

LATORRE, E. & HARRIES, L. W. 2017. Splicing regulatory factors, ageing and age-related disease. *Ageing Res Rev,* 36**,** 165-170.

LAZZERINI DENCHI, E. & HELIN, K. 2005. E2F1 is crucial for E2F-dependent apoptosis. *EMBO Reports,* 6**,** 661-8.

LEVERSON, J. D., KOSKINEN, P. J., ORRICO, F. C., RAINIO, E. M., JALKANEN, K. J., DASH, A. B., EISENMAN, R. N. & NESS, S. A. 1998. Pim-1 kinase and p100 cooperate to enhance c-Myb activity. *Mol Cell,* 2**,** 417-25.

LEZINA, L., AKSENOVA, V., IVANOVA, T., PURMESSUR, N., ANTONOV, A. V., TENTLER, D., FEDOROVA, O., GARABADGIU, A. V., TALIANIDIS, I., MELINO, G. & BARLEV, N. A. 2014. KMTase Set7/9 is a critical regulator of E2F1 activity upon genotoxic stress. *Cell Death and Differentiation,* 21**,** 1889-1899.

LI, J., LUTHRA, S., WANG, X. H., CHANDRAN, U. R. & SOBOL, R. W. 2012. Transcriptional profiling reveals elevated Sox2 in DNA polymerase ß null mouse embryonic fibroblasts. *Am J Cancer Res,* 2**,** 699-713.

LI, X. X., SHI, L., ZHOU, X. J., WU, J., XIA, T. S., ZHOU, W. B., SUN, X., ZHU, L., WEI, J. F. & DING, Q. 2017. The role of c-Myc-RBM38 loop in the growth suppression in breast cancer. *J Exp Clin Cancer Res,* 36**,** 49.

LI, Y. & BLENCOWE, B. J. 1999. Distinct factor requirements for exonic splicing enhancer function and binding of U2AF to the polypyrimidine tract. *J Biol Chem,* 274**,** 35074-9.

LI, Z., STANELLE, J., LEURS, C., HANENBERG, H. & PÜTZER, B. M. 2005. Selection of novel mediators of E2F1-induced apoptosis through retroviral expression of an antisense cDNA library. *Nucleic Acids Research,* 33**,** 2813-2821.

LIBAN, T. J., THWAITES, M. J., DICK, F. A. & RUBIN, S. M. 2016. Structural Conservation and E2F Binding Specificity within the Retinoblastoma Pocket Protein Family. *Journal of Molecular Biology,* 428**,** 3960-3971.

LIMA, C. D. & REVERTER, D. 2008. Structure of the human SENP7 catalytic domain and poly-SUMO deconjugation activities for SENP6 and SENP7. *J Biol Chem,* 283**,** 32045-55.

LIN, J. C., TSAO, M. F. & LIN, Y. J. 2016. Differential Impacts of Alternative Splicing Networks on Apoptosis. *Int J Mol Sci,* 17.

LIN, Z., XU, Y. N., NAMGOONG, S. & KIM, N. H. 2014. JMY functions as actin nucleation-promoting factor and mediator for p53-mediated DNA damage in porcine oocytes. *PLoS One,* 9**,** e109385.

LINDSTROM, M. S. & WIMAN, K. G. 2002. Myc and E2F1 induce p53 through p14ARF-independent mechanisms in human fibroblasts. *Oncogene,* 22**,** 4993-5005.

LIU, H., TANG, X., SRIVASTAVA, A., PECOT, T., DANIEL, P., HEMMELGARN, B., REYES, S., FACKLER, N., BAJWA, A., KLADNEY, R., KOIVISTO, C., CHEN, Z., WANG, Q., HUANG, K., MACHIRAJU, R., SAENZ-ROBLES, M. T., CANTALUPO, P., PIPAS, J. M. & LEONE, G. 2015a. Redeployment of Myc and E2f1-3 drives Rb-deficient cell cycles. *Nat Cell Biol,* 17**,** 1036-48.

LIU, K., LIN, F.-T., RUPPERT, J. M. & LIN, W.-C. 2003. Regulation of E2F1 by BRCT domain-containing protein TopBP1. *Molecular and Cellular Biology,* 23**,** 3287-304.

LIU, K., LUO, Y., LIN, F.-T. & LIN, W.-C. 2004. TopBP1 recruits Brg1/Brm to repress E2F1-induced apoptosis, a novel pRb-independent and E2F1-specific control for cell survival. *Genes & Development,* 18**,** 673-86.

LIU, S., MOZAFFARI-JOVIN, S., WOLLENHAUPT, J., SANTOS, K. F., THEUSER, M., DUNIN-HORKAWICZ, S., FABRIZIO, P., BUJNICKI, J. M., LÜHRMANN, R. & WAHL, M. C. 2015b. A composite double-/single-stranded RNA-binding region in protein Prp3 supports tri-snRNP stability and splicing. *Elife,* 4**,** e07320.

LIU, Y., BEYER, A. & AEBERSOLD, R. 2016. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell,* 165**,** 535-50.

LOFTUS, S. J., LIU, G., CARR, S. M., MUNRO, S. & LA THANGUE, N. B. 2012. NEDDylation regulates E2F-1-dependent transcription. *EMBO Reports,* 13**,** 811-8.

LOGAN, N., GRAHAM, A., ZHAO, X., FISHER, R., MAITI, B., LEONE, G. & LA THANGUE, N. B. 2005. E2F-8: an E2F family member with a similar organization of DNA-binding domains to E2F-7. *Oncogene,* 24**,** 5000-4.

LOPEZ DE SILANES, I., GALBAN, S., MARTINDALE, J. L., YANG, X., MAZAN-MAMCZARZ, K., INDIG, F. E., FALCO, G., ZHAN, M. & GOROSPE, M. 2005. Identification and functional outcome of mRNAs associated with RNA-binding protein TIA-1. *Mol Cell Biol,* 25**,** 9520-31.

LU, Y.-C. 2014. Arginine methylation on E2F1. Thesis (D.Phil.)--University of Oxford, 2015, (Medical Sciences Division - Clinical Pharmacology ; Linacre College.).

LU, Z., BAUZON, F., FU, H., CUI, J., ZHAO, H., NAKAYAMA, K., NAKAYAMA, K. I. & ZHU, L. 2014. Skp2 suppresses apoptosis in Rb1 deficient tumors by limiting E2F1 activity. *Nature communications,* 5**,** 3463-3463.

LUCO, R. F., PAN, Q., TOMINAGA, K., BLENCOWE, B. J., PEREIRA-SMITH, O. M. & MISTELI, T. 2010. Regulation of alternative splicing by histone modifications. *Science,* 327**,** 996-1000.

MAHANIC, C. S., BUDHAVARAPU, V., GRAVES, J. D., LI, G. & LIN, W.-C. 2015. Regulation of E2 Promoter Binding Factor 1 (E2F1) Transcriptional Activity through a Deubiquitinating Enzyme, UCH37. *The Journal of Biological Chemistry,* 290**,** 26508-26522.

MAHESWARAN, E., PEDERSEN, C. B., DITZEL, H. J. & GJERSTORFF, M. F. 2015. Lack of ADAM2, CALR3 and SAGE1 Cancer/Testis Antigen Expression in Lung and Breast Cancer. *PLoS One,* 10**,** e0134967.

MAKAROVA, O. V., MAKAROV, E. M., LIU, S., VORNLOCHER, H. P. & LÜHRMANN, R. 2002. Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6*U5 tri-snRNP formation and pre-mRNA splicing. *EMBO J,* 21**,** 1148-57.

MARANI, M., TENEV, T., HANCOCK, D., DOWNWARD, J. & LEMOINE, N. R. 2002. Identification of novel isoforms of the BH3 domain protein Bim which directly activate Bax to trigger apoptosis. *Mol Cell Biol,* 22**,** 3577-89.

MARJON, K., CAMERON, MICHAEL J., QUANG, P., CLASQUIN, MICHELLE F., MANDLEY, E., KUNII, K., MCVAY, M., CHOE, S., KERNYTSKY, A., GROSS, S., KONTEATIS, Z., MURTIE, J., BLAKE, MICHELLE L., TRAVINS, J., DORSCH, M., BILLER, SCOTT A. & MARKS, KEVIN M. 2016. MTAP Deletions in Cancer Create Vulnerability to Targeting of the MAT2A/PRMT5/RIOK1 Axis. *Cell Reports,* 15**,** 574-587.

MARTINEZ-BALBAS, M. A., BAUER, U. M., NIELSEN, S. J., BREHM, A. & KOUZARIDES, T. 2000. Regulation of E2F1 activity by acetylation. *Embo j,* 19**,** 662-71.

MATHIOUDAKI, K., PAPADOKOSTOPOULOU, A., SCORILAS, A., XYNOPOULOS, D., AGNANTI, N. & TALIERI, M. 2008. The PRMT1 gene expression pattern in colon cancer. *British Journal of Cancer,* 99**,** 2094-2099.

MATHIOUDAKI, K., SCORILAS, A., ARDAVANIS, A., LYMBERI, P., TSIAMBAS, E., DEVETZI, M., APOSTOLAKI, A. & TALIERI, M. 2011. Clinical evaluation of PRMT1 gene expression in breast cancer. *Tumour biology,* 32**,** 575-82.

MATTER, N., HERRLICH, P. & KONIG, H. 2002. Signal-dependent regulation of splicing via phosphorylation of Sam68. *Nature,* 420**,** 691-5.

MATYS, V., KEL-MARGOULIS, O. V., FRICKE, E., LIEBICH, I., LAND, S., BARRE-DIRRIE, A., REUTER, I., CHEKMENEV, D., KRULL, M., HORNISCHER, K., VOSS, N., STEGMAIER, P., LEWICKI-POTAPOV, B., SAXEL, H., KEL, A. E. & WINGENDER, E. 2006. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res,* 34**,** D108-10.

MEISTER, G., EGGERT, C., BÜHLER, D., BRAHMS, H., KAMBACH, C. & FISCHER, U. 2001. Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. *Current Biology,* 11**,** 1990-1994.

MERCATANTE, D. R., MOHLER, J. L. & KOLE, R. 2002. Cellular response to an antisense-mediated shift of Bcl-x pre-mRNA splicing and antineoplastic agents. *J Biol Chem,* 277**,** 49374-82.

MERDZHANOVA, G., EDMOND, V., DE SERANNO, S., VAN DEN BROECK, A., CORCOS, L., BRAMBILLA, C., BRAMBILLA, E., GAZZERI, S. & EYMIN, B. 2008. E2F1 controls alternative splicing pattern of genes involved in apoptosis through upregulation of the splicing factor SC35. *Cell Death Differ,* 15**,** 1815-23.

MILTON, A., LUOTO, K., INGRAM, L., MUNRO, S., LOGAN, N., GRAHAM, A. L., BRUMMELKAMP, T. R., HIJMANS, E. M., BERNARDS, R. & LA THANGUE, N. B. 2006. A functionally distinct member of the DP family of E2F subunits. *Oncogene,* 25**,** 3212-8.

MISTELI, T. & SPECTOR, D. L. 1999. RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Mol Cell,* 3**,** 697-705.

MORALES, Y., CÁCERES, T., MAY, K. & HEVEL, J. M. 2016. Biochemistry and regulation of the protein arginine methyltransferases (PRMTs). *Archives of Biochemistry and Biophysics,* 590**,** 138-152.

MORETTIN, A., BALDWIN, R. M. & CÔTÉ, J. 2015. Arginine methyltransferases as novel therapeutic targets for breast cancer. *Mutagenesis,* 30**,** 177-189.

MORONI, M. C., HICKMAN, E. S., LAZZERINI DENCHI, E., CAPRARA, G., COLLI, E., CECCONI, F., MÜLLER, H. & HELIN, K. 2001. Apaf-1 is a transcriptional target for E2F and p53. *Nature Cell Biology,* 3**,** 552-8.

MOUNIR, Z., KORN, J. M., WESTERLING, T., LIN, F., KIRBY, C. A., SCHIRLE, M., MCALLISTER, G., HOFFMAN, G., RAMADAN, N., HARTUNG, A., FENG, Y., KIPP, D. R., QUINN, C., FODOR, M., BAIRD, J., SCHOUMACHER, M., MEYER, R., DEEDS, J., BUCHWALTER, G., STAMS, T., KEEN, N., SELLERS, W. R., BROWN, M. & PAGLIARINI, R. A. 2016. ERG signaling in prostate cancer is driven through PRMT5-dependent methylation of the Androgen Receptor. *eLife,* 5**,** e13964.

MÜHLETHALER-MOTTET, A., FLAHAUT, M., BOURLOUD, K. B., NARDOU, K., COULON, A., LIBERMAN, J., THOME, M. & GROSS, N. 2011. Individual caspase-10 isoforms play distinct and opposing roles in the initiation of death receptor-mediated tumour cell apoptosis. *Cell Death & Disease,* 2**,** e125.

MÜLLER, H., BRACKEN, A. P., VERNELL, R., MORONI, M. C., CHRISTIANS, F., GRASSILLI, E., PROSPERINI, E., VIGO, E., OLINER, J. D. & HELIN, K. 2001. E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. *Genes & Development,* 15**,** 267-85.

NAGL, N. G., WANG, X., PATSIALOU, A., VAN SCOY, M. & MORAN, E. 2007. Distinct mammalian SWI/SNF chromatin remodeling complexes with opposing roles in cell-cycle control. *EMBO Journal,* 26**,** 752-63.

NAGY, Z. & TORA, L. 2007. Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. *Oncogene,* 26**,** 5341-57.

NAKAKIDO, M., DENG, Z., SUZUKI, T., DOHMAE, N., NAKAMURA, Y. & HAMAMOTO, R. 2015. PRMT6 increases cytoplasmic localization of p21CDKN1A in cancer cells through arginine methylation and makes more resistant to cytotoxic agents. *Oncotarget,* 6**,** 30957-67.

NEAULT, M., MALLETTE, F. A., VOGEL, G., MICHAUD-LEVESQUE, J. & RICHARD, S. 2012. Ablation of PRMT6 reveals a role as a negative transcriptional regulator of the p53 tumor suppressor. *Nucleic Acids Res,* 40**,** 9513-21.

NGUYEN, T. H. D., GALEJ, W. P., BAI, X.-C., SAVVA, C. G., NEWMAN, A. J., SCHERES, S. H. W. & NAGAI, K. 2015. The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature,* 523**,** 47-52.

NICHOLS, R. C., WANG, X. W., TANG, J., HAMILTON, B. J., HIGH, F. A., HERSCHMAN, H. R. & RIGBY, W. F. C. 2000. The RGG Domain in hnRNP A2 Affects Subcellular Localization. *Experimental Cell Research,* 256**,** 522-532.

NIKOLAI, B. C., LANZ, R. B., YORK, B., DASGUPTA, S., MITSIADES, N., CREIGHTON, C. J., TSIMELZON, A., HILSENBECK, S. G., LONARD, D. M., SMITH, C. L. & O'MALLEY, B. W. 2016. HER2 signaling drives DNA anabolism and proliferation through SRC-3 phosphorylation and E2F1-regulated genes. *Cancer research,* 76**,** 1463-1475.

NIRANJANAKUMARI, S., LASDA, E., BRAZAS, R. & GARCIA-BLANCO, M. A. 2002. Reversible cross-linking combined with immunoprecipitation to study RNA-protein interactions in vivo. *Methods,* 26**,** 182-90.

OGARA, M. F., SIRKIN, P. F., CARCAGNO, A. L., MARAZITA, M. C., SONZOGNI, S. V., CERUTI, J. M. & CÁNEPA, E. T. 2013. Chromatin relaxation-mediated induction of p19INK4d increases the ability of cells to repair damaged DNA. *PLoS ONE,* 8**,** e61143-e61143.

OH, T. G., BAILEY, P., DRAY, E., SMITH, A. G., GOODE, J., ERIKSSON, N., FUNDER, J. W., FULLER, P. J., SIMPSON, E. R., TILLEY, W. D., LEEDMAN, P. J., CLARKE, C. L., GRIMMOND, S., DOWHAN, D. H. & MUSCAT, G. E. 2014. PRMT2 and RORgamma expression are associated with breast cancer survival outcomes. *Mol Endocrinol,* 28**,** 1166-85.

OU, C. Y., LABONTE, M. J., MANEGOLD, P. C., SO, A. Y., IANCULESCU, I., GERKE, D. S., YAMAMOTO, K. R., LADNER, R. D., KAHN, M., KIM, J. H. & STALLCUP, M. R. 2011. A coactivator role of CARM1 in the dysregulation of β-catenin activity in colorectal cancer cell growth and gene expression. *Mol Cancer Res,* 9**,** 660-70.

PAL, S., VISHWANATH, S. N., ERDJUMENT-BROMAGE, H., TEMPST, P. & SIF, S. 2004. Human SWI/SNF-associated PRMT5 methylates histone H3 arginine 8 and negatively regulates expression of ST7 and NM23 tumor suppressor genes. *Molecular and Cellular Biology,* 24**,** 9630-45.

PARK, J. H., SZEMES, M., VIEIRA, G. C., MELEGH, Z., MALIK, S., HEESOM, K. J., VON WALLWITZ-FREITAS, L., GREENHOUGH, A., BROWN, K. W., ZHENG, Y. G., CATCHPOOLE, D., DEERY, M. J. & MALIK, K. 2015. Protein arginine methyltransferase 5 is a key regulator of the MYCN oncoprotein in neuroblastoma cells. *Mol Oncol,* 9**,** 617-27.

PARONETTO, M. P., ACHSEL, T., MASSIELLO, A., CHALFANT, C. E. & SETTE, C. 2007. The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x. *The Journal of Cell Biology,* 176**,** 929-939.

PASSOS, D. O., QUARESMA, A. J. C. & KOBARG, J. 2006. The methylation of the C-terminal region of hnRNPQ (NSAP1) is important for its nuclear localization. *Biochemical and Biophysical Research Communications,* 346**,** 517-525.

PHATNANI, H. P., JONES, J. C. & GREENLEAF, A. L. 2004. Expanding the functional repertoire of CTD kinase I and RNA polymerase II: novel phosphoCTD-associating proteins in the yeast proteome. *Biochemistry,* 43**,** 15702-19.

PHILLIPS, A. C., ERNST, M. K., BATES, S., RICE, N. R. & VOUSDEN, K. H. 1999. E2F-1 Potentiates Cell Death by Blocking Antiapoptotic Signaling Pathways. *Molecular Cell,* 4**,** 771-781.

PIERCE, A. M., GIMENEZ-CONTI, I. B., SCHNEIDER-BROUSSARD, R., MARTINEZ, L. A., CONTI, C. J. & JOHNSON, D. G. 1998. Increased E2F1 activity induces skin tumors in mice heterozygous and nullizygous for p53. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 8858-63.

PIERCE, A. M., SCHNEIDER-BROUSSARD, R., GIMENEZ-CONTI, I. B., RUSSELL, J. L., CONTI, C. J. & JOHNSON, D. G. 1999. E2F1 has both oncogenic and tumor-suppressive properties in a transgenic model. *Molecular and Cellular Biology,* 19**,** 6408-14.

POLAGER, S., OFIR, M. & GINSBERG, D. 2008. E2F1 regulates autophagy and the transcription of autophagy genes. *Oncogene,* 27**,** 4860-4864.

PRABHU, L., WEI, H., CHEN, L., DEMIR, O., SANDUSKY, G., SUN, E., WANG, J., MO, J., ZENG, L., FISHEL, M., SAFA, A., AMARO, R., KORC, M., ZHANG, Z. Y. & LU, T. 2017. Adapting AlphaLISA high throughput screen to discover a novel small-molecule inhibitor targeting protein arginine methyltransferase 5 in pancreatic and colorectal cancers. *Oncotarget*.

PRADEEPA, M. M., SUTHERLAND, H. G., ULE, J., GRIMES, G. R. & BICKMORE, W. A. 2012. Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. *PLoS Genet,* 8**,** e1002717.

PRASAD, S. B., YADAV, S. S., DAS, M., MODI, A., KUMARI, S., PANDEY, L. K., SINGH, S., PRADHAN, S. & NARAYAN, G. 2015. PI3K/AKT pathway-mediated regulation of p27(Kip1) is associated with cell cycle arrest and apoptosis in cervical cancer. *Cell Oncol (Dordr),* 38**,** 215-25.

PUNNOOSE, E. A., LEVERSON, J. D., PEALE, F., BOGHAERT, E. R., BELMONT, L. D., TAN, N., YOUNG, A., MITTEN, M., INGALLA, E., DARBONNE, W. C., OLEKSIJEW, A., TAPANG, P., YUE, P., OEH, J., LEE, L., MAIGA, S., FAIRBROTHER, W. J., AMIOT, M., SOUERS, A. J. & SAMPATH, D. 2016. Expression Profile of BCL-2, BCL-XL, and MCL-1 Predicts Pharmacological Response to the BCL-2 Selective Antagonist Venetoclax in Multiple Myeloma Models. *Mol Cancer Ther,* 15**,** 1132-44.

PÜTZER, B. M., TUVE, S., TANNAPFEL, A. & STIEWE, T. 2003. Increased DeltaN-p73 expression in tumors by upregulation of the E2F1-regulated, TA-promoter-derived DeltaN'-p73 transcript. *Cell Death Differ,* 10**,** 612-4.

QIN, X. Q., LIVINGSTON, D. M., KAELIN, W. G. & ADAMS, P. D. 1994. Deregulated transcription factor E2F-1 expression leads to S-phase entry and p53-mediated apoptosis. *Proceedings of the National Academy of Sciences,* 91**,** 10918-10922.

RAN, F. A., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat. Protocols,* 8**,** 2281-2308.

RAUCH, J., MORAN-JONES, K., ALBRECHT, V., SCHWARZL, T., HUNTER, K., GIRES, O. & KOLCH, W. 2011. c-Myc regulates RNA splicing of the A-Raf kinase and its activation of the ERK pathway. *Cancer Res,* 71**,** 4664-74.

REINHOLD, W. C., SUNSHINE, M., LIU, H., VARMA, S., KOHN, K. W., MORRIS, J., DOROSHOW, J. & POMMIER, Y. 2012. CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Res,* 72**,** 3499-511.

RENSHAW, S. A., DEMPSEY, C. E., BARNES, F. A., BAGSTAFF, S. M., DOWER, S. K., BINGLE, C. D. & WHYTE, M. K. 2004. Three novel Bid proteins generated by alternative splicing of the human Bid gene. *J Biol Chem,* 279**,** 2846-55.

ROMEO, K., LOUAULT, Y., CANTALOUBE, S., LOIODICE, I. A., GENEVIÈVE & QUIVY, J.-P. 2015. The SENP7 SUMO-Protease Presents a Module of Two HP1 Interaction Motifs that Locks HP1 Protein at Pericentric Heterochromatin: Cell Reports. *Cell Reports,* 10**,** 771-782.

ROSONINA, E., IP, J. Y., CALARCO, J. A., BAKOWSKI, M. A., EMILI, A., MCCRACKEN, S., TUCKER, P., INGLES, C. J. & BLENCOWE, B. J. 2005. Role for PSF in mediating transcriptional activator-dependent stimulation of pre-mRNA processing in vivo. *Mol Cell Biol,* 25**,** 6734-46.

ROWORTH, A. P., GHARI, F. & LA THANGUE, N. B. 2015. To live or let die – complexity within the E2F1 pathway. *Molecular & Cellular Oncology,* 2**,** e970480.

RUTZ, B. & SERAPHIN, B. 1999. Transient interaction of BBP/ScSF1 and Mud2 with the splicing machinery affects the kinetics of spliceosome assembly. *Rna,* 5**,** 819-31.

SAKAMAKI, J., DAITOKU, H., UENO, K., HAGIWARA, A., YAMAGATA, K. & FUKAMIZU, A. 2011. Arginine methylation of BCL-2 antagonist of cell death (BAD) counteracts its phosphorylation and inactivation by Akt. *Proc Natl Acad Sci U S A,* 108**,** 6085-90.

SANCHEZ-JIMENEZ, C., LUDENA, M. D. & IZQUIERDO, J. M. 2015. T-cell intracellular antigens function as tumor suppressor genes. *Cell Death Dis,* 6**,** e1669.

SCHNEIDER, C., WILL, C. L., MAKAROVA, O. V., MAKAROV, E. M. & LUHRMANN, R. 2002. Human U4/U6.U5 and U4atac/U6atac.U5 tri-snRNPs exhibit similar protein compositions. *Mol Cell Biol,* 22**,** 3219-29.

SCHULZE, A., ZERFASS, K., SPITKOVSKY, D., MIDDENDORP, S., BERGES, J., HELIN, K., JANSEN-DURR, P. & HENGLEIN, B. 1995. Cell cycle regulation of the cyclin A gene promoter is mediated by a variant E2F site. *Proceedings of the National Academy of Sciences,* 92**,** 11264-11268.

SCHWARTZ, S., MESHORER, E. & AST, G. 2009. Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol,* 16**,** 990-5.

SCHWERK, C. & SCHULZE-OSTHOFF, K. 2005. Regulation of Apoptosis by Alternative Pre-mRNA Splicing. *Molecular Cell,* 19**,** 1-13.

SEELAN, R. S., IRWIN, M., VAN DER STOOP, P., QIAN, C., KAELIN, W. G. & LIU, W. 2002. The human p73 promoter: characterization and identification of functional E2F binding sites. *Neoplasia,* 4**,** 195-203.

SHATS, I., DENG, M., DAVIDOVICH, A., ZHANG, C., KWON, J. S., MANANDHAR, D., GORDAN, R., YAO, G. & YOU, L. 2017. Expression level is a key determinant of E2F1-mediated cell fate. *Cell Death Differ,* 24**,** 626-637.

SHAW, N., ZHAO, M., CHENG, C., XU, H., SAARIKETTU, J., LI, Y., DA, Y., YAO, Z., SILVENNOINEN, O., YANG, J., LIU, Z.-J., WANG, B.-C. & RAO, Z. 2007. The multifunctional human p100 protein 'hooks' methylated ligands. *Nature Structural & Molecular Biology,* 14**,** 779-784.

SHEN, L. N., GEOFFROY, M. C., JAFFRAY, E. G. & HAY, R. T. 2009. Characterization of SENP7, a SUMO-2/3-specific isopeptidase. *Biochem J,* 421**,** 223-30.

SHENG, X. & WANG, Z. 2016. Protein arginine methyltransferase 5 regulates multiple signaling pathways to promote lung cancer cell proliferation. *BMC Cancer,* 16**,** 567.

SHILO, K., WU, X., SHARMA, S., WELLIVER, M., DUAN, W., VILLALONA-CALERO, M., FUKUOKA, J., SIF, S., BAIOCCHI, R., HITCHCOCK, C. L., ZHAO, W. & OTTERSON, G. A. 2013. Cellular localization of protein arginine methyltransferase-5 correlates with grade of lung tumors. *Diagnostic pathology,* 8**,** 201.

SHU, L., YAN, W. & CHEN, X. 2006. RNPC1, an RNA-binding protein and a target of the p53 family, is required for maintaining the stability of the basal and stress-induced p21 transcript. *Genes Dev,* 20**,** 2961-72.

SINGH, R. K. & DAGNINO, L. 2017. CDH1 regulates E2F1 degradation in response to differentiation signals in keratinocytes. *Oncotarget,* 8**,** 4977-4993.

SKALSKA, L., BELTRAN-NEBOT, M., ULE, J. & JENNER, R. G. 2017. Regulatory feedback from nascent RNA to chromatin and transcription. *Nat Rev Mol Cell Biol,* 18**,** 331-337.

STEIN, C., RIEDL, S., RÜTHNICK, D., NÖTZOLD, R. R. & BAUER, U. M. 2012. The arginine methyltransferase PRMT6 regulates cell proliferation and senescence through transcriptional repression of tumor suppressor genes. *Nucleic Acids Res,* 40**,** 9522-33.

STEVENS, C., SMITH, L. & LA THANGUE, N. B. 2003. Chk2 activates E2F-1 in response to DNA damage. *Nature Cell Biology,* 5**,** 401-409.

STIEWE, T., THESELING, C. C. & PÜTZER, B. M. 2002. Transactivation-deficient Delta TA-p73 inhibits p53 by direct competition for DNA binding: implications for tumorigenesis. *J Biol Chem,* 277**,** 14177-85.

STREUBEL, G., BOUCHARD, C., BERBERICH, H., ZELLER, M. S., TEICHMANN, S., ADAMKIEWICZ, J., MÜLLER, R., KLEMPNAUER, K.-H. & BAUER, U.-M. 2013. PRMT4 Is a Novel Coactivator of c-Myb-Dependent Transcription in Haematopoietic Cell Lines. *PLOS Genetics,* 9**,** e1003343.

SU, C., ZHANG, C., TECLE, A., FU, X., HE, J., SONG, J., ZHANG, W., SUN, X., REN, Y., SILVENNOINEN, O., YAO, Z., YANG, X., WEI, M. & YANG, J. 2015. Tudor staphylococcal nuclease (Tudor-SN), a novel regulator facilitating G1/S phase transition, acting as a co-activator of E2F-1 in cell cycle regulation. *J Biol Chem,* 290**,** 7208-20.

SUCHANKOVA, J., LEGARTOVA, S., SEHNALOVA, P., KOZUBEK, S., VALENTE, S., LABELLA, D., MAI, A., ECKERICH, C., FACKELMAYER, F. O., SOROKIN, D. V. & BARTOVA, E. 2014.

PRMT1 arginine methyltransferase accumulates in cytoplasmic bodies that respond to selective inhibition and DNA damage. *Eur J Histochem,* 58**,** 2389.

TAO, Y., KASSATLY, R. F., CRESS, W. D. & HOROWITZ, J. M. 1997. Subunit composition determines E2F DNA-binding site specificity. *Mol Cell Biol,* 17**,** 6994-7007.

TAUBERT, S., GORRINI, C., FRANK, S. R., PARISI, T., FUCHS, M., CHAN, H. M., LIVINGSTON, D. M. & AMATI, B. 2004. E2F-dependent histone acetylation and recruitment of the Tip60 acetyltransferase complex to chromatin in late G1. *Molecular and Cellular Biology,* 24**,** 4546-56.

TILGNER, H., NIKOLAOU, C., ALTHAMMER, S., SAMMETH, M., BEATO, M., VALCARCEL, J. & GUIGO, R. 2009. Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol,* 16**,** 996-1001.

TOKI, H., INOUE, M., MINOWA, O., MOTEGI, H., SAIKI, Y., WAKANA, S., MASUYA, H., GONDO, Y., SHIROISHI, T., YAO, R. & NODA, T. 2014. Novel retinoblastoma mutation abrogating the interaction to E2F2/3, but not E2F1, led to selective suppression of thyroid tumors. *Cancer Sci,* 105**,** 1360-8.

TONG, X., DRAPKIN, R., YALAMANCHILI, R., MOSIALOS, G. & KIEFF, E. 1995. The Epstein-Barr virus nuclear protein 2 acidic domain forms a complex with a novel cellular coactivator that can interact with TFIIE. *Mol Cell Biol,* 15**,** 4735-44.

TOPHKHANE, C., YANG, S.-H., JIANG, Y., MA, Z., SUBRAMANIAM, D., ANANT, S., YOGOSAWA, S., SAKAI, T., LIU, W.-G., EDGERTON, S., THOR, A. & YANG, X. 2012. p53 inactivation upregulates p73 expression through E2F-1 mediated transcription. *PLoS ONE,* 7**,** e43564-e43564.

TU, M., LI, H., LV, N., XI, C., LU, Z., WEI, J., CHEN, J., GUO, F., JIANG, K., SONG, G., GAO, W. & MIAO, Y. 2017. Vasohibin 2 reduces chemosensitivity to gemcitabine in pancreatic cancer cells via Jun proto-oncogene dependent transactivation of ribonucleotide reductase regulatory subunit M2. *Molecular Cancer,* 16**,** 66.

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

TYAGI, A., RYME, J., BRODIN, D., OSTLUND FARRANTS, A. K. & VISA, N. 2009. SWI/SNF associates with nascent pre-mRNPs and regulates alternative pre-mRNA processing. *PLoS Genet,* 5**,** e1000470.

UHLEN, M., FAGERBERG, L., HALLSTROM, B. M., LINDSKOG, C., OKSVOLD, P., MARDINOGLU, A., SIVERTSSON, A., KAMPF, C., SJOSTEDT, E., ASPLUND, A., OLSSON, I., EDLUND, K., LUNDBERG, E., NAVANI, S., SZIGYARTO, C. A., ODEBERG, J., DJUREINOVIC, D., TAKANEN, J. O., HOBER, S., ALM, T., EDQVIST, P. H., BERLING, H., TEGEL, H., MULDER, J., ROCKBERG, J., NILSSON, P., SCHWENK, J. M., HAMSTEN, M., VON FEILITZEN, K., FORSBERG, M., PERSSON, L., JOHANSSON, F., ZWAHLEN, M., VON HEIJNE, G., NIELSEN, J. & PONTEN, F. 2015. Proteomics. Tissue-based map of the human proteome. *Science,* 347**,** 1260419.

ULE, J., JENSEN, K. B., RUGGIU, M., MELE, A., ULE, A. & DARNELL, R. B. 2003. CLIP identifies Nova-regulated RNA networks in the brain. *Science,* 302**,** 1212-5.

UNDERWOOD, J. G., BOUTZ, P. L., DOUGHERTY, J. D., STOILOV, P. & BLACK, D. L. 2005. Homologues of the Caenorhabditis elegans Fox-1 protein are neuronal splicing regulators in mammals. *Mol Cell Biol,* 25**,** 10005-16.

VALLE, B. L., D'SOUZA, T., BECKER, K. G., WOOD, W. H., 3RD, ZHANG, Y., WERSTO, R. P. & MORIN, P. J. 2013. Non-steroidal anti-inflammatory drugs decrease E2F1 expression and inhibit cell growth in ovarian cancer cells. *PLoS One,* 8**,** e61836.

VANG, R., SHIH IE, M. & KURMAN, R. J. 2009. OVARIAN LOW-GRADE AND HIGH-GRADE SEROUS CARCINOMA: Pathogenesis, Clinicopathologic and Molecular Biologic Features, and Diagnostic Problems. *Adv Anat Pathol,* 16**,** 267-82.

VERBEEREN, J., NIEMELA, E. H., TURUNEN, J. J., WILL, C. L., RAVANTTI, J. J., LUHRMANN, R. & FRILANDER, M. J. 2010. An ancient mechanism for splicing control: U11 snRNP as an activator of alternative splicing. *Mol Cell,* 37**,** 821-33.

VHUIYAN, M. I., PAK, M. L., PARK, M. A., THOMAS, D., LAKOWSKI, T. M., CHALFANT, C. E. & FRANKEL, A. 2017. PRMT2 interacts with splicing factors and regulates the alternative splicing of BCL-X. *J Biochem*.

VU, N. T., PARK, M. A., SHULTZ, J. C., GOEHE, R. W., HOEFERLIN, L. A., SHULTZ, M. D., SMITH, S. A., LYNCH, K. W. & CHALFANT, C. E. 2013. hnRNP U enhances caspase-9 splicing and is modulated by AKT-dependent phosphorylation of hnRNP L. *J Biol Chem,* 288**,** 8575-84.

WADA, K., INOUE, K. & HAGIWARA, M. 2002. Identification of methylated proteins by protein arginine N-methyltransferase 1, PRMT1, with a new expression cloning strategy. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research,* 1591**,** 1-10.

WALPORT, L. J., HOPKINSON, R. J., CHOWDHURY, R., SCHILLER, R., GE, W., KAWAMURA, A. & SCHOFIELD, C. J. 2016. Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. 7**,** 11974.

WANG, B., MA, A., ZHANG, L., JIN, W.-L., QIAN, Y., XU, G., QIU, B., YANG, Z., LIU, Y., XIA, Q. & LIU, Y. 2015a. POH1 deubiquitylates and stabilizes E2F1 to promote tumour formation. *Nature Communications,* 6**,** 8704.

WANG, E., DIMOVA, N. & CAMBI, F. 2007. PLP/DM20 ratio is regulated by hnRNPH and F and a novel G-rich enhancer in oligodendrocytes. *Nucleic Acids Res,* 35**,** 4164-78.

WANG, L., PAL, S. & SIF, S. 2008. Protein arginine methyltransferase 5 suppresses the transcription of the RB family of tumor suppressors in leukemia and lymphoma cells. *Molecular and Cellular Biology,* 28**,** 6262-77.

WANG, L., ZENG, H., WANG, Q., ZHAO, Z., BOYER, T. G., BIAN, X. & XU, W. 2015b. MED12 methylation by CARM1 sensitizes human breast cancer cells to chemotherapy drugs. *Science Advances,* 1.

WANG, X., LIU, X., FANG, J., LU, Y., HE, J., YAO, X., YAO, Z. & YANG, J. 2010. Coactivator P100 protein enhances STAT6-dependent transcriptional activation but has no effect on STAT1-mediated gene transcription. *Anat Rec (Hoboken),* 293**,** 1010-6.

WANG, Y., MA, M., XIAO, X. & WANG, Z. 2012. Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules. *Nat Struct Mol Biol,* 19**,** 1044-1052.

WILL, C. L. & LÜHRMANN, R. 2011. Spliceosome Structure and Function. *Cold Spring Harbor Perspectives in Biology,* 3**,** a003707.

WILL, C. L., SCHNEIDER, C., HOSSBACH, M., URLAUB, H., RAUHUT, R., ELBASHIR, S., TUSCHL, T. & LUHRMANN, R. 2004. The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome. *Rna,* 10**,** 929-41.

WILL, C. L., SCHNEIDER, C., REED, R. & LUHRMANN, R. 1999. Identification of both shared and distinct proteins in the major and minor spliceosomes. *Science,* 284**,** 2003-5.

WIREN, K. M., ZHANG, X., CHANG, C., KEENAN, E. & ORWOLL, E. S. 1997. Transcriptional up-regulation of the human androgen receptor by androgen in bone cells. *Endocrinology,* 138**,** 2291-300.

WITALISON, E. E., THOMPSON, P. R. & HOFSETH, L. J. 2015. Protein Arginine Deiminases and Associated Citrullination: Physiological Functions and Diseases Associated with Dysregulation. *Curr Drug Targets*.

WOODS, K., THOMSON, J. M. & HAMMOND, S. M. 2007. Direct regulation of an oncogenic micro-RNA cluster by E2F transcription factors. *Journal of Biological Chemistry,* 282**,** 2130-4.

WU, C. L., ZUKERBERG, L. R., NGWU, C., HARLOW, E. & LEES, J. A. 1995. In vivo association of E2F and DP family proteins. *Mol Cell Biol,* 15**,** 2536-46.

WU, J., ZHOU, X. J., SUN, X., XIA, T. S., LI, X. X., SHI, L., ZHU, L., ZHOU, W. B., WEI, J. F. & DING, Q. 2017. RBM38 is involved in TGF-beta-induced epithelial-to-mesenchymal transition by stabilising zonula occludens-1 mRNA in breast cancer. *Br J Cancer,* 117**,** 675-684.

WU, J. Y. & MANIATIS, T. 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell,* 75**,** 1061-70.

WU, L., TIMMERS, C., MAITI, B., SAAVEDRA, H. I., SANG, L., CHONG, G. T., NUCKOLLS, F., GIANGRANDE, P., WRIGHT, F. A., FIELD, S. J., GREENBERG, M. E., ORKIN, S., NEVINS, J. R., ROBINSON, M. L. & LEONE, G. 2001. The E2F1-3 transcription factors are essential for cellular proliferation. *Nature,* 414**,** 457-62.

WU, M., SETO, E. & ZHANG, J. 2015. E2F1 enhances glycolysis through suppressing Sirt6 transcription in cancer cells. *Oncotarget,* 6**,** 11252-11263.

XIE, Q., BAI, Y., WU, J., SUN, Y., WANG, Y., ZHANG, Y., MEI, P. & YUAN, Z. 2011. Methylation-mediated regulation of E2F1 in DNA damage-induced cell death. *J Recept Signal Transduct Res,* 31**,** 139-46.

XIE, W., JIANG, P., MIAO, L., ZHAO, Y., ZHIMIN, Z., QING, L., ZHU, W.-G. & WU, M. 2006. Novel link between E2F1 and Smac/DIABLO: proapoptotic Smac/DIABLO is transcriptionally upregulated by E2F1. *Nucleic Acids Research,* 34**,** 2046-55.

XU, E., ZHANG, J. & CHEN, X. 2013. MDM2 expression is repressed by the RNA-binding protein RNPC1 via mRNA stability. *Oncogene,* 32**,** 2169-78.

XU, M., SHEPPARD, K. A., PENG, C. Y., YEE, A. S. & PIWNICA-WORMS, H. 1994. Cyclin A/CDK2 binds directly to E2F-1 and inhibits the DNA-binding activity of E2F-1/DP-1 by phosphorylation. *Molecular and Cellular Biology,* 14**,** 8420-31.

XU, Y., GAO, X. D., LEE, J. H., HUANG, H., TAN, H., AHN, J., REINKE, L. M., PETER, M. E., FENG, Y., GIUS, D., SIZIOPIKOU, K. P., PENG, J., XIAO, X. & CHENG, C. 2014. Cell type-restricted activity of hnRNPM promotes breast cancer metastasis via regulating alternative splicing. *Genes Dev,* 28**,** 1191-203.

YAMASAKI, L., BRONSON, R., WILLIAMS, B. O., DYSON, N. J., HARLOW, E. & JACKS, T. 1998. Loss of E2F-1 reduces tumorigenesis and extends the lifespan of Rb1(+/-)mice. *Nature Genetics,* 18**,** 360-4.

YAMASAKI, L., JACKS, T., BRONSON, R., GOILLOT, E., HARLOW, E. & DYSON, N. J. 1996. Tumor induction and tissue atrophy in mice lacking E2F-1. *Cell,* 85**,** 537-48.

YAN, F., ALINARI, L., LUSTBERG, M. E., MARTIN, L. K., CORDERO-NIEVES, H. M., BANASAVADI-SIDDEGOWDA, Y., VIRK, S., BARNHOLTZ-SLOAN, J., BELL, E. H., WOJTON, J., JACOB, N. K., CHAKRAVARTI, A., NOWICKI, M. O., WU, X., LAPALOMBELLA, R., DATTA, J., YU, B., GORDON, K., HASELEY, A., PATTON, J. T., SMITH, P. L., RYU, J., ZHANG, X., MO, X., MARCUCCI, G., NUOVO, G., KWON, C. H., BYRD, J. C., CHIOCCA, E. A., LI, C., SIF, S., JACOB, S., LAWLER, S., KAUR, B. & BAIOCCHI, R. A. 2014. Genetic validation of the protein arginine methyltransferase PRMT5 as a candidate therapeutic target in glioblastoma. *Cancer Research,* 74**,** 1752-65.

YANG, I. S., SON, H. & KIM, S. 2016. ISOexpresso: a web-based platform for isoform-level expression analysis in human cancer. *BMC Genomics,* 17**,** 631.

YANG, J., VALINEVA, T., HONG, J., BU, T., YAO, Z., JENSEN, O. N., FRILANDER, M. J. & SILVENNOINEN, O. 2007. Transcriptional co-activator protein p100 interacts with snRNP proteins and facilitates the assembly of the spliceosome. *Nucleic Acids Research,* 35**,** 4485-94.

YANG, Y. & BEDFORD, M. T. 2013. Protein arginine methyltransferases and cancer. *Nat Rev Cancer,* 13**,** 37-50.

YANG, Y., HADJIKYRIACOU, A., XIA, Z., GAYATRI, S., KIM, D., ZURITA-LOPEZ, C., KELLY, R., GUO, A., LI, W., CLARKE, S. G. & BEDFORD, M. T. 2015. PRMT9 is a Type II methyltransferase that methylates the splicing factor SAP145. *Nature communications,* 6**,** 6428-6428.

YAO, G., LEE, T. J., MORI, S., NEVINS, J. R. & YOU, L. 2008. A bistable Rb-E2F switch underlies the restriction point. *Nature Cell Biology,* 10**,** 476-82.

YAO, R., JIANG, H., MA, Y., WANG, L., DU, J., HOU, P., GAO, Y., ZHAO, L., WANG, G., ZHANG, Y., LIU, D. X., HUANG, B. & LU, J. 2014. PRMT7 induces epithelial-to-mesenchymal transition and promotes metastasis in breast cancer. *Cancer Res,* 74**,** 5656-67.

YE, Y., ZHANG, B., MAO, R., ZHANG, C., WANG, Y., XING, J., LIU, Y. C., LUO, X., DING, H., YANG, Y., ZHOU, B., JIANG, H., CHEN, K., LUO, C. & ZHENG, M. 2017. Discovery and optimization of selective inhibitors of protein arginine methyltransferase 5 by docking-based virtual screening. *Org Biomol Chem,* 15**,** 3648-3661.

YOSHIMATSU, M., TOYOKAWA, G., HAYAMI, S., UNOKI, M., TSUNODA, T., FIELD, H. I., KELLY, J. D., NEAL, D. E., MAEHARA, Y., PONDER, B. A., NAKAMURA, Y. & HAMAMOTO, R. 2011. Dysregulation of PRMT1 and PRMT6, Type I arginine methyltransferases, is involved in various types of human cancers. *International Journal of Cancer,* 128**,** 562-73.

YOSHIMOTO, T., BOEHM, M., OLIVE, M., CROOK, M. F., SAN, H., LANGENICKEL, T. & NABEL, E. G. 2006. The arginine methyltransferase PRMT2 binds RB and regulates E2F function. *Experimental Cell Research,* 312**,** 2040-2053.

YU, L., DI, Y., XIN, L., REN, Y., LIU, X., SUN, X., ZHANG, W., YAO, Z. & YANG, J. 2017. SND1 acts as a novel gene transcription activator recognizing the conserved Motif domains of Smad promoters, inducing TGF[beta]1 response and breast cancer metastasis. *Oncogene,* 36**,** 3903-3914.

YU, W., CHAN-ON, W., TEO, M., ONG, C. K., CUTCUTACHE, I., ALLEN, G. E., WONG, B., MYINT, S. S., LIM, K. H., VOORHOEVE, P. M., ROZEN, S., SOO, K. C., TAN, P. & TEH, B. T. 2011. First somatic mutation of E2F1 in a critical DNA binding residue discovered in well-differentiated papillary mesothelioma of the peritoneum. *Genome Biology,* 12**,** R96.

ZHAN, L., ZHANG, Y., WANG, W., SONG, E., FAN, Y. & WEI, B. 2016. E2F1: a promising regulator in ovarian carcinoma | SpringerLink. *Tumour Biology,* 37**,** 2823-2831.

ZHANG, H. S., GAVIN, M., DAHIYA, A., POSTIGO, A. A., MA, D., LUO, R. X., HARBOUR, J. W. & DEAN, D. C. 2000. Exit from G1 and S phase of the cell cycle is regulated by repressor complexes containing HDAC-Rb-hSWI/SNF and Rb-hSWI/SNF. *Cell,* 101**,** 79-89.

ZHANG, J., LIEU, Y. K., ALI, A. M., PENSON, A., REGGIO, K. S., RABADAN, R., RAZA, A., MUKHERJEE, S. & MANLEY, J. L. 2015. Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. *Proc Natl Acad Sci U S A,* 112**,** E4726-34.

ZHANG, J., XU, E., REN, C., YAN, W., ZHANG, M., CHEN, M., CARDIFF, R. D., IMAI, D. M., WISNER, E. & CHEN, X. 2014a. Mice deficient in Rbm38, a target of the p53 family, are susceptible to accelerated aging and spontaneous tumors. *Proc Natl Acad Sci U S A,* 111**,** 18637-42.

ZHANG, J., XU, K., LIU, P., GENG, Y., WANG, B., GAN, W., GUO, J., WU, F., CHIN, Y. R., BERRIOS, C., LIEN, E. C., TOKER, A., DECAPRIO, J. A., SICINSKI, P. & WEI, W. 2016a. Inhibition of Rb Phosphorylation Leads to mTORC2-Mediated Activation of Akt. *Mol Cell,* 62**,** 929-42.

ZHANG, L. & WANG, C. 2006. F-box protein Skp2: a novel transcriptional target of E2F. *Oncogene,* 25**,** 2615-27.

ZHANG, W., JI, W., LIU, X., OUYANG, G. & XIAO, W. 2014b. ELL inhibits E2F1 transcriptional activity by enhancing E2F1 deacetylation via recruitment of histone deacetylase 1. *Mol Cell Biol,* 34**,** 765-75.

ZHANG, X., TRÉPANIER, V., BEAUJOIS, R., VIRANAICKEN, W., DROBETSKY, E. & DESGROSEILLERS, L. 2016b. The downregulation of the RNA-binding protein Staufen2 in response to DNA damage promotes apoptosis. *Nucleic Acids Research,* 44**,** 3695-3712.

ZHANG, X. H. & CHASIN, L. A. 2004. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev,* 18**,** 1241-50.

ZHAO, Y., TAN, J., ZHUANG, L., JIANG, X., LIU, E. T. & YU, Q. 2005. Inhibitors of histone deacetylases target the Rb-E2F1 pathway for apoptosis induction through activation of proapoptotic protein Bim. *Proceedings of the National Academy of Sciences,* 102**,** 16090-5.

ZHENG, S., MOEHLENBRINK, J., LU, Y.-C., ZALMAS, L.-P., SAGUM, C. A., CARR, S., MCGOURAN, J. F., ALEXANDER, L., FEDOROV, O., MUNRO, S., KESSLER, B., BEDFORD, M. T., YU, Q. & LA THANGUE, N. B. 2013. Arginine methylation-dependent reader-writer interplay governs growth control by E2F-1. *Molecular Cell,* 52**,** 37-51.

ZHONG, J., CAO, R. X., LIU, J. H., LIU, Y. B., WANG, J., LIU, L. P., CHEN, Y. J., YANG, J., ZHANG, Q. H., WU, Y., DING, W. J., HONG, T., XIAO, X. H., ZU, X. Y. & WEN, G. B. 2014. Nuclear loss of protein arginine N-methyltransferase 2 in breast carcinoma is associated with tumor grade and overexpression of cyclin D1 protein. *Oncogene,* 33**,** 5546-58.

ZHOU, H.-L., LUO, G., WISE, J. A. & LOU, H. 2014. Regulation of alternative splicing by local histone modifications: potential roles for RNA-guided mechanisms. *Nucleic Acids Research,* 42**,** 701-713.

ZHU, H., REN, S., BITLER, B. G., AIRD, K. M., TU, Z., SKORDALAKES, E., ZHU, Y., YAN, J., SUN, Y. & ZHANG, R. 2015. SPOP E3 Ubiquitin Ligase Adaptor Promotes Cellular Senescence by Degrading the SENP7 deSUMOylase. *Cell Rep,* 13**,** 1183-93.

ZUO, P. & MANIATIS, T. 1996. The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes Dev,* 10**,** 1356-68.

ZURITA-LOPEZ, C. I., SANDBERG, T., KELLY, R. & CLARKE, S. G. 2012. Human protein arginine methyltransferase 7 (PRMT7) is a type III enzyme forming omega-NG-monomethylated arginine residues. *J Biol Chem,* 287**,** 7859-70.

Appendices

Appendix 1. Sequence alignments for CRISPR cell lines

KEY
UPPERCASE: Exon
Lowercase: Intron
Highlighted: potential sgRNA binding site

E2F1 Subject: WT B, Query: WT A

```
Query  79   tctcttctggcctcactcctggttactgggcatcctcccgttttcgccacagGTGAAGCG
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  196  tctcttctggcctcactcctggttactgggcatcctcccgttttcgccacagGTGAAGCG

Query  139  GAGGCTGGACCTGGAAACTGACCATCAGTACCTGGCCGAGAGCAGTGGGCCAGCTCGGGG
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  256  GAGGCTGGACCTGGAAACTGACCATCAGTACCTGGCCGAGAGCAGTGGGCCAGCTCGGGG

Query  199  CAGAGGCCGCCATCCAGGAAAAGgtacctatggagcttgggctgggcagacccccctgtga
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  316  CAGAGGCCGCCATCCAGGAAAAGgtacctatggagcttgggctgggcagacccccctgtga

Query  259  ccaggcccaggttagggatgtgaacagacaaacatgcttgcaagctaacgtctggcttgt
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  376  ccaggcccaggttagggatgtgaacagacaaacatgcttgcaagctaacgtctggcttgt

Query  319  cgggcttgagccaggcctctgaggaccctgatgggcagttggagtggccacctggcttg
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  436  cgggcttgagccaggcctctgaggaccctgatgggcagttggagtggccacctggcttg
```

E2F1 Subject: E2F1$^{-/-}$ C, Query: WT A

```
Query  79   tctcttctggcctcactcctggttactgggcatcctcccgttttcgccacagGTGAAGCG
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  80   tctcttctggcctcactcctggttactgggcatcctcccgttttcgccacagGTGAAGCG

Query  139  GAGGCTGGACCTGGAAACT-GACCATCAGTACCTGGCCGAGAGCAGTGGGCC-A-G-CTC
            |||||||||||||||||| | ||   | ||||||||| | | ||   |||  | | |||
Sbjct  140  GAGGCTGGACCTGGAAACTTGCCCGANATTACCTGGCC-A-A-CACNNGGCAGACGCCTC

Query  195  GGGGCAGAGGCCGCCATCCAGGaaaaggtacctatggagcttgggctggg-cagaccccc
             | ||||| ||||  ||| | ||||| ||||||||| ||      ||||| |  | ||||
Sbjct  197  CGTGCAGAAGCCGGTATCTATGAAAATGTACCTATGCAGACCCC-CTGGGACCNAGCCCC

Query  254  tgtgaccaggcccaggttagggatgtg  280
            ||||||   ||||| |||| ||||||||
Sbjct  256  TGTGACGGAGCCCAAGTTACGGATGTG  282
```

E2F1 Subject: E2F1$^{-/-}$ D, Query: WT A

```
Query  99   ggttactgggcatcctcccgttttcgccacagGTGAAGCGGAGGCTGGACCTGGAAACTG
            ||| | ||||||||||||||||        ||||  |||||||||  |||| ||||||  |
Sbjct  357  ggtnattgggcatcctcccgtnnnnnngncagGGAAAGCGGAGGACGGACATGGAAATTT

Query  159  ACCATCA-GTACCTGGCCGAGAGCAGTGGGCCAGCTCGGGGCAGAGGCCGCCATCCAGGa
            |||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  417  ACCATCATTTACCTGGCCGAGAGCAGTGGGCCAGCTCGGGGCAGAGGCCGCCATCCAGGa

Query  218  aaaggtacctatggagcttgggctgggcagacccccctgtgaccaggcccaggttagggat
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  477  aaaggtacctatggagcttgggctgggcagacccccctgtgaccaggcccaggttagggat

Query  278  gtgaacagacaaacatgcttgcaagctaacgtctggcttgtcgggcttgagccaggcctc
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  537  gtgaacagacaaacatgcttgcaagctaacgtctggcttgtcgggcttgagccaggcctc

Query  338  tgaggaccctgatgggcagttggagtggccacctggcttg  377
            |||||||||||||||||||||||||||||||||||||||
Sbjct  597  tgaggaccctgatgggcagttggagtggccacctggcttg  636
```

DNML1 Subject: WT B, Query: WT A

```
Query  518 TCACATAGAGTCATTGTATTCCACCTGTCCTTCAATTTAGTTTTTTCTGAGCTTCTTTGC
             ||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Sbjct  563   TCACATAGAGTCATTGTATTCCACCTGTCCTTCAATTTAGTTTTTTTTGAGCTTCTTTGC

Query  578   AGCCTTTGATGTGTTTTTAAGAAAGCTGAATGCACAAGAGGATCTGTGACACTGACATGG
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  623   AGCCTTTGATGTGTTTTTAAGAAAGCTGAATGCACAAGAGGATCTGTGACACTGACATGG

Query  638   CTGTGGTGTGCATACTGTGTAGTTACATAGCCCTTCCAATTCTGGGTCCATTTGCACTAG
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  683   CTGTGGTGTGCATACTGTGTAGTTACATAGCCCTTCCAATTCTGGGTCCATTTGCACTAG

Query  698   CAAATTAAAATATGCTTTGATTCATACTTAAACCTGAAAGCAGGAATGCCTACATTAATT
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  743   CAAATTAAAATATGCTTTGATTCATACTTAAACCTGAAAGCAGGAATGCCTACATTAATT

Query  758   CCTACATTAAAAACAGCCATCTACCCTTGATTATCTAGAAAGACTTGGTAATGATGGTCA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  803   CCTACATTAAAAACAGCCATCTACCCTTGATTATCTAGAAAGACTTGGTAATGATGGTCA

Query  818   GTTCCTTTTAGATTTCAGAAAATCAAATGATGACCTAAATTTCCCTTAATTTGCAAATAC
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  863   GTTCCTTTTAGATTTCAGAAAATCAAATGATGACCTAAATTTCCCTTAATTTGCAAATAC

Query  878   AGTAGTAATTAAGGTACATCTCTAAAGTGGAGCACTTACACCAGGCTCTAAGATTCACTT
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  923   AGTAGTAATTAAGGTACATCTCTAAAGTGGAGCACTTACACCAGGCTCTAAGATTCACTT

Query  938   TGAGGTGGAACTTAAAACCAGTGTACTGTATGTATGCATTGGTAATAGCTACTTTTGCTT
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  983   TGAGGTGGAACTTAAAACCAGTGTACTGTATGTATGCATTGGTAATAGCTACTTTTGCTT
```

```
Query  998   CATAGNTCATACCAACAAA  1016
             ||||||||||||||||||
Sbjct  1043  CATAGNTCATACCAACAAA  1061
```

DNML1 Subject: E2F1$^{-/-}$ C, Query: WT A

```
Query  515   TCCTCACATAGAGTCATTGTATTCCACCTGTCCTTCAATTTAGTTTTTTCTGAGCTTCTT
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  529   TCCTCACATAGAGTCATTGTATTCCACCTGTCCTTCAATTTAGTTTTTTCTGAGCTTCTT

Query  575   TGCAGCCTTTGATGTGTTTTTAAGAAAGCTGAATGCACAAGAGGATCTGTGACACTGACA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  589   TGCAGCCTTTGATGTGTTTTTAAGAAAGCTGAATGCACAAGAGGATCTGTGACACTGACA

Query  635   TGGCTGTGGTGTGCATACTGTGTAGTTACATAGCCCTTCCAATTCTGGGTCCATTTGCAC
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  649   TGGCTGTGGTGTGCATACTGTGTAGTTACATAGCCCTTCCAATTCTGGGTCCATTTGCAC

Query  695   TAGCAAATTAAAATATGCTTTGATTCATACTTAAACCTGAAAGCAGGAATGCCTACATTA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  709   TAGCAAATTAAAATATGCTTTGATTCATACTTAAACCTGAAAGCAGGAATGCCTACATTA

Query  755   ATTCCTACATTAAAAACAGCCATCTACCCTTGATTATCTAGAAAGACTTGGTAATGATGG
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  769   ATTCCTACATTAAAAACAGCCATCTACCCTTGATTATCTAGAAAGACTTGGTAATGATGG

Query  815   TCAGTTCCTTTTAGATTTCAGAAAATCAAATGATGACCTAAATTTCCCTTAATTTGCAAA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  829   TCAGTTCCTTTTAGATTTCAGAAAATCAAATGATGACCTAAATTTCCCTTAATTTGCAAA

Query  875   TACAGTAGTAATTAAGGTACATCTCTAAAGTGGAGCACTTACACCAGGCTCTAAGATTCA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  889   TACAGTAGTAATTAAGGTACATCTCTAAAGTGGAGCACTTACACCAGGCTCTAAGATTCA

Query  935   CTTTGAGGTGGAACTTAAAACCAGTGTACTGTATGTATGCATTGGTAATAGCTACTTTTG
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  949   CTTTGAGGTGGAACTTAAAACCAGTGTACTGTATGTATGCATTGGTAATAGCTACTTTTG

Query  995   CTTCATAGNTCATACCAACAAA  1016
             ||||||||||||||||||||||
Sbjct  1009  CTTCATAGNTCATACCAACAAA  1030
```

DNML1 Subject: E2F1$^{-/-}$ D, Query: WT A

```
Query  515   TCCTCACATAGAGTCATTGTATTCCACCTGTCCTTCAATTTAGTTTTTTCTGAGCTTCTT
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  529   TCCTCACATAGAGTCATTGTATTCCACCTGTCCTTCAATTTAGTTTTTTCTGAGCTTCTT

Query  575   TGCAGCCTTTGATGTGTTTTTAAGAAAGCTGAATGCACAAGAGGATCTGTGACACTGACA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  589   TGCAGCCTTTGATGTGTTTTTAAGAAAGCTGAATGCACAAGAGGATCTGTGACACTGACA
```

```
Query  635   TGGCTGTGGTGTGCATACTGTGTAGTTACATAGCCCTTCCAATTCTGGGTCCATTTGCAC
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  649   TGGCTGTGGTGTGCATACTGTGTAGTTACATAGCCCTTCCAATTCTGGGTCCATTTGCAC

Query  695   TAGCAAATTAAAATATGCTTTGATTCATACTTAAACCTGAAAGCAGGAATGCCTACATTA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  709   TAGCAAATTAAAATATGCTTTGATTCATACTTAAACCTGAAAGCAGGAATGCCTACATTA

Query  755   ATTCCTACATTAAAAACAGCCATCTACCCTTGATTATCTAGAAAGACTTGGTAATGATGG
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  769   ATTCCTACATTAAAAACAGCCATCTACCCTTGATTATCTAGAAAGACTTGGTAATGATGG

Query  815   TCAGTTCCTTTTAGATTTCAGAAAATCAAATGATGACCTAAATTTCCCTTAATTTGCAAA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  829   TCAGTTCCTTTTAGATTTCAGAAAATCAAATGATGACCTAAATTTCCCTTAATTTGCAAA

Query  875   TACAGTAGTAATTAAGGTACATCTCTAAAGTGGAGCACTTACACCAGGCTCTAAGATTCA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  889   TACAGTAGTAATTAAGGTACATCTCTAAAGTGGAGCACTTACACCAGGCTCTAAGATTCA

Query  935   CTTTGAGGTGGAACTTAAAACCAGTGTACTGTATGTATGCATTGGTAATAGCTACTTTTG
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  949   CTTTGAGGTGGAACTTAAAACCAGTGTACTGTATGTATGCATTGGTAATAGCTACTTTTG

Query  995   CTTCATAGNTCATACCAACAAA  1016
             |||||||||||||||||||||
Sbjct  1009  CTTCATAGNTCATACCAACAAA  1030
```

NR0B1 Subject: WT B, Query: WT A

```
Query  80    CTCATGCTTGAGCTGGCCCAGGACCGCTTGCAGTTCGAGACTGTGGAAGTCTCGGAGCCC
             |||||||||||| |||||||||||||||||||| |||||||||||||| | |||||||
Sbjct  79    CTCATGCTTGAGNTGGCCCAGGACCGCTTGCAGTTTGAGACTGTGGAAGTTTTGGAGCCC

Query  140   AGCATGCTGCAGAAGATCCTCACCACCAGGCGGCGGGAGACCGGGGGCAACGAGCCACTG
             |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Sbjct  139   AGCATGCTGCAGAAGATCCTCACCACCCGGCGGCGGGAGACCGGGGGCAACGAGCCACTG

Query  200   CCCGTGCCCACGCTGCAGCACCATTTGGCACCGCCGGCGGAGGCCAGGAAGGTGCCCTCC
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  199   CCCGTGCCCACGCTGCAGCACCATTTGGCACCGCCGGCGGAGGCCAGGAAGGTGCCCTCC

Query  260   GCCTCCCAGGTCCAAGCCATCAAGTGCTTTCTTTCCAAATGCTGGAGTCTGAACATCAGT
             |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Sbjct  259   GCCTCCCAGGTCCAAGCCATCAAGTGCTTTCTTTCCAAATGCTGGAGTTTGAACATCAGT

Query  320   ACCAAGGAGTACGCCTACCTCAAGGGGACCGTGCTCTTTAACCCGGgtaagggtactggc
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  319   ACCAAGGAGTACGCCTACCTCAAGGGGACCGTGCTCTTTAACCCGGgtaagggtactggc

Query  380   cttaggcgccggcttttcccagctcacaaaagcatcgggcagtgcctatctaggggcgcg
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  379   cttaggcgccggcttttcccagctcacaaaagcatcgggcagtgcctatctaggggcgcg

Query  440   ggcagtaacgagttttcagtgatcaggagagtgtcggggcaaaggtgaagaaatcgtgac
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct  439  ggcagtaacgagttttcagtgatcaggagagtgtcggggcaaaggtgaagaaatcgtgac

Query  500  taactcagcagagttggggtggggagctccaggaaccacttctgctgggtgggctgctat
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  499  taactcagcagagttggggtggggagctccaggaaccacttctgctgggtgggctgctat
```

NR0B1 Subject: E2F1$^{-/-}$ C, Query: WT A

```
Query  71   GCGTCCCTGCTCATGCTTGAGCTGGCCCAGGACCGCTTGCAGTTCGAGACTGTGGAAGTC
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  69   GCGTCCCTGCTCATGCTTGAGCTGGCCCAGGACCGCTTGCAGTTCGAGACTGTGGAAGTC

Query  131  TCGGAGCCCAGCATGCTGCAGAAGATCCTCACCACCAGGCGGCGGGAGACCGGGGGCAAC
            |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Sbjct  129  TCGGAGCCCAGCATGCTGCAGAAGATCCTCACCACCCGGCGGCGGGAGACCGGGGGCAAC

Query  191  GAGCCACTGCCCGTGCCCACGCTGCAGCACCATTTGGCACCGCCGGCGGAGGCCAGGAAG
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  189  GAGCCACTGCCCGTGCCCACGCTGCAGCACCATTTGGCACCGCCGGCGGAGGCCAGGAAG

Query  251  GTGCCCTCCGCCTCCCAGGTCCAAGCCATCAAGTGCTTTCTTTCCAAATGCTGGAGTCTG
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Sbjct  249  GTGCCCTCCGCCTCCCAGGTCCAAGCCATCAAGTGCTTTCTTTCCAAATGCTGGAGTTTG

Query  311  AACATCAGTACCAAGGAGTACGCCTACCTCAAGGGGACCGTGCTCTTTAACCCGGgtaag
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  309  AACATCAGTACCAAGGAGTACGCCTACCTCAAGGGGACCGTGCTCTTTAACCCGGgtaag

Query  371  ggtactggccttaggcgccggcttttcccagctcacaaaagcatcgggcagtgcctatct
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  369  ggtactggccttaggcgccggcttttcccagctcacaaaagcatcgggcagtgcctatct

Query  431  aggggcgcgggcagtaacgagttttcagtgatcaggagagtgtcggggcaaaggtgaaga
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  429  aggggcgcgggcagtaacgagttttcagtgatcaggagagtgtcggggcaaaggtgaaga

Query  491  aatcgtgactaactcagcagagttggggtggggagctccaggaaccacttctgctgggtg
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  489  aatcgtgactaactcagcagagttggggtggggagctccaggaaccacttctgctgggtg
```

NR0B1 Subject: E2F1$^{-/-}$ D, Query: WT A

```
Query  75   CCCTGCTCATGCTTGAGCTGGCCCAGGACCGCTTGCAGTTCGAGACTGTGGAAGTCTCGG
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  74   CCCTGCTCATGCTTGAGCTGGCCCAGGACCGCTTGCAGTTCGAGACTGTGGAAGTCTCGG

Query  135  AGCCCAGCATGCTGCAGAAGATCCTCACCACCAGGCGGCGGGAGACCGGGGGCAACGAGC
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  134  AGCCCAGCATGCTGCAGAAGATCCTCACCACCAGGCGGCGGGAGACCGGGGGCAACGAGC

Query  195  CACTGCCCGTGCCCACGCTGCAGCACCATTTGGCACCGCCGGCGGAGGCCAGGAAGGTGC
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  194  CACTGCCCGTGCCCACGCTGCAGCACCATTTGGCACCGCCGGCGGAGGCCAGGAAGGTGC
```

```
Query  255  CCTCCGCCTCCCAGGTCCAAGCCATCAAGTGCTTTCTTTCCAAATGCTGGAGTCTGAACA
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  254  CCTCCGCCTCCCAGGTCCAAGCCATCAAGTGCTTTCTTTCCAAATGCTGGAGTCTGAACA

Query  315  TCAGTACCAAGGAGTACGCCTACCTCAAGGGGACCGTGCTCTTTAACCCGGgtaagggta
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  314  TCAGTACCAAGGAGTACGCCTACCTCAAGGGGACCGTGCTCTTTAACCCGGgtaagggta

Query  375  ctggccttaggcgccggcttttcccagctcacaaaagcatcgggcagtgcctatctaggg
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  374  ctggccttaggcgccggcttttcccagctcacaaaagcatcgggcagtgcctatctaggg

Query  435  gcgcgggcagtaacgagttttcagtgatcaggagagtgtcggggcaaaggtgaagaaatc
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  434  gcgcgggcagtaacgagttttcagtgatcaggagagtgtcggggcaaaggtgaagaaatc

Query  495  gtgactaactcagcagagttggggtggggagctccaggaaccacttctgctgggtgggct
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  494  gtgactaactcagcagagttggggtggggagctccaggaaccacttctgctgggtgggct
```

Clonal difference: GTC to GTT: both code for Valine, no expected differences in protein

MACROD2 Subject: WT A, Query: WT B

```
Query  140  tcagggactgtggtctttttttttccagCTCCAGGCGAGGACACACCTAGGATGCCTGG
            ||||||||||||||| |||||||||||||||||| |||||||||||||||||||||||||
Sbjct  141  tcagggactgtggtttttttttttttccagCTCCCGGCGAGGACACACCTAGGATGCCTGG

Query  200  GAAAAGTGAAGGCTCCAGTGACCTAGAAAATACTCCAGGTCCTGATGgtaaggtttgag
            ||||| |||||| |||||||||||||||||| |||||||||||||||||||||||||||
Sbjct  201  GAAAAATGAAGGGTCCAGTGACCTAGAAAATAATCCAGGTCCTGATGgtaaggtttgag

Query  260  ctattggagcccatggaaactacaaatttggcatattttctatgtgtaatttaggattaa
            |||||||||||| |||||| |||||||||||||||||| |||||||||||||||||||||
Sbjct  261  ctattggagcccctggaaaatacaaatttggcatatttttatgtgtaatttaggattaa

Query  320  gggaactatctgttctataaagcaacatatttggttagaacacagtaaatttcaaaaaca
            ||||| ||| |||| |||||||||||||||||||||||||||||||||||||||||||||
Sbjct  321  gggaaatatttgttttataaagcaacatatttggttagaacacagtaaatttcaaaaaca

Query  380  attttgtgctacaaactaataaaaatggcaacctcttatttaaaaaaaaacctgtttct
            ||||||||| |||||| |||||||||||||||||| |||||||||||||||||||||||||
Sbjct  381  attttgtgntacaaaataataaaaatggcaacctttttatttaaaaaaaaaacctgtttct

Query  440  agctgttttaaaacagcatatgtcctggtttaatcatatcttataatactgaggactcaa
            |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Sbjct  441  agctgttttaaaacagcatatgtcctggtttaatcatattttataatactgaggactcaa

Query  500  tagatactttcaagattaaactgaactgaagtctatctagtgaactttatggaagcatta
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  501  tagatactttcaagattaaactgaactgaagtctatctagtgaactttatggaagcatta
```

WT A shows differences from WT B. WT B is the canonical sequence

```
WT B  1   PGEDTPRMPGKSEGSSDLENTPGPD  25
          PGEDTPRMPGK EGSSDLEN PGPD
WT A  1   PGEDTPRMPGKNEGSSDLENNPGPD  25
```

MACROD2 Subject: E2F1$^{-/-}$ C, Query: WT B

```
Query  79   aaaagggcaatggagggtgacagaaatgattggccctcaggctgttataattctttattc
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Sbjct  79   aaaagggcaatggagggtgacagaaatgattggccctcaggctgttataattctctattc

Query  139  atcagggactgtggtctttttttttccagCTCCAGGCGAGGACACACCTAGGATGCCTG
            |||||||||||||||||||| | ||||||||||||||||||||||||||||||||||||
Sbjct  139  atcagggactgtggtcttttttctcttccagCTCCAGGCGAGGACACACCTAGGATGCCTG

Query  199  GGAAAAGTGAAGGCTCCAGTGACCTAGAAAATACTCCAGGTCCTGATGgtaaggtttga
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Sbjct  199  GGAAAAGTGAAGGCTCCAGTGACCTAGAAAATACTCCAGGTCCTGATGgtaaggttctga

Query  259  gctattggagcccatggaaactacaaatttggcatattttctatgtgtaatttaggatta
            |||||||||||||||||||||||||| |||||||||||||||||||| ||||||||
Sbjct  259  gctattggagcccatggaaactacaaatctggcatattttctatgtgtaatctaggatta

Query  319  agggaactatctgttctataaagcaacatatttggttagaacacagtaaatttcaaaaac
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  319  agggaactatctgttctataaagcaacatatttggttagaacacagtaaatttcaaaaac

Query  379  aattttgtgctacaaactaataaaaatggcaacctcttatttaaaaaaaaaacctgtttc
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  379  aattttgtgctacaaactaataaaaatggcaacctcttatttaaaaaaaaaacctgtttc

Query  439  tagctgttttaaaacagcatatgtcctggtttaatcatatcttataatactgaggactca
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  439  tagctgttttaaaacagcatatgtcctggtttaatcatatcttataatactgaggactca

Query  499  atagatactttcaagattaaactgaactgaagtctatctagtgaactttatggaagcatt
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  499  atagatactttcaagattaaactgaactgaagtctatctagtgaactttatggaagcatt
```

MACROD2 Subject: E2F1$^{-/-}$ D, Query: WT B

```
Query  79   aaaagggcaatggagggtgacagaaatgattggccctcaggctgttataattctttattc
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Sbjct  81   aaaagggcaatggagggtgacagaaatgattggccctcaggctgttataattctctattc

Query  139  atcagggactgtggtctttttttttccagCTCCAGGCGAGGACACACCTAGGATGCCTG
            |||||||||||||||||||| | ||||||||||||||||||||||||||||||||||||
Sbjct  141  atcagggactgtggtcttttttctcttccagCTCCAGGCGAGGACACACCTAGGATGCCTG

Query  199  GGAAAAGTGAAGGCTCCAGTGACCTAGAAAATACTCCAGGTCCTGATGgtaaggtttga
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Sbjct  201  GGAAAAGTGAAGGCTCCAGTGACCTAGAAAATACTCCAGGTCCTGATGgtaaggttctga

Query  259  gctattggagcccatggaaactacaaatttggcatattttctatgtgtaatttaggatta
            |||||||||||||||||||||||||| |||||||||||||||||||| ||||||||
Sbjct  261  gctattggagcccatggaaactacaaatctggcatattttctatgtgtaatctaggatta
```

```
Query  319  agggaactatctgttctataaagcaacatatttggttagaacacagtaaatttcaaaaac
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  321  agggaactatctgttctataaagcaacatatttggttagaacacagtaaatttcaaaaac

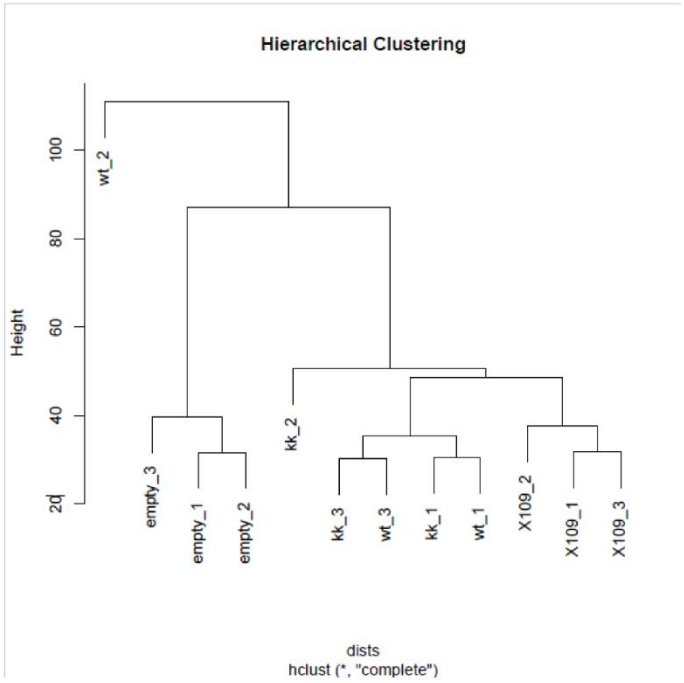
Query  379  aattttgtgctacaaactaataaaaatggcaacctcttatttaaaaaaaaaacctgtttc
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  381  aattttgtgctacaaactaataaaaatggcaacctcttatttaaaaaaaaaacctgtttc

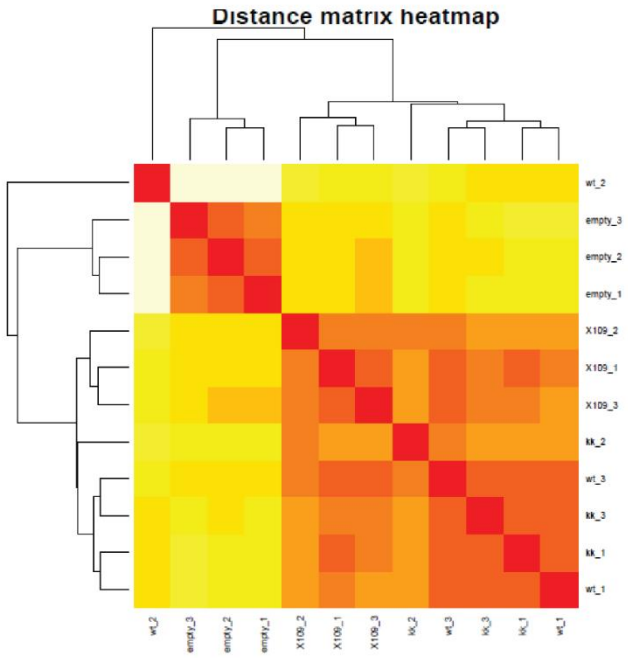
Query  439  tagctgttttaaaacagcatatgtcctggtttaatcatatcttataatactgaggactca
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  441  tagctgttttaaaacagcatatgtcctggtttaatcatatcttataatactgaggactca

Query  499  atagatactttcaagattaaactgaactgaagtctatctagtgaactttatggaagcatt
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  501  atagatactttcaagattaaactgaactgaagtctatctagtgaactttatggaagcatt
```

No changes in coding sequence between WT B and either E2F1$^{-/-}$ cell line.

Appendix 2. Hierarchical clustering of RNA-seq data





Appendix 2. Hierarchical clustering of RNA-seq analysis. Expression was compared between samples to determine similarity.