Investigating the Impact of PAD4-mediated Citrullination on the E2F Genomic Landscape

Koyo Harada

St Catherine's College

University of Oxford

A thesis submitted for the degree of *Doctor of Philosophy* in Oncology

Hilary Term 2023

Table of Contents

Investigating the Impact of PAD4-mediated Citrullination on the E2F Genomic Landscape

Koyo Harada

St Catherine's College
University of Oxford

A thesis submitted for the degree of *Doctor of Philosophy* in Oncology

Abstract

E2F1 is a master transcription factor that plays a pivotal role in regulating cell cycle progression and proliferation, as well as mediating other cell fates including apoptosis and the inflammatory response. These diverse biological outcomes of E2F1 activity are determined in part by post-translational modifications, including arginine methylation mediated by protein arginine methyltransferases (PRMT) 1 and 5, and citrullination mediated by peptidyl arginine deiminase (PAD) 4. The activity of PRMT5 and PRMT1 are antagonistic to one another; the former symmetrically methylates R111/R113 of E2F1 and favours proliferation, whilst the latter promotes E2F1-dependent apoptosis via asymmetric methylation at R109. On the other hand, PAD4 channels E2F1 into a pro-inflammatory pathway by citrullinating its target arginine residues, which also include R109. Recently, PRMT5-driven methylation events at R111/R113 and recruitment of the reader protein p100/TSN were found to permit E2F1 to regulate many genes at the level of alternative splicing. However, despite mounting evidence suggesting that many RNA-binding proteins are substrates for citrullination, the link between PADs and RNA processing remains unclear. In this study, by performing RNA sequencing, we demonstrated that E2F1 knockdown and PAD4 inhibition exhibit significant global impacts on the human cancer transcriptome at the level of alternative RNA splicing. Moreover, we uncovered that this regulation involves the citrullination of E2F1, which enhances its interaction with p100/TSN and the splicing factor SRSF3. Using RNA immunoprecipitation methods, we additionally elucidated that PAD4 plays an essential role in regulating the RNA binding affinity of SRSF3 in an E2F1-dependent manner, which underpins the molecular mechanism by which PAD4 and E2F1 together regulate alternative splicing. In summary, we discover a new function of PAD4 as a splicing regulator in cancer cells and propose a mechanism by which this alternative splicing is mediated by the novel interplay between E2F1 and SRSF3.

(296 words)

Statement of Authorship

My thesis contains parts that were completed in collaboration as follows:

- RNA-seq bioinformatics analyses (Chapter 3) were conducted in collaboration with Dr Anastasia Samsonova and Dr Alexander Kanapin (Centre for Genome Bioinformatics, St. Petersburg State University).

- Immunoblotting for E2F1 and PAD4 (Chapter 3: Figure 3-1A) and SRSF3 citrullination experiment (Chapter 5: Figure 5-2A) were performed in collaboration with Dr Amit Shrestha in our research group (Department of Oncology, University of Oxford).

- Immunoprecipitation-mass spectrometry (IP-MS) experiment to investigate the E2F1 interactome (Chapter 5: Table 7-1 and Table 7-2) was performed in collaboration with Dr Iolanda Vendrell (Target Discovery Institute, University of Oxford) and Dr Simon Carr in our research group (Department of Oncology, University of Oxford).

(All declared in the respective sections within the thesis.)

I confirm that the other parts of the thesis I am submitting are wholly my own work, and no part has been accepted or is currently being submitted for any other degree or diploma in this university or elsewhere.

Koyo Harada

April 2023

Acknowledgement

Foremost, I would like to thank my supervisor, Professor Nick La Thangue, for his unfailing support and guidance throughout my DPhil project. I would also like to thank my fellow members of the NLT lab; especially my mentor Dr Simon Carr, who taught and helped me a lot with his invaluable scientific insight.

Additionally, I am sincerely grateful to Ezoe Memorial Recruit Foundation for their very generous funding; this project would not be achievable without their gratified support. I would also like to thank my collaborators; Dr Anastasia Samsonova and Dr Alexander Kanapin for bioinformatic analysis, Dr Iolanda Vendrell for mass spectrometry, and Professor Sarah Blagden for some insightful advices as my second supervisor.

In addition, special thanks go to my family (Itsumo no Minasama). I would especially like to thank my parents, Hideaki and Makiko, and my brother, Yuki, as well as my grandparents, for their endless love and support for the past four years and throughout my life. I would not be able to finish this work without you all. My sincere thanks also go to Megumi, for her patience, care, and love during tough times, and for always being my source of motivation and inspiration. I would also like to thank many other friends, whose unwavering support made my journey in Oxford more exciting and enjoyable than I could have ever imagined.

Last but not least, I would like to express my deepest gratitude to my favourite baseball team, Saitama Seibu Lions. Their successful match outcomes demonstrated a significant positive correlation with my productivity, and thus they made a noteworthy contribution to the success of this research by winning the NPB Pacific League title back-to-back in 2018 and 2019. Regrettably, no mention shall be made of their performance in subsequent seasons.

List of Figures

List of Tables

Abbreviations

A3SS	Alternative 3’ splice sites
A5SS	Alternative 5’ splice sites
ACPA	Anti-citrullinated protein antibody
AML	Acute myeloid leukaemia
AS	Alternative splicing
cDNA	Complementary DNA
ChIP	Chromatin Immunoprecipitation
CLIP	Crosslinking IP
CRISPR	Clustered regularly interspaced short palindromic repeats
CTD	C-terminal domain
DBD	DNA-binding domain
DEG	Differentially expressed genes
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DP	Dimerization Partner
E2F	E2 promoter binding factor
E2F1cr	E2F1 CRISPR-knockout
ECM	Extracellular matrix
EMT	Epithelial-mesenchymal transition
ES cells	Embryonic stem cells
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
FACS	Florescence-activated cell sorting
FDR	False discovery rate
GO	Gene Ontology
HA	Hemagglutinin
HDAC	Histone Deacetylase

hnRNP	Heterogeneous nuclear ribonucleoprotein particle
I/E ratio	Inclusion/Exclusion ratio
IL	Interleukins
IP	Immunoprecipitation
IP-MS	Immunoprecipitation-Mass spectrometry
ISE	Intronic splicing enhancer
ISS	Intronic splicing silencer
LZ	Leucine zipper
MHC	Major histocompatibility complex
MXE	mutually exclusive exon
NET	Neutrophil extracellular trap
NLS	Nuclear localization signal
PAD	Peptidyl arginine deiminase
PCR	Polymerase chain reaction
PI	Propidium iodide
pRB	Retinoblastoma protein
PRMT	Protein arginine methyltransferases
PSI	Percent spliced in
PTM	Post-translational modification
qPCR	Quantitative PCR
RA	Rheumatoid arthritis
RBP	RNA-binding protein
RGG	Arginine-glycine-rich
RI	Retained intron
RIP	RNA-immunoprecipitation
RNA	Ribonucleic acid
RNAP II	RNA polymerase II
RNA-seq	RNA sequencing
RNP	Ribonucleoprotein particle

RRM	RNA-recognition motifs
RS	Arginine-Serine-rich
RT	Room temperature
RT-qPCR	Quantitative Reverse Transcription PCR
SD	Standard deviations
SE	Skipped exon
SFPQ	Splicing factor proline- and glutamine-rich
SN	Staphylococcal nuclease
SR	Serine-Arginine-rich
SRE	Splicing regulatory element
SRSF	Serine-Arginine-rich splicing factor
SS	Splice site
SUMO	Small ubiquitin-like modifier
Tet-On	Tetracycline inducible
TRIBE	Targets of RNA-binding proteins Identified By Editing
TSN	Tudor staphylococcal nuclease
UV	Ultraviolet
WT	Wild type

Chapter 1

Introduction

1.1 Hallmarks of Cancer

Cancer is one of the most devastating diseases in the world. According to the World Health Organization, it constitutes the second leading cause of death worldwide, accounting for almost 10 million deaths, or nearly one in six deaths, in 2020 alone (Ferlay et al., 2020). The burden of cancer incidence and mortality has grown in recent years, not only reflecting the ageing and growth of the population but also being affected by some socio-economic factors (Sung et al., 2021). This is exemplified by a recent outbreak of the coronavirus disease 2019 (COVID-19), which adversely impacted the diagnosis and treatment of cancer worldwide (Siegel et al., 2023).

Cancer is a broad term and collectively refers to a large number of diseases that are often characterised as, but are not limited to, uncontrolled cell growth and invasion of healthy cells in the body. The key features or hallmarks found in cancer were originally summarised in a seminal paper by Hanahan and Weinberg (Hanahan & Weinberg, 2000). They initially described six hallmarks constituting the complexities of neoplastic disease: self-sufficiency in growth signals, insensitivity to anti-growth signals, evasion of apoptosis, tissue invasion and metastasis, limitless replicative potential and sustained angiogenesis (Hanahan & Weinberg, 2000), and later added another four properties: avoiding immune destruction, deregulating cellular energetics, tumour promoting inflammation, and genome instability and mutation (Hanahan & Weinberg, 2011). The authors consider the latter two characteristics (inflammation and genome instability) as ‘enabling characteristics’, to distinguish the molecular and cellular mechanisms through which cancer cells acquire the other hallmark capabilities. Most recently, four new enabling characteristics were added by Hanahan, reflecting the expanded understanding of human cancer in the last decade; unlocking phenotypic plasticity, non-mutational

epigenetic reprogramming, polymorphic microbiomes, and senescent cells (Hanahan, 2022). Here, each of these eight hallmarks and six enabling characteristics of cancer will be briefly reviewed, to conceptually understand the complex biology of cancer (Figure 1-1).

Self-Sufficiency in Growth Signals

For normal cells to re-enter the cell cycle from a quiescent state, mitogenic growth signals must be received. Cancer cells instead become independent of these signals and acquire a chronic proliferative state via various mechanisms. For example, cancer cells may produce the growth signalling molecules to stimulate themselves in an autocrine manner, or stimulate nearby normal cells to secrete the growth factors in a paracrine fashion. Additionally, deregulated expression or mutation of signalling components, such as the growth factor receptors on the cell surface or downstream proteins involved in intracellular signalling (e.g. B-Raf is mutated in up to 40% of human melanomas (Davies & Samuels, 2010)), could further enforce the mitogenic independence of cancer cells. Conversely, defects in the pathways antagonising growth signals also attribute to the proliferative state of the cells. (Hanahan & Weinberg, 2000, 2011)

Insensitivity to Anti-Growth Signals

As well as inducing and maintaining pro-growth signals for positive regulation, cancer cells have to overcome pathways and factors negatively regulating cell proliferation. Indeed, tumour suppressor genes involved in such inhibitory mechanisms are often mutated in tumour cells. The p53 protein and the retinoblastoma protein (pRB) are amongst the most commonly mutated or deactivated tumour suppressors (Rivlin et al., 2011; Vélez-Cruz & Johnson, 2017). In addition, cancer cells are also known to evade

'contact inhibition', which is usually observed in a dense population of normal cells where cell-to-cell contact negatively impacts further cell growth. (Hanahan & Weinberg, 2000, 2011)

Evasion of Apoptosis

Continuous exposure to pro-growth and stress stimuli found in the tumour environment usually activates apoptotic pathways in cells, and cancer cells need to evade these physiological defence mechanisms to progress to tumorigenesis. In most cases, tumour cells up-regulate the expression of anti-apoptotic factors (e.g. BCL-2, BCL-xL, BCL-w, and Mcl-1) and down-regulate pro-apoptotic factors (e.g. Bax, Bak, and Bim) (Adams & Cory, 2007; Labi et al., 2006). (Hanahan & Weinberg, 2000, 2011)

Tissue Invasion and Metastasis

Tumour cells often invade the nearby environment and eventually migrate from the primary tumour to another distant location to form a new, or secondary tumour. The latter process is referred to as metastasis and is often regarded as the most deadly aspect of cancer (Fares et al., 2020). The epithelial-to-mesenchymal transition (EMT), a physiological process involved in development and wound healing (Yilmaz & Christofori, 2009), is exploited by cancer cells; cellular adhesion-related proteins like E-cadherin are deactivated, whilst the enhanced expression of N-cadherin instead promotes cell migration, allowing the cancer cells to escape from their local environment (Curto et al., 2007). (Hanahan & Weinberg, 2000, 2011)

Limitless Replicative Potential

Normal cells can undertake only a limited number of cell divisions due to telomere shortening at the end of each round of DNA replication (Shammas, 2011). Cancer cells

overcome this restriction and can divide indefinitely by sustaining the length of chromosomal telomeres (Blasco, 2005; Shay & Wright, 2000). This is usually mediated by the expression of an enzyme called telomerase, which is usually inactivated in differentiated cells under physiological conditions. (Hanahan & Weinberg, 2000, 2011)

Sustained Angiogenesis

In solid tumours with a growing mass, enhanced levels of neovascularisation are required to supply the cells in the centre of the tumour with sufficient amounts of nutrients and oxygen, as well as provide the capability to remove metabolic wastes and carbon dioxide. Cancer cells, therefore, release some pro-angiogenic factors including vascular endothelial growth factor, whilst suppressing the expression of anti-angiogenic factors such as thrombospondin 1 (Kazerounian et al., 2008), to help with the chronic growth and sprouting of vasculature which is otherwise in a quiescent state. (Hanahan & Weinberg, 2000, 2011)

Deregulating Cellular Energetics

Deregulation in energy metabolism is another key hallmark in cancer, given the increased demand for 'fuel' from the rapidly growing and dividing tumour cells. Unlike normal cells which partly utilise the citric acid cycle in aerobic conditions, the metabolism of cancer cells is reprogrammed to hugely depend on glycolysis (Schiliro & Firestein, 2021), a process that is less efficient in ATP production but requires much less oxygen. Increased dependency on glycolysis results in the channelling of glycolytic intermediates into cellular biosynthetic pathways, which further promotes cell proliferation and growth by supporting the biosynthesis of nucleosides and amino acids, and thus of macromolecules and organelles. (Hanahan & Weinberg, 2000, 2011)

Avoiding Immune Destruction

Normally, components of the immune system, including $CD8^+$ cytotoxic T lymphocytes and natural killer cells, target abnormal or damaged cells and destroy them before they can develop into cancer (R. Kim et al., 2007; Teng et al., 2008). Tumour cells need to evade this targeted destruction, and they do this by up-regulating the expression of the immune checkpoint inhibitor PDL1 to become less immunogenic against T cells (C. Sun et al., 2018), or they may directly secrete small inhibitory molecules against the immune response, such as TGF-β. (Hanahan & Weinberg, 2000, 2011)

Tumour Promoting Inflammation

In addition to avoiding immune destruction, cancer cells can also hijack the immune system to instead promote their growth and survival. This pro-tumour inflammatory response typically involves the secretion of bioactive signalling molecules to the nearby environment affecting other hallmark-facilitating pathways; such as growth or survival factors to promote proliferation or inhibit apoptosis, pro-angiogenic factors to enhance angiogenesis, and extracellular matrix modifying factors to alter the EMT profile. (Hanahan & Weinberg, 2000, 2011)

Genome Instability and Mutation

As described above, the acquisition of cancer hallmarks often involves mutation or deactivation in the genome. Successive genomic alterations do not usually occur in normal cells thanks to surveillance mechanisms including DNA damage repair, DNA polymerase proofreading and proapoptotic signals (Jackson & Bartek, 2009; Kastan, 2008; Sigal & Rotter, 2000). Defects in these defence systems would therefore lead to

substantial genomic instability with high mutation rates (Hanahan & Weinberg, 2000, 2011).

Unlocking Phenotypic Plasticity

During organogenesis, a series of developmental and differentiation steps occur in cells which permit them to form functional and organised tissues. In the course of this process, many precursor cells become terminally differentiated, whereby they become fully committed to their cell fates in their respective tissues and permanently lose the capability of cell division (Yuan et al., 2019). In tumours, however, cells can undergo molecular and phenotypic changes to unlock this restriction and gain proliferative potential as well as broader phenotypic plasticity beyond the spectrum defined by the tissue of origin. There exist several ways to escape from the state of terminal differentiation. Dedifferentiation refers to the process whereby a mature cell returns to a progenitor-like state (Barker et al., 2009; Perekatt et al., 2018; Shih et al., 2001). Blocked differentiation, on the other hand, occurs when progenitor cells are prevented from advancing further into a non-proliferative state during differentiation and instead retain their proliferative ability(de Thé, 2018; He et al., 1999; Warrell Jr et al., 1993); Finally, transdifferentiation describes the phenomenon whereby a progenitor cell that has already committed to one differentiation path switches into a different developmental lineage (Yuan et al., 2019). (Hanahan, 2022)

Non-mutational Epigenetic Reprogramming

In addition to the mutations across the human genome, epigenetically regulated changes in gene expression can also contribute to the acquisition of cellular functional capabilities during tumour progression. Potentially induced by microenvironmental factors such as

hypoxia, reprogramming in the cancer cell epigenetic profile may result in intra-tumoral heterogeneity and deregulation of accessory and stromal cells (Hanahan, 2022).

Polymorphic Microbiomes

In recent years, a growing body of evidence suggests that microbiota residing in various tissues have a profound influence on a multitude of biological and cellular events in both physiological and pathological conditions (Thomas et al., 2017). Cancer is no exception, and a close association between the microbiological profile and cancer phenotype has been reported (Dzutsev et al., 2017; Helmink et al., 2019). Although there remain a lot of unanswered questions, microbiomes seem to play both a pro- and anti-tumour role by modulating tumour cell activity. For example, the gut microbiome was shown to have a substantial impact on the pathogenesis of colon cancer, whereby bacterial-derived toxic molecules induce the DNA damage response, or disruption of damage checkpoint factors, leading to genome instability in colon epithelial cells (Okumura et al., 2021; Pleguezuelos-Manzano et al., 2020; Sears & Garrett, 2014). Bacteria can also produce factors that influence the adaptive and innate immune systems, which would result in an activated immune response with increased expression of various chemokines and cytokines. (Hanahan, 2022)

Senescent Cells

Cell senescence is the irreversible and stable process of cells becoming unable to proliferate despite optimal growth conditions. This is long regarded as part of a surveillance mechanism against tumour initiation and progression, where stress stimuli induce cells with existing DNA damage to undergo senescence, before they have a chance to develop into neoplasia (Birch & Gil, 2020; Gorgoulis et al., 2019; S. Lee & Schmitt,

2019). An accumulating body of evidence indicates, however, that senescent cells can also promote cancer development and malignancy progression in some specific cases (Kowald et al., 2020; S. Lee & Schmitt, 2019; B. Wang et al., 2020). The major mechanism of this tumour-promoting effect is referred to as the senescence-associated secretory phenotype (SASP) where; in a paracrine manner, senescent cells can provide factors related to proliferative signalling, avoidance of apoptosis, induction of angiogenesis, stimulation of invasion and metastasis, and suppression of tumour immunity (B. Wang et al., 2020). (Hanahan, 2022)

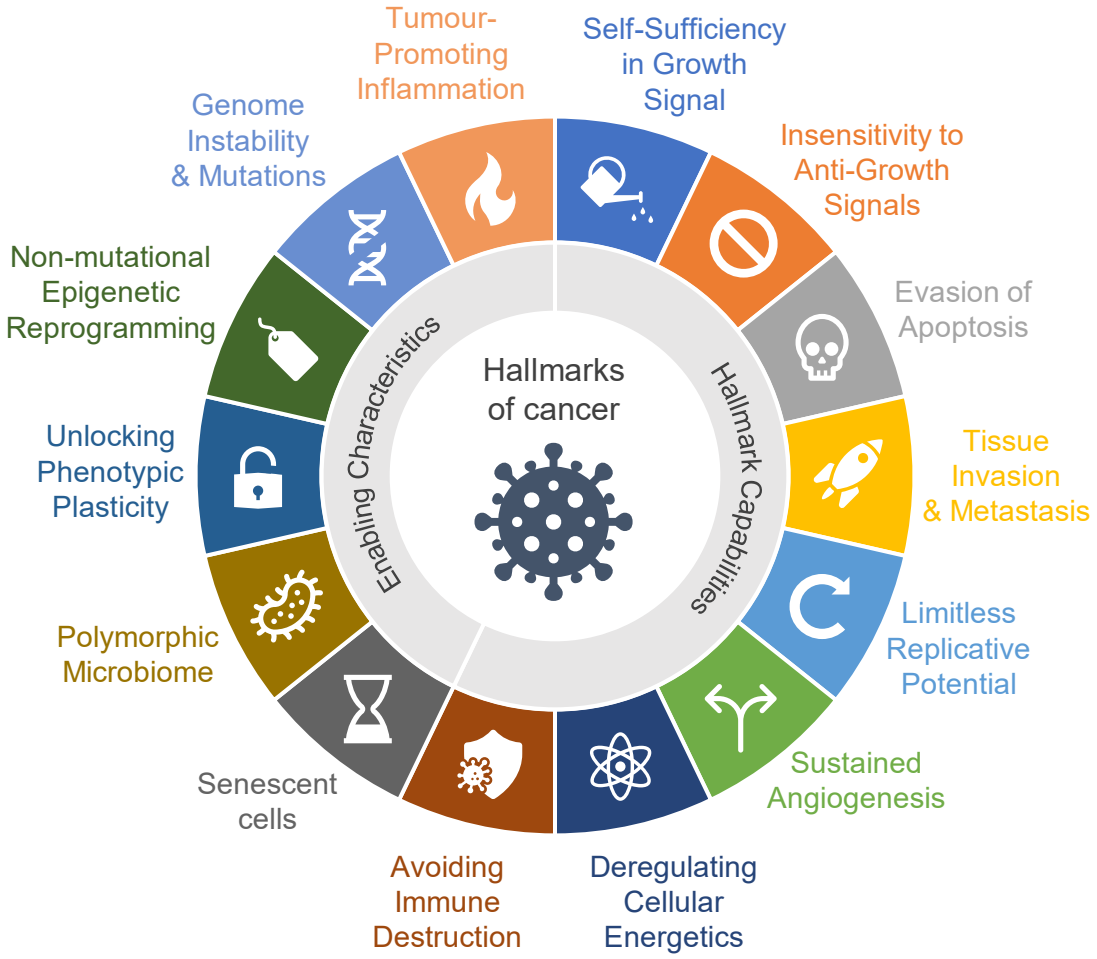


Figure 1-1 Hallmarks of Cancer

The illustration outlines eight hallmark capabilities and six enabling characteristics which together conceptualise the complexity of cancer phenotypes and genotypes. Adapted from Hanahan, 2022; Hanahan & Weinberg, 2000, 2011.

1.2 Biological Functions of E2F1

1.2.1 Properties of the E2F family

The E2F family of transcription factors consists of eight members. They are known to regulate the expression of diverse sets of genes and play pivotal roles in a range of biological processes, including cell growth and cell cycle progression, apoptosis, differentiation, DNA damage repair, metabolism and angiogenesis (Biswas & Johnson, 2012; Blanchet et al., 2011; DeGregori & Johnson, 2006; Ren et al., 2002; Weijts et al., 2012). In general, the E2F proteins can be divided into two subfamilies according to their transcription regulation functions: activators (E2F1-3a/b) and repressors (E2F4-8) (Roworth et al., 2015). The former group of E2F members, when over-expressed, can potently promote transcription, and their localisation to the promoter regions of E2F-target genes usually coincides with transcriptional activation in the G_1/S phase of the cell cycle (Iaquinta & Lees, 2007). On the other hand, binding of the latter E2F family members to their target genes was correlated with transcriptional repression during the G_0/G_1 phase, and over-expression did not result in transcriptional activation (Iaquinta & Lees, 2007).

There are multiple evolutionarily conserved domains possessed by the E2F family of proteins (Figure 1-2). The highest level of homology is found in the winged-helix DNA binding domain (DBD); E2F1-6 have a single DBD, whilst E2F7 and E2F8 contain two distinct DBDs, which is the reason why these two family members are regarded as ‘atypical’ (DeGregori & Johnson, 2006; Morgunova et al., 2015). E2Fs, as transcription factors, can bind to specific DNA motifs in the proximal promoter regions of their target genes via the DBD. The consensus E2F-binding motif of TTTSSCGC (S is either G or

C) was identified *in vitro* by the CASTing experiment which involves repetitive immuno-precipitation and PCR (Tao et al., 1997). However, with recent technological innovations in genome-wide profiling, including ChIP-seq, it has become apparent that this motif is not as strict *in vivo* as previously demonstrated *in vitro* (Bieda et al., 2006; Cao et al., 2011; X. Xu et al., 2007).

In addition to the DBD, E2F1-6 contain a dimerization partner (DP) binding domain, which consists of a leucine zipper (LZ) and a marked box (MB) domain and is located downstream of the DBD (Figure 1-2). The DP domain allows E2F1-6 to form a heterodimeric complex with one of three DP proteins (DP1, DP2/3, or DP4), generating a functional transcription factor complex capable of efficiently associating with the promoters of target genes (L. R. Bandara et al., 1993; DeGregori & Johnson, 2006; Milton et al., 2006; Ormondroyd et al., 1995; Y. Zhang & Chellappan, 1995). DP1 and DP2/3 are both known as transcriptional activators, whereby their dimerization with an E2F family member usually results in increased gene expression (L. Bandara et al., 1994; de la Luna et al., 1996; Girling et al., 1993; Ormondroyd et al., 1995). DP4 reportedly has a distinct function to suppress transcription by the E2F/DP complex (Ingram et al., 2011). The functional differences of DP isozymes in this complex are not fully understood. For example, with E2F1 it was initially observed that switching between DP1 and DP2 as a dimeric partner had little effect on the DNA-binding site specificity of the E2F1/DP complex (Tao et al., 1997). However, a more recent study has revealed that depletion of DP1 inhibited the E2F1-dependent activation of the pro-growth gene CDC6, and prevented cell cycle progression, whilst the expression of the pro-apoptotic ARF gene and apoptotic profile of cells was not affected (Komori et al., 2018). This indicates a potential role of DP proteins in channelling E2F1 into different signalling pathways.

The activity of the E2F1-5/DP heterodimeric complex is regulated by the pocket proteins (also called "retinoblastoma family proteins") including pRB, p107 and p130. The binding site for the pocket proteins can be found within the C-terminal transactivation domain of E2F1-5 (Figure 1-2), meaning that pocket protein binding to E2Fs can suppress their transcriptional activities by physically blocking key residues in this domain (Iaquinta & Lees, 2007; D. Xie et al., 2021). Moreover, pRB and other pocket proteins are known to facilitate the recruitment of co-repressor complexes to the E2F-target promoter region (Iaquinta & Lees, 2007; D. Xie et al., 2021). The components of this inhibitory complex are typically chromatin modifiers and remodelling proteins, such as BRM/BRG1, histone deacetylases (HDACs), RBP1, DNMT1, CtIP, CtBP, HPC2, mSin3B, SUV39H, and PRMT5 (DeGregori & Johnson, 2006). Furthermore, several studies support the view that the pocket proteins selectively bind to different E2F members; the activating E2Fs (E2F1-3) are exclusively associated with pRB whilst the others (E2F4 and E2F5) can also interact with p107 and p130 *in vitro* (Liban et al., 2016, 2017), although *in vivo* this observation has been questioned (Toki et al., 2014). pRB is known as a master regulator of the cell cycle and is an archetypal tumour suppressor, which is reported to be mutated or functionally inactivated in most types of cancer (Vélez-Cruz & Johnson, 2017).

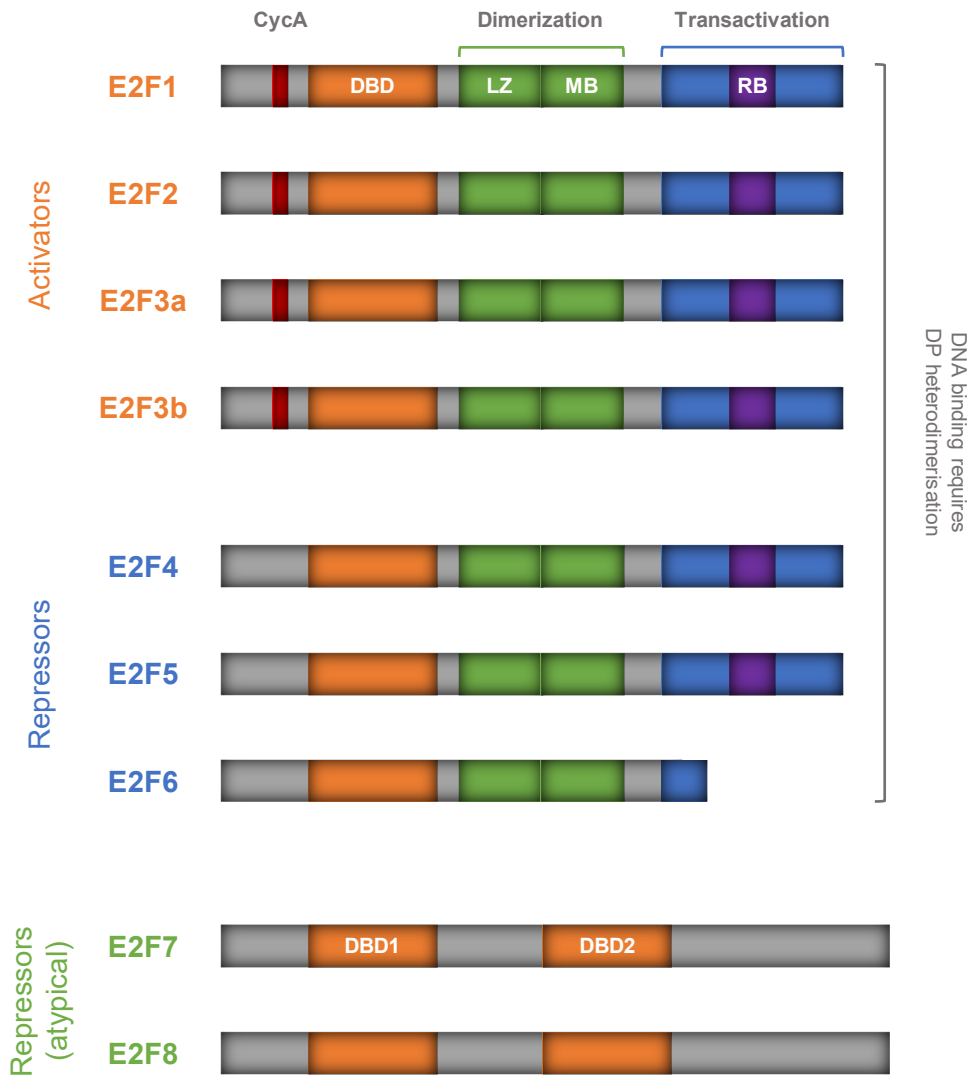


Figure 1-2 Properties of the E2F Family of Transcription Factor.

The schematic diagram illustrating the structures of the eight members of the E2F transcription factor family. DBD: DNA binding domain, LZ: Leucine zipper, MB: Marked box, RB: pRB binding site, CycA: Cyclin A binding site.

1.2.2 E2F1 and cell cycle control

E2F1 is the first discovered member of the E2F family of transcription factors, when it was cloned as a result of screening for pRB binding partner proteins (Helin et al., 1992; Kaelin et al., 1992; Shan et al., 1992). The biological function of E2F1 has been most characterised in the cell cycle regulation (Figure 1-3), especially at the restriction point (R-point) during the G1 phase. After this checkpoint, the cell becomes fully committed to the cell cycle and no longer requires growth factors or mitogens to complete the remaining cell cycle stages (Roworth et al., 2015). In the G_1 phase, mitogenic signalling stimulates the PI3K and/or MAPK pathways, leading to the activation of MYC (J. Zhu et al., 2008). This transcription factor then promotes the expression of various cell cycle-related genes including E2F1-3, cyclin D and E, CDK4, and CDC25 (Mateyak et al., 1999; Santoni-Rugiu et al., 2000). Subsequently, cyclin D-CDK4/6 and cyclin E-CDK2 phosphorylate pRB at multiple threonine and serine residues in a sequential manner (Malumbres & Barbacid, 2001; Rubin, 2013). This phosphorylation is the key mechanism causing pRB to be released from the E2F/DP heterodimer and permits activation of E2F1. Phosphorylation of pRB at T356/373 and S788/795 reportedly decreases its binding to E2F1 by 10-fold (Burke et al., 2014). Activated E2F1 then promotes the expression of its target genes to progress the cell cycle. Up-regulated genes include E2F1 itself, E2F2-3, cyclin E, and MYC (Araki et al., 2003). This positive feedback loop is thus believed to play an important role in accumulating active E2F1 at late G_1 and marking an irreversible commitment to cell division (Roworth et al., 2015).

When cells enter S phase, the activity of E2F1 needs to be switched off to retain the directionality of the cell cycle and ensure a single replication of DNA per cycle. This is mediated by a negative feedback loop coordinated by the delayed expression of a subset

of E2F1 target genes (Roworth et al., 2015). Briefly, E2F1 up-regulates the expression of cyclin A at the beginning of S phase, which allows the cyclin A/CDK2 complex to bind to E2F1/DP1 and phosphorylate DP1. This phosphorylation event significantly reduces the DNA binding affinity of DP1, resulting in the dissociation of the E2F1/DP1 from its target promoter and subsequent transcriptional suppression (Schulze et al., 1995; M. Xu et al., 1994; H. S. Zhang et al., 2000). Cyclin A/CDK2 can further enforce the negative feedback loop by promoting the degradation of its inhibitor p27 (Bertoli et al., 2013). In addition, E2F1 up-regulates the expression of the SCF regulatory subunit SKP2, which in turn regulates the activity of ubiquitin ligase and subsequent degradation of E2F1 itself (Schulze et al., 1995; M. Xu et al., 1994; H. S. Zhang et al., 2000). Furthermore, the expression of repressor E2F family members E2F7 and E2F8 is up-regulated during S phase and further contributes to the suppression of E2F-target genes (Logan et al., 2005).

The expression of genes involved in this negative feedback loop is carefully controlled in a timely manner. For example, the expression of cyclin E peaks at late G_1/S, whilst cyclin A expression begins during S phase and peaks at G_2 (Roworth et al., 2015). Our understanding of the underlying mechanism in this 'delayed' gene expression is still limited, but it was suggested that genes with delayed expression, such as cyclin A, contain atypical E2F1 binding sites in their promoter region, which makes them less sensitive to gene activation by E2F1 (Schulze et al., 1995; M. Xu et al., 1994). In addition, pRB facilitates the recruitment of a distinct set of repressor proteins for the 'delayed' genes; a pRB-SWI/SNF repressor complex at the promoter of these genes is more stable and less responsive to pRB phosphorylation status in comparison to an HDAC-pRB-hSWI/SNF complex at the promoter of genes peaking at late G_1/S (H. S. Zhang et al., 2000). After entering S phase, BRG1, a component of the SWI/SNF complex, is phosphorylated which

finally causes release of the complex from the promoter, resulting in the delayed expression of genes like cyclin E (H. S. Zhang et al., 2000).

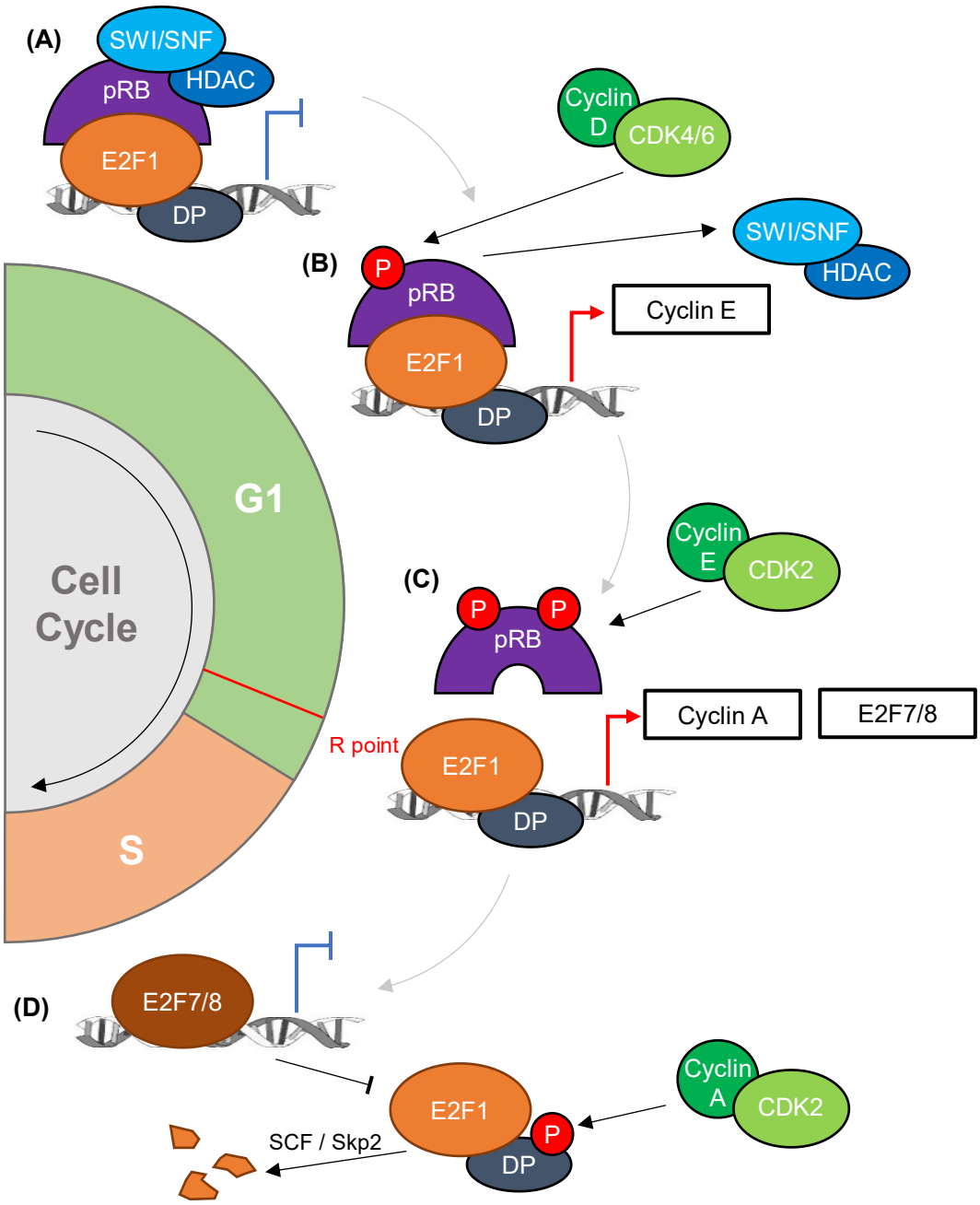


Figure 1-3 Cell Cycle and E2F1.

Diagram shows the role that E2F1 plays during the cell cycle progression from G1 to S phase. (A) In the early G1 phase, the pRB firmly binds to E2F1/DP and recruits the suppressor complex containing histone modifier or chromatin remodelling factors (e.g. SWI/SNF and HDACs) to inhibit the expression of pro-growth genes. (B) Upon the mitogenic stimuli, CDK4/6-cyclin D phosphorylates pRB and dissociates several pRB-binding factors, resulting in the relieved suppression for the genes such as cyclin E. (C) Cyclin E in turn binds to CDK2 to phosphorylate pRB further, leading to the release of E2F1/DP. More genes are expressed to drive the cell cycle and pass through the restriction point into the S phase. (D) In S phase, E2F1 is no longer required and becomes degraded, usually via the phosphorylation by CDK2-Cyclin A and subsequent ubiquitination-dependent pathway.

1.2.3 E2F1 and apoptosis

In addition to its crucial role in cell cycle regulation, E2F1 has also been implicated as a critical component regulating cell death by apoptosis. Prolonged or over-expression of E2F1 reportedly induces G1 checkpoint activation and apoptosis in response to various cellular stress and DNA damage, both *in vitro* and in mouse models (Holmberg et al., 1998; Pierce, Fisher, et al., 1998; Pierce, Gimenez-Conti, et al., 1998; Pierce et al., 1999; X. Q. Qin et al., 1994; Shan & Lee, 1994). In most cases, such E2F1-mediated apoptosis can be linked to the activation of p53 in response to DNA damage. For example, several studies have demonstrated that the transcriptional activation of the $p14^{ARF}$ gene is regulated by E2F1 (Bates et al., 1998; Robertson & Jones, 1998). The $p14^{ARF}$ protein is a known co-activator of p53 and can inhibit the Mdm2-dependent degradation of p53. E2F1 is also thought to contribute to the accumulation of p53 by regulating the phosphorylation-mediated stabilisation of this protein through the ATM/Nbs1/Chk2 pathway (Powers et al., 2004; Rogoff et al., 2004). In addition, another study identified several co-factor genes of p53 including ASPP1, ASPP2, JMY, and TP53INP1 which were E2F1-targets, and whose expression was directly governed by the transcriptional activity of E2F1 (Hershko et al., 2005). Moreover, the weight of evidence suggests that E2F1 can also promote apoptosis in a p53-independent fashion. Here, E2F1 regulates the expression of its target genes, such as p73, APAF1, Caspases, and BH3-only, that subsequently activate the pro-apoptotic pathways without help from p53 activity (Furukawa et al., 2002; Hershko & Ginsberg, 2004; Irwin et al., 2000; Moroni et al., 2001; Nahle et al., 2002; Stiewe & Pützer, 2000).

1.2.4 Post-translational modifications of E2F1

The seemingly opposing roles that E2F1 undertakes to regulate cell proliferation and apoptosis were discussed above. In addition, cumulative evidence also points to the relevance of E2F1 activity in an even wider spectrum of biological processes, such as differentiation, DNA damage repair, inflammation, and metabolism (Biswas & Johnson, 2012; Blanchet et al., 2011; Ghari et al., 2016; Ren et al., 2002). Given such intricate nature and significance of E2F1 biology, it would be of immense importance to understand the regulatory mechanisms of switching E2F1 activity into diverse cellular functions. Post-translational modifications (PTMs) are central to the epigenetic regulation of proteins and affect many aspects of target protein biology from subcellular localization to protein-protein interactions, sequence-specific DNA binding, protein stability, and transcriptional regulatory activity (G. Duan & Walther, 2015; Ramazi & Zahiri, 2021). As a result, PTMs are integral to various cellular processes as well as a number of pathogenic conditions. Transcription factors are widely regulated by a range of PTMs, and E2F1 is no exception; overwhelming evidence supports the idea that various PTMs act as functional switches of E2F1 activity (Figure 1-4) (Carr et al., 2015; Roworth et al., 2015). Some of these E2F1 modifications and their biological outcomes are discussed below.

Phosphorylation

Phosphorylation of E2F1 has been most studied in the context of the DNA damage response, where E2F1 rapidly becomes phosphorylated after DNA damage detection by multiple sensor kinases (Carr et al., 2015; Roworth et al., 2015). Two distinct serine (S) residues were identified as target sites for phosphorylation after DNA damage; S31 is targeted by the DNA damage sensing kinases ATM/ATR, and S364 is targeted by the

tumour suppressor kinase CHK2 (Stevens et al., 2003; B. Wang et al., 2004). The former mark can reportedly be recognised by the 14-3-3τ protein, which stabilises E2F1 by inhibiting its ubiquitination (B. Wang et al., 2004). The modification mediated by CHK2 was shown to positively regulate the stability of the E2F1/pRB complex by increasing the binding affinity of E2F1 to pRB (Carnevale et al., 2012). Interestingly, this E2F1/pRB complex, which usually suppresses E2F transcriptional activity, seems to act to promote the expression of pro-apoptotic genes including p73 and NOXA in the context of a DNA damage response (Carnevale et al., 2012). Later, it was revealed that caspase-dependent cleavage of pRB underpins the key mechanism, where the association between E2F1 and the remaining p68 fragment of pRB remains intact and together activates the expression of these genes (Bertin-Ciftci et al., 2013). This exemplifies that PTMs can act as an additional layer of regulation to control E2F1 biology, as well as highlighting the context-dependent outcomes of the E2F1-pRB interaction.

In addition to the DNA damage response, there is a growing body of research suggesting that phosphorylation of E2F1 may play a role in apoptotic pathways in cancer cells too. Another target residue of E2F1 for phosphorylation, S375, was discovered in colorectal cancer cells (Morris et al., 2008; Zhao et al., 2013). This CDK8-dependent modification can directly modulate the transcriptional activity of E2F1, resulting in the inhibition of E2F1-induced apoptosis (Zhao et al., 2013). Furthermore, a recent study has proposed that the inhibition of CDK8 was shown to sensitise colorectal cancer cells to radiotherapy in an E2F1-dependent manner (B. Chen et al., 2020), highlighting the potential significance of E2F1 phosphorylation in cancer therapy.

Acetylation

In addition to phosphorylation, DNA damage can induce another PTM of E2F1, acetylation, at several lysine residues (K117, K120, and K125). This is mediated by multiple acetyl-transferases including p300/CBP, PCAF, and Tip60 (Budhavarapu et al., 2012; Galbiati et al., 2005; Nagy & Tora, 2007; Taubert et al., 2004). These target lysine residues of E2F1 are in close proximity to its DBD and thus the acetylation of E2F1 can reportedly modulate its binding affinity to the promoter of target genes including the pro-apoptotic p73 gene, highlighting the crosstalk between phosphorylation and acetylation in E2F1-mediated apoptosis (Roworth et al., 2015). A recent study further revealed that acetylated E2F1, in association with p300/CBP via the bromodomain, can be directly recruited to the site of DNA double-strand breaks, where p300/CBP catalyses the acetylation of various lysine residues of histone H3, potentially extending the regulatory environment under E2F1 control to chromatin remodelling (Manickavinayaham et al., 2019).

SUMOylation

A previous study established that E2F1 has a unique role in metabolic regulation, where the E2F1-pRB complex was identified as a transcriptional repressor for various genes involved in oxidative metabolism (Blanchet et al., 2011). More recently, it was claimed that E2F1 can respond to oxidative stress by promoting cell cycle arrest at the G_1/S phase checkpoint, and this response was attributed to the direct SUMOylation of lysine 266 in E2F1 (Graves et al., 2020). Such modification is mediated by an enzyme called SUMO2, and in unperturbed cells, this reaction is inhibited by the deSUMOylating enzyme SENP3. Increased levels of oxidative stress can release E2F1 from SENP3, leading to the accumulation of SUMOylated E2F1 with a unique profile of transcriptional activity,

specifically promoting the expression of the cell cycle inhibitors CDKN1A/B (Graves et al., 2020). In addition, another study demonstrated that E2F1 can also be SUMOylated by the SUMO E1 enzyme, which enhances the E2F1 binding affinity to the EZH2 promoter and promotes increased expression of the gene (Du et al., 2020). It is intriguing that the same PTM (namely SUMOylation) on the same target protein (namely E2F1), when mediated by different enzymes, can result in completely distinct biological outcomes, though the interplay between these two SUMOylating enzymes still needs to be investigated.

Ubiquitination

After the G_1 to S phase transition, the activity of E2F1 needs to be switched off as part of a pivotal mechanism to ensure the directionality and commitment to the cell cycle. This is mainly regulated by the expression of the ubiquitin ligase subunit SKP2, whose expression peaks during S phase. SKP2 contributes to the ubiquitination of E2F1 as a component of the SCF^{skp2} (Skp2 (S phase kinase binding protein 2)-CDC53 (Cullin)-F-box) E3 ligase complex, and E2F1 ubiquitination subsequently results in proteasome-mediated degradation of the protein (Marti et al., 1999). The temporal expression of SKP2 is known to be regulated by the repressor E2F family members E2F7 and E2F8, which suppress SKP2 gene expression prior to cells entering S phase (Roworth et al., 2015). Ubiquitination of E2F1 is also involved in the DNA damage response and apoptosis. One study demonstrated that the E3-ubiquitin ligase cIAP mediates the K63-poly-ubiquitination of E2F1 at lysine residues K161 and K164 under conditions of DNA damage (Glorian et al., 2017). In contrast to the SKP2-mediated reaction, this K63-linked ubiquitination facilitates the stabilisation and transcriptional activation of E2F1, and site-specific mutagenesis of the K161/164 sites caused enhanced expression of genes

including CCNE, TP73 and APAF1 (Glorian et al., 2017). Interestingly, inhibition or silencing of the methyltransferase PRMT1 was shown to reduce this cIAP-mediated ubiquitination, indicating a potential interplay between different modifications during the DNA damage response.

Methylation

Protein methylation is one of the most common PTMs in the human proteome, with both lysine and arginine methylation of E2F1 previously reported. For lysine methylation, the K185 residue of E2F1 was shown to be the primary target of the methyltransferase SET9/SETD7, which is counteracted by the activity of lysine-specific demethylase 1 (LSD1) (Kontaki & Talianidis, 2010; Q. Xie et al., 2011). Here, the SET9-dependent lysine methylation of E2F1 was demonstrated to suppress the phosphorylation and acetylation events on E2F1 in response to DNA damage in p53-/- H1299 lung carcinoma cell line (Kontaki & Talianidis, 2010). Instead, SET9 over-expression was found to increase the ubiquitination of E2F1, which inhibits the accumulation of E2F1 through the ubiquitin proteasome pathway. As a result, E2F1 transactivation activity on its pro-apoptotic target genes like p73 is significantly reduced, leading to the suppression of E2F1-dependent apoptosis (Kontaki & Talianidis, 2010). Paradoxically, another study later claimed that, in response to genotoxic agents, SET9-dependent lysine methylation at the same K185 residue was linked to an increased level of apoptosis (Q. Xie et al., 2011). Here, the authors demonstrated that K185 methylation can stabilise E2F1 protein levels in WT and p53-/- U2OS cells lines, resulting in the enhanced expression of E2F1-target genes including p73 and Bim, and promoting apoptosis (Q. Xie et al., 2011). These seemingly conflicting results probably arose from the difference in cell types, p53 state, and DNA damage reagents, but still constitute the object of future studies.

E2F1 is also methylated at arginine residues (R109, R111, and R113) via the activity of two protein arginine methyltransferases (PRMTs); PRMT1 targets R109 for asymmetric di-methylation, whilst PRMT5 symmetrically di-methylates R111 and R113 (Cho et al., 2012; Zheng et al., 2013). Of particular importance is that these two modifications result in the engagement of E2F1 into distinct biological pathways with opposing outcomes. The asymmetric methylation at R109 mediated by PRMT1 was found to increase in response to DNA damage. This not only increases E2F1 levels by stabilising the protein, but also results in the enhanced expression of downstream pro-apoptotic E2F1-target genes including p73 and APAF1 (Zheng et al., 2013). In accordance with these findings, DNA damage-induced growth inhibition was not observed in the cells treated with PRMT1 siRNA or expressing the PRMT1-driven methylation defective mutant of E2F1 (R109K: R109 was replaced by lysine) (Zheng et al., 2013). In contrast, PRMT5-dependent symmetric dimethylation of E2F1 arginine residues R111 and R113 are reported to favour cell proliferation and increase the turnover of this transcription factor (Cho et al., 2012; Zheng et al., 2013). Cells depleted of PRMT5 or expressing the PRMT5-driven methylation defective mutant of E2F1 (R111/R113K: R111/R113 were replaced by lysine) were shown to have reduced cell growth and increased sensitivity to DNA damage agents, suggesting PRMT5-mediated methylation may ensure cell cycle commitment through regulating the activity of E2F1 (Zheng et al., 2013). Interestingly, PRMT5- and PRMT1-mediated methylation events occur in close proximity and were demonstrated to be mutually exclusive, whereby methylation at one site seems to inhibit the binding of the other PRMT to E2F1 (Cho et al., 2012; Zheng et al., 2013). In addition, in cells undergoing an unperturbed cell cycle, cyclin A was found to directly interact with E2F1, resulting in the inhibition of PRMT1 binding and augmenting PRMT5-regulated

methylation at R111/R113 (Cho et al., 2012; Zheng et al., 2013). E2F1 activity in its downstream biological processes of proliferation or apoptosis is therefore regulated by the precise balance of activity between PRMT5 and PRMT1.

Whilst no reader proteins for the PRMT1-mediated E2F1 methylation mark have been identified to date, the transcriptional co-activator protein, p100/Tudor staphylococcal nuclease (TSN), was identified to specifically recognise symmetrically di-methylated arginine residues of E2F1 targeted by PRMT5 (Zheng et al., 2013). p100/TSN is an evolutionarily conserved protein which is characterised by an N-terminal tandem repeat of four staphylococcal nuclease (SN)-like domains (SN domains) and a C-terminal fusion of a Tudor domain with a partial SN domain (Gutierrez-Beltran et al., 2016; Su et al., 2015). Whilst initially discovered as a transcriptional co-activator, the role of p100/TSN has been expanded into other aspects of gene expression regulation, including RNA splicing, interference, stability, and editing, as well as in the regulation of protein and lipid homeostasis (Ochoa et al., 2018). In the context of E2F1 biology, p100/TSN was originally shown to co-localise with symmetrically methylated E2F1 in a PRMT5-dependent manner, where this protein acts as a transcriptional co-factor to activate the expression of proliferation-related E2F1-target genes (Zheng et al., 2013). However, the pro-growth effect of PRMT5-mediated methylation of E2F1 is not limited solely to cell proliferation. Indeed, PRMT5-dependent methylation of E2F1 was later shown to extend the gene regulatory activity of the protein to RNA processing, and this role was also dependent on p100/TSN (Roworth et al., 2019). Here, the p100/TSN protein facilitates the recruitment of a large body of RNA and components of the splicing machinery to E2F1, and thereby allows E2F1 to regulate the expression of a subset of its target genes at the level of alternative splicing (Roworth et al., 2019). In addition, another study

demonstrated that PRMT5 can upregulate the transcription of genes related to cell migration and invasion in an E2F1-dependent manner in cancer cells (Barczak et al., 2020). More recently, the E2F1-PRMT5 axis was also established to regulate the expression of numerous long non-coding (lnc) RNA species in cancer cells, some of which encode for MHC class I associated peptides and can be presented to the immune system (Barczak et al., 2023). These studies emphasise the significance of arginine methylation in substantially expanding the genomic landscape under E2F1 control.

Citrullination

Not only can arginine residues in proteins be modified by methylation, but they can also undergo citrullination, another form of PTM which involves a direct conversion from arginine to citrulline (Harada et al., 2023). E2F1 is one of many target proteins that undergo this modification, as evidenced by Ghari et al.; PAD4 directly interacts with E2F1 and mediates citrullination of this transcription factor both *in vitro* and in cells (Ghari et al., 2016). Predominant sites of E2F1 citrullination were identified by mass spectrometry as arginine R109 and R127, whilst data from the use of site-specific mutants indicated R111 and R113 could also be involved (Ghari et al., 2016). Given the connection of citrullination and PAD4 with various settings of inflammation (Ciesielski et al., 2022), it was unsurprising that this modification was implicated to drive E2F1 activity towards an inflammatory pathway in cells (Ghari et al., 2016). Citrullinated E2F1 was demonstrated to bind to the promoters of pro-inflammatory genes like TNFα and IL-1β and promote their expression (Ghari et al., 2016). The authors also suggested that this citrullination mark had an important role in regulating acetylation at neighbouring lysine residues, and the combination of these two marks facilitated the recruitment of the inflammatory-related bromodomain protein BRD4 to E2F1 (Ghari et al., 2016). Given

this interplay between citrullination and acetylation, as well as the fact that PAD4 shares the same target arginine residue (R109) with PRMT1, citrullination may act to interconnect between multiple PTMs and regulate biological outcomes together in an orchestrated manner.

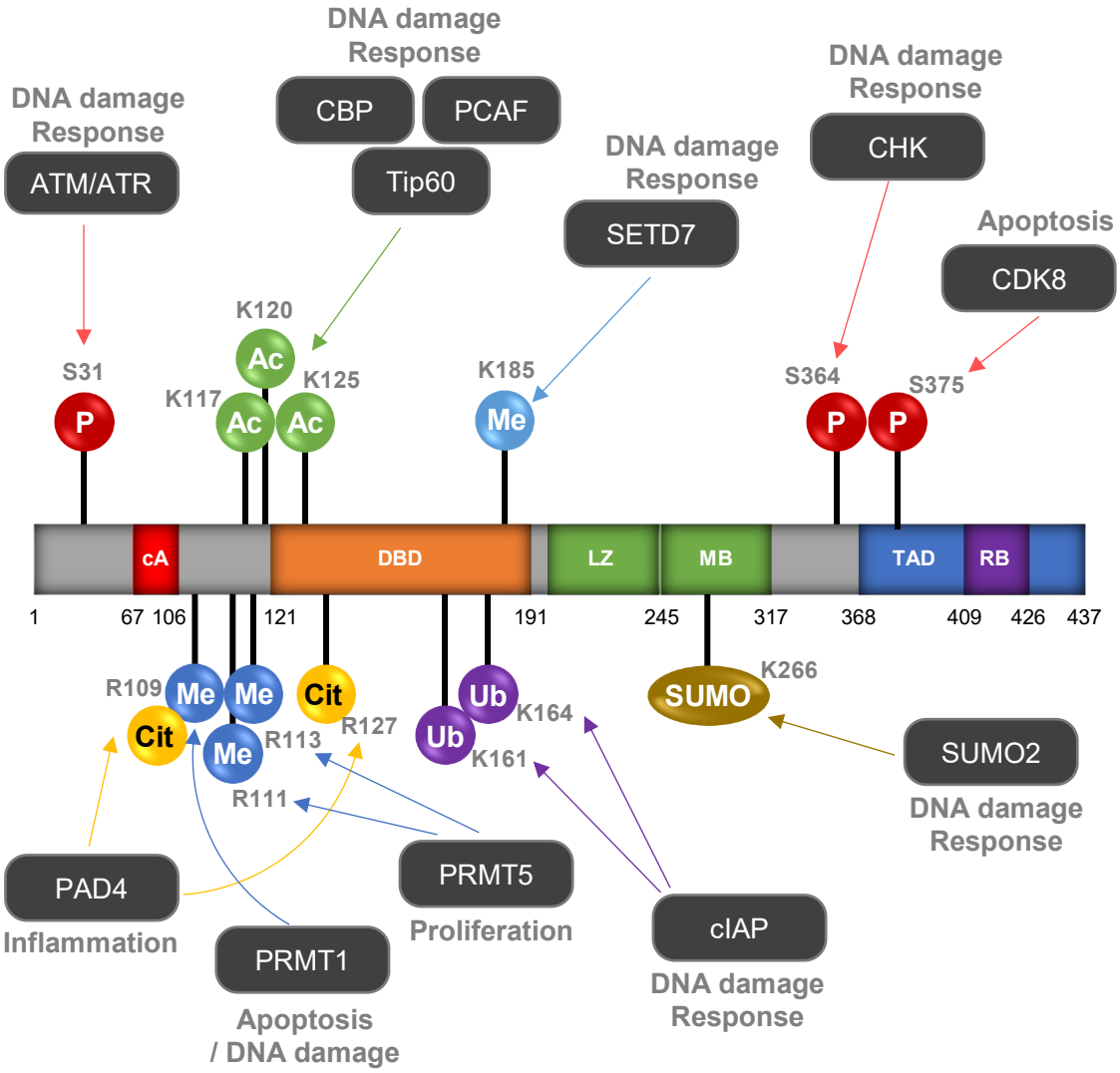


Figure 1-4 Post-Translational Modifications of E2F1.

Diagram of E2F1 showing its post-translational modifications. P: phosphorylation, Ac: acetylation, Me: methylation, Cit: citrullination, Ub: ubiquitination, SUMO: SUMOylation. Factors delivering PTMs shown in the dark-grey boxes, with assigned biological outcomes. DBD: DNA binding domain, LZ: Leucine zipper, MB: Marked box, RB: pRB binding site, cA: Cyclin A binding site, TAD: Transactivation domain. S: serine, K: lysine, R: arginine.

1.3 Biological Functions of PAD4

1.3.1 Peptidyl arginine deiminases (PADs)

In contemporary research, citrullination has gained significant attention as a crucial post-translational modification among over 200 modifications of proteins (G. Duan & Walther, 2015). This modification is believed to play a critical role in numerous cellular processes including immunity, development, and homeostasis (Christophorou, 2022). Moreover, there is mounting evidence indicating its association with several diseases, such as cancer and neurodegenerative diseases (Christophorou, 2022; Harada et al., 2023; Yuzhalin, 2019). What makes citrullination unique is that, unlike other PTMs which add a small chemical moiety such as acetyl or methyl group, this modification instead involves the direct structural conversion of a peptidyl arginine residue into citrulline (Figure 1-5). During the process of citrullination, a guanidinium group of the arginine residue is hydrolysed, resulting in the production of urea as well as the loss of positive charge and two potential hydrogen bond donors (Slade et al., 2014). This may lead to a multitude of biological consequences, where citrullination has been reported to influence the target protein's activity, structure, stability, localisation, and interaction with other protein(s) and nucleic acids (Christophorou, 2022).

Protein citrullination is catalysed by a family of enzymes called peptidyl arginine deiminases (PADs or PADIs). In the human genome, there are five members of the PAD family identified, including PAD1, PAD2, PAD3, PAD4, and PAD6, and these genes are clustered at the chromosomal position around 1p36.13 (Vossenaar et al., 2003). Despite the high degree of structural homology observed in the catalytic domain of the enzyme family, their tissue distributions do not appear to mirror these genomic and structural

similarities. Rather, each member exhibits a distinct and unique expression pattern in tissues (Table 1-1). The activities of PAD1-4 are known to be tightly regulated by calcium, with the presence of calcium ions increasing enzymatic activity by >10,000 fold (Arita et al., 2004; Slade et al., 2015). PAD6 is the only member without Ca^{2+} binding sites and its amino acid sequence in the active site is also not conserved (Raijmakers et al., 2007). As a result, the enzymatic activity of PAD6 remains unclear; it was shown to be inactive *in vitro* (Raijmakers et al., 2007), although α-tubulin in oocytes has been identified as a potential substrate (Esposito et al., 2007). To date, enzymes acting to remove or return peptidyl citrulline to native arginine residue have not been discovered, and therefore citrullination is considered an irreversible PTM.

Table 1-1 Tissue distribution and abnormal cancerous expression of PADs.

PAD isozymes	Tissue-specific expression	Examples of known substrates	Deregulated over-expression in Cancer
PAD1	Immune cells, keratinocytes, hair follicle, epidermis, uterus	Keratin, Filaggrin, S100A3, MEK1-ERK1/2-MMP2 signalling enzymes	N/A
PAD2	Immune cells, Salivary gland, brain, bone marrow, spinal cord, uterus, skin, spleen, pancreas, kidney, skeletal muscle	MPB, CXCL10, CXCL11, Vimentin, Actin, GFAP, S100A3, Histone H3, Histone H4	Gastric (tissue and blood), liver (tissue and blood), colorectal (tissue and blood, though decreased expression in tissue also reported), oesophagus; (tissue and blood), breast (tissue and blood), skin
PAD3	Immune cells, keratinocytes, hair follicles, nerves	Filaggrin, Vimentin, Trichohyalin, S100A3	N/A
PAD4	Immune cells (neutrophils, monocytes, macrophages), brain, uterus, joints, bone marrow	NFC1, NCF2, S100A3, Collagen Type I, HAT p300, NPM1, GSK3β, ING4, RPS2, FUS, EWS,0 TAF15, ADAMTS13, E2F1, Histone H1 (R54Cit), Histone H2A (R3Cit), Histone H3 (R17Cit, R26Cit, R2Cit, R8Cit), Histone H4 (R23Cit, R3Cit)	Gastric (tissue and blood), liver (tissue and blood), ovarian (tissue and blood), colon, endometrial, hepatocellular, lung, bladder
PAD6	Ovary, early embryo, testicles, ovum, oocyte, thymus	α-tubulin	N/A

Adapted from Yuzhalin et al. 2019 and Harada et al. 2023.

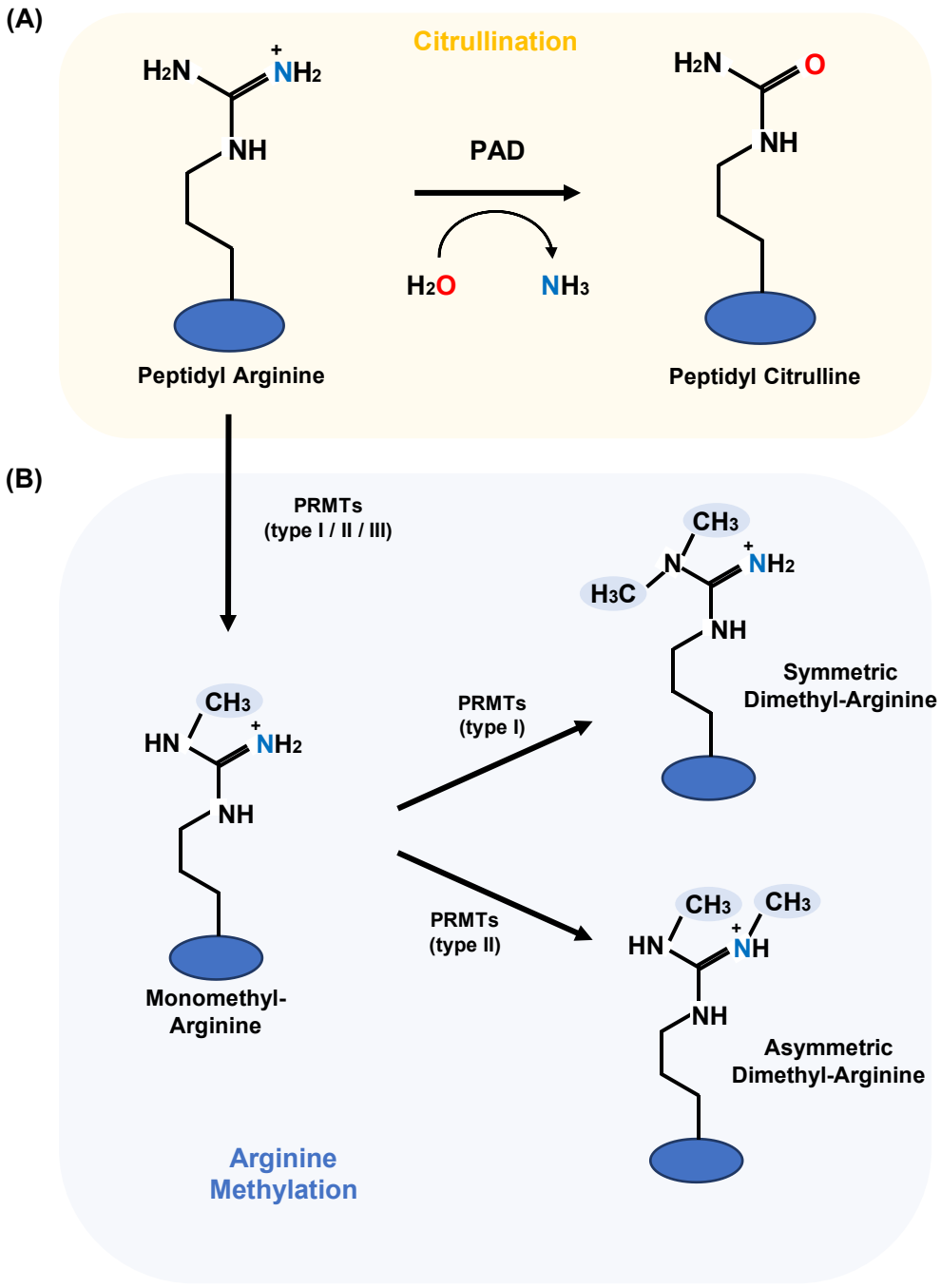


Figure 1-5 Citrullination and Arginine Methylation.

(A) Arginine is converted into citrulline by PADs through deimination / citrullination. The reaction results in the release of ammonia in return for H_2O consumption. (B) Arginine di-methylation is catalysed by PRMTs via a mono-methylated intermediate. The initial mono-methylation is catalysed by type I/II/III, and subsequently Type I mediates the symmetric and Type II mediates the asymmetric demethylation. Adapted from Harada et al. 2023.

1.3.2 Properties of PAD4

PAD4 is the most studied member of the PAD family. This enzyme is also known as the only PAD member with a classical monopartite nuclear localization signal (NLS) (56-PPAKKKST-63) and is thus considered to be predominantly nuclear localised (Nakashima et al., 2002). The structure of PAD4 consists of two main domains: the N-terminal domain (M1-P300) with two immunoglobulin-like domains, and the C-terminal domain (N301-P6663) with the catalytic site (Figure 1-6). As previously mentioned, PAD4 is a calcium-dependent enzyme, and five Ca^{2+}-binding sites were identified in a crystal-structure study (Arita et al., 2004); two sites are located nearby the active site cleft in the C-terminal domain, and the other three in the N-terminal domain. The two Ca^{2+}-binding sites at the C-terminal domain appear to play a crucial role in the dependency of PAD4 activity on calcium. This is because when Ca^{2+} binds to these sites, it causes a change in the enzyme's structural conformation and helps to stabilize its active site. This is necessary for the enzyme to recognize its target substrate and carry out its catalytic function. (Arita et al., 2004). Almost all of the residues forming these two sites were discovered to be conserved amongst PAD1-4, suggesting that the activities of other Ca^{2+}-dependent PADs may also be regulated by similar mechanisms (Arita et al., 2004). The other three sites in the N-terminal domain seem to stabilise the protein structure during the folding process and ensure the correct and active conformational change of PAD4 (Arita et al., 2004; Y.-L. Liu et al., 2017). Interestingly, even though they are positioned away from the C-terminal catalytic site of PAD4, a recent site-directed mutagenesis study demonstrated that Ca^{2+} binding to these sites are also critical for PAD4's enzymatic activity by inducing a conformational change of the C-terminal active site (Y.-L. Liu et al., 2017). The crystal structure of PAD4 suggests that in the presence of Ca^{2+}, active

PAD4 forms a homodimer in a head-to-tail manner, whereby several residues including Y435, R441, D465, V469 and W548 were identified as being essential for dimerization (Arita et al., 2004; C.-Y. Lee et al., 2017).

Under physiological conditions, the intracellular concentration of calcium ions is tightly maintained at around 100 nM (Clapham, 2007). PAD4 activation in turn requires a much higher Ca^{2+} concentration (10^3-10^4 nM) (Arita et al., 2004). Indeed, the role of PAD4 and citrullination has been described in multiple biological events in which the Ca^{2+} concentration exceeds typical physiological levels, such as occurs during apoptosis or terminal differentiation (Asaga et al., 1998; György et al., 2006). Conversely, even in cells like neutrophils, which express abundant levels of PAD4, citrullination of substrates cannot be detected until the cells are stimulated with calcium ionophore or inflammatory molecules such as TNFα and lipopolysaccharide (LPS) (Neeli et al., 2008). Therefore, some consider that PAD4 is catalytically inactive in a normal physiological condition. However, multiple reports also identified substrates of PADs under physiological conditions; for example, about 10% of histones in HL-60 granulocytes are citrullinated by PAD4 (Hagiwara et al., 2002), and another proteomic revealed some PAD4 substrates in cells without ectopic PAD4 expression nor Ca^{2+} ionophore (Tanikawa et al., 2018). This indicates that there may be a yet-unknown compensatory mechanism that activates PAD4 in a Ca^{2+}-independent manner, or that there could exist intracellular supra-physiological pockets of high calcium concentration.

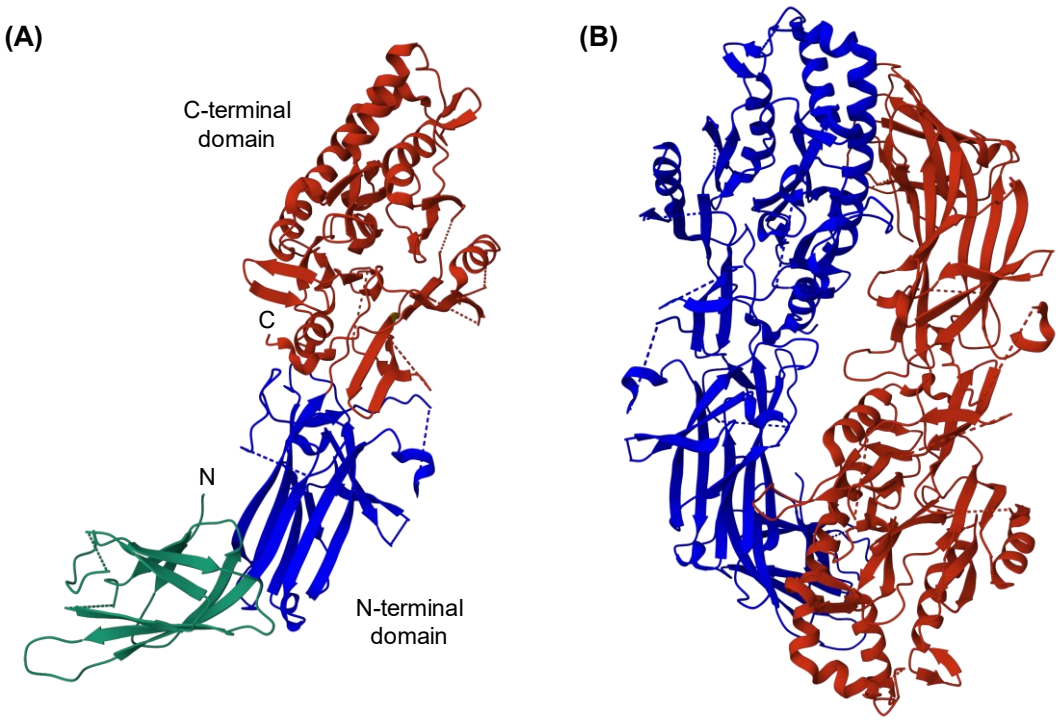


Figure 1-6 Crystal Structure of PAD4.

(A) The crystal structure of PAD4 monomer, displaying the C-terminal domain (red) and the two immunoglobulin-like subdomains in the N-terminal domain (subdomain 1 in green, 2 in blue). (B) The crystal structure of PAD4 dimer in a head-to-tail fashion. Both images taken from the RCSB PDB (https://www.rcsb.org/) of PDB ID 1WD8 (Arita et al., 2004).

1.3.3 PAD4 in the immune system

First discovered in myeloid leukaemia HL-60 cells differentiating to granulocytes (Nakashima et al., 1999), PAD4 has been most heavily studied in the context of an immune response. The expression of PAD4 is widely found in neutrophils and other leucocytes, as well as in bone marrow, lymphoid tissues, and blood (Uhlen et al., 2019). The physiological roles of PAD4 in these immune cells have only just been delineated in the last decade. For example, PAD4 is involved in inflammatory signal transduction via modulation of cytokine levels, whereby PAD4-dependent citrullination of nuclear factor-κB (NFκB) p65 subunit leads to its enhanced interaction with importin α3 and nuclear localisation. This results in the transcriptional activation of genes like TNFα and interleukin-1β (IL-1β) (B. Sun et al., 2017).

Furthermore, PAD4 is known to be involved in the formation of neutrophil extracellular traps (NETs) in immune cells including neutrophils, monocytes, and macrophages (Mutua & Gershwin, 2021; Vorobjeva & Chernyak, 2020). During the process of NET formation (often referred to as 'NETosis'), both the nuclear and cytoplasmic membranes of cells become degraded, leading to the extracellular release of citrullinated chromatin along with bactericidal and lysosomal proteins such as neutrophil elastase, myeloperoxidase, cathelicidin, and cathepsin G (Folco et al., 2018; Papayannopoulos et al., 2010; Wong et al., 2018). NETosis is therefore regarded as a unique form of programmed cell death as part of an immune response against pathogens such as bacteria and viruses (Mutua & Gershwin, 2021; Vorobjeva & Chernyak, 2020). PAD4 is considered to be a crucial regulator of NETosis, where histone citrullination (H3, H4 and H2A) mediated by PAD4 plays a key role in chromatin decondensation through neutralisation of positive charges on arginine residues, which precedes the subsequent

release of chromatin into the extracellular environment (Y. Wang et al., 2009). Indeed, neutrophils isolated from PAD4 knockout mice were found to be deficient in forming NETs (P. Li, Li, et al., 2010), and the chemical inhibition of PAD4 was shown to be sufficient to suppress NETosis in both mouse and human cells (Lewis et al., 2015). Despite considerable progress in unravelling the molecular mechanisms of PAD4's involvement in NET formation, many questions remain unanswered, particularly with regard to its emerging role in various diseases, including cancer (Mutua & Gershwin, 2021; Vorobjeva & Chernyak, 2020). This is a highly active area of research, and in the forthcoming section, we will examine recent findings on the relevance of NETosis to carcinogenesis.

In addition to its physiological roles in the immune system, deregulated activity and abnormal expression of PAD4 have been documented in a multitude of autoimmune or inflammatory diseases, including rheumatoid arthritis (RA), systemic lupus erythematosus, Alzheimer's diseases, and cancers (Valesini et al., 2016). Notably, the inappropriate activity of PAD4 has been extensively studied in RA. This autoimmune disease is characterised by chronic inflammation at the synovial joints and infiltration of blood-derived cells, typically macrophages. Abnormal levels of citrullinated proteins are also often detected in the synovial fluid of RA patients (Kurowska et al., 2017; Pruijn et al., 2010). The presence of these modified proteins is believed to lead to the production of anti-citrulline peptide antibodies (ACPA), which are a major hallmark of RA found in around 70 % of patients and have been used for diagnostic purposes for many years (Kurowska et al., 2017). The presence of ACPA is often linked to disease progression and joint damage in human patients, though the underlying mechanisms of how ACPA contribute to the development of RA are not fully understood (Kurowska et al., 2017). It

was suggested by multiple *in vitro* studies that ACPA are involved in multiple inflammatory responses including TNF-α production, NET induction, and promoting osteoclastogenesis, but this work has yet to be confirmed *in vivo* (Kurowska et al., 2017).

1.3.4 PAD4 in cancer

In recent years, multiple lines of evidence have illustrated a deep, intricate association between PAD4-dependent citrullination and cancer progression. Abnormal expression of PAD4 has been found in numerous cancer types relative to their respective normal tissue counterparts (Table 1-1) (X. Chang et al., 2009; X. Chang & Han, 2006). In contrast, benign tumours and non-tumour inflamed tissues were not shown to express high levels of PAD4, supporting a potential role for this PAD enzyme in cancer progression (X. Chang et al., 2009; X. Chang & Han, 2006). In addition, PAD4 expression was identified in the blood of cancer patients (X. Chang et al., 2009). Deregulated expression and activity of this PAD member have been linked to various cellular processes in tumours, such as cell signalling, transcription, NETosis, and Epithelial-to-Mesenchymal transition (EMT) (Ghari et al., 2016; P. Li et al., 2008; Stadler et al., 2013; Tanikawa et al., 2012; L. Yang et al., 2020; Yao et al., 2008; Yuzhalin et al., 2018; X. Zhang et al., 2011), which will be reviewed respectively.

PAD4 and gene regulation in cancer

A multitude of research suggests that PAD4 regulates gene expression, particularly at the level of transcription, in various cancer cells (Ghari et al., 2016; P. Li et al., 2008; Stadler et al., 2013; Tanikawa et al., 2012; L. Yang et al., 2020; Yao et al., 2008; Yuzhalin et al., 2018; X. Zhang et al., 2011). Typically, this is mediated by the citrullination of histones; several arginine residues of histone H3 and H4 have been described as targets of PAD4

(P. Li et al., 2008; P. Li, Wang, et al., 2010; G.-Y. Liu et al., 2006; Tanikawa et al., 2012; Yao et al., 2008). In addition, several transcription factors and co-factors are known to be substrates of PAD4, such as ING4, p300, E2F1, and ELK-1 (Ghari et al., 2016; Q. Guo & Fast, 2011; Y.-H. Lee et al., 2005; X. Zhang et al., 2011).

A pivotal and multifaceted role of PAD4-mediated signal transduction is exemplified with the interplay between PAD4 and p53. PAD4 is known to act as a suppressor of p53-target genes; In a p53-dependent manner, PAD4 is localised to the promoters of p53-target genes, such as OKL38, p21, CIP1 and WAF1, and catalyses histone citrullination which results in down-regulation of these target genes (P. Li et al., 2008; Yao et al., 2008). In addition, PAD4 can directly citrullinate a protein called inhibitor of growth protein 4 (ING4) and inhibit its interaction with p53 (Q. Guo & Fast, 2011). ING4 is known to enhance p53 transcriptional activity, in part by promoting acetylation of p53 through the acetyltransferase p300-dependent pathway (Shiseki et al., 2003). Citrullinated ING4, which dissociates from p53, can longer recruit p300 to p53. This leads to the suppression of p53-target genes including p21 (P. Li, Wang, et al., 2010). Together, these findings suggest that PAD4 regulates the transcription activity of p53 through the citrullination of both histone and non-histone proteins.

In a follow-up study, the recruitment of PAD4 to p53 target gene promoters was found to be significantly reduced in response to UV-induced DNA damage, resulting in the transactivation of p53-target genes including p21, GADD45, and PUMA (P. Li, Wang, et al., 2010). This study also identified HDAC2 as a co-suppressor of PAD4 and demonstrated that PAD4 inhibitor Cl-amidine treatment and/or HDAC inhibitor treatment could suppress cancer cell growth (P. Li, Wang, et al., 2010). Consistent with this finding,

over-expression of PAD4 was demonstrated to induce apoptosis through p21-dependent cell cycle arrest and via the Bax-dependent pathway in human leukaemia HL-60 cells and acute T leukaemia Jurkat cells (G.-Y. Liu et al., 2006). These results together suggest that PAD4 has an important regulatory role in p53-dependent apoptosis, by either providing a repressive histone modification mark at p53-target genes or targeting the non-histone substrates involved in the regulation of p53 activity. However, a seemingly contradictory result was provided by Tanikawa et al., who demonstrated that increased levels of histone H4 citrullination were apparent after DNA damage, and *PAD4*$^{-/-}$ mice exhibited resistance to radiation-induced apoptosis in the thymus (Tanikawa et al., 2012). This discrepancy may result from the fact that Adriamycin (ADR) or γ-rays were used to induce DNA damage in this report, rather than using UV like in the study by Li et al. (P. Li, Wang, et al., 2010; Tanikawa et al., 2012).

The role of PAD4 as a transcriptional co-activator has also been reported (Ghari et al., 2016; X. Zhang et al., 2011). For example, one study demonstrated that the transcription factor ELK-1 is a direct substrate of PAD4 in breast cancer cells (X. Zhang et al., 2011). PAD4-dependent citrullination facilitates the subsequent phosphorylation of ELK-1 by the kinase ERK2, in response to Epidermal Growth Factor (EGF) stimulation. This causes ELK-1 to form a complex with the histone acetyltransferase p300 (X. Zhang et al., 2011). This protein complex subsequently acetylates histones at the promoter region of target genes, including the known proto-oncogene c-FOS, resulting in transactivation (X. Zhang et al., 2011).

In addition, PAD4 also acts as a transcriptional co-activator to E2F1, as we described earlier in this introduction (Ghari et al., 2016). E2F1 is citrullinated by PAD4 at arginine

residues R109 and R127, which promotes the subsequent interaction of BRD4 with acetylated E2F1. The E2F1/BRD4 complex can then be recruited to the promoters of pro-inflammatory genes such as TNFα and IL-1β, and promote their expression as part of an inflammatory response (Ghari et al., 2016).

The critical roles of PAD4 in gene regulation in different cancer-related pathways have been emphasized in the preceding paragraphs. Nevertheless, many questions remain unresolved concerning the precise mechanisms by which PAD4 affects cancer progression, and whether it is feasible to target this pathway in cancer therapy. For example, it would be of interest to characterise the context-dependent molecular mechanisms that underlie the seemingly opposing roles of PAD4 in the apoptotic pathway induced by diverse types of DNA damage.

PAD4 and NETosis in cancer

The vital role that PAD4 plays during NETosis was described in an earlier section. There is growing evidence to suggest that this immune response to pathogens is also involved in various pathological conditions, including cancer. Notably, NETs have been implicated in promoting metastasis and proliferation in several types of tumours (Houghton et al., 2010; Martins-Cardoso et al., 2020; Masucci et al., 2020; Wolach et al., 2018). For example, it was recently demonstrated that abnormal levels of NETs were discovered in the liver metastases of breast cancer patients (L. Yang et al., 2020). This study showed that NETs can act as a chemotactic signal to enhance cancer metastases; mechanistically, DNA contained in NETs was detected by the transmembrane protein CCDC25, leading to the activation of the ILK-β-parvin pathway and thus enhanced cell motility (L. Yang et al., 2020).

PAD4 and epithelial-to-mesenchymal transition in cancer

Epithelial-to-Mesenchymal Transition (EMT) is a process through which epithelial cells with cell polarity and adhesion lose such characteristics and instead gain a mesenchymal phenotype, such as migratory capacity, invasiveness, elevated resistance to apoptosis, and greatly increased production of extracellular matrix (ECM) components (Brabletz et al., 2018). EMT appears to be of importance in many biological processes and several pathological conditions, and most notably plays a critical role in cancer metastasis (Brabletz et al., 2018).

Several studies proposed that PAD4 might be involved in EMT signal transduction in cancer cells. For example, Stadler et al. revealed that depletion of PAD4 in breast cancer cells leads to enhanced cell motility and invasiveness, with increased levels of vimentin (a mesenchymal marker) and decreased levels of E-cadherin and β-catenin (epithelial markers) (Stadler et al., 2013). Mechanistically, PAD4 was shown to directly citrullinate glycogen synthase kinase-3β (GSK3β) and positively influence its nuclear localisation, leading to the transactivation of multiple genes. Conversely, silencing of PAD4 instead translocated GSK3β out of the nucleus and induced TGF-β signalling and subsequent epithelial-to-mesenchymal transition (Stadler et al., 2013). Consistent with this finding, another study established that over-expression of PAD4 in lung cancer cells can suppress EMT by negatively regulating the expression of ELK-1 (Q. Duan et al., 2016).

However, more recently, it was demonstrated that PAD4, through citrullination of the extracellular matrix (ECM), has a key role to play in promoting cell adhesion and inhibiting motility in colorectal cancer cells (Yuzhalin et al., 2018). Interestingly, in contrast to previous findings (Q. Duan et al., 2016; Stadler et al., 2013), this study

indicated that the inhibition of PAD4 with BB-Cl-amidine treatment was sufficient to disrupt liver metastatic growth and alter the EMT profile with concomitantly decreased expression of mesenchymal markers (Yuzhalin et al., 2018). This result suggests that PAD4 activity regulates the ECM in cancer cells both intra- and extracellularly, as well as indicating the potential of PAD4 as a therapeutic target in liver metastases, though the functional discrepancy between the EMT-promoting and suppressing roles of PAD4 requires further investigation.

PAD4 and pluripotency

Citrullination is often linked to chromatin decondensation, as this modification results in the removal of positive charges from the arginine residues of histone. Given that an open chromatin structure is required in pluripotent cells to maintain unrestricted developmental potential, the roles of PAD4 in pluripotency have been investigated. Indeed, the enhanced expression of this PAD member was discovered in embryonic stem (ES) cells and induced pluripotent stem (iPS) cells (Christophorou et al., 2014). This increased levels of expression and activity (histone H3 citrullination) of PAD4 was closely correlated with the up-regulated expression of pluripotency factors such as Nanog (Christophorou et al., 2014). Interestingly, the authors used SILAC labelling to show that histone H1 is a substrate of PAD4 in ES cells and its citrullination significantly reduced its binding affinity to nucleosomes, leading to global chromatin decondensation (Christophorou et al., 2014).

1.3.5 Citrullination and arginine methylation

Post-translational modifications rarely work alone, but often functionally interact with each other in different combinations to delicately regulate biological outcomes.

Citrullination is no exception, and a multitude of studies have demonstrated the crosstalk between citrullination and other protein modifications, such as methylation, acetylation, and phosphorylation, on numerous substrate proteins, both histone and non-histones. The most established and obvious crosstalk is found between citrullination and arginine methylation, as both PTMs can theoretically target the same arginine residue of peptides and proteins (Fuhrmann & Thompson, 2016).

During the process of arginine methylation, the guanidino group of a peptidyl arginine residue is subject to the sequential addition of methyl groups; first to generate mono-methylated (Rme1) and then di-methylated arginine (Rme2), and the latter can be sub-categorised into symmetric (Rme2s) and asymmetric (Rme2a) methylation depending on the location of the second methyl group (Figure 1-5) (Blanc & Richard, 2017; Q. Wu et al., 2021). Arginine methylation is catalysed by a family of enzymes called the protein arginine methyltransferases (PRMTs). There exist nine members of PRMTs, and they are often classified into three sub-groups depending on their catalytic ability to mediate Rme2a (Type I: PRMT1/2/3/4/6/8), Rme2s (Type II: PRMT5/9), or only Rme1 (Type III: PRMT7) (Blanc & Richard, 2017; Q. Wu et al., 2021). Citrullination mediated by PADs involves the deimination or the hydrolysis of the guanidino group of an arginine residue to convert it to citrulline (Figure 1-5). This should effectively block the recognition and addition of a methyl group to the respective arginine residue by PRMTs (Cuthbert et al., 2004; Hagiwara et al., 2005). There is some controversy over whether or not PADs can catalyse the conversion of methylated arginine to citrulline. According to some early reports, increased levels of histone citrullination in HL-60 cells were correlated with a decrease in arginine methylation levels (Cuthbert et al., 2004; Hagiwara et al., 2005; Y. Wang et al., 2004). However, other *in vitro* studies demonstrated that arginine residues

with meR1, meR2s, or meR2a modifications are completely blocked for conversion to citrulline, or are converted at rates several thousand-fold slower as compared to the native peptidyl arginine (Hidaka et al., 2005; Kearney et al., 2005; Raijmakers et al., 2007). It is therefore most likely that arginine methylation antagonises citrullination. Some examples of citrullination-methylation crosstalk, and the biological outcomes linked to this interplay are summarised in Table 1-2.

Table 1-2 Examples of Citrullination-Methylation Crosstalk.

Substrate	Site for crosstalk	Biological consequences
Histone	H3(R2/8/17/26) and H4(R3) targeted by PRMT1/4 for Me and PAD4 by Cit.	Me and Cit target histones in the promoters of genes such as pS2, CTCF, and OKL38, to regulate gene expression positively or negatively.
p300	C-terminal R2142 residue targeted by PRMT4 for Me and PAD4 for Cit.	Me promotes and Cit inhibits the interaction between p300 and GRIP, affecting the transcriptional co-activator function of this complex.
SFPQ	Several N-terminal R residues targeted by PRMT1 for Me and PAD4 for Cit.	Me promotes and Cit inhibits the association between SFPQ and its target mRNA.
RPS2	R residues in N-terminal RG-repeat region targeted by PRMT3 for Me and PAD4 for Cit.	Me of RPS2 may regulate the biogenesis of ribosomes in yeast.
E2F1	R109 targeted by PRMT1 for Me and PAD4 for Cit. R111/113 targeted by PRMT5 for Me and antagonise R109 met.	Each PTM channels E2F1 into a distinct biological pathway; R109Me to apoptosis, R111/113Me to proliferation, and R109Cit to pro-inflammatory gene expression.

Adapted from Harada et al. 2023.

1.3.6 PAD inhibitors in pre-clinical studies

As reviewed above, the regulation of PAD4 has been closely linked to multiple cellular processes contributing to the occurrence and progression of cancer. Not surprisingly, there exists considerable interest in targeting PAD4 for therapeutic intervention, and a significant effort has been devoted to the development of potent PAD4-specific inhibitors over the last decade. Indeed, some inhibitors have already shown promising anti-tumour effects in several pre-clinical models (Deng et al., 2022; Kholia et al., 2015; Kosgodage et al., 2018; McElwee et al., 2012; H. Qin et al., 2017; Y. Wang et al., 2012; Wei et al., 2021). Here, a few examples of pan-PAD and PAD4-specific inhibitors and their potential value in cancer treatment will be discussed (Figure 1-7).

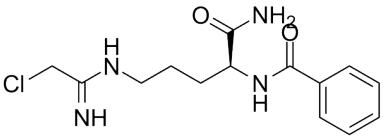
Derived from benzoyl-arginine compounds, the irreversible pan-PAD inhibitor Cl-amidine and its modified version BB-Cl-amidine are perhaps the most frequently used compounds to study PAD function both *in vitro* and *in vivo,* and are therefore often used as a benchmark to assess other compounds' efficacy (Kawalkowska et al., 2016; Ledet et al., 2018; Mondal & Thompson, 2019). These compounds have been shown to inhibit histone citrullination and NET formation (Biron et al., 2017), and also significantly reduced the severity of various murine models of inflammatory diseases including collagen-induced arthritis (CIA), colitis, and lupus (Chumanevich et al., 2011; Knight et al., 2015; Willis et al., 2011), as well as in several types of cancer (breast, prostate, and glioblastoma) (Kholia et al., 2015; Kosgodage et al., 2018; McElwee et al., 2012; H. Qin et al., 2017; Y. Wang et al., 2012).

GSK484 and GSK199 are the most efficient reversible PAD4-specific small molecule inhibitors. Both compounds have been extensively used *ex vivo* and *in vivo* to investigate

PAD4 biology in physiological and pathological conditions and were shown to block histone citrullination as well as disrupt NETosis in both mouse and human cells (Lewis et al., 2015; Mondal & Thompson, 2019). A recent study has indicated that triple-negative breast cancer cells of a mouse xenograft model became sensitised against radiotherapy after GSK484 treatment (Wei et al., 2021). More recently, Wang et al. reported that GSK484 can sensitise HCT116 cells to irradiation and reduce migration and invasion (B. Wang et al., 2023). In addition, another PAD4-specific inhibitor JBI-589 was recently demonstrated to inhibit primary tumour growth in mouse models by negatively regulating the expression of chemokine receptor CXCR2, which reduced the accumulation of neutrophils in the tumour microenvironment (Deng et al., 2022).

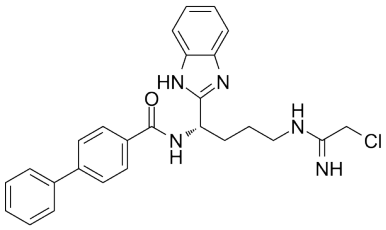
Despite these promising anti-tumour activities in pre-clinical disease models, no pan-PAD or PAD4-specific inhibitors have been clinically approved to date. Successful use of these inhibitors may still require further understanding of PAD biology which often involves context- or tissue-dependent activities. Mechanism-led approaches towards the development of more potent, clinically safe compounds should contribute to their therapeutic use in various cancer types.

Irreversible Inhibitors



Cl-Amidine

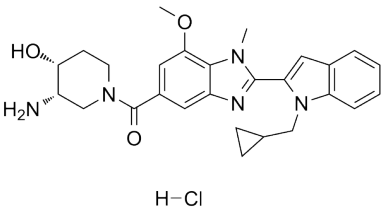
IC_{50} = 0.005 mM
k_{inact}/K_I (M^{-1} min^{-1})
PAD1: 37,000
PAD2: 1,200
PAD3: 2,000
PAD4: 13,000



BB-Cl-Amidine

IC_{50} = 0.008 mM
k_{inact}/K_I (M^{-1} min^{-1})
PAD1: 16,000
PAD2: 5,000
PAD3: 6,000
PAD4: 14,000

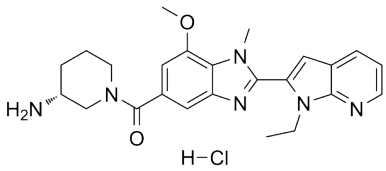
Reversible Inhibitors



GSK484

IC_{50} PAD4 (mM)
0 mM Ca^{2+} = 5 * 10^{-5}
0.2 mM Ca^{2+} = 8 * 10^{-5}
2 mM Ca^{2+} = 2.5 * 10^{-4}

K_{is} (mM)
PAD1: 0.108
PAD2: 0.107
PAD3: 2.09
PAD4: 0.0068



GSK199

IC_{50} PAD4 (mM)
0 mM Ca^{2+} = 2 * 10^{-4}
0.2 mM Ca^{2+} = 2.5 * 10^{-4}
2 mM Ca^{2+} = 1 * 10^{-3}

K_{is} (mM)
PAD1: 0.57
PAD2: 1.58
PAD3: 0.925
PAD4: 0.016

Figure 1-7 Examples of PAD4 inhibitors.

Four examples of irreversible and reversible PAD / PAD4 inhibitors with IC50 values and kinetic profiles. Kinact is the rate of enzyme inactivation and KI is the inhibition constant (concentration at 50% Kinact), where the ratio (kinact/KI) is often used to analyse the time-dependent IC50 values (Fuhrmann et al., 2015). Kis is the dissociation constant for the enzyme-inhibitor complex.

1.4 Alternative Splicing

The genetic information is stored within DNA and passed down through successive generations. In the lifecycle of a cell, this DNA data is accessed in the form of RNA. DNA, due to its strong chemical stability, functions as an ideal repository for genetic information. On the other hand, RNA, being chemically more reactive and less stable, does not typically serve as a storage medium for genetic information, but instead acts as an intermediary between DNA and proteins. However, the notion of RNA as a direct copy of genetic information is oversimplified, as the RNA transcripts derived from DNA frequently undergo processing.

The first evidence of messenger RNA (mRNA) splicing was discovered in the 1970s, whereby some parts of viral sequences were removed from pre-mRNA and the remaining segments were joined together (Berget et al., 1977; Chow et al., 1977). Later, almost all mammalian transcripts mediated by RNA polymerase II (RNAP II) were revealed to undergo this process; during which all the introns, non-coding regions of a transcribed pre-RNA, will be removed and the remaining coding regions called exons are adjoined together to produce a mature mRNA (Kelemen et al., 2013).

Additionally, there is evidence that, in more than 90 % of human genes, different sets of exons and introns can be removed or included to produce more than one transcript from a single gene (E. T. Wang et al., 2008). This mechanism is referred to as alternative splicing, and is typically categorised into several basic types including exon skipping, alternative 5'-splice site, alternative 3'-splice site, mutually exclusive exons, and intron retention (Figure 1-8). Alternative splicing can greatly enhance the complexity of transcriptome and diversity of proteome from a limited number of genes, and thus

modulate various aspects of the resulting RNA transcripts, including stability, translation efficiency, and even functions of coded proteins through the inclusion or exclusion of key domains (Kelemen et al., 2013; Marasco & Kornblihtt, 2022; Stamm et al., 2012). Unsurprisingly, aberrant splicing activity is considered as an emerging hallmark of numerous diseases including cancer; whereby alternative splicing is found to play key roles in a variety of important biological processes such as cellular homeostasis, development, differentiation, and stress response, and (Kalsotra & Cooper, 2011; Kelemen et al., 2013; Marasco & Kornblihtt, 2022; Stamm et al., 2012). In addition, alternative splicing may greatly increase the phenotypic diversity extremely rapidly, which could allow cancer cells to service in a hostile microenvironment, metastasise to remote locations, and acquire resistance mechanisms against conventional drugs (Biamonti et al., 2020). Hence, it is of paramount importance to elucidate the molecular mechanisms and functionalities of alternative splicing within the context of cancer; such understanding holds the potential to enhance both the diagnosis and treatment of this disease.

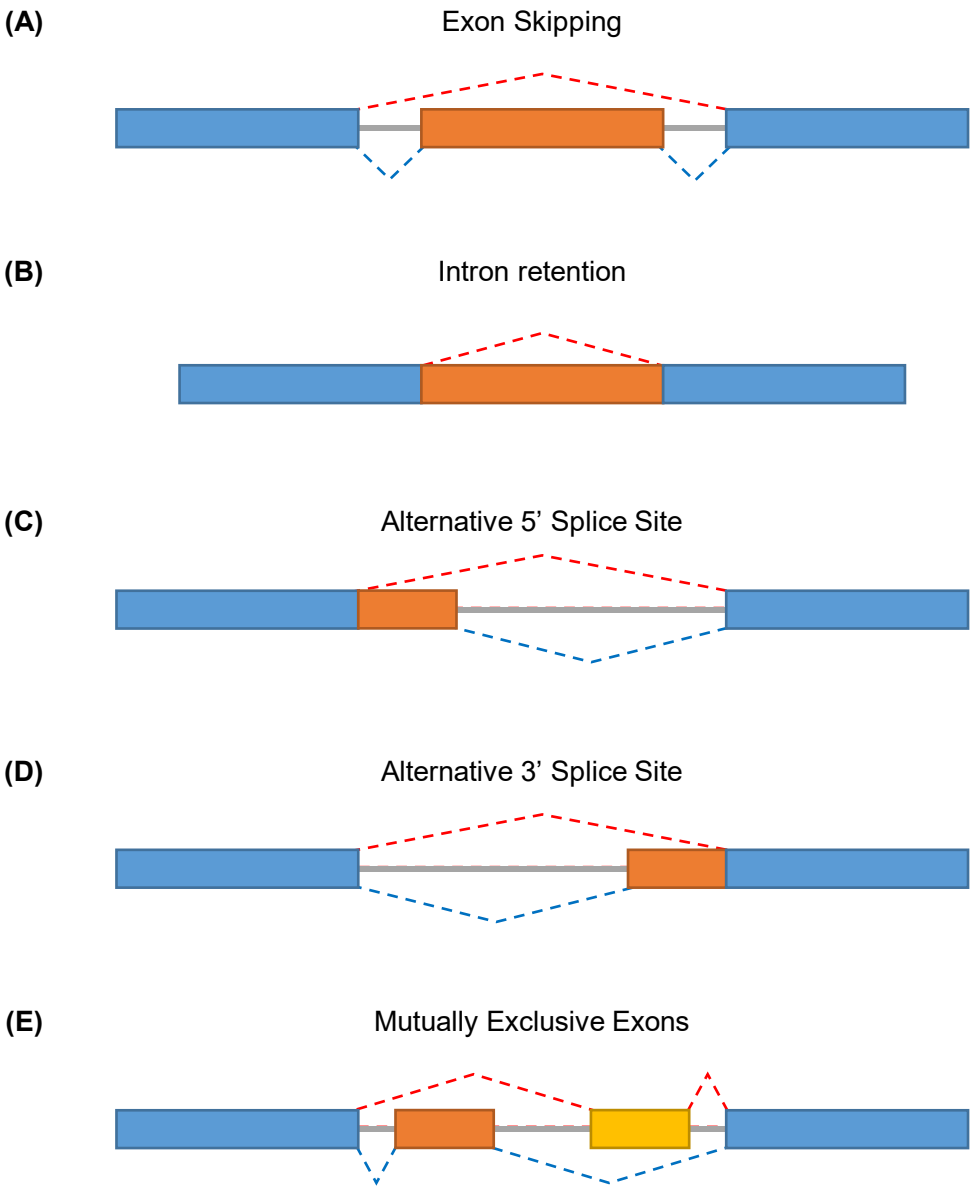


Figure 1-8 Different Types of Alternative Splicing.

Diagram showing the exon-intron structures explaining different types of alternative splicing events. (A) Exon skipping, (B) intron retention, (C) alternative 5' splice site, (D) alternative 3' splice site, and mutually exclusive exons. Blue boxes represent constitutive exons, whereas orange or yellow boxes represent alternatively spliced exons/introns.

1.4.1 Basic mechanism of RNA splicing

Pre-mRNA splicing is mediated by a macromolecular ribonucleoprotein (RNP) complex termed the spliceosome (Will & Lührmann, 2011). Most splicing events are catalysed by the major U2-dependent spliceosome, which is conserved amongst eukaryotes, whilst in rare cases, another minor or U12-dependent spliceosome may act to process removal of specific classes of introns (Stamm et al., 2012). The major spliceosome consists of several snRNPs (U1, U2, U4/U6 and U5) and other non-snRNA splicing factors. The snRNPs are the main components of this protein machinery and are typically composed of an snRNA molecule, seven Sm proteins (B/B', D1, D2, D3, E, F, and G), and a few particle-specific factors (Will & Lührmann, 2011). Both RNA and protein units of the snRNPs are known to have essential roles in multiple facets of the splicing cycle, such as splice-site recognition, the assembly and activation of the spliceosome machinery, and the catalysis of splicing itself (Stamm et al., 2012). In addition, there are some fundamental motifs in the pre-mRNA sequences that are required for mRNA splicing, such as 5' and 3' splice sites, a branch point, and a polypyrimidine tract.

The molecular mechanism of mRNA splicing is underpinned by the sequential assembly and dynamic arrangement of the spliceosome complex (Figure 1-9) (Gehring & Roignant, 2021; Stamm et al., 2012). This process is initiated by the binding of the U1 snRNP to the 5' splice site (SS), resulting in the formation of the E (early) complex. The branch point (located downstream of the 5' SS) and the 3' SS is then recognised by the U2 snRNP and U2 auxiliary factor (U2AF) heterodimer, and subsequently form a stable association in an ATP-dependent manner (Gehring & Roignant, 2021; Stamm et al., 2012). This is now referred to as complex A. Next, U4/U5/U6 snRNPs are recruited to the target pre-mRNA in the form of a tri-snRNP complex, marking complex B. The complex B of the

spliceosome undergoes a major conformation and compositional rearrangement, leading to the release of U1 and U4 snRNPs, U6 replacing the former at the 5' SS, and yielding the active complex B^{act} (Gehring & Roignant, 2021; Stamm et al., 2012). After further remodelling catalysed by the RNA helicase Prp2 in an ATP-dependent fashion, this complex is transitioned to another one called B*. The B* complex performs the first step of splicing; the hydroxyl (OH) of the adenosine at the branch point attacks the 5' SS to break the phosphodiester bond and form a 'lariat'-like intermediate of the intron-3' exon and the free 5' exon (Gehring & Roignant, 2021; Stamm et al., 2012). Another set of rearrangements then occurs aided by Prp16. This involves a conformational change of the U2 snRNP and repositioning of the RNA intermediate within the spliceosome catalytic centre, resulting in the formation of the C complex. This complex finally completes the second step of RNA splicing; the OH group of the free 5' exon attacks the 3' SS to break the phosphodiester bond and form another phosphodiester bond between 5' SS and 3' SS. As a result, the two exons are successfully adjoined and the lariat intron is released with the attached U2, U5 and U6 snRNPs that are recycled in the next round of splicing (Gehring & Roignant, 2021; Stamm et al., 2012).

The molecular mechanism of mRNA splicing is virtually independent of the process of transcription. However, evidence suggests that the majority of pre-mRNA splicing events in humans occur co-transcriptionally (Sousa-Luís & Carmo-Fonseca, 2022). This means that the target splice sites are defined and spliced while the rest of the RNA transcript is still being transcribed by RNA polymerase II. Recent advances in sequencing instrumentation have contributed to the growth of this field. For example, Sousa-Luis et al. has recently developed a technique called POINT (polymerase intact nascent transcript) to specifically analyse the nascent RNA transcripts, whereby splice site

recognition and splicing can happen very shortly after the respective sequences of the RNA were transcribed (Sousa-Luís et al., 2021). In addition, another study using a technique called CoLa-seq (co-transcriptional sequencing) developed by Zeng et al. indicated that the majority of intron splicing can be completed prior to the adjacent downstream exon being transcribed (Zeng et al., 2022). These recent studies collectively highlight that co-transcriptional splicing is a common event in human cells and is organised in a very rapid fashion.

Given that co-transcriptional splicing is a common event in human cells, there is growing interest in the crosstalk between splicing and transcription as another layer of gene regulation. During co-transcriptional splicing, not all of the SSs and splicing regulatory elements (SREs) are available and therefore the efficiency of splicing is expected to be largely influenced by other transcriptional factors or chromatin structures. For example, several studies have suggested that the speed of transcription is a major regulatory factor for splice site selection (de la Mata et al., 2003; Hollander et al., 2016; Kessler et al., 1993; S. W. Kim et al., 2017). In addition, the splicing machinery was shown to interact directly with the C-terminal domain (CTD) of RNAP II, which acts to assist the recruitment of some splicing machinery components (de la Mata & Kornblihtt, 2006). Additionally, it has been recently revealed that splicing is hugely impacted by the co-transcriptional chemical modification of pre-mRNA; for example, the splicing regulator hnRNP G (/RBMX) specifically binds to N^6-methyladenosine (m^6A) marks on the RNA as well as the RNAP II CTD and regulates splicing events in proximity to the modification (K. I. Zhou et al., 2019). Finally, some transcription factors, such as nuclear receptors, T-Box proteins, and E2F1, have been reported to play a direct role in splicing regulation (Rambout et al., 2018; Roworth et al., 2019).

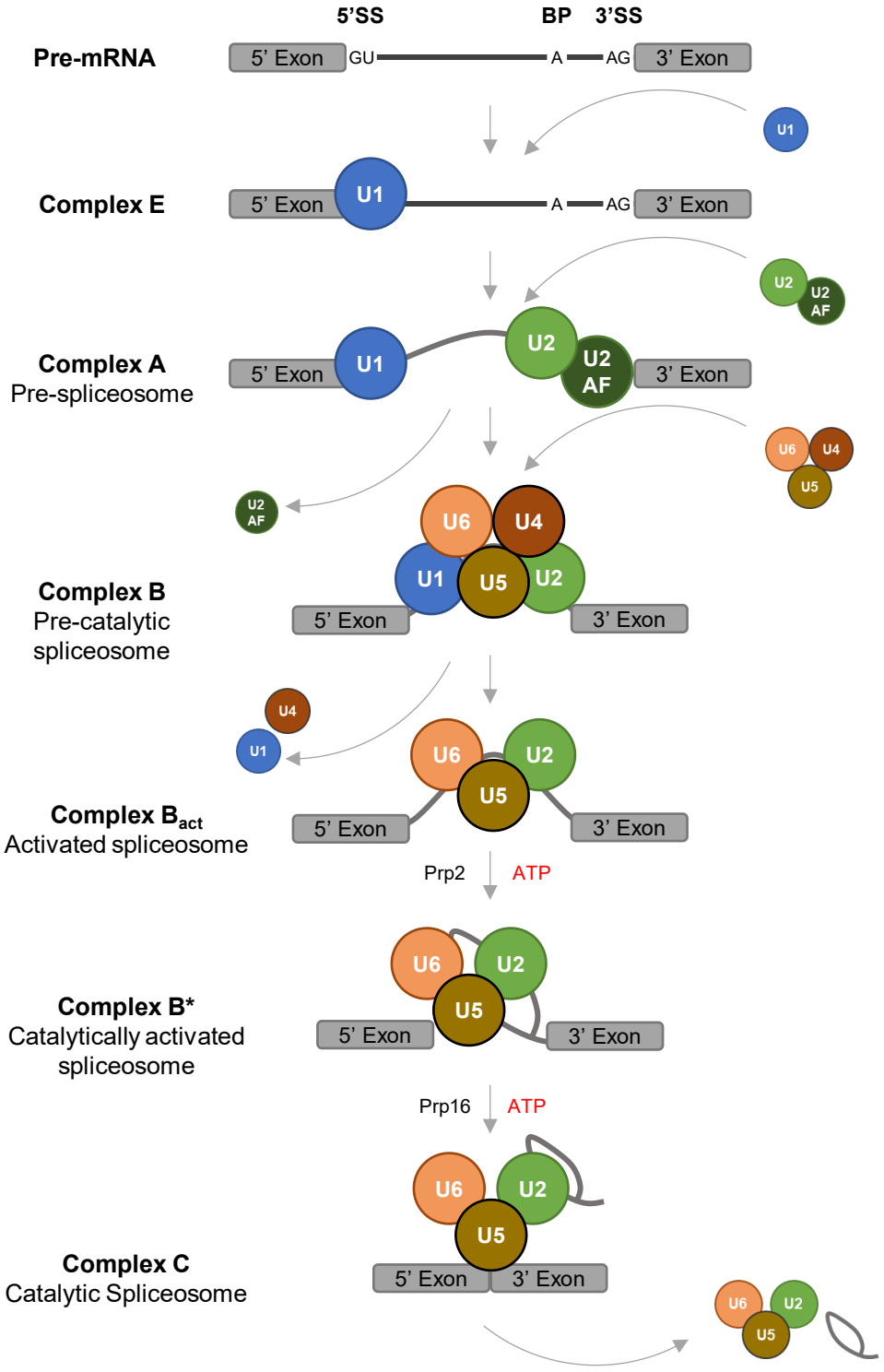


Figure 1-9 Mechanism of Pre-mRNA Splicing.

Diagram showing the mechanism of mRNA splicing by the sequential assembly and conformational rearrangement of the spliceosome complex (Stamm et al., 2012). SS: Splice site. BP: Branch point.

1.4.2 Alternative splice site selection

As highlighted above, the correct definition of exons and introns by the 5' and 3' SS are essential for mRNA splicing. In mammals, the consensus sequences for the 5' and 3' SS are known as YAG/GURAGU and NYAG/G respectively (Y = pyrimidine, R = A or G, / = splice site) (H. Sun & Chasin, 2000; M. Q. Zhang, 1998). However, the majority of splice sites in the human genome do not match these consensus sequences perfectly, and many potential splice sites with similar sequences are not used and form so-called 'pseudoexons' (H. Sun & Chasin, 2000; X. H.-F. Zhang et al., 2003, 2005). Indeed, the recognition of splice sites is known to be influenced by a multitude of factors in addition to the splice site consensus sequences; such as exon size, spatial proximity between 5' and 3' SS, RNA secondary structure and auxiliary splicing regulatory elements (SRE) (Stamm et al., 2012). Alternatively spliced exons are often defined by weaker splice sites, which deviate in sequence from the consensus. Therefore, the recognition of alternative SS is largely regulated by cis-acting determinants, especially SREs (Chasin, 2007). These motif sequences can be either exonic or intronic and act as splicing enhancers or silencers (ESE/S, ISE/S), which work together with numerous RNA-binding proteins (RBPs) that directly recognise the SREs. The RBPs, often described as *trans*-acting splicing regulators, can be broadly categorised into two groups: one containing arginine-serine-rich (RS) domains and RNA recognition motifs (RRMs), and another without RS domains. The major family of the former group is the Serine-Arginine-rich proteins (SR proteins) (Graveley, 2000; Zhong et al., 2009), whilst the heterogeneous nuclear ribonucleoprotein particles (hnRNPs) and other RBPs (e.g. NOVA and FOX) are included in the latter (Dreyfuss et al., 1993; Geuens et al., 2016).

The SR protein family is characterised by the possession of one or two RNA-recognition motif(s) (RRMs) in their N-terminus, and the Arginine-Serine-rich (RS) domain in their C-terminus (Shepard & Hertel, 2009; Stamm et al., 2012). As a general principle, SR proteins recognise RNA at the RRM, and usually bind splicing enhancer elements (Shepard & Hertel, 2009). Conversely, the RS domain plays a role as a splicing activator domain, typically by modulating the protein-protein or protein-RNA interactions (Shepard & Hertel, 2009). There are 12 members of SR proteins identified in the human genome, termed SR splicing factor (SRSF) 1-12 respectively. The roles that SRSFs play in both constitutive and alternative splicing have been reported, as well as their function in other types of RNA processing, such as transcriptional, co-transcriptional, and post-transcriptional regulation pathways (Jeong, 2017; Shepard & Hertel, 2009). Reflecting their involvement in various biological processes, deregulated activity and abnormal expression of the SR proteins have been described in numerous diseases, including cancer (Kalsotra & Cooper, 2011; Kelemen et al., 2013; Marasco & Kornblihtt, 2022; Shepard & Hertel, 2009; Z. Zhou et al., 2020).

The heterogeneous nuclear ribonucleoproteins (hnRNP) family of proteins contains one or two RNA-binding domains (RBDs): the RNA recognition motif (RRM), the quasi-RRM, a glycine-rich domain constituting an RGG box and a KH domain (Geuens et al., 2016). Instead of the RS domain in SR proteins, the hnRNPs have some auxiliary domains, including proline-, glycine- or acid-rich domains, which typically act to alter protein-protein interactions. The activity of most hnRNPs, including the most studied member hnRNP A1, is often associated with, but not limited to, inhibition of splicing events (Stamm et al., 2012). The expression level of hnRNPs is known to be altered in various

types of cancer as well as neurodegenerative diseases including amyotrophic lateral sclerosis (ALS), and Alzheimer's disease (Geuens et al., 2016).

The SR proteins and hnRNPs are involved in multiple steps of splice site recognition and the splicing cycle, resulting in the regulation of alternative splicing in a positive or negative fashion. Typically, the SR protein family members were reported to promote the formation of the E complex by assisting the recruitment of U1 snRNP to the 5' SS or U2AF to the branch point of the target pre-mRNA, as well as activating the subsequent recruitment of U4/U5/U6 to mark complex B in the later step (Roscigno & Garcia-Blanco, 1995; Staknis & Reed, 1994; J. Y. Wu & Maniatis, 1993). The underlying molecular mechanisms of SRSF activity to promote such recruitment of the spliceosome components remain to be elucidated. Several studies reported the direct interaction between SR proteins and $U2AF^{35}$, U1-70K, and U4/5/6 tri-snRNPs, alongside the recognition of regulatory elements of pre-mRNA sequences at the RRMs (Roscigno & Garcia-Blanco, 1995; Staknis & Reed, 1994; J. Y. Wu & Maniatis, 1993). On the other hand, the inhibitory hnRNPs can antagonise the activity of SR proteins during the formation of the E complex (Cáceres et al., 1994). The underpinning mechanism here seems much simpler; for example, hnRNP I (also known as PTB) can inhibit the recruitment of $U2AF^{65}$ by competitively binding to the polypyrimidine tract (Lin & Patton, 1995; Singh et al., 1995). However, it is noteworthy that this splicing regulatory model is overly simplified, as both activator roles for hnRNPs and suppressor roles for SR protein have been documented (Buratti et al., 2007; Charlet-B et al., 2002; Förch et al., 2002; Gromak et al., 2003; Kanopka et al., 1996; Shin & Manley, 2002).

1.4.3 Alternative splicing in cancer

In the past decade, alternative splicing has been closely linked to tumour occurrence and progression, and thus splicing perturbations have been increasingly substantiated as another potential hallmark of cancer (Bonnal et al., 2020; Ladomery, 2013; Y. Zhang et al., 2021). In most cases, somatic mutations in the cancer genome lead to the disruption of splice sites and other cis-acting regulatory elements, or directly alter the expression or function of spliceosome components and splicing regulators (Bonnal et al., 2020; Ladomery, 2013; Y. Zhang et al., 2021).

Aberrant alternative splicing has been reported to contribute to the regulation of numerous events during cancer progression, such as proliferation, apoptosis, metastasis, and angiogenesis (Brown et al., 2011; David et al., 2010; Kuan et al., 2000; Wagner et al., 2019). For example, it has been demonstrated that the well-known oncoprotein c-MYC can up-regulate the expression of some hnRNPs including polypyrimidine tract-binding protein (PTB), hnRNP A1, and hnRNP A2 (David et al., 2010). Over-expression of such splicing regulators results in the preferential expression of PKM2, the embryonic isoform of pyruvate kinase (PKM), which is commonly re-expressed in many types of cancer and promotes tumour growth and proliferation by allowing cancer cells to undergo aerobic glycolysis (Christofk et al., 2008; David et al., 2010). In addition, some key regulators of apoptotic pathways are also influenced by alternative splicing. The most-studied example is B-cell lymphoma (BCL-x), which has two isoforms as a result of differential 5’ splice site selection in exon 2. The long isoform (BCL-xL) and short isoform (BCL-xS) are known to exhibit antagonistic biological outcomes, whereby BCL-xL can inhibit apoptosis, whilst BCL-xS has a pro-apoptotic activity (Kuan et al., 2000).

SRSF3, a member of the auxiliary splicing regulator SR protein family, represents another example of a splicing factor that plays a critical role in tumour-related pathways (Z. Zhou et al., 2020). The fundamental activities of this smallest SR protein member have been described not only in RNA splicing but also in other facets of RNA biology such as the regulation of RNA export, translation and polyadenylation (Bedard et al., 2007; Lou et al., 1998; Müller-McNicoll et al., 2016). In reflection of this cellular multi-functionality, the aberrant expression of SRSF3 has been reported in many types of cancer, such as colorectal, gastric, breast cancers, and skin cancers (Dewaele et al., 2016; Ke et al., 2018; J.-L. Wang et al., 2020; S. Zhu et al., 2016). Indeed, there exists a frequent correlation between SRSF3 over-expression and poor prognosis in patients, with an increasing body of evidence suggesting its relevance to numerous cellular events related to cancer progression (Dewaele et al., 2016; Ke et al., 2018; J.-L. Wang et al., 2020; S. Zhu et al., 2016). The potential oncogenic role of SRSF3 has been well documented in several studies in colorectal cancer cells, whereby SRSF3 regulates alternative splicing of genes including ArhGAP30, Caspase-2, PKM, and B7-H3 (Jang et al., 2014; J.-L. Wang et al., 2020; Z. Wang et al., 2012; C. Zhang et al., 2021). By favouring specific isoforms (i.e. S-ArhGAP30, Casp-2S, PKM2, and exon 4 retention of B7-H3), SRSF3 can act to suppress apoptosis (ArhGAP30 and Caspase-2); promote invasion (ArhGAP30), aerobic glycolysis and autophagy (PKM); and assist immune evasion (B7-H3), all which together contribute to the growth and proliferation of cancer cells (Jang et al., 2014; J.-L. Wang et al., 2020; Z. Wang et al., 2012; C. Zhang et al., 2021). Interestingly, over-expression of SRSF3 was found to induce autoregulation in cancer cells; SRSF3 can induce the inclusion of its own alternative exon 4 with an early stop codon, facilitating nonsense-mediated decay (J. Guo et al., 2015). This auto-splicing regulation is considered as part

of the regulation of cellular homeostasis, but given the oncogenic role of SRSF3, it may also act as a defensive mechanism against SRSF3-dependent tumorigenesis. Indeed, two hnRNP proteins, PTBP1 and PTBP2, can inhibit this AS event of SRSF3 and facilitate the expression of this SR protein at full length, indicating a mechanism which can be exploited by cancer cells (J. Guo et al., 2015).

With such deep relevance to cancer-related biological events, SRSF3 is emerging as an attractive potential therapeutic target. For example, a recent study provided evidence that indicated colorectal cancer cells treated with the SRSF3-specific small molecule inhibitor SFI003 had a significant suppression in their pro-growth and metastatic profiles (Y. Zhang et al., 2022). The molecular functions of this splicing factor in the context of cancer biology still require extensive analysis. Nonetheless, further understanding of the regulatory mechanisms of the SRSF3-mediated signalling pathways may contribute to the establishment of novel drug candidates and therapeutic strategies for the treatment of cancer.

1.4.4 Splicing, citrullination, and arginine methylation

With recent developments in techniques and analytical approaches to mass spectrometry, our understanding of post-translational modifications (PTMs) and their substrates has been greatly expanded. Notably, it is becoming evident that alternative splicing is significantly regulated by the PTMs on splicing factors, especially via citrullination and methylation at RG/RGG motifs. This sequence element is prevalent in RNA-binding proteins (RBPs) and is thought to play a key role in mediating protein-protein or protein-RNA interactions (Rajyaguru & Parker, 2012; Thandapani et al., 2013). Citrullination and methylation at the RGG motif are expected to modulate such interactions; by removing

the positive charges from arginine residues, or by steric hindrance and increased hydrophobicity, respectively (Christophorou, 2022; Pahlich et al., 2006).

A collection of recent large-scale proteomics studies has successfully profiled a large number of RGG-containing RBPs as being targets for both citrullination and arginine methylation (C. Chen et al., 2011; Fong et al., 2019; C.-Y. Lee et al., 2018; Lewallen et al., 2015; W.-J. Li et al., 2021; Lim, Lee, et al., 2020; Lim, Park, et al., 2020; Musiani et al., 2019; Radzisheuskaya et al., 2019; Tanikawa et al., 2018). Furthermore, direct crosstalk between the two modifications on target arginine residues has been identified for several splicing factors. For example, Splicing factor proline- and glutamine-rich (SFPQ; also known as PTB-associated splicing factor PSF) has been demonstrated as a target of PAD4 for citrullination and PRMT1 for asymmetric methylation within its N-terminal domain which contains three RGG repeats (Snijders et al., 2015). These modifications seem to antagonise each other and exhibit opposite molecular outcomes; arginine methylation of SFPQ leads to an increased association with mRNA, whilst citrullination results in mRNA dissociation (Snijders et al., 2015). Another example can be found in the study by Guo et al. The authors demonstrated that the human ribosomal protein S2 (RPS2) was citrullinated by PAD4 at the N-terminal arginine and glycine-rich repeat domain (Q. Guo et al., 2011), and the same region of this RGG-containing RBP was previously established as a target site of PRMT3 (Bachand & Silver, 2004; Swiercz et al., 2005, 2007). In summary, the interplay between PADs and PRMTs on target substrates is not an uncommon phenomenon for RNA-binding splicing factors, although to date, no one has provided direct evidence of mRNA splicing events being regulated by citrullination.

1.5 Research Objectives

Given that many RNA-binding factors have been discovered to be direct substrates of both PADs and PRMTs, it is conceivable that PAD4-dependent citrullination, and its crosstalk with arginine methylation, may play a profound role in the regulation of RNA processing. However, to date, there is little evidence demonstrating the genome-wide influence of citrullination beyond the level of transcription.

E2F1 is a known substrate for PAD4-driven citrullination, where this modification, together with PRMT-mediated arginine methylation, plays a crucial role in directing distinct biological outcomes by modulating the activity of this transcription factor (Cho et al., 2012; Ghari et al., 2016; Zheng et al., 2013). Previous work in our lab has revealed that citrullination of E2F1 can activate this transcription factor on pro-inflammatory genes, by assisting its protein-protein interaction with BRD4 (Ghari et al., 2016). Interestingly, one of the PAD4-targeted arginine residues of E2F1 (R109) is demonstrably a target of PRMT1 for arginine methylation (Zheng et al., 2013). This asymmetric arginine methylation event, which drives E2F1 activity towards the pro-apoptotic pathway, is not only expected to inhibit citrullination at the same residue, but will also antagonise arginine methylation at the neighbouring R111 and R113 residues of E2F1 by PRMT5 (Cho et al., 2012; Zheng et al., 2013). These symmetric methylation marks can instead channel E2F1 to drive the expression of genes involved in proliferation (Cho et al., 2012; Zheng et al., 2013). Additionally, it has been recently uncovered that the PRMT5-dependent modification can expand the activity of E2F1 beyond its conventional role as a transcription factor, transforming it into a regulator of alternative RNA splicing for its target genes. (Roworth et al., 2019). This is mediated by the multifunctional protein p100/TSN, which specifically recognises the arginine methylation marks at R111 and

R113 and allows E2F1 to interact with the splicing machinery (Roworth et al., 2019). Given the potential functional crosstalk between PAD4 and PRMTs over E2F1 activity, we reasoned that investigating the genome-wide impact of PAD4-mediated E2F1 citrullination would provide evidence for citrullination-regulated RNA processing events in cancer cells.

In this study, by performing an RNA sequencing experiment, we have demonstrated that the chemical inhibition of PAD4 and knockout of E2F1 can result in significant global impacts on the human cancer transcriptome at the level of alternative RNA splicing. Furthermore, by examining the splicing events validated by RT-qPCR, we have highlighted the mechanism by which this alternative splicing is regulated by the E2F1/PAD4 axis. This regulation involves the citrullination of E2F1, which may enhance its interaction with p100/TSN and the SR family splicing regulator SRSF3. Using RNA immunoprecipitation methods, we additionally uncovered that PAD4 plays a critical role to regulate the RNA binding affinity of SRSF3 in an E2F1-dependent manner. This may potentially underpin the molecular mechanism by which PAD4 and E2F1 together influence alternative splicing, by directing SRSF3 into distinct binding sites on target RNA and thus modulating its activity. In summary, we demonstrated a novel function of PAD4 to regulate alternative splicing in cancer cells, and propose a mechanism by which this alternative splicing is mediated by the interplay between E2F1 and SRSF3.

Chapter 2

Materials and Methods

2.1 Cell Lines, Culture, and Compound Treatment

All cell lines used in this study were maintained at 37 °C and 5 % CO_2 in a humidified incubator. HCT116 p53+/+ and p53-/- cell lines were maintained in Dulbecco's modified Eagle medium (DMEM) (Sigma), supplemented with 10 % (v/v) foetal calf serum (FCS) (PAN biotech) and 1 % (v/v) Penicillin-Streptomycin (Pen-Strep) (Gibco). Tet-On inducible U2OS cells were maintained in DMEM medium supplemented with 10 % (v/v) tetracycline-free FCS, 1 % (v/v) Pen-Strep, 100 µg/ml Geneticin (G418) (Santa Cruz Biotechnology), and 150 µg/ml hygromycin B (TOKU-E). Suspension HL60 cells were maintained in RPMI-1640 medium (Sigma) supplemented with 10 % (v/v) FCS and 1 % (v/v) Pen-Strep. Adherent cells were usually passaged every 3-4 days at a confluency of 70-90 % by washing with sterile PBS buffer and then digesting with Trypsin (Lonza). HCT116 p53+/+ and p53-/- CRISPR/Cas9 E2F1 knockout (E2F1cr) cells were generated from cell lines (ATCC) by a previous member in our lab as previously outlined (Ran et al., 2013). E2F1 CRISPR single guide RNA sequence used is 5'-GCATTCTTCTTCTG GCTGGG-3'.

For PAD4 inhibition, cells were treated with 10 µM GSK484 (Sigma) for 72 hours unless otherwise stated. GSK484 is dissolved in Dimethyl sulfoxide (DMSO) (Sigma), and thus control groups were treated with equal volumes of DMSO alone. For the Tet-on cell system, cells were treated with 1 µg/ml doxycycline for 24 hours to induce expression of Tet-responsive PAD4. All drugs were directly added to the culture medium. To modulate intracellular calcium ion concentration, cells were treated with 5 µM calcium ionophore (A23187) (Sigma-Aldrich) for 30 minutes shortly before harvesting for immunoblotting.

2.2 Antibodies

The list of primary and secondary antibodies used in this study were shown below in Table 2-1.

Table 2-1 List of antibodies used in this study.

Antigen	Name (Catalogue no.)	Species	Supplier
E2F1	KH-95 (sc-251)	Mouse	Santa Cruz Biotech
E2F1	(3742S)	Rabbit	Cell Signaling
p53	DO-1 (sc-126)	Mouse	Santa Cruz Biotech
pRB	4H1 (9390)	Mouse	Cell Signaling
PAD4	ab128086	Mouse	Abcam
PAD4	ab50247	Rabbit	Abcam
SRSF3	7B4 (33-4200)	Mouse	Invitrogen
SND1 (p100/TSN)	(A302-883A)	Rabbit	Bethyl
DDX39A	H-6 (sc-271395)	Mouse	Santa Cruz Biotech
B-actin	AC-74 (A2228)	Mouse	Sigma-Aldrich
Anti-Citrullined (modified)	(17-347B)	Rabbit	Upstate
Histone H3	ab8896	Rabbit	Abcam
Histone H3 (citrulline R17+R2+R8)	ab5103	Rabbit	Abcam
Sm	Y12 (NB600-546)	Mouse	Novus
FLAG	M2 (F1840)	Mouse	Sigma-Aldrich
HA	16B12 (MMS-101P)	Mouse	Enzo
HRP-conjugated secondary IgG	AP160P / AP187P	Mouse / Rabbit	Merck
TrueBlot® IgG HRP	18-8817-33 / 18-8816-33	Mouse / Rabbit	Rockland
Polyclonal IgG (for IP)	I5381 / I5006	Mouse / Rabbit	Sigma-Aldrich

2.3 DNA Plasmid Transformation and Isolation

Approximately 100 ng of plasmid DNA was mixed with 25 µl of BL21(DE3)pLysS Competent Cells (Agilent Technologies) and incubated on ice for 20-30 minutes. The mixture was then heat-shocked by being placed in a 42 °C water bath for 45 seconds, followed by a 2-minute incubation on ice. 225 µl LB (Sigma) media without antibiotics was added to the competent cell/DNA mixture, and they were incubated at 37 °C for an hour to allow the bacterial cells to metabolically recover and produce the antibiotic resistance proteins encoded in the plasmid backbone. The transformation mixture was spread onto a 100 mm LB agar (Sigma) plate containing the appropriate antibiotics (100 µg/ml ampicillin for the plasmids used in this study) and incubated at 37 °C overnight. The plasmid DNA was purified using Maxiprep kits (Invitrogen) according to the manufacturer's instructions.

2.4 DNA Plasmid Transfection

GeneJuice® Transfection Reagent (Merck) and plasmid DNA of interest (1-2 µg per 100 mm dish and 2-4 µg per 150 mm dish) were separately mixed with reduced serum medium Opti-MEM® (Gibco) and incubated for 5 minutes at room temperature (RT). 3 µl of GeneJuice was used per 1 µg plasmid DNA. The transfection reagent/medium mixture was then added to the plasmid/medium mixture, mixed well, and incubated at RT for 15 minutes. The final mixture was then added dropwise to the cells, which were typically transfected at 30-50 % confluency and right after seeded to a new dish. The cells were harvested 36-72 hours after transfection. The DNA plasmids used in this study was summarised in Table 2-2.

Table 2-2 List of plasmids used in this study.

Plasmid	Tag
pcDNA 3.1 E2F1 (WT)	HA / Flag
pcDNA 3.1 E2F1 (R4K)	HA
pcDNA 3.1 PAD4	HA / Flag
pcDNA 3.1 SRSF3	Flag
pcDNA 3.1 empty vector	N/A

2.5 Small Interfering RNA (siRNA) Transfection

Oligofectamine™ Transfection Reagent (Invitrogen) and siRNA (1-2 μg per 100 mm dish and 2-4 μg per 150 mm dish) were separately mixed with reduced serum Opti-MEM medium and incubated at RT for 5 minutes. The mixture was then added to the plasmid/medium mixture, mixed well, and incubated at RT for 15 minutes. The final mixture was then added dropwise to the cells, which were typically transfected at 30-50 % confluency and right after seeded to a new dish. The cells were harvested 36-72 hours after transfection. The siRNA sequences used in this study were summarised in Table 2-3.

Table 2-3 List of siRNAs used in this study.

siRNA	Sequence (sense strand) 5'-3'
siSRSF3	Dharmacon on-target plus SMART
siPAD4	GGUCCUGCUACAAACUGUUTT
siGFP (Negative control)	AGCUGACCCUGAAGUUCUU

2.6 RNA Isolation

1 ml TRIzol™ reagent (Invitrogen) was added to 30-50 mg of cells, mixed well until homogenised, and incubated for 5 minutes at RT to permit a complete dissociation of nucleoprotein complex. 200 µl of chloroform per 1 ml TRIzol was then added for the lysate, and vigorously hand-shaken for 15 seconds. After a 2-minute incubation, the mixture was centrifuged at 13,000 rpm for 15 minutes. The top aqueous phase was then transferred to a new tube (typically 50-60 % of the volume of TRIzol added) and mixed with 500 µl 2-Propanol per 1 ml TRIzol used for lysis. For isolation of RNA with small quantities (e.g. RNA immunoprecipitation), 3-5 µl of GlycoBlue™ co-precipitant (Invitrogen) was added to increase the visibility of the RNA pellets in the following steps. The mixture was then incubated, either at RT for 10 minutes (when the amount of RNA was expected to be large enough), or at -20 °C overnight (for small quantities of RNA isolated). The precipitating solution was then centrifuged at 13,000 rpm at 4 °C for 15 minutes, resulting in an RNA pellet visible at the bottom of tube. The supernatant was discarded with a pipette, and the pellet was washed with 1 ml of 70 % ethanol per 1 ml TRIzol used in lysis. After a centrifugation at 7,500 rpm at 4 °C for 15 minutes, the supernatant was removed with a pipette and the pellet was air-dried for 5-10 minutes. For RNA sequencing and RT-qPCR, the RNA pellet was resuspended in 20-50 µl of RNase-free water by pipetting up and down, and concentration was determined using Nano-drop. For RNA immunoprecipitation, the pellet was resuspended in 11 µl RNase free water and directly used for reverse transcription.

2.7 RNA Sequencing (RNA-seq)

Wildtype E2F1 (WT-E2F1) and CRISPR-knockout E2F1 (E2F1cr) HCT116 cells treated with 10 μM PAD4 inhibitor GSK484 (Sigma-Aldrich) or DMSO for 72 hours, and total RNA was isolated using TRIzol as described in section 2.6. RNA-seq was performed by BGI Genomics (https://www.bgi.com) (Figure 2-1). RNA samples (biological triplets from each of four experimental conditions) were quality-controlled using Agilent 2100 Bioanalyzer (Agilent RNA 6000 Nano Kit) for RNA concentration, RNA integrity number (RIN) value, 28S/18S and the fragment length distribution. mRNA was enriched and purified from total RNA samples using Oligo dT selection, and then fragmented for first-strand / second-strand cDNA synthesis. This results in generating partially double-stranded DNA with single-stranded ends. Partial single strands on the 5' end of DNA fragments were end-repaired by polymerase, and those at the 3' end were removed by exonuclease and added with a single adenine. Processed samples were then ligated with the sequencing adapters and amplified in polymerase chain reaction (PCR) with Agilent 2100 Bioanalyzer and ABI StepOnePlus Real-Time PCR System. Finally, RNA-seq was performed using Illumina HiSeq Platform, resulting in 5.12 Gb reads generated per sample.

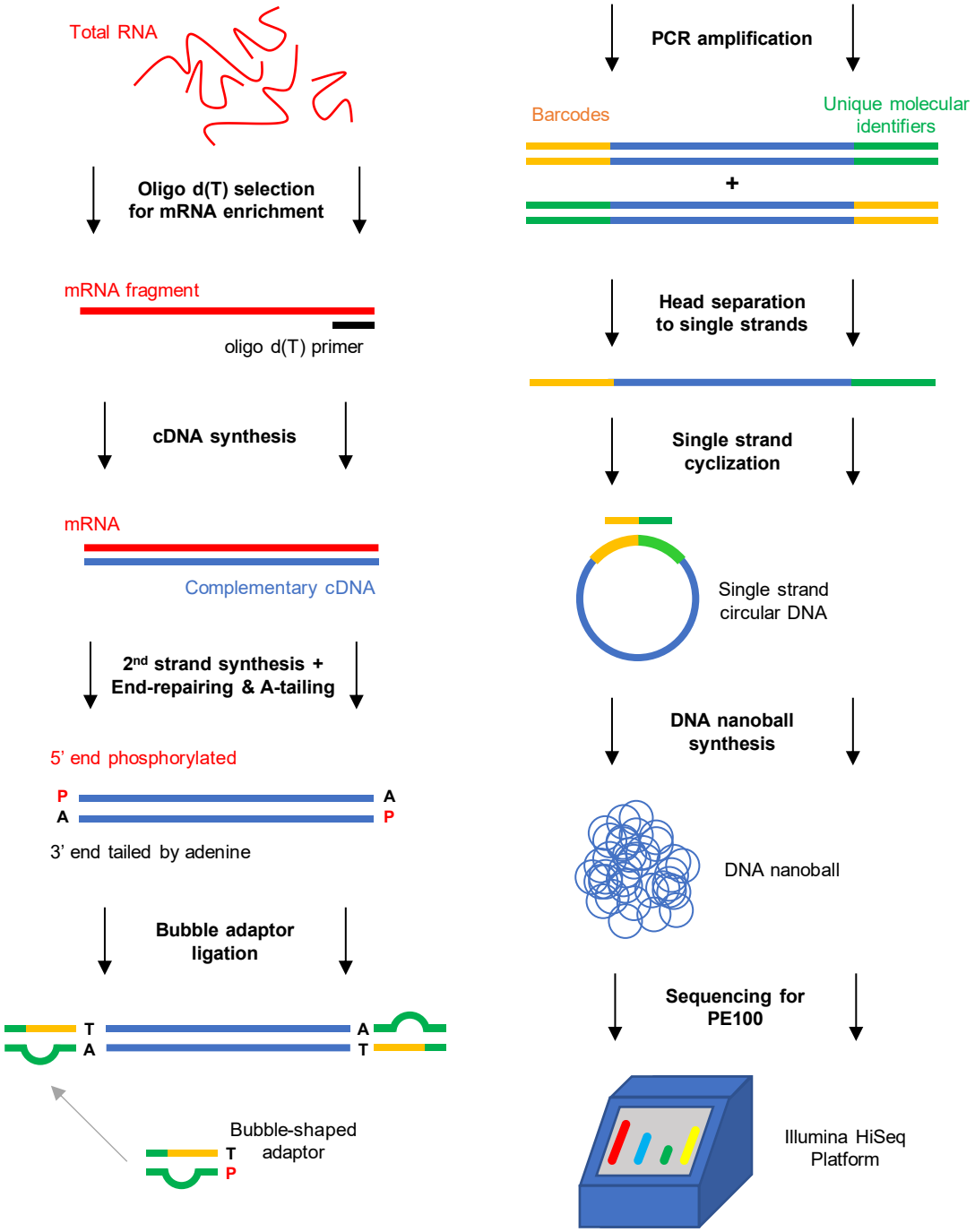


Figure 2-1 RNA-seq pipeline.

Diagram showing the experimental pipeline of RNA sequencing performed by BGI Genomics. mRNA fragments were enriched by oligo d(T) selection, and reverse transcribed to complementary DNA (cDNA). Double-stranded cDNA was then end-repaired and tailed by adenine, followed by ligation of bubble-shaped adaptors. PCR is then carried out to amplify these adaptor-ligated cDNA. Amplified double-stranded cDNA was heat-shocked to be dissociated to single-stranded, and then formed into ss circular DNA. The DNA nanoball was then formed from sscDNA and sequenced with the Illumina HiSeq platform.

2.8 RNA-seq Data Analysis

FASTQ files were generated from biological triplets of WT-E2F1 and E2F1cr HCT116 cells treated with either GSK484 or DMSO (control). The raw data was processed using TrimGalore (ver.0.4.3) to remove adaptors and low-quality base readings. The trimmed reads were then aligned to the human genome build hg19 with STAR aligner (ver.2.7) with two mismatched allowed. For all the analyses, three treatment conditions (WT-GSK484, E2F1cr-DMSO, and E2F1cr-GSK484) were analysed with respect to the control condition (WT-DMSO) unless otherwise stated. The quality of the sequence data was analysed by visualising read assignment and distance matrix heatmap (Figure 3-4) using the DESeq2 R Bioconductor package (ver.1.25.17).

Differential gene expression analysis based on the negative binomial distribution was carried out using the DESeq2 R Bioconductor package (ver.1.25.17) with the aligned reads. Statistical significance was assessed by calculating the adjusted *P*-value using the Benjamini-Hochberg procedure, in order to reduce the false discovery rate (FDR). The genes were considered to be differentially expressed when their adjusted *P*-value was less than 0.01, and further filtered for two-fold change in the absolute expression level. Alternative splicing analysis was performed with rMATS software (ver.4.0.1) (https://rnaseq-mats.sourceforge.net/). For each splicing event, $\Delta\psi$ (delta PSI) was calculated to show a difference in percent spliced in between conditions (drug-treated group against a control group). The raw data was filtered for the FDR to be less than 0.01.

For the lists of genes identified through DESeq and rMATS analyses, the parametric Gene Set Enrichment Analysis (pGSEA) was carried out with the R PGSEA package (ver.1.58) for the GO vocabulary from the Molecular Signatures Database (ver.6.2) of the Broad Institute, as previously described (Barczak et al., 2020). A linear model was applied to

employ the limma package (v.3.44.0) followed by empirical Bayesian analysis to determine concepts associated with significant differences between treated and untreated samples. Differences were considered significant if the adjusted *p*-value, calculated using the Benjamini-Hochberg method, to minimize the false discovery rate, was < 0.005.

For ENCODE ChIP-seq data analysis, ChIP-seq data for E2F1 tracks from the ENCODE project (http://genome.ucsc.edu/ENCODE/) was analysed for the 'Txn factor ChIP track', 'ENCODE 3 TFBS Track', 'Uniform TFBS track' and 'SYDH TFBS track' to display E2F1 ChIP-seq peaks or signals as appropriate.

2.9 Reverse Transcription and cDNA Synthesis

1 µg RNA was added to a nuclease-free microcentrifuge tube with 1 µl of appropriate primers (oligo(dT)$_{20}$ (Invitrogen) for mRNA and random hexamer primers (Invitrogen) for RNA immunoprecipitation), 1 µl of 10 mM dNTP mix (10 mM each of dATP, dGTP, dCTP and dTTP), and nuclease-free water to make the total volume to 13 µl. The mixture was incubated at 65 °C for 5 minutes and then cooled down on ice for at least 1 minute.

The contents of tubes were collected by brief centrifuge, and then mixed with 4 µl 5X First-Strand buffer (Invitrogen), 1 µl 0.1M DTT (Invitrogen), 1 µl RNaseOUT™ RNase inhibitor (Invitrogen), and 1 µl SuperScript™ III Reverse Transcriptase (Invitrogen). Temperature cycling was then carried out as per the manufacturer's instruction for Moloney murine leukaemia virus reverse transcriptase (MMLV-RT). The synthesised cDNA was diluted in nuclease-free water and used for quantitative PCR.

2.10 Quantitative PCR and Data Analysis

RNA was isolated as described in sections 2.6 and 2.16, and cDNA was prepared as described in section 2.9. Alternatively, chromatin was isolated in ChIP as described in

section 2.15. Technical triplets (unless otherwise stated) from each biological repeat were used in quantitative reverse-transcription PCR (RT-qPCR) or quantitative PCR (qPCR) with the Brilliant III Ultra-Fast SYBR® Green QPCR Master Mix (Agilent) and appropriate primer pairs on the AriaMX real-time qPCR system (Agilent) as per the manufacturer's instructions.

qPCR data analysis was carried out using the $\Delta\Delta C_t$ method where raw C_t (cycle threshold) values from the experiment were normalised against C_t values for housekeeping genes (typically GAPDH, for RT-qPCR) or IgG (for ChIP and RIP), followed by another normalisation against input (for ChIP and RIP) or control sample (for RT-qPCR). Subsequently, $\Delta\Delta C_t$ values were converted and presented as average (mean) fold enrichment from three biological repeats (technical triplets within each repeat). Error bars indicate standard errors (SE). The formula for the calculation of Fold Enrichment were described below.

RT-qPCR primers used in this study were listed in Table 2-4. These primers, as well as ChIP primers (Table 2-5) and RIP primers (Table 2-6) were all designed using the Primer-BLAST tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) to avoid GC-rich regions and potential secondary structures (Ye et al., 2012). Primers were designed so that PCR products will have a size around 100 nt and melting temperature at around 60 °C.

Equation: Calculation of Fold Enrichment for RT-qPCR

$\Delta C_t = C_t$(Gene of Interest) - C_t(Housekeeping gene e.g. GAPDH)

$\Delta\Delta C_t = \Delta C_t$(Treated sample) - ΔC_t(Control sample)

Fold Enrichment = $2^{-\Delta\Delta Ct}$

Table 2-4 List of RT-qPCR primers used in this study.

RT-qPCR primer	Forward primer sequence 5'-3'	Reverse primer sequence 5'-3'
EXOC4 exon 7 (inclusion)	ACCAACCAAGGTTGCTTCT	CAGGTATCCCAGGACCACAG
EXOC4 exon 7 (exclusion)	AGAACCAACCAAGATGCTATTAACT	GTCCAGTGCTGGCATAGCTT
RBM25 exon 2 (inclusion)	AGTGGAGCGCACTCGTAAC	AAGACCCCGCCGCTGA
RBM25 exon 2 (exclusion)	AGTGGAGCGCACTCGTAAC	CTGCAGCAGTCCGCCG
SPIN1 exon 3 (inclusion)	TGGGATTACAGGAAGTTTATCTCG	ACTGAAAAACTGCAGCACAGA
SPIN1 exon 3 (exclusion)	GGTCAGCAGGAAGTTTATCTCG	ACTGAAAAACTGCAGCACAGA
SNAP23 intron 3 (inclusion)	GCCATTGAGGTAAGAAAATGTTAG	CAGCCAGAAGTATTAGGATCGC
SNAP23 intron 3 (exclusion)	TGGGTTTAGCCATTGAGTATTGGA	CACACTTTAACATAGCAGCCCCC
EXOC4 (total expression)	GGGGTCCTGATGACAACTTAAT	GACGGTAATGTTTCTGGCTCC
RBM25 (total expression)	CTGCTCCAACTGTCTTAGTACCC	GGATGATCCTTTCTTGCGCC
SPIN1 (total expression)	TCCGCGGCGAGGGAG	CCGCCGTCGCTCGTTC
SNAP23 (total expression)	CAAAAGGAACAACTAAACCGCA	CACAGCATTTGTTGAGTTCTGT
GAPDH	CCATCAATGACCCCTTCATTGACC	GAAGGCCATGCCAGTGAGCTTCC

2.11 MTT Assay

5 mg/ml MTT solution was prepared in PBS buffer, filter-sterilised, and stored at -20 °C. 10 µl MTT solution was added to cells in each well of a 96-well plate and incubated at 37 °C for 10 minutes. After incubation, cells were treated with 110 µl MTT solvent (20 % (v/v) sodium dodecyl sulphate (SDS), 50 % (v/v) dimethylformamide) for 15 minutes at RT, and the absorbance was measured at OD = 590 nm using the Sunrise™ plate reader (Tecan).

2.12 Bradford Assay

A calibration curve was constructed by dispensing various concentrations (0-10 µg) of Bovine Serum Albumin (BSA) (Sigma) in Bradford reagent (Sigma) and measuring absorbance at OD = 595 nm with the Sunrise plate reader. Relative reading was plotted against the concentration of standard and was used as a reference for subsequent protein concentration measurements.

2.13 Immunoblotting

For adherent HCT116 cells, the media was removed, and cells were washed with PBS buffer once. Cells were then treated with trypsin at 37 °C for 5-10 minutes and collected. The cell pellet was washed once with PBS and lysed in TNN buffer (50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1 mM EDTA, 1 % (v/v) IGEPAL® CA-630 (Sigma-Aldrich), supplemented with 1 mM NaF, 1 mM Na_3VO_4, 0.1% Proteinase inhibitor cocktail [1 M PMSF, 1 µg/ml leupeptin, 1 µg/ml aprotinin, 1 µg/ml pepstatin A], 0.05% AEBSF) on ice for 30 minutes. The lysate was then centrifuged at 13,000 rpm for 15 minutes at 4 °C to remove the cell debris. The supernatant containing intercellular materials was transferred to a fresh tube. The protein concentration was then determined by Bradford assay (Sigma).

The appropriate volume was mixed with sodium dodecyl sulphate (SDS) loading buffer (62.5 mM Tris-HCl (pH 6.8), 2 % (v/v) SDS, 8% glycerol, 0.1% bromophenol blue, 300mM β-mercaptoethanol) so that equal amounts of proteins (typically 30-100 ug) were loaded for each experiment. The mixture was boiled at 95 °C for 5 minutes to denature proteins.

Samples were loaded to SDS-polyacrylamide gel and proteins were separated by size with electrophoresis (SDS-PAGE). Gels consist of acrylamide (Serva) of different percentages (8-14 %), 375 mM Tris-HCl (pH 8.8), 0.1 % (v/v) SDS, 0.15 % (v/v) ammonium persulphate (APS) (Sigma), 0.0 7% (v/v) tetramethyl ethylenediamine (TEMED) (Sigma), and top stacking gel of 5 % (v/v) acrylamide,125 mM Tris-HCl (pH 6.8), 0.1 % (v/v) SDS, 0.15 % (v/v) APS, 0.28 % (v/v) TEMED. The protein ladder (Thermo Scientific) was loaded alongside the samples to estimate the molecular weight of proteins and monitor the progress of SDS-PAGE. Separated proteins in the gel were then transferred onto PVDF membrane (Merck), which was activated by being soaked in 100 % methanol for a few seconds, using Mini Trans-Blot® Cell (Bio-Rad) at 300 mA for 1 hour or at 150 mA overnight. Membranes were stained with Ponceau S to make sure the successful transfer of protein, and then blocked in 5 % (v/v) skimmed milk in PBS-Tween 0.3 % (PBST) for 1 hour at RT to prevent further non-specific protein binding.

Blocked membranes were incubated with primary antibody diluted in a blocking solution at 4 °C overnight and washed three times with PBST (intervals of 10 minutes). Membranes were then incubated with appropriate secondary antibody diluted in a blocking solution for 1 hour at RT, followed by another set of PBST washing three times (10-minute intervals). Membranes were incubated with enhanced chemiluminescence (ECL) substrate (2.5 mM luminol, 396 μM coumaric acid, 100 mM Tris-HCl (pH 8.0),

0.02 % (v/v) hydrogen peroxide) (Bio-Rad) for 2-3 minutes, and the signals were exposed to Fuji Medical X-ray films (Fujifilm) and developed and visualised with the CP1000 automatic developer (AGFA). For quantification of protein expression, the band intensity was measured using ImageJ software (Schneider et al., 2012).

2.14 Co-Immunoprecipitation (IP)

Cells were harvested and lysed as described in section 2.13. 5-10 % of cell lysate was kept for a loading input. Standardised samples were incubated with primary antibody (1-3 µg per 1 ml sample) or appropriate control IgG antibody at 4 °C overnight on a tube rotator, followed by incubation with 50 µl TrueBlot beads (Rockland) at 4 °C for 2 hours. Beads bound with proteins were then washed three times with lysis buffer, and resuspended and boiled in 50-100 µl SDS loading buffer at 95 °C for 5 minutes to denature and dissociate proteins from beads. Proteins in the SDS loading buffer, together with the input samples collected earlier, were analysed by immunoblotting as described in section 2.13.

2.15 Chromatin Immunoprecipitation (ChIP)

Cells were harvested as previously described. The cell pellet was then suspended in 3 ml of PBS and incubated with 200 µl of ethylene glycol bis-succinimidyl succinate (EGS) (Thermo Scientific) (10 mg/ml in DMSO) for 30 minutes at RT to crosslink proteins. The mixture was then mixed with 90 µl of formaldehyde (Sigma) for further cross-linking. After a 15-minute incubation, the reaction was stopped by adding 1.1 ml of 0.5 M glycine and being incubated for another 5 minutes at RT. Crosslinked cells were washed with cold PBS twice, and resuspended in 1 ml Lysis Buffer I (10 mM Tris (pH 8.0), 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, and 0.1 % (v/v) proteinase inhibitor cocktail).

Following a 10-minute incubation at RT, the cells were again collected by centrifugation and resuspended in 2 ml of Lysis Buffer II (10 mM Tris (pH 8.0), 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.5 % (v/v) sodium lauroyl sarcosinate, 0.1 % (v/v) sodium deoxycholate, and 0.1 % (v/v) proteinase inhibitor cocktail). The lysate was then sonicated using Bioruptor Pico sonication device (Diagenode) for 14 cycles of 10 seconds on / 10 seconds off at 4 °C. The sonicated sample was then mixed with 110 µl of 10 % IGEPAL® CA-630 and centrifuged at 13,000 rpm for 15 minutes at 4 °C to remove debris. The supernatant was collected and pre-cleared at 4 °C for 1-2 hours with appropriate IgG and A/G beads (pre-blocked with 1 mg/ml BSA and 400 µg/ml salmon sperm DNA).

The protein concentration of the pre-cleared sample was determined by Nanodrop and 5-10 % of the sample was collected for input. The normalised amounts (typically 1-2 mg and made up to 1 ml with lysis buffer) were incubated with 3 µg antibody or control IgG overnight at 4 °C, followed by another 2-hour incubation with pre-blocked A/G beads at 4 °C. Proteins-bound beads were washed with low-salt buffer (20 mM Tris (pH 8.0), 150 mM NaCl, 2 mM EDTA, 1 % (v/v) Triton, 0.1 % (v/v) SDS, and 0.1 % (v/v) protease inhibitor cocktail) twice, LiCl buffer (10 mM Tris (pH 8.0), 250 mM LiCl, 1 mM EDTA, 2 % (v/v) IGEPAL® CA-630, 2 % (v/v) sodium deoxycholate, and 0.1 % (v/v) protease inhibitor cocktail) four times, and TE buffer (10 mM Tris (pH 8.0), 1mM EDTA, and 0.1 % (v/v) protease inhibitor cocktail) twice. Protein-DNA complexes were eluted from beads by being incubated in 500 µl Elution Buffer (1 % SDS, 0.1 M sodium bicarbonate) at 65 °C for 30 minutes. The elute was collected and added with 20 µl 5M NaCl, 10 µl 0.5 M EDTA, 20 µl 1M Tris (pH 6.5), and 1 µl RNase A before incubated at 55 °C for 3 hours. The sample was then reverse-crosslinked at 65 °C overnight, followed by another 3-hour incubation at 55 °C with 10 µl of 10 mg/ml proteinase K to digest proteins. DNA

from the sample was then purified using QIAquick PCR Purification Kit (Qiagen) as per the manufacturer's protocol. ChIP primers used in this study are listed in Table 2-5.

Table 2-5 List of ChIP primers used in this study.

ChIP primer	Forward primer sequence 5'-3'	Reverse primer sequence 5'-3'
EXOC4	AACTTGCTCCCTTAGTCCCG	CGCAGGAGTCTCCACCAATG
SPIN1	GCGGACCTTGTAAGTTGCAG	GAGACGGCGCATCAGGAAAG
SNAP23	CGAACGGGAAGTGAGCAGG	CAACTCGGACACCCCAACA
CDC6	GGCCTCACAGCGACTCTAAGA	CTCGGACTCACCACAAGC
TK	TCCCGGATTCCTCCCACGAG	TGCGCCTCCGGGAAGTTCAC
CDC25A	TCTGCTGGGAGTTTTCATTGACCTC	TTGGCGCCAAACGGAATCCACCAATC

2.16 RNA Immunoprecipitation (RIP)

Cells were harvested as previously described and lysed in NETN Buffer (25 mM Tris (pH 8.0), 150 mM NaCl, 5 mM EDTA, 0.5 % (v/v) NP-40, supplemented with 0.1 % (v/v) RNaseOUT). The lysates were either kept on ice for 30-60 minutes for immediate immunoprecipitation or frozen at -80 °C until ready to process further. The lysate was sonicated as described in section 2.15, and centrifuged at 13,000 rpm for 15 minutes to remove cell debris. 5-10 % of the sample was kept separately for input. The standardised amounts of sample (made up to 1 ml) were incubated with 5 µg antibody of interest or appropriate control IgG at 4 °C overnight on a rotator, followed by incubation with 25 µl Pierce™ Protein A/G Magnetic Beads (Thermo Scientific) at 4 °C for 2-3 hours on a rotator. Protein-bound beads were washed with NETN Buffer four times. A small fraction (typically 5-10 %) of beads was taken from each sample and proteins were eluted in the SDS loading buffer to confirm a successful immunoprecipitation of target protein(s). The remaining beads and input solution were directly added by 1 ml TRIzol, and RNA was extracted as described in section 2.6. Purified RNA was reverse-transcribed to cDNA as

described in section 2.9, and analysed in qPCR as described in section 2.10. RIP primers used in this study are listed in Table 2-6.

Table 2-6 List of RIP primers used in this study.

RIP primer	Forward primer sequence 5'-3'	Reverse primer sequence 5'-3'
EXOC4 (set 1)	GGTGGCAGACAGTGGCTATC	CACTCCCACCTACCTTGGTT
EXOC4 (set 2)	TTGTTTATGCCACTAGGTTGCT	GCAGCGGCTACAGCATTAAA
EXOC4 (set 3)	TGATTTATCTGGTTGTACAGATGC	CTGATGGTTCAGAGGCCGTA
EXOC4 (set 4)	GGGATACCTGCAGGACACTG	TCTTCACCCATACATCTGCCA
U1 snRNA	GGCAGGGGAGATACCATGATC	CCCCCACTACCACAAATTATGC
U4 snRNA	CTTTGCGCAGTGGCAGTATC	CAGTGCCGACTATATTGCAAGTC
U5 snRNA	ATACTCTGGTTTCTCTTCAGATCGC	CTTGCCAAAGCAAGGCCTC
PKM	AATACGACTCACTATAGGGATAGC TCGTGAGGCTGAGGC	GCTCGATCGAGGCGTGCTAGG GGAGCAACATCCGTC

2.17 Flow Cytometry

FACS cell cycle analysis was performed using propidium iodide (PI) (Abcam) as per the manufacturer's protocol. Cells were harvested as previously described and washed twice with cold PBS Buffer. The cell pellet was then re-suspended in 3 µl 70 % ethanol (in PBS Buffer), mixed well, and fixed overnight at -20 °C. The mixture of 450 µl of 1 mg/ml PI and 50 µl RNase A in 10 ml PBS was prepared, and cells were treated with 350 µl of this PI mixture for 45 minutes at RT. The stained samples were run and analysed using BD Accuri™ C6 Plus Flow Cytometer (BD Biosciences).

2.18 Immunoprecipitation-Mass Spectrometry

HCT116 p53-/- cells were treated with DMSO or GSK484 (10 µM, 72 hours) and HCT116 p53-/- E2F1cr cells were treated with DMSO. The lysates from 20 x 150 mm dishes for each condition were prepared in TNN buffer and pre-cleared for 1 hour with pre-blocked A/G beads and 3ug mouse IgG. The extracts were then immunoprecipitated twice to ensure E2F1 pull-down; 1st IP with 5 µg anti-E2F1 antibody (KH-95) and 100 µl

protein A/G beads, and flowthrough then immunoprecipitated a second time with a further 5 µg anti-E2F1 antibody and 100ul protein A/G beads. Beads bound with proteins were then washed four times with the TNN lysis buffer, and eluted in 100 µl SDS buffer (4% SDS with DTT and Tris (pH 6.8)) at 95 °C for 10 minutes to denature and dissociate proteins from beads. Elutes were combined and concentrated to up to 30 µl volume in a speed-vac for use on S-Trap columns following the manufacturer's instructions.

Tryptic peptides were analysed by liquid chromatography tandem mass spectrometry (LC-MS/MS) using the Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (Thermo Fisher) connected to Ultimate 3000 UHPLC system (Thermo Fisher). Briefly, the samples were loaded and separated on a trap column (PepMapC18 (Thermo Fisher); 300µm x 5mm, 5µm particle size) using a gradient from 2 to 35% (v/v) solvent B in 5 % DMSO in 60 mins. The spectra were acquired with cycle times of 1 second. The MS1 spectra were acquired using the Orbitrap at 120K resolution, with a scan time from 400 to 1500 m/z, an automatic gain control (AGC) target value of 4e5, and using s-lens RF30. The MS2 spectra were acquired using the Orbitrap at 30k resolution, quadrupole isolation of 1.6, AGC target value of 5e4, maximum injection time of 54 ms.

The data was searched and labelled using the PEAKS-X (Bioinformatics Solutions Inc.) to infer peptide and protein IDs, and the label-free quantitation (LFQ) analysis was performed using the Progenesis QI software (Nonlinear Dynamics). After the normalisation process against all proteins with an identity, protein enrichment in the treatment condition (WT cells treated with GSK484) was analysed with reference to the background condition (E2F1cr cells) as well as the control condition (WT cells treated with DMSO), by relative abundances (log2). Those from which more than a single peptide was enriched, with two-sample Student's T-test (combined with Permutation FDR) *p*

values < 0.05 (-log p > 1.301), were filtered. Subsequently, Gene Ontology (GO) analyses were performed for these enriched proteins using the ShinyGO bioinformatic tool (http://bioinformatics.sdstate.edu/go/) (S. X. Ge et al., 2020) for GO molecular function terms and KEGG pathway terms.

2.19 Statistical Analysis

Statistical analyses were performed using a two-tailed, unpaired Student's *t*-test with Excel software (Microsoft) and GraphPad Prism 9 (GraphPad Software). Sample size of ≥ 3 was tried to be obtained where possible, unless otherwise stated. Data were displayed as means with standard deviations (SD) unless otherwise stated. *P*-values < 0.05 were considered significant and are labelled by *p < 0.05, **p < 0.01, ***p < 0.001, and ****p ≤ 0.0001.

Chapter 3

Genome-wide Analysis of PAD4 Inhibition and E2F1 Knockdown via RNA-seq

3.1 Introduction

Previous work in our research group has established that the multifaceted activity of E2F1 is largely regulated by its post-translational modifications (PTMs) including arginine methylation and citrullination (Cho et al., 2012; Ghari et al., 2016; Zheng et al., 2013). Arginine methylation is mediated by a family of proteins called protein arginine methyltransferases (PRMTs), and E2F1 is a known substrate of two PRMT members; PRMT1 asymmetrically di-methylates E2F1 at the arginine residue R109, whilst PRMT5 symmetrically di-methylates R111 and R113 (Cho et al., 2012; Zheng et al., 2013). These two types of methylation events catalysed by distinct PRMTs were discovered to be mutually exclusive and result in antagonistic outcomes. On the one hand, PRMT1-dependent methylation blocks methylation by PRMT5 and channels E2F1 into the apoptotic pathway. On the other hand, PRMT5-driven methylation can promote cell growth and proliferation while antagonising PRMT1-mediated methylation (Cho et al., 2012; Zheng et al., 2013).

In addition to methylation, another work from our research group has elucidated that E2F1 is also targeted for citrullination (Ghari et al., 2016). This modification is catalysed by another family of proteins termed peptidyl arginine deiminases (PADs), and the described study revealed that several arginine residues of E2F1, including R109 targeted by PRMT1, can also be citrullinated by PAD4 *in vivo* and in cells (Ghari et al., 2016). This citrullination appears to play a critical role in enhancing the chromatin association of E2F1, especially to the promoters of pro-inflammatory cytokine genes. The authors demonstrated that this localisation was mediated by bromodomain-containing protein 4 (BRD4), which primarily recognises the acetylated lysine residues of E2F1. Citrullination of E2F1 can assist acetylation to augment this protein-protein interaction, leading to the

transactivation of inflammatory genes (Ghari et al., 2016). This study uncovered a novel function of the E2F1/PAD4 axis in inflammatory pathways as well as indicated a potential interplay between PAD4 and PRMT1 sharing the same target arginine residue. However, the effect of PAD4-mediated citrullination was only examined within the context of the conventional E2F1 activity in a transcription-based mechanism.

Interestingly, it was recently demonstrated that E2F1 may also undertake a pivotal role in gene regulation at the level of RNA processing, beyond its classical function as a transcription factor (Roworth et al., 2019). Mechanistically, PRMT5-driven methylation event allows E2F1 to interact with the p100/TSN protein, whose activity has been heavily implicated in the regulation of mRNA splicing by facilitating the sequential assembly of the spliceosome complex (Gutierrez-Beltran et al., 2016; Zheng et al., 2013). Through p100/TSN, methylated E2F1 can be associated with some components of the spliceosome on the target pre-mRNA, resulting in alternative splicing (AS) perturbations (Roworth et al., 2019). Given the potential crosstalk between methylation and citrullination, and the relevance of E2F1 and PAD4 in cancer biology (Christophorou, 2022; Fang et al., 2020; Roworth et al., 2015; Yuzhalin, 2019), it seemed plausible that PAD4-mediated citrullination of E2F1 may have a similar impact on the cancer transcriptome through RNA splicing. Therefore, we decided to conduct RNA sequencing to investigate the genome-wide influence of PAD4 inhibition and E2F1 knockdown in cancer cells.

3.2 Optimising RNA-seq condition

To begin with, we decided to perform an RNA sequencing (RNA-seq) experiment to assess the genome-wide impact of PAD4-mediated citrullination on E2F1 target genes in cancer cells. This sequencing technique has been developed as a powerful tool to visualise the intricate nature of the cellular transcriptome, including transcriptional and splicing diversities, as well as the influence of experimental treatments. With higher degrees of experimental resolution and ability to discover novel genomic variants, RNA-seq is increasingly becoming the first choice in transcriptomic and splicing analysis over other techniques like microarray (Marguerat & Bähler, 2010). We therefore reasoned that the RNA-seq experiment should allow us to identify novel genes and pathways regulated by the E2F1/PAD4 axis, as well as to examine their impact on the splicing profile of the cancer genome.

It was determined that our RNA-seq was to be performed in the colorectal cancer cell line HCT116. This conclusion was based on several advantages as follows; (1) PAD4 has been heavily implicated in inflammation, a known important risk factor for the development of colon cancer. Indeed, abnormal expression of PAD4 has been reported in colorectal cancer cells (X. Chang et al., 2009); (2) A previous study in our research group regarding PRMT-mediated E2F1 methylation was performed in HCT116 cells (Roworth et al., 2019), and thus using the same cell line may allow us to undertake comparative analysis about the citrullination-methylation crosstalk; (3) E2F1 CRISPR/cas9 knockout cells of HCT116 cell lines were previously made in our lab (Barczak et al., 2020), which could facilitate the study of the impact of E2F1 knockdown on cells in a more convenient manner. The expression of E2F1 and PAD4 was confirmed by immunoblotting (Figure 3-1A).

The reversible PAD4-specific inhibitor GSK484 was selected as a tool to modulate the intracellular activity of PAD4 in the RNA-seq experiment. This small molecule drug has been widely used to investigate the PAD4 activity both *in vitro* and *in vivo* (Lewis et al., 2015; Mondal & Thompson, 2019; Wei et al., 2021). The GSK484 treatment was reported to inhibit histone citrullination as well as disrupt NETosis in both mouse and human cells (Lewis et al., 2015; Mondal & Thompson, 2019), and increasing evidence indicates its potential anti-tumour effect (Wei et al., 2021). Mechanistically, GSK484 binds to an inactive, low-calcium form of PAD4 to inhibit its enzymatic activity. The low clearance rate, good drug distribution, and long half-life together make GSK484 considered a potent drug for *in vivo* use (Lewis et al., 2015).

Before proceeding with the RNA-seq experiment, the drug concentration and incubation time were carefully titrated and controlled. First, an MTT assay was carried out to examine the metabolic activities of cells under the GSK484 treatment. MTT is chemically stable as a yellow tetrazolium salt and can be converted to a purple formazan upon reduction. Since viable cells contain active NAD(P)H-dependent oxidoreductase, the colour change of this formazan is regarded as a measurement for cell proliferation, viability, or cytotoxicity. Usually, the colour is measured as a change in absorbance at 570 nm with a plate reading spectrophotometer, and the darker the purple MTT solution is, the more viable, metabolically active cells that are present.

To begin with, an MTT assay was performed for different numbers of wildtype (WT) HCT116 p53+/+ cells. Between 1000 and 16000 cells were seeded in each well of a 96-well plate and incubated for 72 hours. The result indicated that 3000 cells per well are optimum for the cell line, as it did not reach the maximum confluency in the tested incubation periods (Figure S-1). Next, another set of MTT assays was carried out in WT

and E2F1cr HCT116 cells treated with different concentrations of GSK484 for various time periods, with 3000 cells per well seeded at the beginning as determined above. The relative percentage of cell viability was calculated against negative control, in which cells were not treated with GSK484. The results demonstrated that HCT116 cells were viable even at relatively high concentrations of GSK484 up to 10 μM (Figure 3-1B). We did not see a major impact of different incubation lengths (from 24 hours to 72 hours) on cell viability here. It was decided to incubate cells with GSK484 at 10 μM concentration for 72 hours in the subsequent experiments, to make sure that sufficient amounts of GSK484 reach the targets in nucleus and cytoplasm regardless of rounds of cell divisions and protein turnovers. A caveat is that we did not examine the potential of side effects caused by this relative long incubation time; another experiment using a parallel method (e.g. siRNA) to deplete PAD4 activity/expression would have been advantageous.

Next, to adjust the dose of GSK484, we examined intracellular citrullination levels by immunoblotting after treatment with different concentrations of this PAD4 inhibitor. There are multiple ways to detect global citrullination levels including anti-peptidyl citrulline antibodies and anti-modified citrulline kits. However, these techniques reportedly have limitations, and their quality and reproducibility are still questioned. We therefore decided to simply use one of the most established antibodies for citrullinated protein, ab5103 (Abcam) against histone H3 protein with citrullination at R2/R8/R17, as a marker for PAD4 activity in cells. PAD4 is the only nuclear member of the PAD family and has been demonstrated to target H3 for citrullination, and indeed many have used this as a maker for PAD4 activity (Mastronardi et al., 2006; Rumble et al., 2017). We treated HCT116 WT cells with a variety of concentrations of GSK484 for 72 hours and with 5 μM calcium ionophore (A23187) for 30 minutes shortly before harvesting cells and

visualising PAD4-mediated citrullination of H3 by immunoblotting. The results showed a reduction in histone H3 citrullination with increased concentrations of GSK484 (Figure 3-1C), whilst the level of total histone H3 was not hugely impacted. The quantitative analysis of this blot demonstrated that the amount of citrullinated H3 was reduced by 5-fold at 10 μM GSK484 treatment compared to the negative control level. We therefore concluded from this result and MTT assay that GSK484 was able to successfully inhibit PAD4-mediated citrullination of substrate proteins in cells without significantly altering cellular metabolic activities, at the concentration of 10 μM for the incubation period of 72 hours.

Finally, to verify the titrated experimental condition of GSK484 treatment with an alternative technique, the fluorescence-activated cell sorting (FACS) analysis was performed to examine the cell cycle profile of HCT116 cells treated with this PAD4-specific inhibitor. The FACS cell cycle analysis utilises a DNA dye called propidium iodide (PI) to measure the DNA content which is varied throughout different stages of the cell cycle. WT and E2F1cr HCT116 cells were fixed with ethanol and their DNA was stained with PI followed by flow cytometry analysis. Figure 3-2 illustrated the average proportions of cells in each cell cycle stage. The result indicated that both WT and E2F1cr cells did not experience significant alteration of cell cycle states with up to 10 μM concentration of GSK484 treatment, and sub G1 levels, which indicates the cells undergoing apoptosis, remained relatively low throughout different concentrations of GSK484 in both cell lines (less than 3 %).

Overall, MTT assay, immunoblotting, and FACS analysis together indicated that GSK484 treatment at 10 μM for 72 hours can successfully inhibit the enzymatic activity of PAD4 in HCT116 cells without significantly altering their cellular physiology and

viability. We therefore decided to use this GSK484 treatment condition which should minimise the influence of deregulated non-specific cellular events on the RNA-seq result.

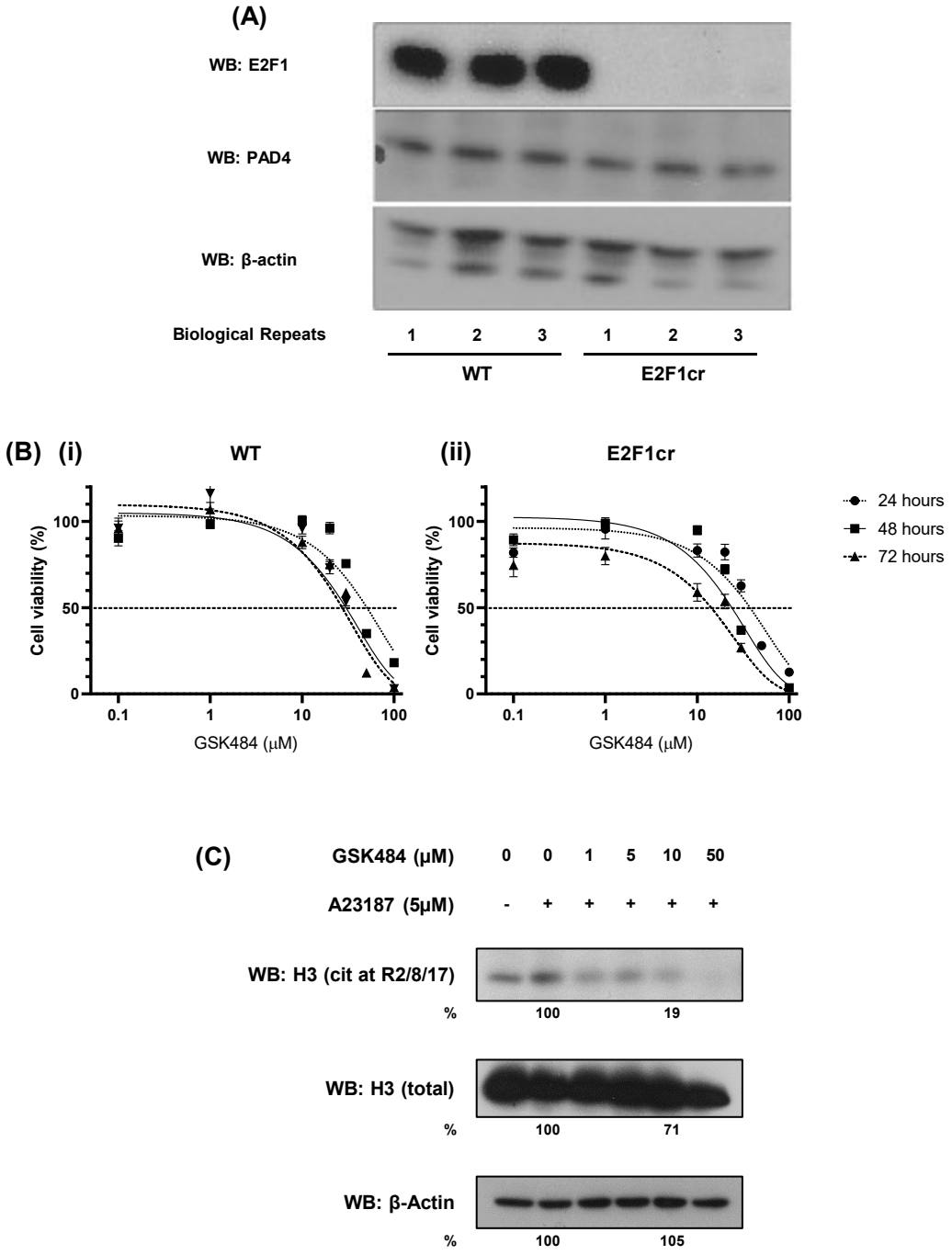


Figure 3-1 Optimisation of RNA-seq experimental condition.

(A) Immunoblot to show the expression of E2F1, PAD4 and β-Actin in HCT116 (p53+/+) WT and E2F1cr cells. *N = 3*. This experiment was performed in collaboration with Amit Shrestha in our research group. (B) (i) WT and (ii) E2F1cr HCT116 cells were treated with various concentrations of GSK484, and the variability was analysed by MTT assays. Shown in percentage relative to the untreated population. *N* = 3. (± S.D.) (B) WT HCT116 cells were treated with various concentrations of GSK484 for 72 hours and 5 μM A23814 (Calcium Ionophore) for 30 minutes shortly before harvesting. The lysates were immunoblotted using anti-citrullinated H3, anti-H3 (total), and anti-β-Actin antibodies. Image J was used to measure the area of the band size at 0 and 10 μM to estimate the protein amounts. *N = 2*.

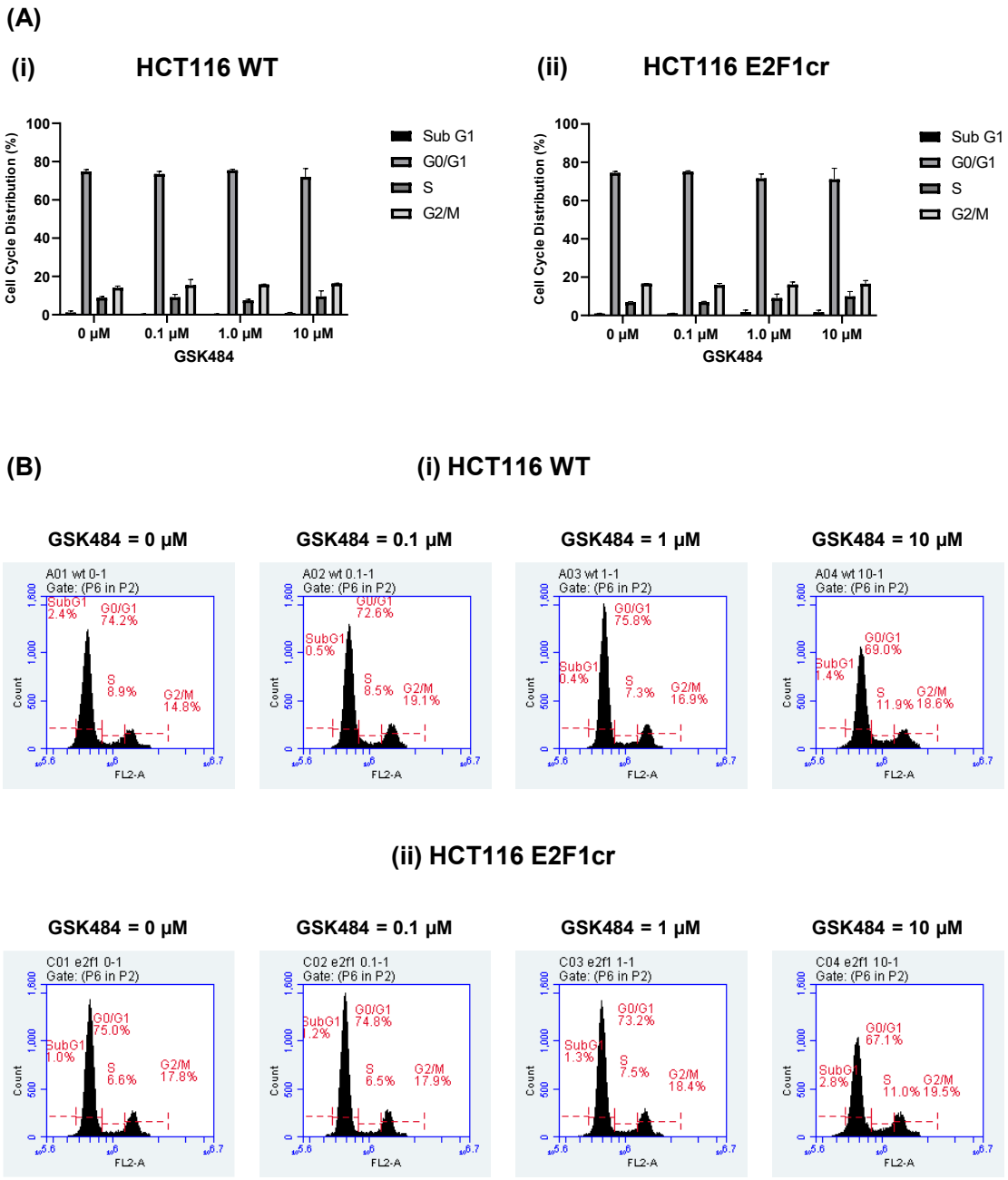


Figure 3-2 Cell cycle analysis by flow cytometry to examine the effect of GSK484.

Flow cytometric analysis using propidium iodide to check the cell cycle states of HCT116 cells treated with various concentrations of the PAD4 inhibitor GSK484. (A) Cell cycle distributions for (i) WT and (ii) E2F1cr cells. *N* = 3. (± S.D.) (B) FACS cell cycle profiles for (i) WT and (ii) E2F1cr cells. A representative example from *N* = 3

3.3 RNA-seq and Quality Control

To explore the influence of the E2F1/PAD4 axis on the cancer transcriptome and spliceome, an RNA-seq experiment was designed and performed. WT and E2F1cr HCT116 p53+/+ cells were treated with 10 μM PAD4-specific inhibitor GSK484 or DMSO (negative control) for 72 hours as determined in the aforementioned section. Biological triplicates of cells in four experimental conditions (WT-DMSO, WT-GSK484, E2F1cr-DMSO, and E2F1cr-GSK484) were prepared, and RNA sample from each was purified using TRIzol reagent. The RNA quality was analysed using Nanodrop by measuring the absorbances at 230, 260 and 280 nm to calculate 260/280 and 260/230 ratios. The former ratio was controlled between 1.8 and 2.0 so that the sample is free from protein contamination, whereas the latter indicates contamination caused by organic compounds or chaotropic agents and was controlled between 2.0 and 2.2 (Figure 3-3A). An immunoblot also confirmed the identities of WT and E2F1cr cells and the successful inhibition of PAD4 activity via the detection of citrullinated H3 (Figure 3-3B). The RNA-seq was performed by BGI Genomics as detailed in Figure 2-1.

With the RNA-seq data returned from the company, we carried out bioinformatics analyses in collaboration with Anastasia Samsonova and Alexander Kanapin (Centre for Genome Bioinformatics, St. Petersburg State University). The read assignment showed that all samples but the third repeat for E2F1cr-DMSO have over 20 million mapped reads, indicating a good overall sequencing coverage of the human genome and low contamination (Figure 3-4A). In addition, the distance matrix heatmap illustrated that each treatment group has good reproducibility over three biological repeats (Figure 3-4B).

(A)

Treatment	Repeat	A260/280	A260/230
WT-DMSO	1	1.97	2.20
	2	1.97	2.27
	3	1.96	2.07
WT-GSK484	1	1.98	2.20
	2	2.01	2.26
	3	1.94	2.18
E2F1cr-DMSO	1	1.96	2.19
	2	1.97	2.27
	3	1.94	2.04
E2F1cr-GSK484	1	1.96	2.19
	2	2.00	2.26
	3	1.95	2.20

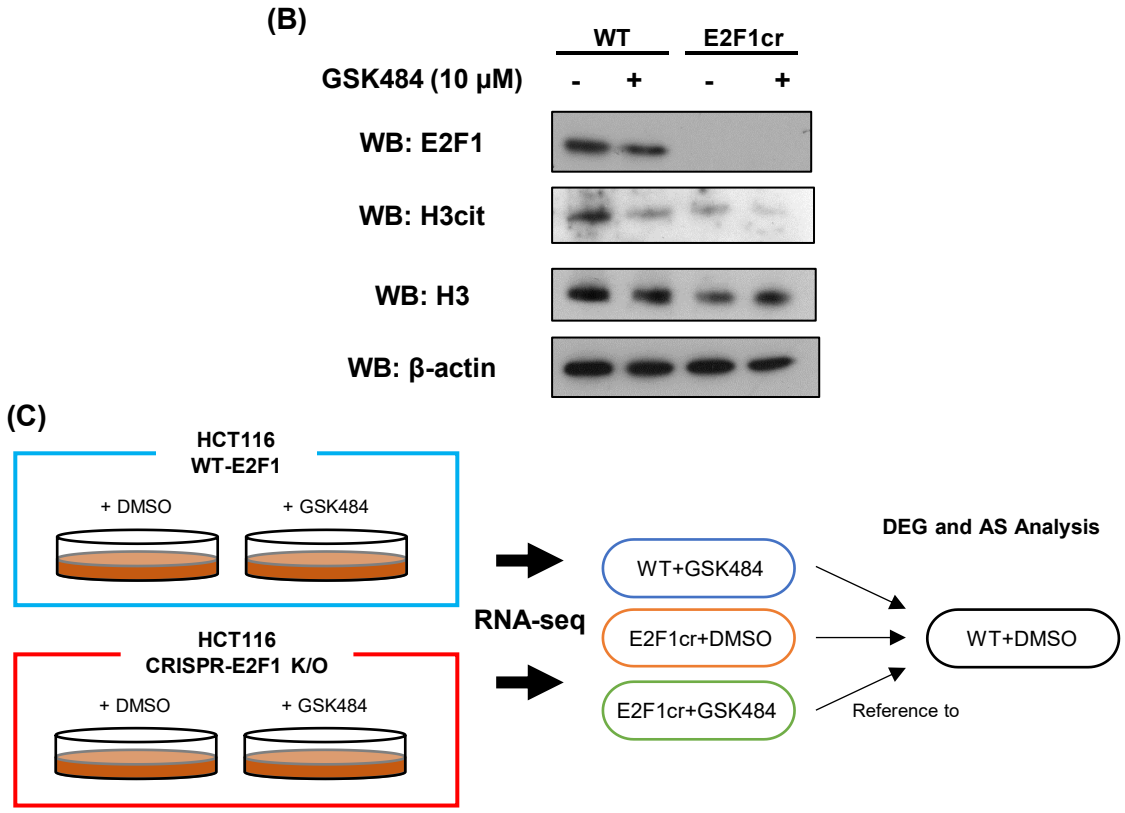


Figure 3-3 RNA-seq quality control and setup.

(A) RNA quality was analysed using Nanodrop by measuring the absorbances at 230, 260 and 280 nm to calculate A260/280 and A260/230 ratios to check the protein and organic/chaotropic contaminations, respectively. Roughly, the former was controlled to be between 1.8 and 2.0, and the latter between 2.0 and 2.2. (B) Immunoblot to show the successful PAD4 inhibition in WT and E2F1cr HCT116 cells treated with 10 μM GSK484 for 72 hours. *N* = 3. (C) Schematic representation of RNA-seq setup for this study. WT and E2F1cr HCT116 cells were treated with 10 μM GSK484 or DMSO for 72 hours, and the three experimental groups were analysed with reference to the control WT-DMSO condition.

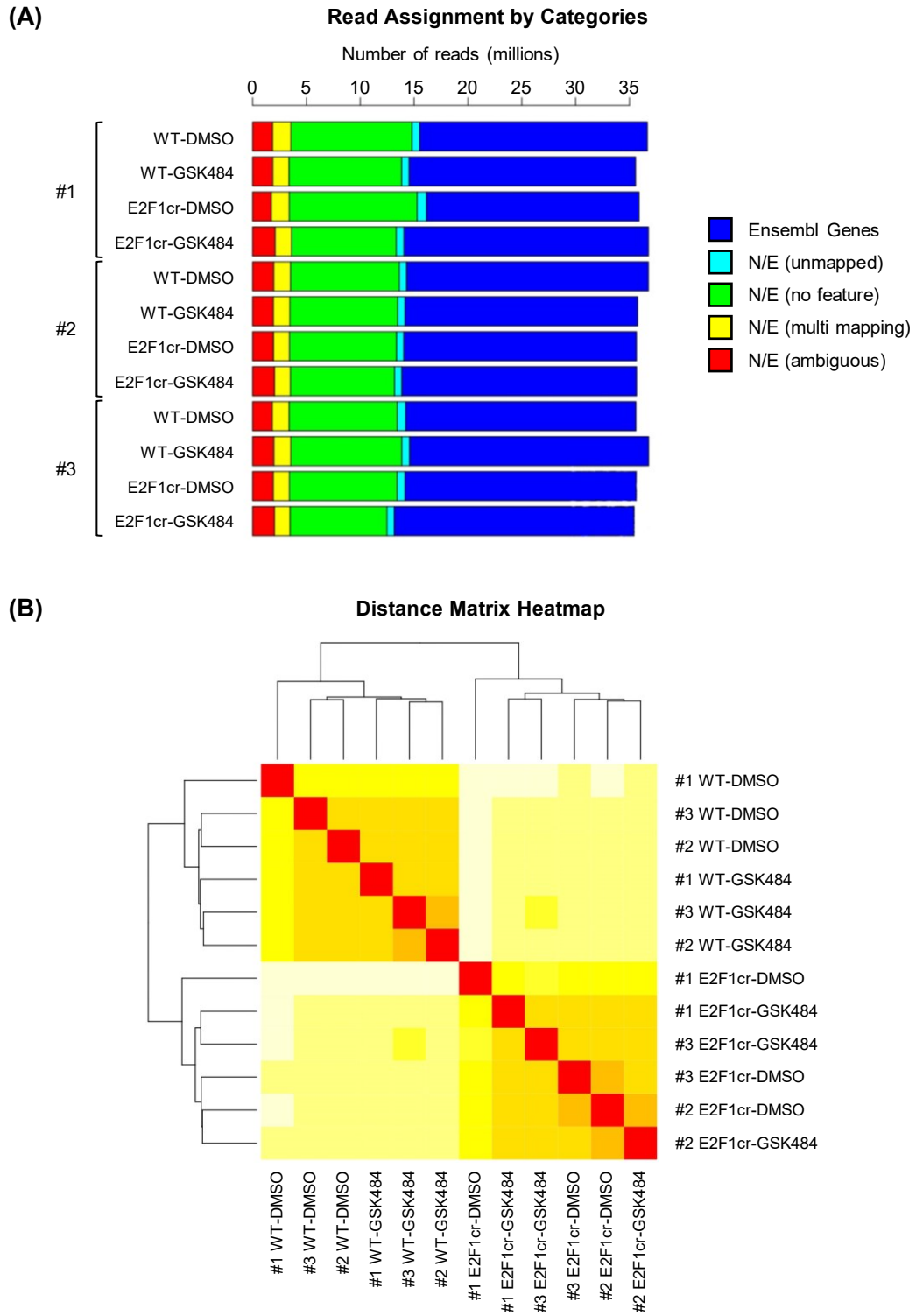


Figure 3-4 RNA-seq quality control.

(A) Read assignment by categories for RNA-seq reads aligned to hg19 human genome build. Obtained with DESeq 2 package. Blue reads are corresponding to Ensembl genes, and the other colours to non-Ensembl (N/E) reads as indicated in the legend. (B) Distance matrix heatmap for triple biological replicates of each experimental condition. Obtained with DESeq 2 package.

3.4 Bioinformatics Analysis of Differentially Expressed Genes

First, the RNA-seq data were examined for differentially expressed genes (DEGs) using the DESeq2 R Bioconductor package. Here, the data from three 'treatment conditions' (WT-GSK484, E2F1cr-DMSO, and E2F1cr-GSK484) were processed and analysed with reference to the 'control condition' (WT-DMSO) unless otherwise stated (Figure 3-3C). The genes were considered to be differentially expressed when their significance (adjusted *P*: P_{adj}) value was less than 0.01 and their expression level change was two-fold or greater ($|\log_2| \geq 1.0$) (Figure 3-5A).

Interestingly, the inhibition of PAD4 did not have a significant impact on gene expression in HCT116 cells, as no genes in the WT-GSK484 condition made the two-fold change cut-off over the control group (WT-DMSO) (Figure 3-5B). Even removing this cut-off, only 11 DEGs have statistically significant P_{adj} values (< 0.01) (Table S-1), suggesting that PAD4's role in transcription regulation was limited under this setting in HCT116 cells. This is somewhat surprising, as PAD4 reportedly plays a role as a transcription co-factor by mediating citrullination of histones or other transcription factors including E2F1 in various cancer cell lines (Fang et al., 2020; Q. Guo et al., 2011; Roworth et al., 2015). It should be taken into account, however, that most of these experiments have been conducted under conditions where PAD4 activity was augmented by over-expression of the enzyme or modulation of calcium ion influx. The result described here may instead reflect the potential physiological activity of PAD4 under low-to-medium expression, which is a relatively less-studied area of PAD biology.

Besides, most of the DEGs (10 out of 11) in WT-GSK484, which have significant P_{adj} values but did not make the 2-fold change cut-off, were identified as potential E2F1-targets through the analysis of E2F1 chromatin immunoprecipitation sequencing (ChIP-

seq) data published as part of the ENCODE project (http://genome.ucsc.edu/ENCODE/) (Table S-1 and Figure S-2). This indicates that PAD4 may undertake a minor role as a transcription co-factor for E2F1-mediated gene regulation.

In contrast to the WT-GSK484 condition, the E2F1cr-DMSO and E2F1cr-GSK484 conditions demonstrated large numbers of significant DEGs over the control condition (WT-DMSO) (Figure 3-5B). It was fascinating to observe that the majority of these DEGs (419 upregulated and 244 downregulated) were shared between both E2F1cr-DMSO and E2F1cr-GSK484 conditions. This shared population of DEGs comprises 74.1 % of the total DEGs found in E2F1cr-DMSO and 81.0 % in E2F1cr-GSK484. Differential expression of these genes should be mainly regulated by E2F1 and not influenced by GSK484 treatment significantly. The rest of the DEGs, which were discovered uniquely in either E2F1cr-DMSO or E2F1cr-GSK484 group, are likely affected by the E2F1 knockdown and/or PAD4 inhibition. Since no significant DEGs were found in WT-GSK484 group, the impact of the PAD4-specific inhibitor treatment on the transcription-based gene expression seems to be augmented in the absence of E2F1. PAD4 itself did not appear in the list of DEGs in this experiment, therefore those ‘unique’ DEGs observed were not a result of E2F1-mediated differential expression of PAD4.

Additionally, we analysed the difference in average fold changes between the ‘shared’ and ‘unique’ populations of DEGs (Figure 3-6). Interestingly, the shared populations of up-regulated and down-regulated DEGs were shown to have significantly greater (higher and lower respectively) average fold change with wider distributions than unique populations in both E2F1cr-DMSO and E2F1cr-GSK484 conditions. Overall, the findings from these analyses may collectively reflect the importance of E2F1 as a master transcription factor; E2F1 has much greater genome-wide influence in terms of both the

number of its target genes and the magnitude of differential expression, as compared to PAD4 whose function was limited to a smaller number of genes and at a smaller scale.

Next, we analysed up-regulated and down-regulated DEGs separately. Here, it was revealed that the total number of up-regulated DEGs was greater than that of down-regulated DEGs in this RNA-seq (604 genes over 328) (Figure 3-5B). The same trend was found in the population shared between E2F1cr-DMSO and E2F1cr-GSK484, where 419 genes were up-regulated compared to 244 down-regulated genes. On the other hand, when looking at DEGs uniquely found in E2F1cr-GSK484, there is not an apparent difference like other DEG populations and only slightly more DEGs were found to be down-regulated (87 genes) over up-regulated (69 genes) (Figure 3-5B).

The list of DEGs was also mined to examine whether the affected genes were known to be bound by E2F1, using E2F1 ChIP-seq data deposited in the ENCODE project. The analysis revealed that only around 45 % of the whole population of DEGs were known E2F1-target(/bound) genes (Figure 3-5C and Figure 3-7A). Even for the DEGs shared between E2F1cr-DMSO and E2F1cr-GSK484, whose expressions should be primarily affected by the loss of E2F1 and not by PAD4 inhibition, the proportions of known E2F1-targets were only 46.5 % in up-regulated and 34.8 % in down-regulated DEGs. For a more detailed analysis, we compared the intensity of differential expression between the known E2F1-targets and the other genes (Figure 3-7B and C). In both E2F1cr-DMSO and E2F1cr-GSK484 conditions, known E2F1-target up-regulated DEGs were found to have smaller average fold changes. There was no significant difference in down-regulated DEGs in both conditions.

Overall, more than half of the DEGs discovered in the RNA-seq were not identified as classical E2F1 target genes through the ChIP-seq analysis. The expression of such genes was likely regulated by E2F1 indirectly, for example, the loss of E2F1 may affect the expression of other transcription factor(s) which in turn results in these DEGs. However, this experiment alone could not eliminate the possibility that these genes were novel E2F1-targets, which requires to be confirmed by further characterisation. Nevertheless, the present result highlights the genome-wide impact and importance of E2F1-mediated transcription regulation, which appears to be extended beyond its target genes. Of great interest is that the known E2F1-target DEGs have significantly smaller average fold change in expression in up-regulated DEGs but not in down-regulated DEGs (Figure 3-7).

Finally, we performed Gene Ontology (GO) analysis for DEGs identified in this RNA-seq experiment to investigate biological processes and pathways affected by E2F1 knockdown and/or PAD4 inhibition. In total, 42 pathways were enriched with the DEGs in two conditions (E2F1cr-DMSO and E2F1cr-GSK484) (Figure 3-8). It is interesting that the 'positive regulation of cell motility/migration/adhesion' pathways were enriched with up-regulated DEGs from E2F1cr-DMSO and E2F1cr-GSK484 conditions, since this finding is consistent with our recent finding that the loss of E2F1 resulted in dysregulated cell migration, invasion, and cell adhesion (Barczak et al., 2020). In this previous study, however, it was highlighted that E2F1-mediated regulation of cancer cell migration and invasion was dependent on PRMT5-mediated methylation. On the contrary, the present GO analysis showed this pathway was enriched in both E2F1cr-DMSO and E2F1cr-GSK484, indicating that E2F1 alone can potentially modulate cell motility without help from PRMTs. There are also some pathways enriched specifically in each condition. For example, genes that were upregulated in E2F1cr-GSK484 condition were enriched for a

series of nucleoside/ribonucleoside catabolic processes, suggesting a novel involvement of the E2F1/PAD4 axis in cellular metabolic activities. It is important to note that GO analysis does not necessarily make a direct conclusion about biological consequences from the RNA-seq data, as RNA levels do not always correlate with protein levels. However, it is a useful tool to visualise how treatments or conditions influence different biological processes and can give a functional overview which helps us to narrow down candidates for affected pathways.

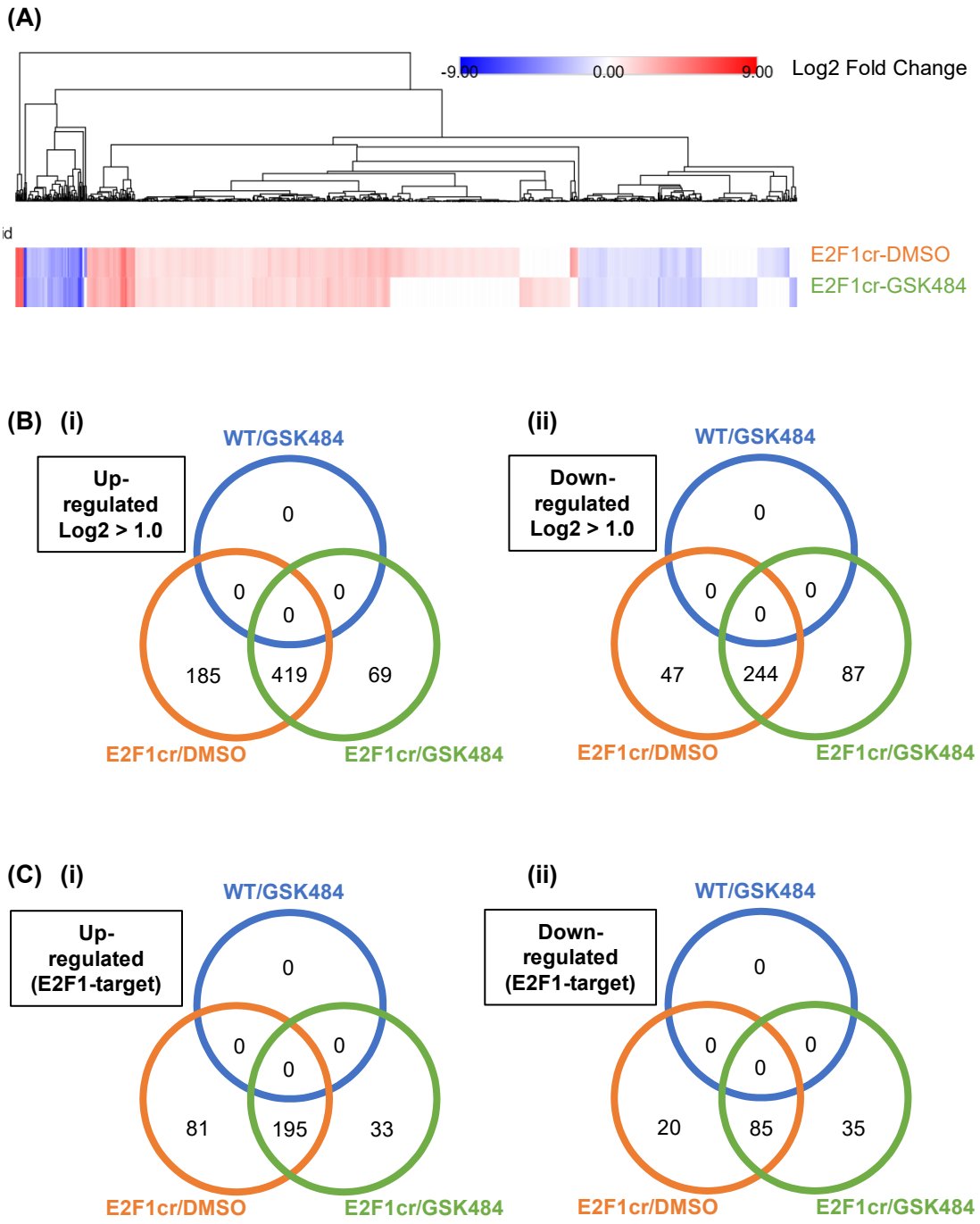


Figure 3-5 RNA-seq analysis for differentially expressed genes.

(A) Heatmap to show the intensity of DEGs found in E2F1cr-DMSO and E2F1cr-GSK484, with adjusted P value < 0.01 and absolute fold change cut-off of log2 >1. No DEGs were found in WT-GSK484 passing both of these cut-offs. (B) Venn diagrams showing the crossover of (i) up-regulated and (ii) down-regulated DEGs found in different treatment conditions with respect to the control WT-DMSO condition, with adjusted P value < 0.01 and absolute fold change cut-off of log2 >1. (B) Venn diagrams showing the crossover of known E2F1-target (i) up-regulated and (ii) down-regulated DEGs with adjusted P value < 0.01 and absolute fold change cut-off of log2 >1.

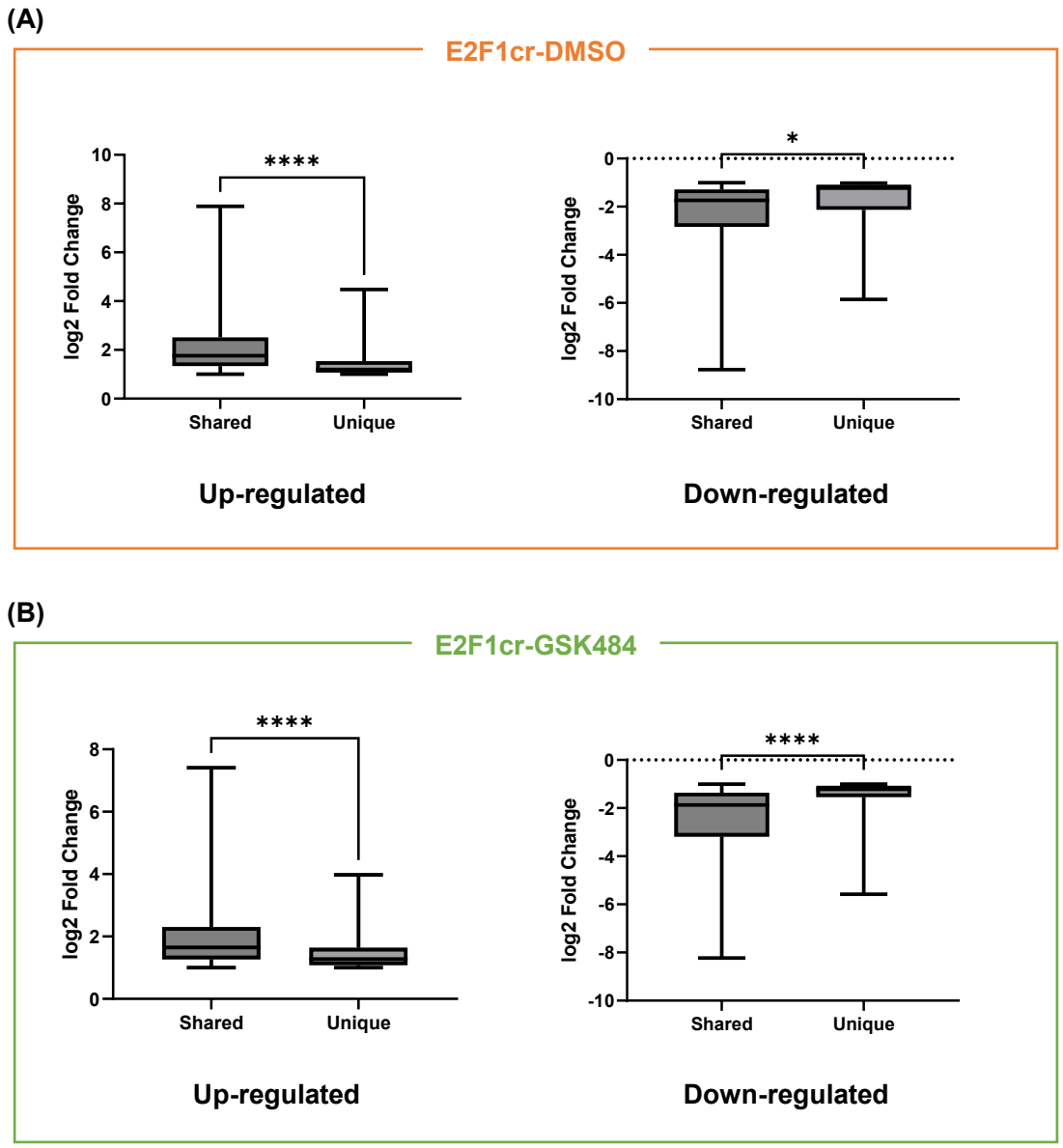


Figure 3-6 Difference between unique and shared DEGs.

(A) Box plots comparing the log2 fold changes between DEGs discovered uniquely in E2F1cr-DMSO group and those shared between E2F1cr-DMSO and E2F1cr-GSK484. (B) Box plots comparing the log2 fold changes between DEGs discovered uniquely in E2F1cr-GSK484 group and those shared between E2F1cr-DMSO and E2F1cr-GSK484. For all plots, a box represents the interquartile range with a median value as a vertical line. Whiskers extended from each quartile show the minimum and maximum values. (* $p < 0.05$, **** $p < 0.0001$).

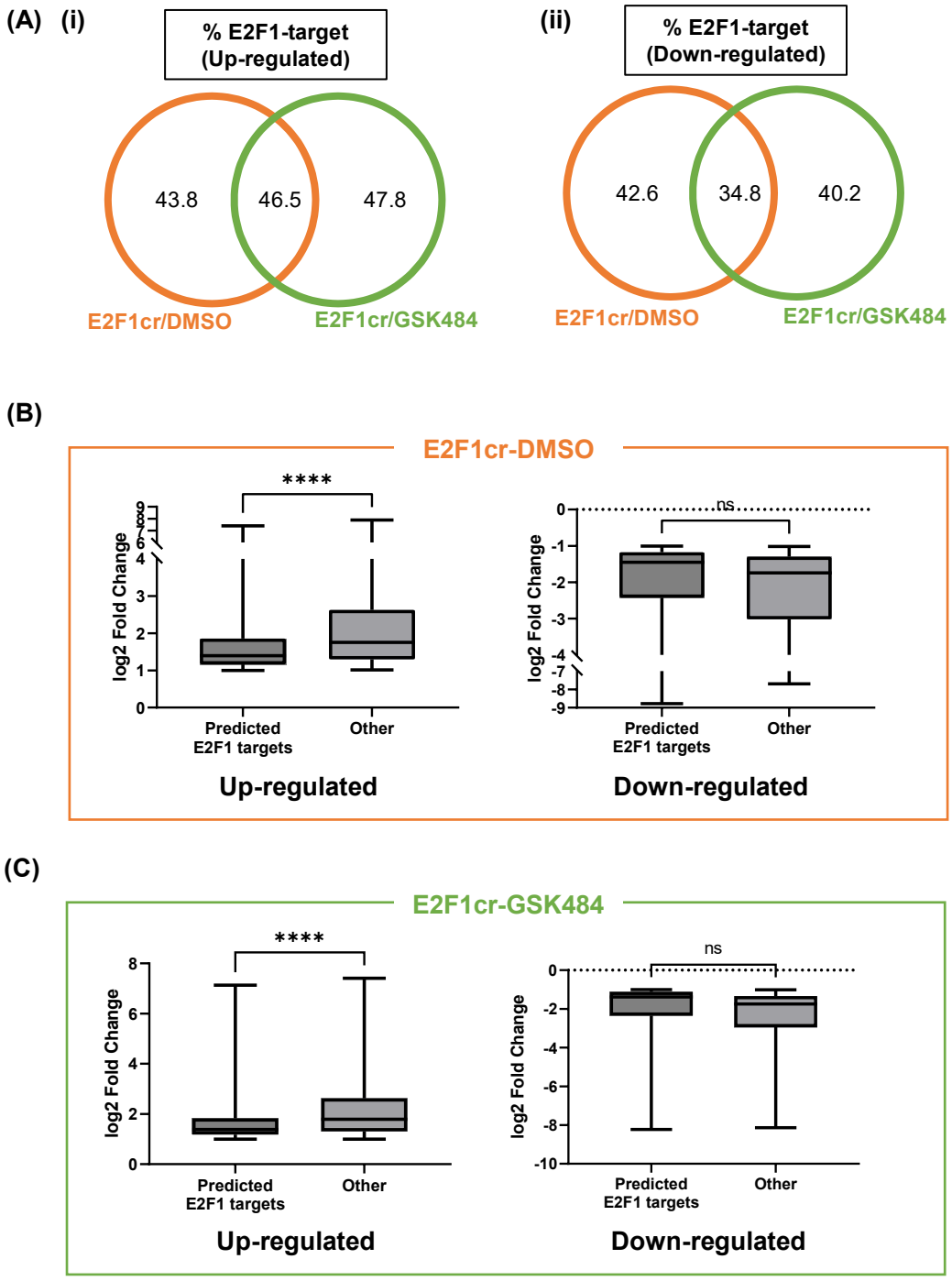


Figure 3-7 Difference between predicted E2F1-targets and other DEGs.

(A) Venn diagrams showing the proportions of predicted E2F1-target genes in (i) up-regulated and (ii) down-regulated DEGs found in E2F1cr-DMSO and E2F1cr-GSK484. (B) Box plots comparing the log2 fold changes of DEGs between predicted E2F1-targets and other genes in E2F1cr-DMSO and (C) in E2F1cr-GSK484. (ns not significant, $p > 0.05$, **** $p < 0.0001$).

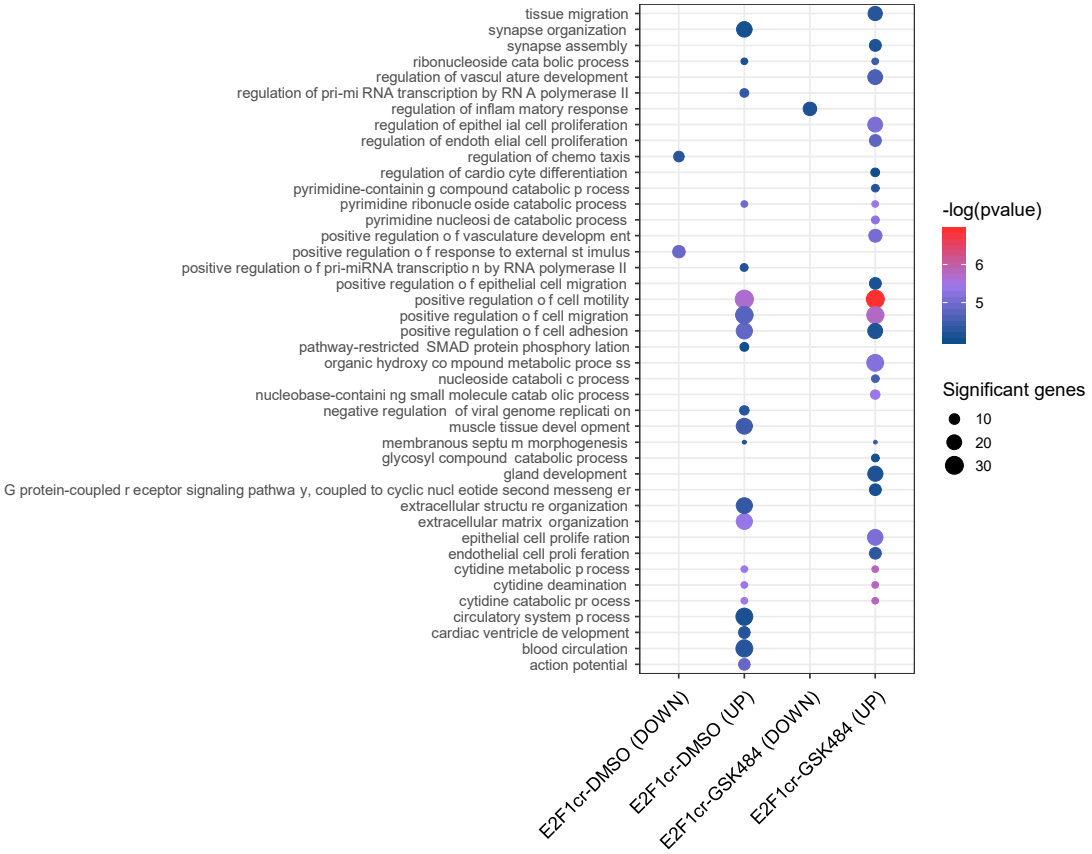


Figure 3-8 Gene Ontology analysis of DEGs.

Up-regulated and down-regulated DEGs in E2F1cr-DMSO and E2F1cr-GSK484 conditions from RNA-seq were analysed using RPGSEA package for Gene Ontology terms. The size of the circles represents the number of significant DEGs enriched in each pathway, and the colour of the circle indicates the p-value of the enrichment, both as described in the legend.

3.5 Bioinformatics Analysis of Alternatively Spliced Genes/Events

PAD4 mediates the citrullination of E2F1 arginine residues, which were also targeted by PRMT1/PRMT5 for methylation (Ghari et al., 2016; Zheng et al., 2013). In a recent study, we have established that PRMT5-mediated methylation allows E2F1 to regulate alternative splicing (AS) of its target transcripts in combination with p100/TSN protein (Roworth et al., 2019). In addition, mounting evidence suggests that a number of RNA-binding proteins were included among the substrates of PAD2 or PAD4 for citrullination in cells (Lewallen et al., 2015; Tanikawa et al., 2018). We therefore questioned whether PAD4-mediated citrullination of E2F1 also exhibited any impact on AS and decided to mine the present RNA-seq data with replicate MATS (rMATS) software, which allows one to analyse differential AS events from RNA-seq data (S. Shen et al., 2014). rMATS can discriminate the following five types of splicing events: exon skipping, intron retention, mutually exclusive exons, alternative 5' splice sites, and alternative 3' splice sites (Figure 1-8). For each splicing event, Δψ (delta PSI) was calculated to show a difference in splicing activity (percent spliced in) from experimental conditions (WT-GSK484, E2F1cr-DMSO and E2F1cr-GSK484) against the control condition (WT-DMSO). The splicing events with FDR less than 0.01 were considered to be statistically significant.

The numbers of significant AS events and genes identified from rMATS analysis were displayed in a heatmap and Venn diagrams by different treatment conditions in Figure 3-9. We observed a major impact of both PAD4 inhibition and E2F1 knockdown at the level of AS; in total 898 AS events derived from 721 genes were detected. Of great interest was the fact that PAD4 inhibition in WT cells (WT-GSK484) alone has a major influence on the splicing of genes. There were 230 genes targeted for AS in WT-GSK484

treatment, and 167 of them were unique in this treatment and not detected in E2F1cr-DMSO and E2F1cr-GSK484. This contrasts with the DEG result where none of the genes was significantly differentially expressed in WT-GSK484 condition and suggests that PAD4 may regulate gene expression principally at the RNA processing level rather than transcription. In addition, when looking at numbers by different treatment conditions more globally, it was clear that the vast majority of splicing events were discovered uniquely in one of the three treatment conditions and only 131 AS events out of 898 were found in overlapping sets between two or more conditions. It is apparent that AS of many RNA transcripts was dependent on the E2F1/PAD4 axis, but their interplay appears to be intricate and context-dependent. This bioinformatic analysis itself cannot conclude whether this splicing regulation was dependent on the citrullination of E2F1 itself, and further biochemical experiments were essential to elucidate its mechanistic details.

Next, alternatively spliced genes were filtered onto E2F1 targets by analysing ENCODE E2F1 ChIP-seq data (Figure 3-10A). Interestingly, 78 % of total AS genes from the RNA-seq were identified as known E2F1-target/bound genes (Figure 3-10B). Even in WT-GSK484 where the E2F1 level was not modulated, 73 % of spliced genes were still identified as known E2F1 targets. Considering only less than half of the DEGs resulting from E2F1 knockdown were known E2F1-targets, this percentage is significantly higher and may reflect the general importance of E2F1 in the AS regulation at the genome-wide scale.

In addition, we compared the lists of alternatively spliced genes and DEGs from the same RNA-seq experiment (Figure 3-11). Interestingly, 97 % of the AS genes did not make the *P*-value and 2-fold change cut-off for differential expression, leaving only 3 % of genes that were significantly regulated at both transcription and alternative splicing. A similar

phenomenon was observed in our E2F1 arginine methylation study, in which we described that genes targeted by E2F1/PRMT5 for AS were often poor targets for transcription regulation (Roworth et al., 2019). Together, E2F1 seems to regulate completely different subsets of its targets via independent mechanisms; one group of genes whose expression is heavily dependent on E2F1 transcriptional activity, and another group containing genes that represent poor transcription targets and thus are regulated principally through AS.

Next, to quantitatively analyse the impact of each treatment condition on AS, we classified the significant AS events into three sub-groups based on their $|\Delta\Psi|$ values: 'strong' with $|\Delta\Psi| > 0.5$, 'moderate' with $0.5 > |\Delta\Psi| > 0.3$, and 'weak' with $0.3 > |\Delta\Psi| > 0.1$. Splicing events with $|\Delta\Psi| < 0.1$ were considered 'insignificant' and removed for this analysis. The numbers and proportions of AS events by these categories were shown in the bar charts in Figure 3-12A. When looking at the total numbers of significant splicing events (including 'strong', 'moderate', and 'weak'), the E2F1cr-DMSO condition had the highest number and was followed by E2F1cr-GSK484 then WT-GSK484. Interestingly, no significant difference was observed between these conditions in terms of the number of 'strong' events, and this means that WT-GSK484 has the highest proportion of 'strong' splicing events which is followed by E2F1cr-GSK484 then WT-GSK484. This trend is also apparent in the $|\Delta\Psi|$ distribution graph (Figure 3-12B), where the mean $|\Delta\Psi|$ value for WT-GSK484 was significantly higher than those of E2F1cr-DMSO and E2F1cr-GSK484. This analysis revealed that PAD4 inhibition had a greater impact on the strength of AS, while the loss of E2F1 instead affect larger numbers of targets.

One of the advantages of rMATS analysis is that this software can discriminate various types of splicing events. Figure 3-13A illustrates the proportion of the spliced genes by

five different splicing types; skipped exon (SE), retained intron (RI), mutually exclusive exons (MXE), alternative 5' splice sites (A5SS), and alternative 3' splice sites (A3SS). For all three treatments, SE, which is known to be the most common splicing event in mammals, composes the majority of AS events. The E2F1cr-DMSO treatment had higher percentages of MXE and A3SS than those of the other two groups, while the E2F1cr-GSK484 treatment had the highest proportion of A5SS among all. The highest number of RI was found in WT-GSK484. Figure 3-13B showed the box plots of $|\Delta\Psi|$ distribution by different types of splicing events in each treatment condition and total. It is apparent RI events hold the highest mean $|\Delta\Psi|$ values in all conditions; the highest amongst five splicing types in E2F1cr-DMSO and E2F1cr-GSK484, and the second highest in WT-GSK484 where A5SS recorded the highest. MXE in turn has the lowest $|\Delta\Psi|$ values in WT-GSK484 and E2F1cr-DMSO groups. When all the spliced events in three conditions were combined, two splicing event types, RI and MXE, have a significant difference in $|\Delta\Psi|$ compared to the mean of the total (Figure 3-13B (iv)).

Finally, Gene Ontology (GO) analysis was performed for the alternatively spliced genes identified through rMATS analysis to uncover biological pathways potentially impacted by splicing regulation mediated through the E2F1/PAD4 axis. The analysis revealed several pathways enriched in each treatment group through the GO analysis, as shown in Figure 3-14. Since only a small number of genes were spliced in more than one treatment condition, there were no pathways over-represented in more than one treatment. Interestingly, classic E2F1-regulated pathways such as cell cycle regulation were not found to be enriched in this RNA-seq analysis, and enriched pathways for splicing GO analysis were quite distinct from GO analysis using the DEGs from the same RNA-seq. This again supports the hypothesis that E2F1 regulates two independent groups of genes

through transcription and AS respectively. Outcomes of E2F1-mediated transcription regulation and splicing regulation seem to be independent of each other and complex, and further analysis is needed. There are some important biological pathways enriched with spliced genes in each condition. For example, AS genes in WT-GSK484 condition over-represent pathways like the regulation of the immune system process, metabolism of some amino acids, and lysosome transport. As previously mentioned, GO analysis is still a loose term and needs to be carefully assessed and validated at biochemical levels. The result nonetheless indicates a potential activity of PAD4 in a wide spectrum of biological processes in cancer cells through the regulation of AS, whereas future study still needs to prove such biological outcome.

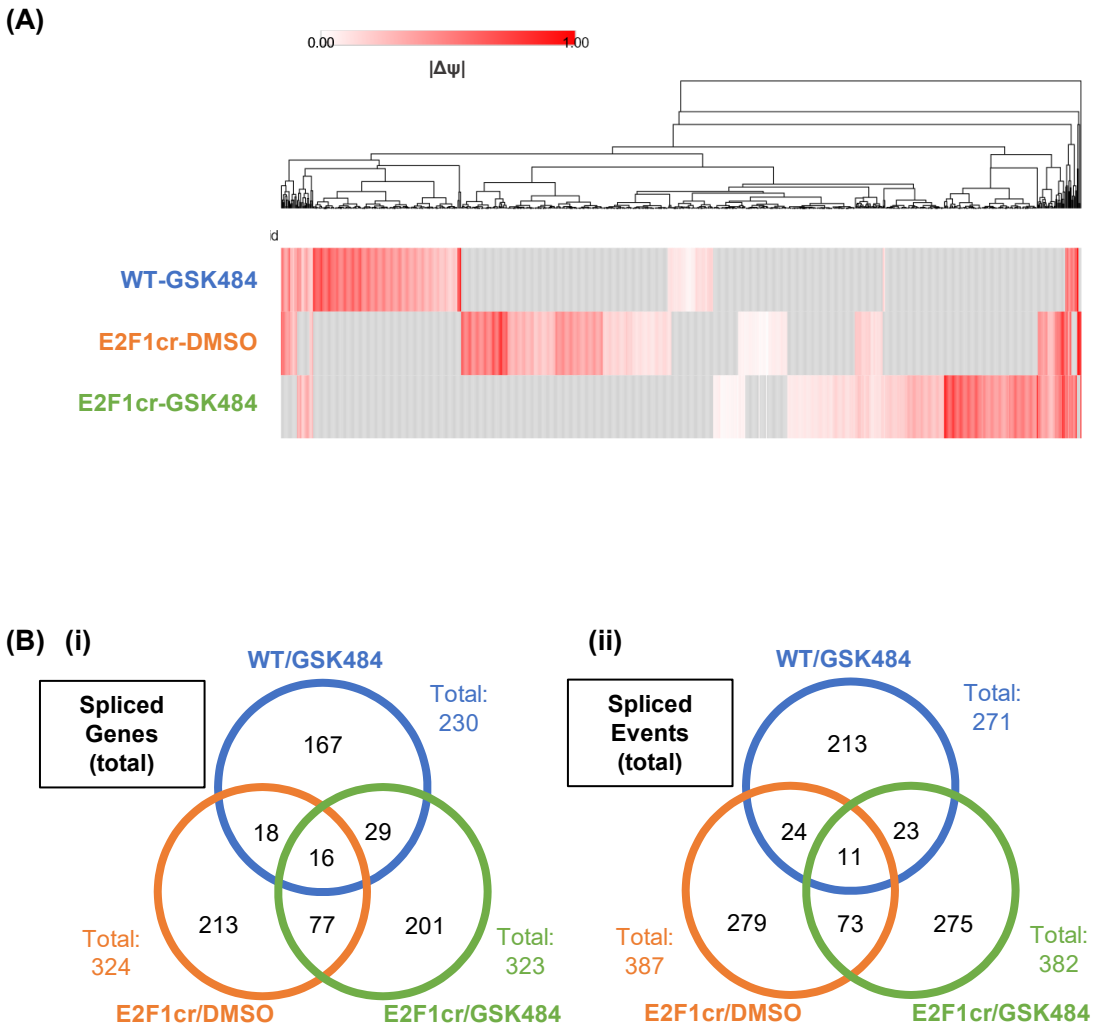


Figure 3-9 RNA-seq analysis for alternative splicing (AS) of transcripts.

(A) Heatmap showing absolute Δψ values of significant AS events (FDR < 0.01) in the three treatment conditions with reference to the control WT-DMSO group. Stronger splicing events were indicated with darker red colour, and weaker ones were shown with paler pink/white colour. (B) Venn diagrams showing significant (FDR < 0.01) (i) AS genes and (ii) AS events grouped by the treatment conditions with respect to WT-DMSO.

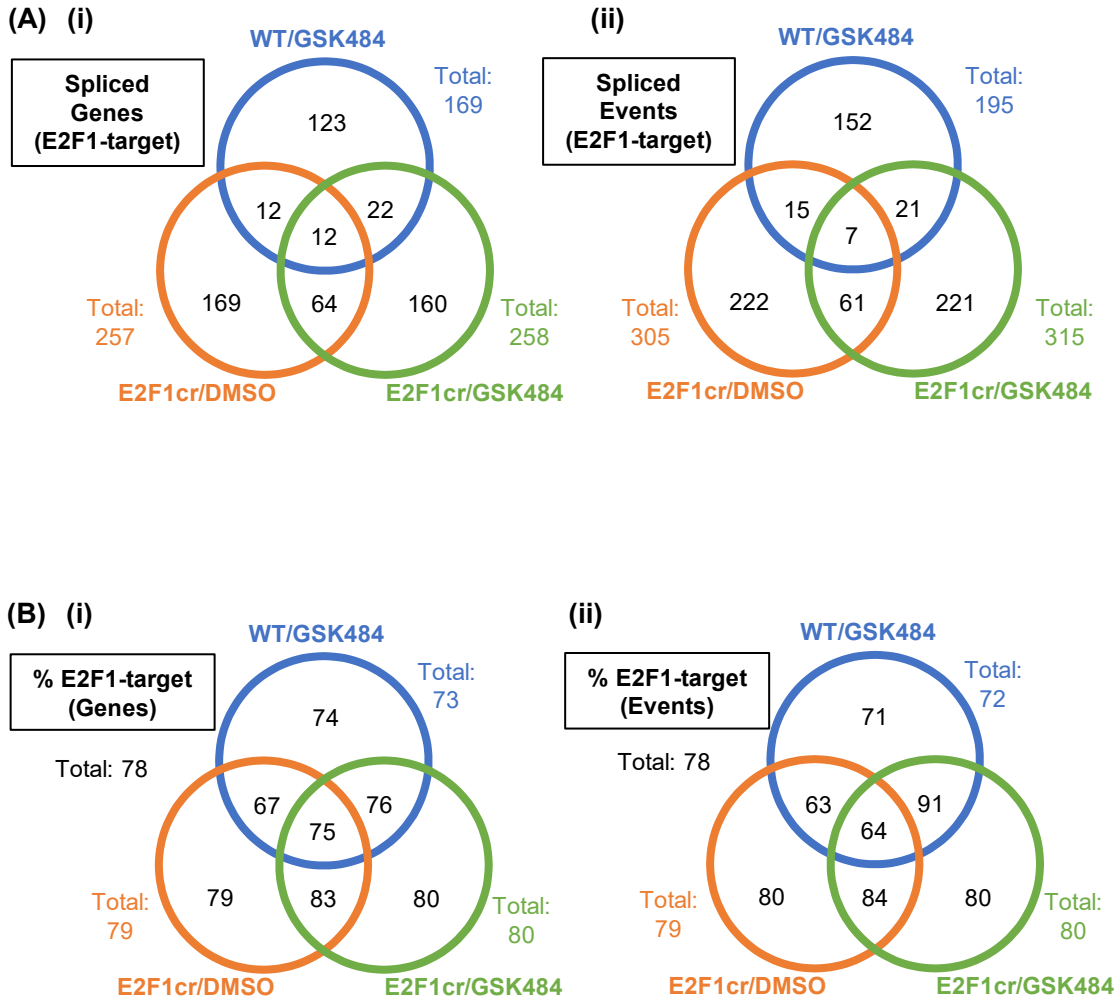


Figure 3-10 E2F1 ChIP-seq analysis for AS genes from RNA-seq.

(A) Venn diagrams illustrating the overlaps of significant AS (i) E2F1-target genes and (ii) E2F1-target events grouped by treatment conditions with respect to WT-DMSO. (FDR < 0.01) (B) Venn diagrams showing the percentages of E2F1-target genes from which alternative splicing events were derived, grouped by three treatment groups with respect to the control WT-DMSO group.

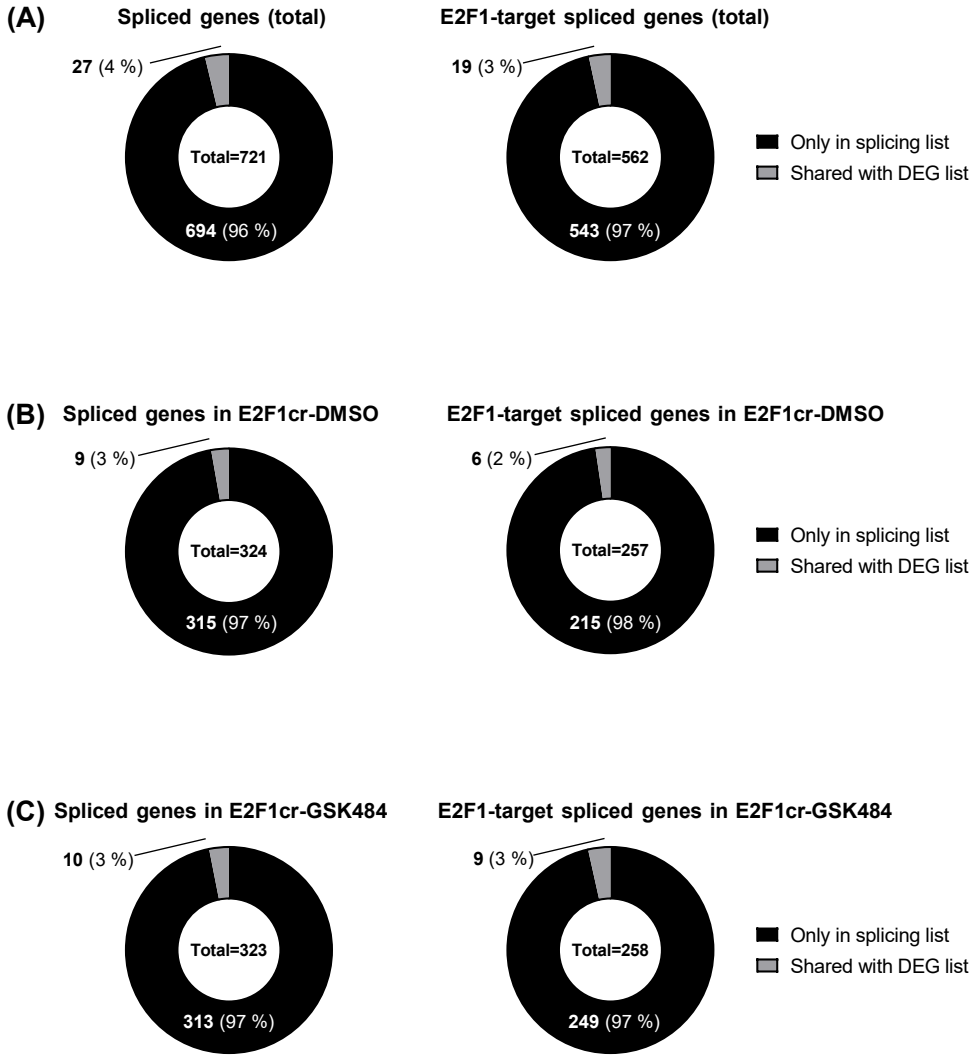


Figure 3-11 Comparison between AS genes and DEGs.

Comparison between alternative spliced genes from rMATS analysis and DEGs from DESeq analysis. (A) The proportions of alternatively spliced genes from a total population which also appeared as DEGs in RNA-seq. (B) The proportions of alternatively spliced genes from E2F1cr-DMSO which also appeared as DEGs in RNA-seq. (C) The proportions of alternatively spliced genes from E2F1cr-GSK484 which also appeared as DEGs in RNA-seq.

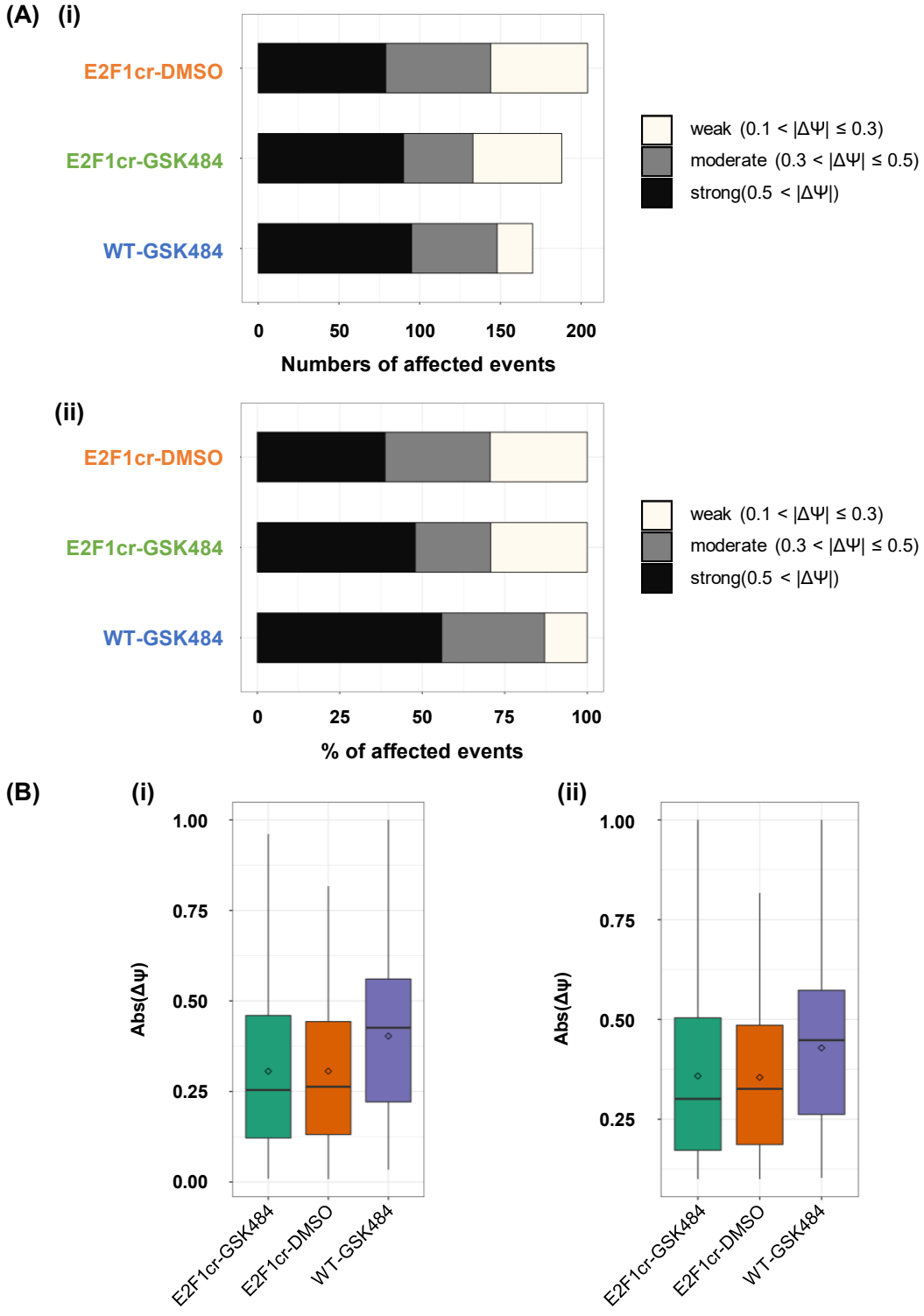


Figure 3-12 Quantitative analysis of AS events by treatments.

(A) The bar graphs showing (i) the numbers and (ii) proportions of significant alternative splicing events (FDR < 0.01, |ΔΨ| > 0.1) by three quantitative categories based on |ΔΨ| (as indicated in legends). (B) The box plots showing the distribution of |ΔΨ| in each treatment group. Box represents the interquartile range with the median value as a vertical line. Whiskers extended from each quartile show the minimum and maximum values.

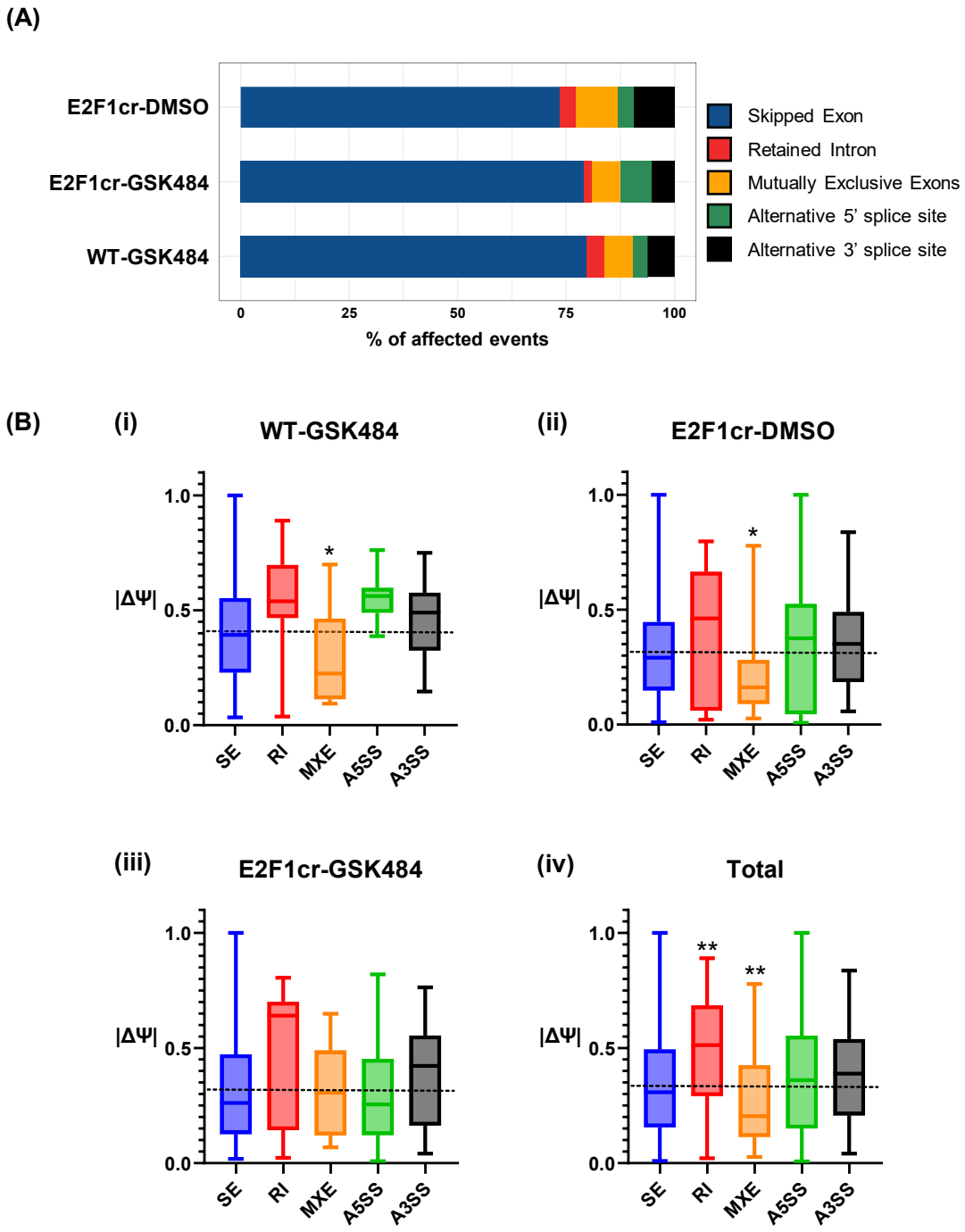


Figure 3-13 Alternative splicing analysis by different splicing types.

(A) The bar graphs showing the proportions of different types of alternative splicing events (indicated in legend) in each treatment group. (B) The box plots showing the |ΔΨ| distribution in five types of splicing events in three treatment groups ((i) WT-GSK484, (ii) E2F1cr-DMSO, and (iii) E2F1cr-GSK484) and (iv) total of three. Box represents the interquartile range with a median value as a vertical line. Whiskers show the minimum and maximum values. Significance of the difference between the mean |ΔΨ| of each splicing type and that of the total population (indicated as the dot vertical line) was determined by *t*-test and indicated at the top of respective boxes (* $p < 0.05$, ** $p < 0.005$).

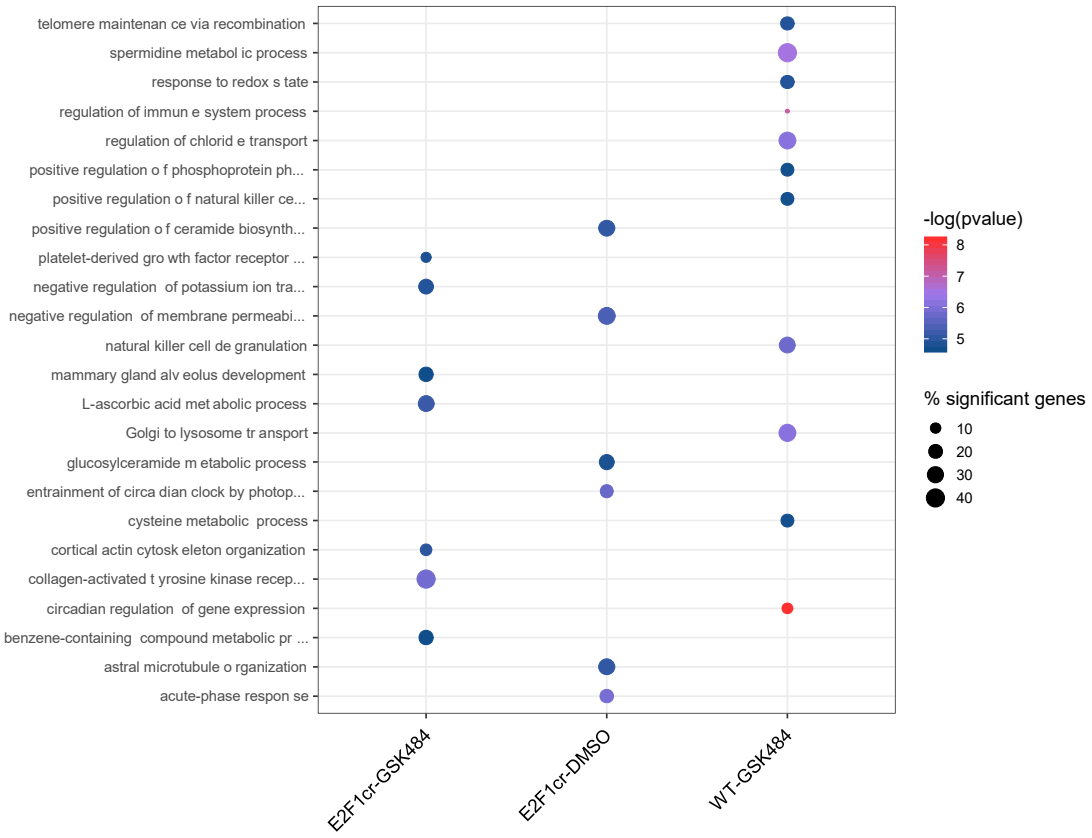


Figure 3-14 Gene Ontology analysis of AS targeted transcripts.

Genes from which significant alternatively spliced transcripts were derived (FDR < 0.01) were analysed with the RPGSEA package for Gene Ontology terms. The size of the circle represents the number of significant spliced genes enriched in each pathway, and the colour of the circle shows the P-value of the enrichment, both as indicated in the legend.

3.6 Chapter Summary

After carefully controlling the RNA quality of biological triplets (Figure 3-3A), an RNA-seq experiment was successfully performed in HCT116 WT and E2F1cr cells with the PAD4-specific inhibitor GSK484 treatment, and we examined the genome-wide influence of the E2F1/PAD4 interplay. The obtained RNA-seq data was first mined for differentially expressed genes (DEGs) (Figure 3-5). Interestingly, PAD4 inhibition in WT cells did not exhibit a significant impact on gene expression. This contrasts with several previous reports indicating that PAD4 can regulate the transcription of various genes through the citrullination of histones and some transcription factors/co-factors (Cuthbert et al., 2004; Ghari et al., 2016; Y. Wang et al., 2004; X. Zhang et al., 2011). This is probably because these studies were mostly performed in conditions with the ectopic or over-expression of PAD4 or the treatment of calcium ionophore, and our result here suggests that PAD4 without abnormal or over-expression might not be sufficient to exert its function as a transcription co-factor.

On the contrary to WT-GSK484 treatment, there were large numbers of DEGs discovered in the E2F1cr-DMSO and E2F1cr-GSK484 conditions (Figure 3-5). The majority of these DEGs were shared between both conditions, most likely reflecting the significant role that E2F1 plays as a master transcription regulator. It is also noteworthy that some DEGs were uniquely found in either E2F1cr-DMSO or E2F1cr-GSK484, and these DEGs should be influenced by E2F1 and PAD4 activity. Given that no significant impact at the level of differential expression was found in WT-GSK484, and PAD4 was not amongst the DEGs, there might exist a yet-unknown mechanism by which E2F1 masks PAD4's activity as a transcriptional co-factor, potentially by modulating the protein-protein or epigenetic interactions.

It has been established that E2F1 can function as either a transcriptional activator or suppressor depending on the cellular contexts (Fang et al., 2020; Roworth et al., 2015). Our RNA-seq discovered a larger number of up-regulated DEGs as compared to down-regulated DEGs (Figure 3-5B), indicating that E2F1's function as a transcription suppressor is more susceptible to genomic knockdown than its activator role. Given that E2F family members reportedly have functional redundancy to one another (Fang et al., 2020; Roworth et al., 2015), this result may suggest that other E2F members may represent greater levels of redundancy for the E2F1's transactivation activity and it might be more difficult to compensate its gene suppressing activity. Interestingly, this tread was not apparent for DEGs uniquely found in E2F1cr-GSK484, whereby only slightly greater number of DEGs were discovered to be down-regulated (87 genes) over up-regulated (69 genes) (Figure 3-5B). PAD4-mediated transcription regulation has been mainly studied through histone citrullination (Cuthbert et al., 2004; Y. Wang et al., 2004), where both activating and silencing effects were reported. Since this population of DEGs should be influenced by both E2F1 and PAD4, the result may reflect this versatile function of PAD4 as a transcription regulator.

More than half of the DEGs found in this RNA-seq did not appear to be direct E2F1-targets through ENCODE E2F1 ChIP-seq analysis (Figure 3-7). Subsequent analysis has revealed that this population of E2F1-regulated genes was affected by the E2F1 loss at an even greater magnitude than the classical E2F1-targets (Figure 3-7), highlighting the significant impact of the E2F1 knockdown even on the genes whose expression is not directly modulated by E2F1. We had previously hypothesized that the activity of E2F1 as a transcriptional repressor might not be as effectively substituted by other E2Fs as its role as an activator. The present finding further supports this hypothesis and potentially

provides a mechanistic insight; E2F1, acting as a transcriptional repressor, may regulate the expression of a larger number of genes with greater magnitude through indirect mechanisms than as an activator. As this regulation is indirect, other E2F members may not exhibit a significant degree of functional redundancy.

Inspired by our previous study regarding E2F1 methylation, as well as the potential interplay between methylation and citrullination (Ghari et al., 2016; Roworth et al., 2019), the RNA-seq data was also analysed for alternative splicing (AS) of RNA transcripts. We uncovered that hundreds of AS events were deregulated by PAD4 inhibition and E2F1 knockdown (Figure 3-9). Most of them were detected in one of the treatment conditions and only a limited number of transcripts were alternatively spliced significantly in two or more conditions, indicating an intricate nature of the E2F1/PAD4 in this AS regulation.

Unlike DEG analysis, this rMATS analysis identified a large number of significant AS events in the WT-GSK484 treatment (Figure 3-9). Current understanding of the intracellular PAD4 activity in normal physiology remains quite limited, and some may even consider this Ca^{2+}-dependent enzyme inactive in such conditions due to the low concentrations of Ca^{2+} ions in normal cells. Having said that, this study instead indicates a potential physiological role that PAD4 with low-to-medium expression undertakes in principally regulating gene regulation at the level of mRNA splicing rather than in a transcription-based mechanism.

In addition, E2F1 ChIP-seq analysis has elucidated that around 78 % of total AS events regulated by PAD4 were identified to be derived from known E2F1-target genes (Figure 3-10). This proportion is much higher than that of DEGs identified in the RNA-seq (Figure 3-7) and thus emphasises the significance of the E2F1 activity in AS regulation.

The result also provides an implication that the direct association between E2F1 and the chromatin of target genes might be integral to the E2F1-dependent AS perturbations, for example via the recruitment of the spliceosome complex or splicing factors, and emphasises the need for future investigation. This would be consistent with our previous finding that DNA-binding domain was required for E2F1 to interact with components of the splicing machinery (Roworth et al., 2019).

We also examined the impact of each treatment condition by a quantitative analysis, and demonstrated that the mean $|\Delta\Psi|$ value of AS events in the WT-GSK484 condition was significantly higher than those of E2F1cr-DMSO and E2F1cr-GSK484 (Figure 3-12). On the contrary, the E2F1cr-DMSO condition had the highest total numbers of significant splicing events and was followed by E2F1cr-GSK484 then WT-GSK484 (Figure 3-12). This analysis revealed that PAD4 inhibition had a greater impact on the strength of AS, while the loss of E2F1 instead affect larger numbers of targets. This finding potentially reflects the difference in roles for AS regulation between PAD4 and E2F1; PAD4 could play a more direct mechanistic role to control AS, and E2F1 may be involved with a more indirect process such as target selection and recruitment of other proteins.

Furthermore, the vast majority of alternatively spliced transcripts were not differentially expressed (Figure 3-11). E2F1 is therefore likely to regulate two independent subsets of genes via distinct mechanistic pathways; one group is consisting of genes regulated in the transcription-based mechanism, and the other group of genes are poor transcription targets and primarily regulated at the level of AS.

Overall, we successfully observed a significant genome-wide influence of PAD4 inhibition and E2F1 knockdown in cancer cells. Most noteworthy, we have established

that the E2F1/PAD4 axis plays an important novel role in regulating AS, and many target RNA transcripts were derived from known E2F1-targets. However, the bioinformatic analysis alone is insufficient to fully elucidate the mechanistic details of this regulation. Further biochemical experiments are required to answer key questions, such as whether the regulation is directly dependent on PAD4-mediated citrullination of E2F1, as well as which downstream splicing factors are involved.

Chapter 4

Validation of RNA-seq Result by RT-qPCR

4.1 Introduction

In Chapter 3, we performed an RNA-seq experiment with the PAD4-specific inhibitor GSK484 treatment in WT and E2F1cr HCT116 cells and successfully demonstrated a significant genome-wide impact of E2F1 knockout and PAD4 inhibition. A large number of genes were differentially expressed in E2F1cr-DMSO and E2F1cr-GSK484 (Figure 3-5). Interestingly, only less than half of them were predicted E2F1 targets through E2F1 ChIP-seq analysis, indicating that many genes were indirectly influenced by the E2F1 activity potentially through the gene regulation of other transcription factors or co-factors (Figure 3-7). In addition, the rMATS analysis revealed that the E2F1/PAD4 axis regulated the expression of numerous E2F1-target genes through alternative splicing (AS) (Figure 3-9). Interestingly, most of the genes targeted for splicing were not significantly differentially expressed, suggesting E2F1 regulates different subsets of its target genes in transcription-based and AS-based mechanisms respectively (Figure 3-11).

However, it is important to acknowledge that RNA-seq experiments and their subsequent analyses may be susceptible to various sources of bias or erroneous signals, potentially leading to false positives. For example, RNA fragmentation might produce length biases, and the subsequent library amplification step could be affected by a primer bias of random hexamers. Some, therefore, argue that it would be crucial to validate RNA-seq results by a parallel method, such as Quantitative Reverse Transcription PCR (RT-qPCR). RT-qPCR is a modified protocol of PCR which allows one to quantitatively measure RNA content. The RNA sample is first reverse transcribed into complementary DNA (cDNA) by a reverse transcriptase enzyme, and the subsequent qPCR experiment uses this cDNA as a template to estimate the amount of the original RNA. Unlike RNA-seq, RT-qPCR is only capable of detecting a single gene with known sequences per experiment, and thus

is not suitable for a large-scale screening experiment to discover novel target genes. Nevertheless, to confirm or measure the expression of a relatively small number of known genes, this method becomes the first choice as it would be much more cost-effective and less time-consuming. It is worth noting, however, that the need for RNA-seq validation remains under debate (Coenye, 2021). Those who disagree with its necessity often claim that RT-qPCR results have higher variances, making them difficult to reproduce, while the normalization process during RNA-seq is sufficient to remove any potential biases, if any.

4.2 Primer Design for RT-qPCR Validation.

Given the splicing perturbations observed in our RNA-seq experiment that are dependent on the E2F1/PAD4 axis, our next focus was to investigate the underlying mechanistic details. We hence concluded that it would be ideal to validate some of the E2F1/PAD4-regulated AS events by RT-qPCR, which should permit downstream experiments to characterise the interplay between E2F1 and PAD4. A series of RT-qPCR validation experiments were therefore performed the AS events identified through the rMATS analysis of the RNA-seq dataset.

Figure 4-1 explains how primers were designed for the AS validation using the Primer-BLAST tools (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Two pairs of primers were designed for each AS event so that they can separately detect the two variants of an RNA transcript resulting from a single AS event (e.g. inclusion or exclusion of an exon of interest). Generally, each primer pair is designed to generate a PCR product spanning a specific exon-exon or exon-intron junction that is uniquely found in the target RNA variant. For example, Figure 4-1A shows the primer design for the exon 2 skipping of the SPIN1 gene. Here, the inclusion pair recognises a junction between the exon 1 and 2. On the other hand, the exclusion pair's product would span the junction between exon 1 and 3, which would only exist in the transcripts undergoing exon 2 skipping. Since these junctions are possessed by different splicing transcripts in a mutually exclusive manner, the ratio of such inclusion/exclusion variants should correspond to the magnitude of the respective splicing event and provide a good statistical measurement. Ideally, these inclusion and exclusion pairs share either forward or reverse primers (Figure 4-1B) in order to avoid potential unwanted biases resulting from different annealing affinities of primers. When this is not possible, different primers were used for inclusion and exclusion

pairs (Figure 4-1B), but the PCR product size was controlled to be around 100 nt. Primers for a retained intron event were designed in a similar fashion, as shown in Figure 4-1C.

We selected some AS events from the RNA-seq and designed primers for these events to be validated by RT-qPCR by focusing on skipped exons (SE), which consists of the majority of the spliced events, and retained introns (RI), which often inserts early stop codons thus likely has a greater impact on the resultant protein structure than other splice events. Additionally, the subsets of splicing events with higher delta psi values were favourably selected for validation as a greater change in exon or intron usage should theoretically be easier to detect by RT-qPCR. The list of selected spliced genes is shown in Figure 4-2.

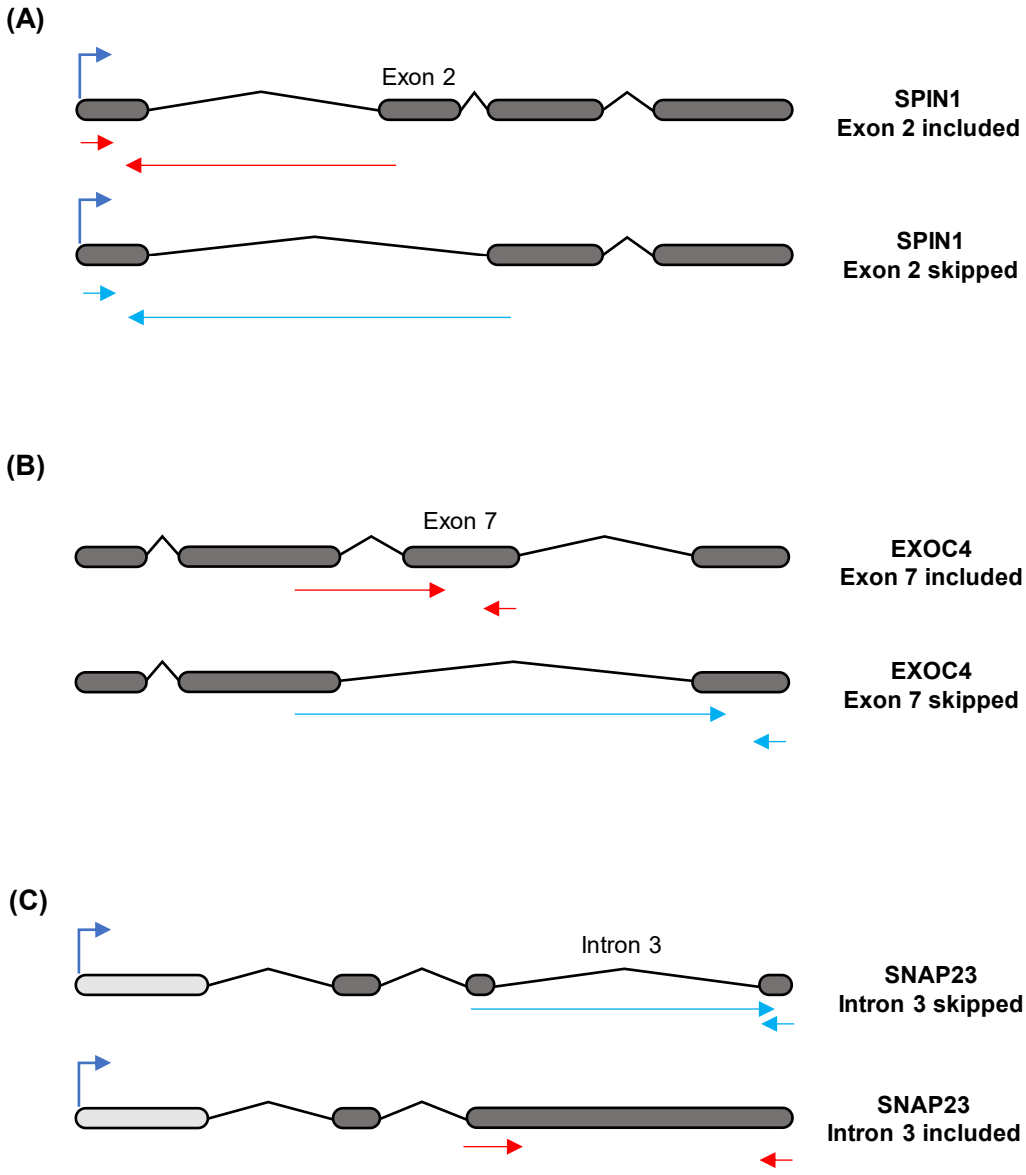


Figure 4-1 Primer design for RNA-seq validation experiment by RT-qPCR.

A diagram showing structure and splice variants of AS target genes with an indication of primers for their RT-qPCR validation. Shaded boxes represent the translated exons, and unshaded boxes illustrate untranslated exons. Primers to detect inclusion of exon/intron were shown in red arrows, and those for exclusion in blue ones. Long arrows represent primers that span the indicated exon-exon or exon-intron junctions. (A) SPIN1 exon 2 skipping, (B) EXOC4 exon 7 skipping, and (C) SNAP23 intron 3-4 inclusion.

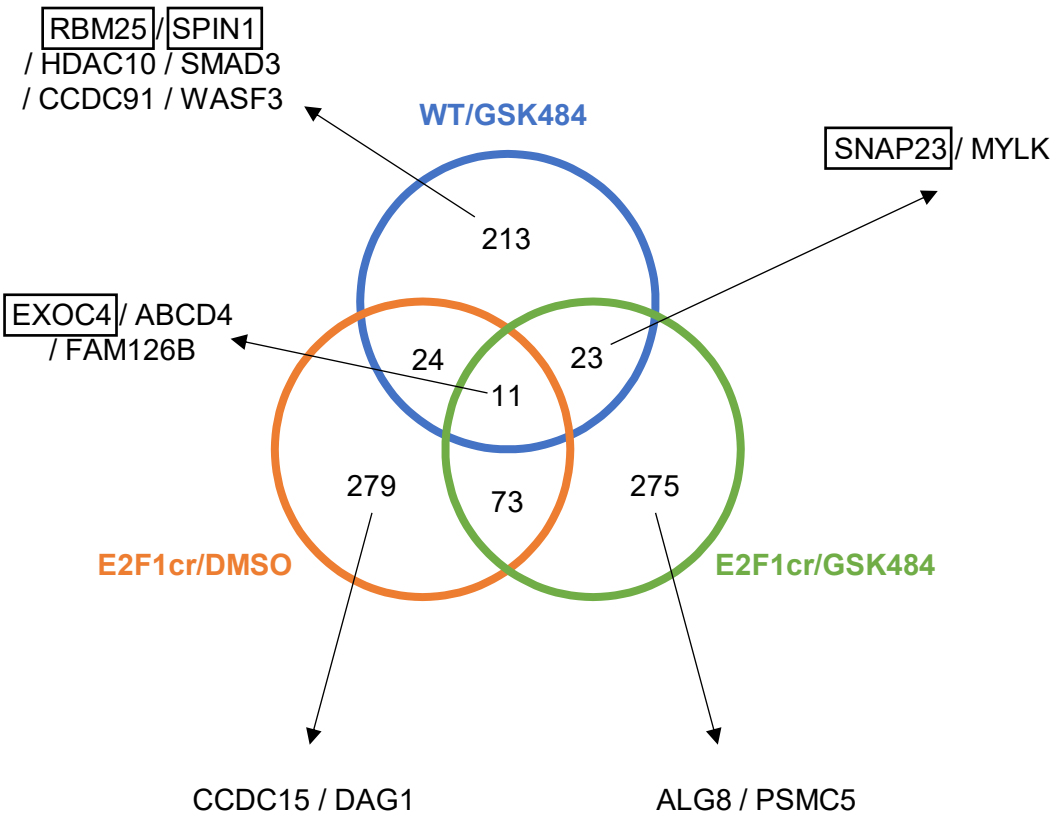


Figure 4-2 Examples of alternatively spliced genes identified in RNA-seq.

Venn diagram showing the number of significant (FDR < 0.01) AS genes in different treatment conditions with reference to the WT-DMSO control, as well as displaying some example genes. Genes in the box (RBM25, SPIN1, SNAP23 and EXOC4) were examples for which RT-qPCR validation results were shown in Figure 4-3.

4.3 RT-qPCR Validation of AS events identified in RNA-seq.

We performed RT-qPCR experiments to validate AS events identified through the rMATS analysis of the RNA-seq data. cDNA templates were freshly prepared from the original sequenced RNA samples derived from HCT116 WT and E2F1cr cells treated with GSK484 or DMSO. An RT-qPCR experiment using these cDNA samples then measured the delta threshold cycle (ΔC_T) value for each splice variant over a reference gene (GAPDH). The inclusion/exclusion (I/E) ratio for exon or intron of interest was calculated in $\Delta\Delta C_T$ (difference in ΔC_T values between inclusion and exclusion transcript variants) as a quantitative measurement of each AS event, and presented in the form of fold enrichment as $2^{-\Delta\Delta Ct}$. Figure 4-3 shows the result of RT-qPCR validation for four example genes (EXOC4, RBM25, SPIN1, SNAP23). Even though we were able to see a trend of changes in their I/E ratios consistent with what we observed in the RNA-seq, large variations between biological repeats and small differences between treatments meant that the result did not pass the significance threshold despite multiple rounds of repeat experiments.

As a potential way to improve the RT-qPCR results, we decided to perform the same experiment in the background of p53 knockout (HCT116 p53-/- cells). The p53 protein is a transcription factor and is most famously known as a tumour suppressor (Zilfou & Lowe, 2009). Functionally, p53 suppresses the expression of various genes related to cell cycle progression, and in response to DNA damage, can activate apoptotic pathways by up-regulating genes including BAX, PUMA and NOXA (Riley et al., 2008). Not surprisingly, this tumour suppressor gene is mutated in more than half of human cancers, and correspondingly many of the cancer cell lines do not also have active p53 proteins in their genomes. HCT116 cell line expresses wildtype p53 protein, but HCT116 p53-/- cell

line with p53 knockdown is also known and used as a colorectal cancer model in multiple studies including previous work from our research group (Barczak et al., 2020). The authors demonstrated the impact of PRMT5-dependent methylation on the E2F1-regulated transcriptome. Interestingly, this similar RNA-seq in HCT116 p53-/- cells with the PRMT inhibitor treatment detected a much larger number of AS events; for example, when examining the E2F1cr-DMSO condition with respect to WT-DMSO condition, almost three times as many AS events were discovered HCT116 p53-/- cells compared to the present RNA-seq in HCT116 p53+/+ cells (Figure S-3). Considering that p53 plays a significant role in inducing apoptosis, one can hypothesise that p53-deficient cells might be less sensitive to AS disturbance-induced apoptosis, and thus a greater level of splicing changes might be observed in HCT116 p53-/- cells. We therefore reasoned to continue investigating whether E2F1/PAD4-regulated AS events could be validated by RT-qPCR in HCT116 p53-/- cells.

HCT116 p53-/- WT and E2F1cr cells, which were previously generated in the lab (Barczak et al., 2020), were treated with 10 μM GSK484 or DMSO (negative control) for 72 hours, matching the condition in which the RNA-seq experiment was carried out in HCT116 p53+/+ cells. The expression of E2F1 and PAD4 was confirmed by immunoblotting (Figure S-4), and the FACS analysis confirmed that the PAD4-specific inhibitor at this concentration and incubation time did not significantly affect the cell cycle states of HCT116 p53-/- cells (Figure S-5). Biological triplicates of RNA samples were purified and reverse-transcribed to cDNA templates, which were then used in RT-qPCR to validate the AS events from the RNA-seq. This switch of cell line indeed improved the results, and we successfully validated several AS events to be regulated by the E2F1/PAD4 axis in HCT116 p53-/- cells (Figure 4-4). One of the validated AS events

were skipped exon 7 (SE7) of EXOC4. In the RNA-seq, this transcript was discovered to be alternatively spliced in all three treatment conditions with positive delta psi (percent spliced in) values (Figure 4-4A (ii)), meaning that both PAD4 and E2F1 up-regulate the inclusion of this AS-target exon. The RT-qPCR experiment here confirmed that the I/E ratio of this exon was significantly increased in all three conditions over the control WT-DMSO condition in HCT116 p53-/- cells (Figure 4-4A (iii)), therefore validated the RNA-seq result in accord. However, this is not the case for all the validated AS genes. For example, the exon 2 inclusion of RBM25 was increased significantly only in the WT-GSK484 treatment in the RNA-seq, but the RT-qPCR result suggested that this was also up-regulated in the E2F1cr-DMSO and E2F1-GSK484 treatments (Figure 4-4B). For SPIN1, the RNA-seq indicated the psi value for its skipped exon 3 was negative in WT-GSK484 condition, meaning this treatment condition promoted the exclusion of this exon. However, under RT-qPCR the I/E ratio was verified to be increased in all three treatments.

The trend we found in changes in each AS event was the same between HCT116 p53+/+ and p53-/- cell lines, but the latter demonstrated it with greater magnitudes and smaller variations (Figure 4-3 and Figure 4-4). Even though we cannot disregard the relevance of the p53 protein which should affect cell physiology in many ways, it should be emphasised that our current objective is to elucidate the molecular mechanism underlying how PAD4 and E2F1 regulate AS, and the purpose of the RT-qPCR validation experiment was to find the E2F1/PAD4-regulated AS events which can be manipulated in downstream experiments. For this basis, we reasoned to continue using HCT116 p53-/- cells where we successfully validated several AS events with good reproducibility.

For 4 splicing-target genes (EXOC4, RBM25, SPIN1, and SNAP23) with good validation results, the total levels of RNA transcripts were measured by RT-qPCR. The Ensembl

genome database (CRCh37) was mined to find a constitutive exon shared amongst all RNA transcripts from each gene, against which a primer pair for RT-qPCR was designed. As illustrated in Figure 4-5, the total expression of these 4 genes was not significantly impacted under any of the treatment conditions. This result was consistent with the RNA-seq analysis of DEGs and also confirmed the AS changes of these transcripts were not an indirect result of differential expression.

We also mined the database of the National Library of Medicine (https://www.ncbi.nlm.nih.gov/gene) to investigate the known functions and annotations of proteins and splice variants derived from the validated genes (EXOC4, RBM25, SPIN1, and SNAP23) (Table 4-1). EXOC4 is probably the most characterised gene out of the four and seems to be involved in many biological pathways as a component of the Exocyst complex (Martin-Urdiroz et al., 2016; Tanaka et al., 2017). RBM25 is an RNA binding protein which reportedly has a role in alternative splicing, indicating a potential feedback mechanism through splicing of splicer (Carlson et al., 2017; Y. Ge et al., 2019). Even though the database did not identify annotations for specific splice variants from each gene, the analysis highlighted that all validated genes have some roles in major cellular pathways and are reportedly deregulated in multiple diseases including cancers (Table 4-1). Given such biological and pathological relevance, we decided to continue using these four validated genes for subsequent experiments.

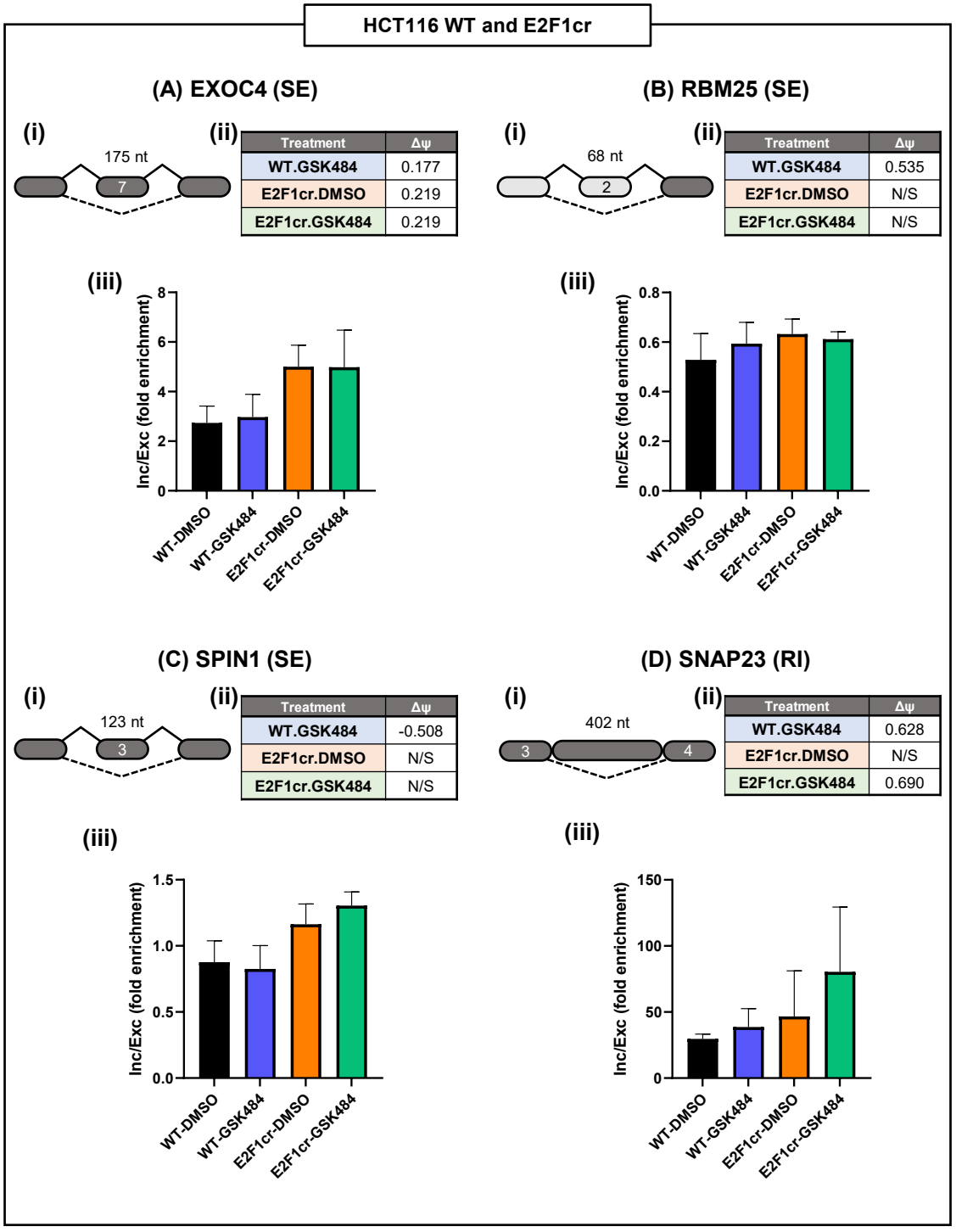


Figure 4-3 RT-qPCR validation for AS genes in HCT116 cells.

Several AS events identified in the RNA-seq were analysed by RT-qPCR using specific primers. *N* = 3 (the same RNA samples sent for the RNA-seq). (A) EXOC4 SE, (B) RBM25 SE, (C) SPIN1 SE, and (D) SNAP23 RI. (i) gene stricture, (ii) psi values from the RNA-seq, and (iii) the inclusion/exclusion ratio (absolute fold enrichment) by RT-qPCR. (± S.D.).

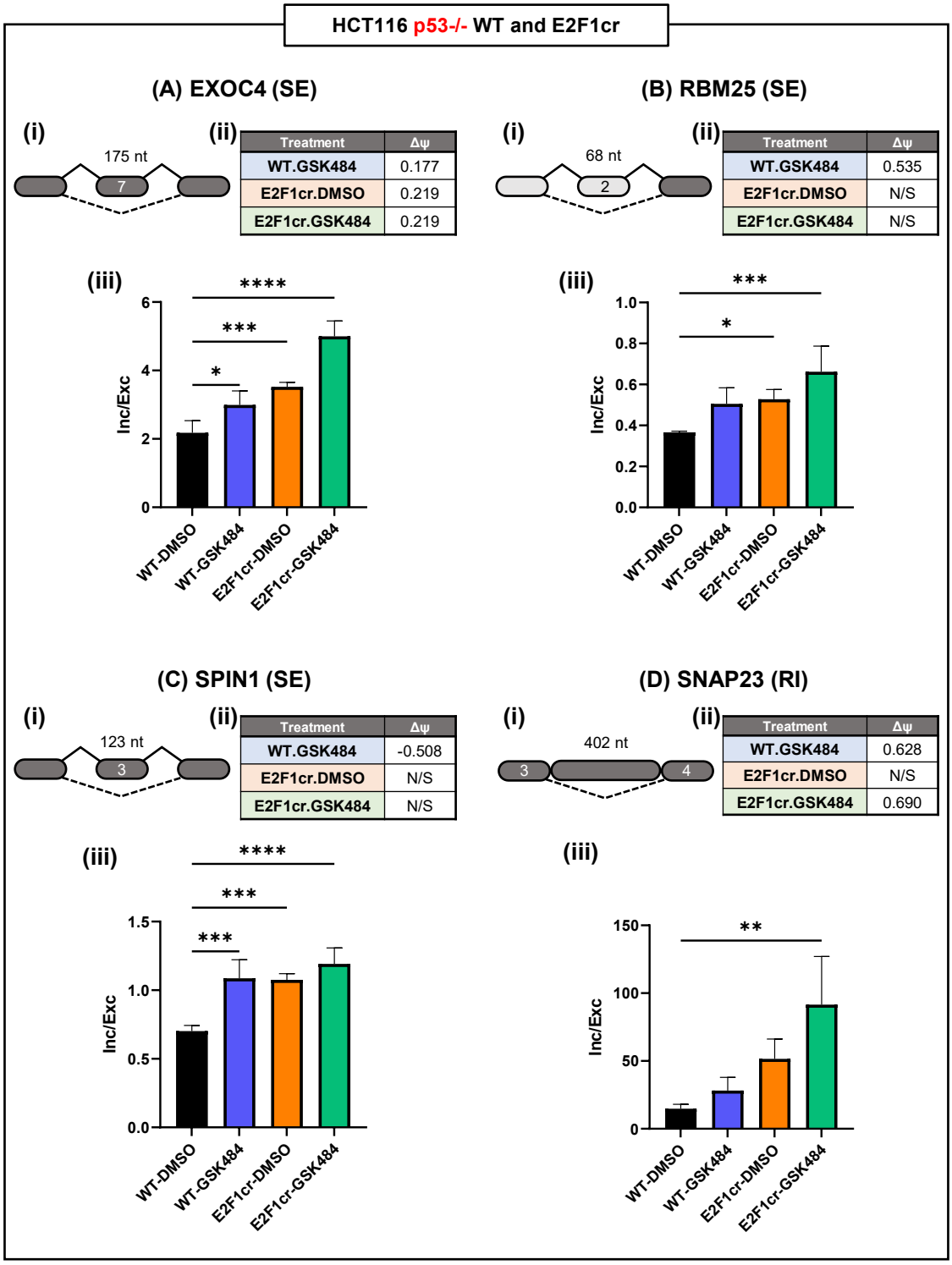


Figure 4-4 RT-qPCR validation for AS genes in HCT116 p53-/- cells.

WT and E2F1cr HCT116 p53-/- cells were treated with 10 μM GSK484 (or DMSO) for 72 hours, and RNA was isolated. Several AS events identified in the RNA-seq were analysed by RT-qPCR using specific primers. *N* = 3. (A) EXOC4 SE, (B) RBM25 SE, (C) SPIN1 SE, and (D) SNAP23 RI. (i) gene stricture, (ii) psi values from the RNA-seq, and (iii) the inclusion/exclusion ratio (absolute fold enrichment) by RT-qPCR. (± S.D. * $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$, **** $p < 0.0001$).

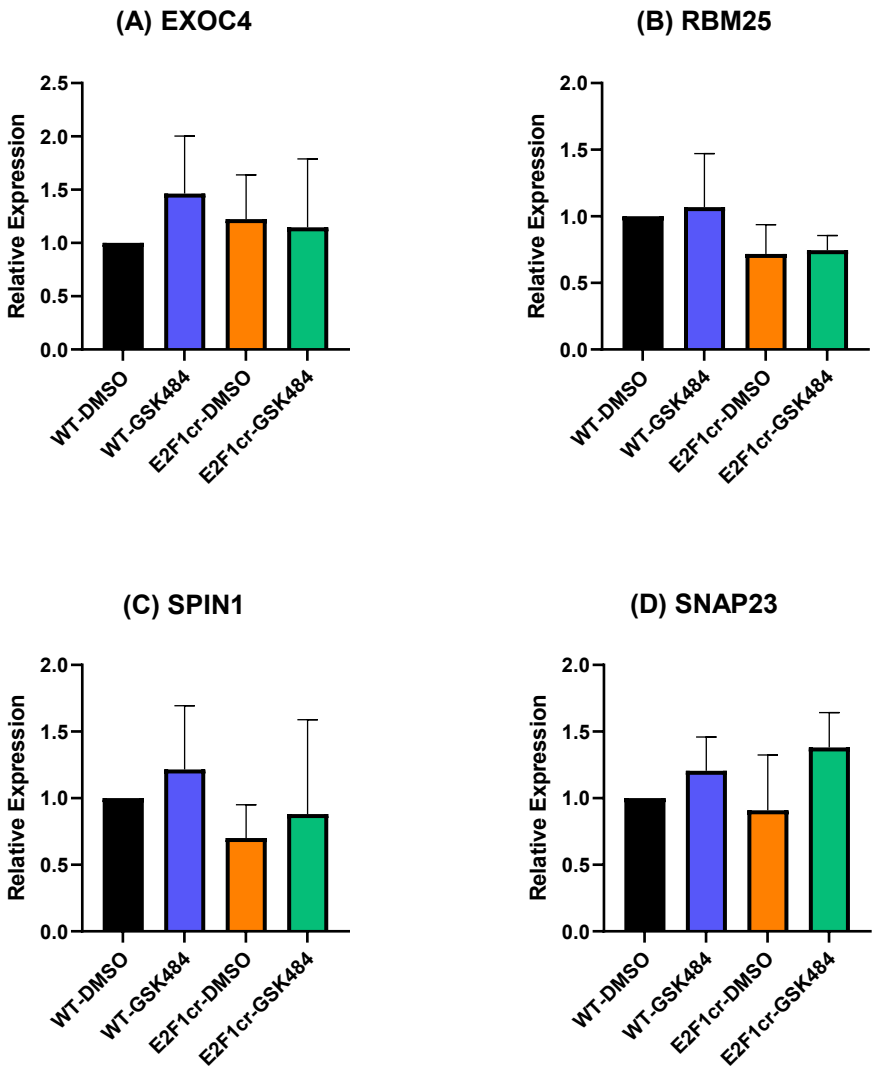


Figure 4-5 Total expression of AS genes in HCT116 p53-/- cells.

The RNA samples of HCT116 (p53-/- cells) used in Figure 4-4 were again analysed by RT-qPCR using primers specific to each constitutive exon of the(A) EXOC4, (B) RBM25, (C) SPIN1, and (D) SNAP23 genes, to measure the total expression (relative to WT-DMSO) of each gene. *N* = 3. (± S.D.)

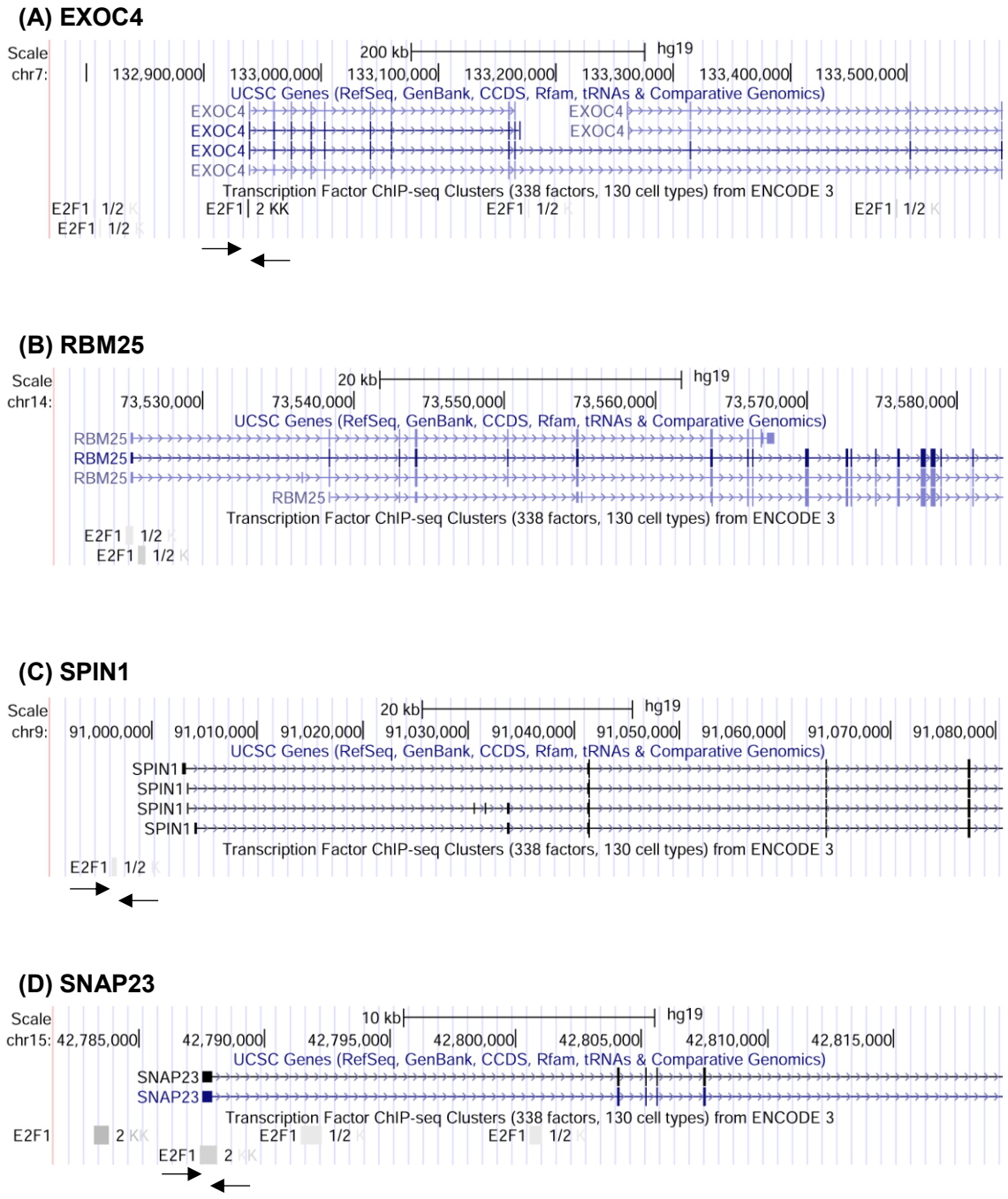


Figure 4-6 The validated AS genes were E2F1 targets.

Screenshots from UCSC Genome Browser (http://genome.ucsc.edu, GRCh37/h19 assembly) (Kent et al., 2002) illustrating the gene structures of (A) EXOC4, (B) RBM25, (C) SPIN1, and (D) SNAP23, with indicated E2F1 ChIP-seq peaks (grey boxes) from the data deposited on the ENCODE project (http://genome.ucsc.e du/ENCODE/). The indicated binding sites were used to design primers for the ChIP experiment (Figure 4-7 and Figure 4-8); shown as arrows below each binding sites.

Table 4-1 Functions of validated AS genes from the RNA-seq.

Gene	Function	Identified Splicing Event
EXOC4 (Sec8) Exoxyst Complex component 8	A component of the 'exocyst' complex. Target secretory vesicles to the intercellular membrane compartment before fusion with the plasma membrane. Regulates many cellular activities (neurotransmission / cell polarity / insulin secretion) by mediating exocytosis. Also reported to be involved with cell cycle regulation and apoptosis. K/O of this gene is embryonic lethal in mice.	Skipped exon 7 No annotation available in exon 7 according to ENSEMBL.
RBM25 RNA-Binding Motif protein 25	RNA-binding protein, playing a role to regulate mRNA alternative splicing. Controls apoptosis through a regulation of alternative splicing of BCL2L1 gene.	Skipped exon 2 Exon 2 is located in untranslated region.
SPIN1 Spindlin 1	Binds histone H3 by recognising trimethylated at Lys -4 and asymmetrically dimethylated at Arg-8, and acts to activate Wnt signalling pathway downstream of PRMT2. Promotes cell proliferation in cancer cells via activation of Wnt signalling. Overexpression of the gene result in metaphase arrest and chromosome instability.	Skipped exon 3 No annotation available in exon 7 according to ENSEMBL.
SNAP23 Synaptosomal-associated protein 23	Essential component of the high affinity receptor for the general membrane fusion machinery and an important regulator of transport vesicle docking and fusion.	Retain intron 3

For each of the validated AS genes (EXOC4, RBM25, SPIN1, and SNAP23), the known physiological and pathological and their splicing events identified in RNA-seq with annotations were described. Functional annotations were provided by the RefSeq database from the National Centre for Biotechnology Information (NCBI) website (https://www.ncbi.nlm.nih.gov/refseq/) (O'Leary et al., 2016).

4.4 E2F1 binds to the promoters of its AS target genes.

All of the four validated AS genes (EXOC4, RBM25, SPIN1, and SNAP23) were identified as E2F1-targets through the ENCODE ChIP-seq analysis (Figure 4-6). We decided to confirm this by performing E2F1 ChIP experiments in HCT116 p53-/- cells using primers targeting the E2F1-binding sites proposed in this ChIP-seq data (Figure 4-6). First, to examine whether E2F1 binds to the AS target promoters at all, an E2F1 ChIP experiment was performed in untreated HCT116 p53-/- cells, and the immunoprecipitated chromatin was analysed by qPCR for the indicated E2F1-binding sites for the AS genes (EXOC4, SPIN1, and SNAP23) and CDC6 (an established E2F1-target for transcriptional regulation) as a positive control. The result demonstrated the association of E2F1 with the EXOC4 and CDC6 promoters, whilst no significant binding was observed at the SPIN1 and SNAP23 promoters (Figure 4-7).

Since the ENCODE project contained a collection of E2F1 ChIP-seq experiments from multiple tissue sources, E2F1 may only bind to the SNAP23 and SPIN1 genes in a tissue-dependent manner and not in HCT116 p53-/- cells. Conversely, given the transcriptional activity of E2F1 widely regulated by various mitogenic factors and different types of post-translational modifications (PTMs) (Dubrez, 2017; Roworth et al., 2015), it would not be inconceivable that the E2F1 binding to its spliced targets is also influenced by similar stimuli. Citrullination is among the PTMs that affect the interaction between E2F1 and chromatin, where citrullinated E2F1 was reportedly localised at the pro-inflammatory genes and promoted their expression (Ghari et al., 2016). We thus questioned whether PAD4 plays any role in modulating the E2F1-chromatin association at the AS target promoters and carried out another round of E2F1 ChIP with different levels of the PAD4 activity.

A group of HCT116 p53-/- cells were transfected with HA-PAD4 plasmids to increase the intracellular expression of PAD4 (Figure 4-8B). Another group is treated with 10 μM GSK484 for 48 hours after the HA-PAD4 transfection, and these two experimental groups were analysed with reference to the control group of untreated cells expressing an empty plasmid vector. Cells from each condition were lysed and incubated with either HA, E2F1, or IgG antibodies, and the IP-ed chromatin was analysed by qPCR. The result for a positive control gene CDC6 showed a good binding of E2F1 to its promoter region in all conditions, proving the ChIP experiment successfully pulled down E2F1-associated chromatins (Figure 4-8A). We observed the E2F1 enrichment at the EXOC4 promoter as well as SNAP23 this time in the untreated cells, but only with a small magnitude.

Of profound interest was that the E2F1-chromatin interaction was significantly enhanced by the ectopic expression of PAD4. This impact was found at the EXOC4, SNAP23, and CDC6 promoters, but not at SPIN1. Importantly, we further demonstrated that the subsequent GSK484 treatment in PAD4 overexpressing cells can impair this enhanced interaction between E2F1 and target promoters, whereby the E2F1 enrichment at the EXOC4 and SNAP23 was decreased by almost 3-fold. This result clearly suggests that PAD4-driven citrullination is a crucial factor in the regulation of E2F1 association with its AS target genes.

Interestingly, the effect of GSK484 destabilising the E2F1-chromatin interaction was not observed for the CDC6 promoter. We thus hypothesised that the citrullination of E2F1 may have different degrees of impact for its transcription targets from those targeted for AS-based regulation. To test this, we checked the binding of E2F1 to the promoters of additional transcriptionally regulated targets, including CDC25A and TK. A similar result to the CDC6 promoter was found for the TK gene, where the GSK484 treatment did not

exhibit an effect to reduce the E2F1 enrichment enhanced by the HA-PAD4 expression (Figure 4-9). However, at the CDC25A promoter, E2F1 binding was significantly reduced and returned to the control level after PAD4 activity was inhibited (Figure 4-9). This result was probably insufficient to draw a conclusion, but it at least indicates the model of stronger interactions at the transcription targets and weaker interactions at the AS target might be too simplified.

In addition to E2F1, we also conducted an HA-PAD4 ChIP experiment. Even though there was a remarkable influence of dynamic PAD4 activity observed in the E2F1-chromatin interaction, the present result indicated no significant enrichment of HA-tagged PAD4 at the target promoters over IgG in any treatment conditions (Figure 4-8).

Overall, it has been demonstrated that PAD4 plays a pivotal role in modulating the interaction of E2F1 with the AS target promoters, which may be a key component of the mechanisms underlying the E2F1/PAD4-dependent AS perturbations. However, it is noteworthy that no E2F1 binding was observed at the SPIN1 promoter, indicating that E2F1 and PAD4 can also regulate splicing of their target genes through other pathways, and there should be at least a mechanism independent of the E2F1-chromatin association.

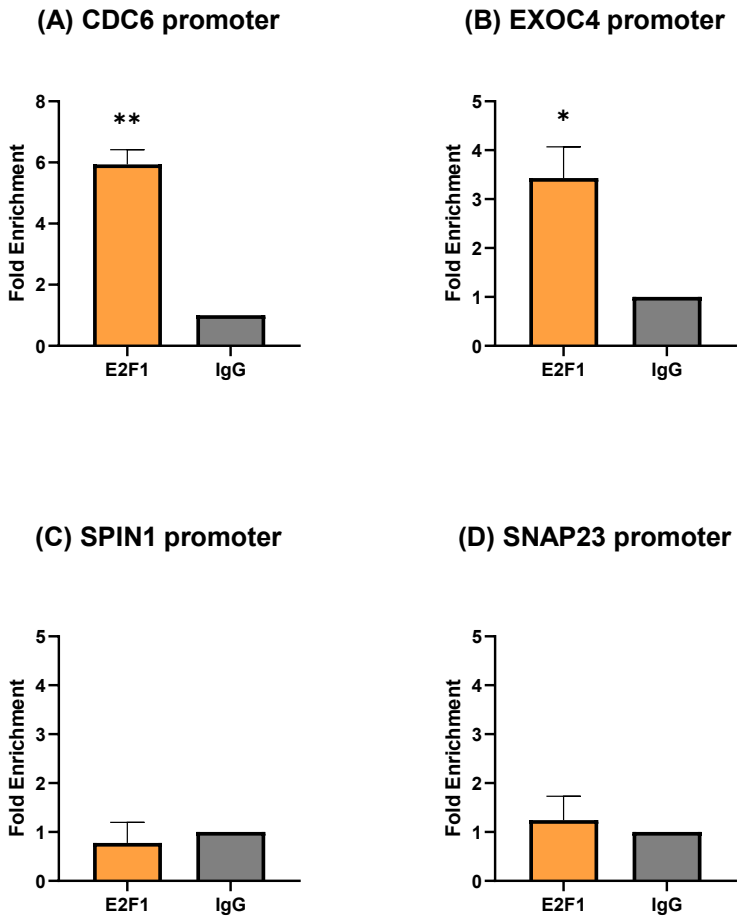


Figure 4-7 E2F1 binds to AS target promoters.

Untreated HCT116 cells were immunoprecipitated using an anti-E2F1 antibody, and analysed by qPCR using the specific primers for the E2F1 binding sites indicated in the UCSC Genome Browser (Figure 4-6). The results are displayed in the form of relative fold enrichment over IgG level. (A) CDC6 (positive control, a known transcriptional E2F1-target), (B) EXOC4, (C) SPIN1, and (D) SNAP23. *N* = 3. (± S.D. T-test was performed to analyse the statistical significance of E2F1 enrichment signal over the IgG signal and the result was indicated above the E2F1 bar. * $p < 0.05$, ** $p < 0.005$)

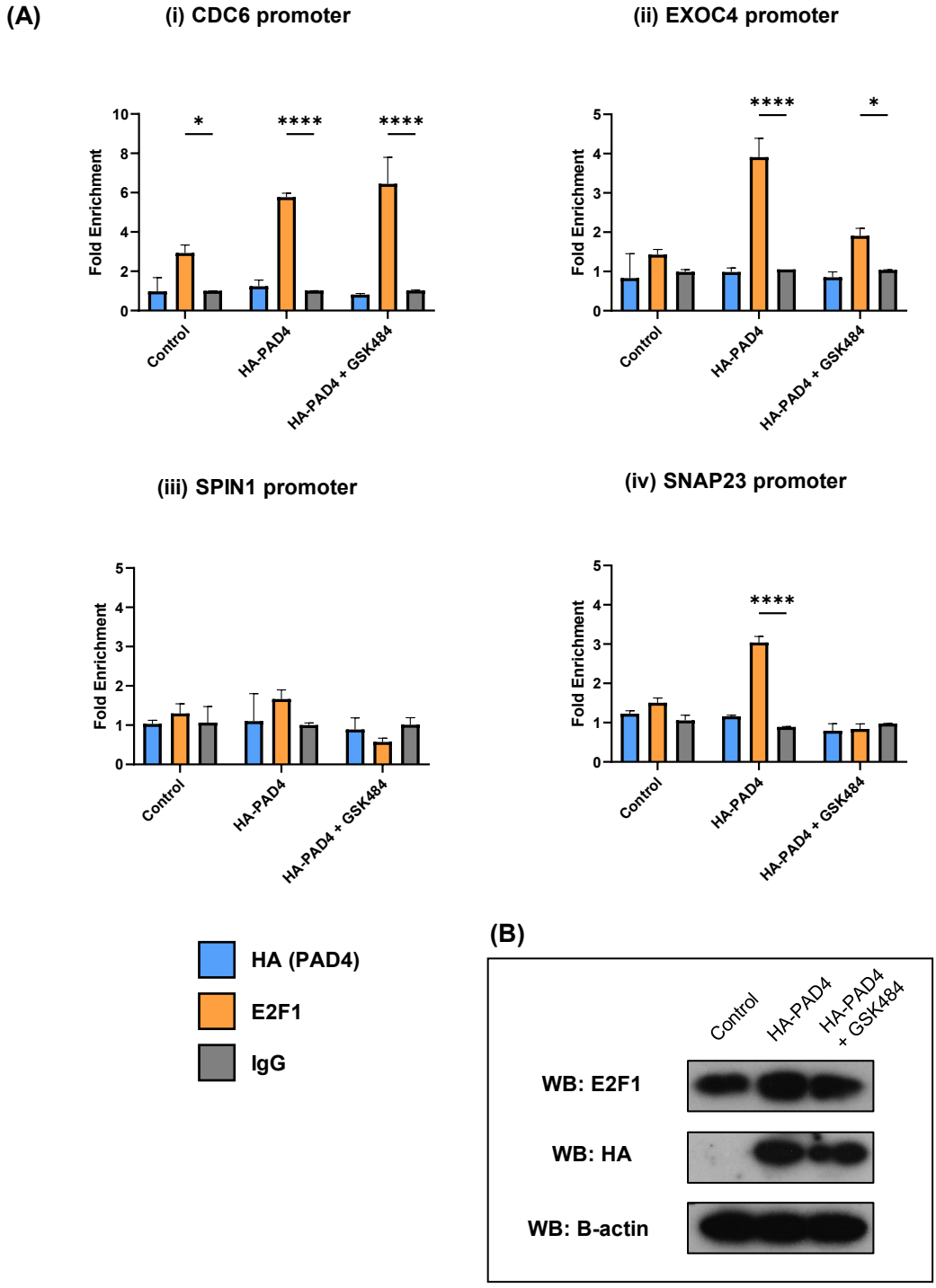


Figure 4-8 E2F1 binding to AS target promoter is modulated by ectopic PAD4.

HCT116 p53-/- cells were transfected with 3 μg HA-PAD4 and treated with 10 μM GSK484 or DMSO. (A) ChIP was performed using HA, E2F1 and IgG antibodies, and analysed by qPCR in the form of relative fold enrichment over IgG level, which showed E2F1 binding was modulated by PAD4 activities. (i) CDC6 (positive control, a known transcriptional E2F1-target), (ii) EXOC4, (iii) SPIN1, and (iv) SNAP23. *N* = 3. (± S.D. * $p < 0.05$, ** $p < 0.005$) (B) Immunoblots showing a successful ectopic expression of HA-PAD4.

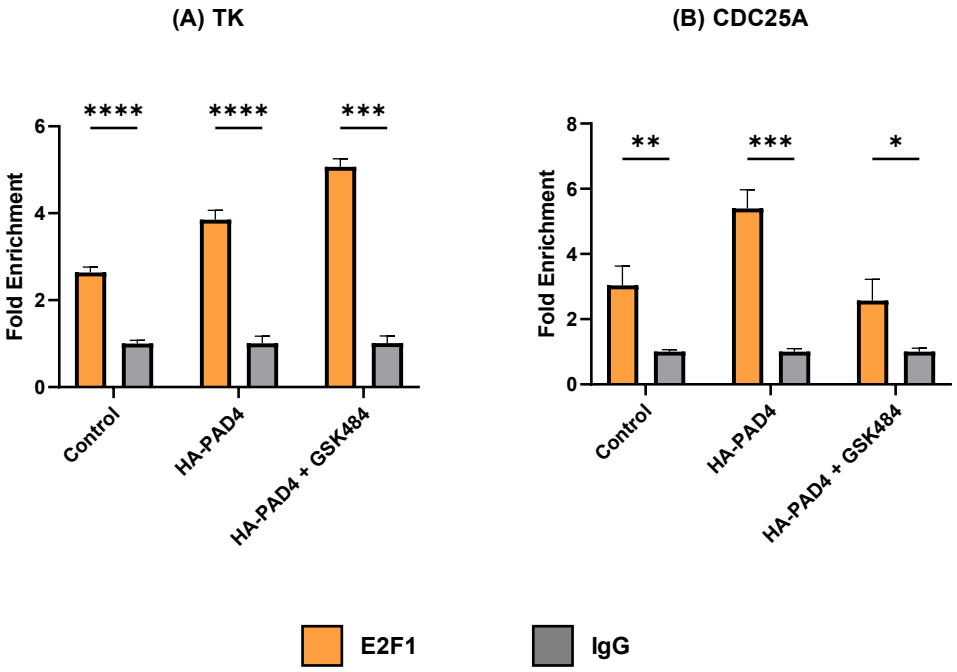


Figure 4-9 E2F1 binding to transcription target promoter is less affected by PAD4.

The ChIP samples used in Figure 4-8 were again analysed by qPCR in the form of relative fold enrichment over IgG level, using primers specific to two additional E2F1 transcriptional targets: (A) TK and (B)CDC25A *N* = 3. (± S.D. * $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$, **** $p < 0.0001$)

4.5 EXOC4 AS is dependent on citrullination of E2F1.

The interplay between E2F1 and PAD4 appears to trigger AS perturbation, and its underlying mechanism was indicated to partly involve the PAD4-dependent citrullination of E2F1 augmenting its chromatin interaction at the promoter of target genes. However, the question of whether this splicing regulation is dependent on the citrullination of E2F1 itself remains unanswered. To specifically address this query, we decided to utilise the citrullination-defective mutant of E2F1, which was generated through site-specific mutagenesis in a previous work from our research group (Ghari et al., 2016). In the described study, PAD4 was discovered to principally target R109 and R127 arginine residues, but the data indicated R111 and R113 residues could also be involved. In this citrullination-defective mutant, these four arginine (R) residues of E2F1 have been replaced by lysine (K) (Figure 4-10A), which should effectively block the PAD4-catalysed citrullination at these target sites. The resultant mutant ($E2F1^{R4K}$) was indeed shown to be less efficiently citrullinated by PAD4 while retaining a similar activity as a transcription factor (Ghari et al., 2016).

Here, HCT116 p53-/- WT cells were transfected with either HA-$E2F1^{WT}$ (wild type), HA-$E2F1^{R4K}$, or pcDNA empty plasmid vector (Figure 4-10B(ii)), and then treated with 10 μM GSK484 or DMSO (negative control). RNA samples were isolated from biological triplicates of each treatment condition and cDNA was synthesised. With the cDNA templates, an RT-qPCR experiment was performed to examine the ability of the E2F1 citrullination defective mutant to regulate alternative splicing. Figure 4-10B shows the I/E ratio of EXOC4 skipped exon 7 in three different conditions with reference to the DMSO-treated group within each treatment. GSK484 treatment was found to promote the inclusion of this exon in control and $E2F1^{WT}$-expressing cells, which was expected and

consistent with the RT-qPCR validation result (Figure 4-3). Most strikingly, this effect of GSK484 was reduced when cells were transfected with the E2F1^{R4K} mutant. This result strongly suggests that the E2F1/PAD4-mediated regulation of AS, at least EXOC4 SE7, is directly dependent on the PAD4-mediated citrullination of E2F1, highlighting an initial insight into the mechanistic details underlying E2F1/PAD4-regulated AS perturbations.

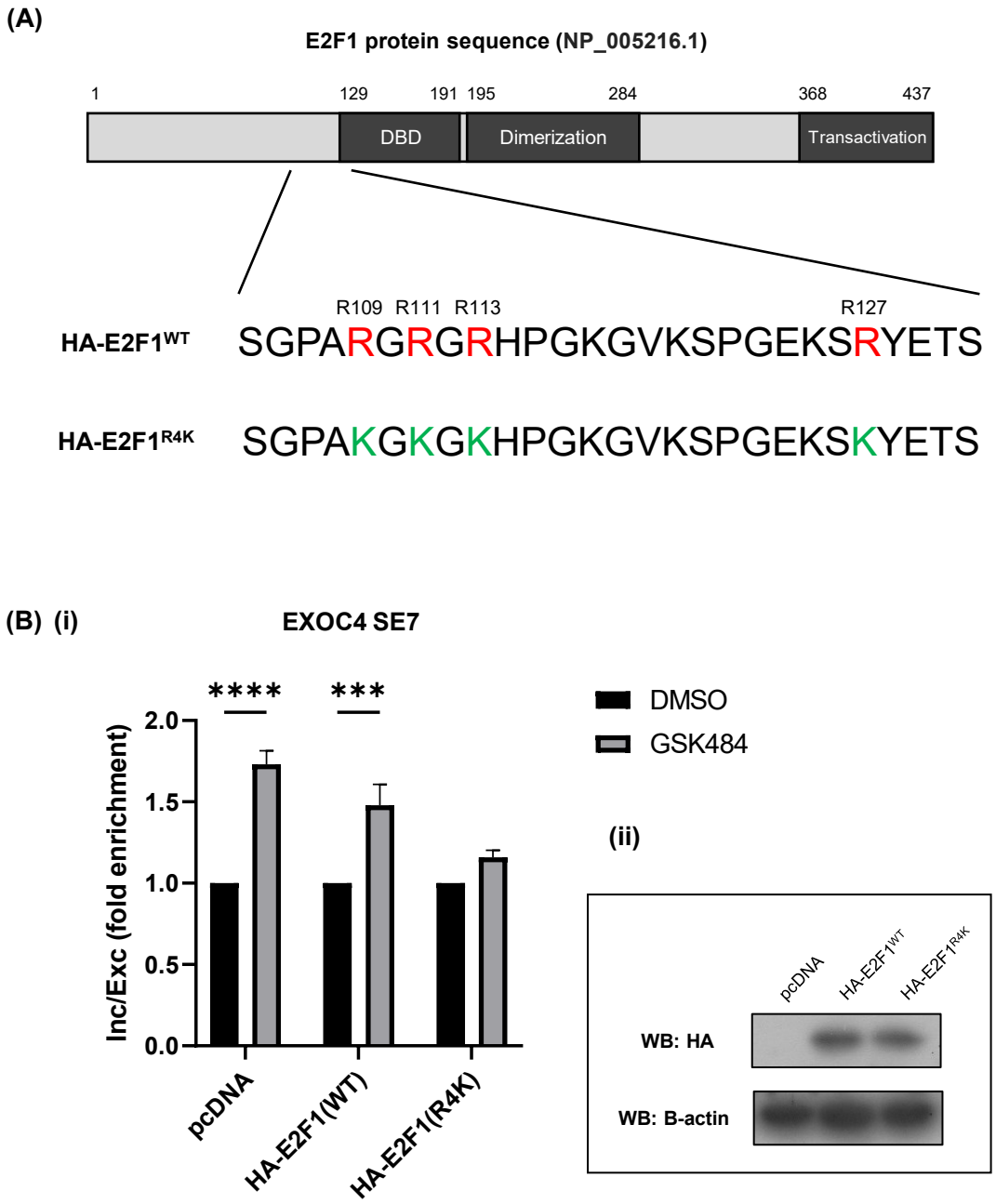


Figure 4-10 Citrullination of E2F1 plays a key role in AS regulation.

(A) Diagram illustrating the E2F1 protein structure (NP_005216.1) (Dubrez, 2017) and partial sequences of wildtype ($E2F1^{WT}$) and citrullination-defective mutants ($E2F1^{R4K}$) used in this study. PAD4-target arginine residues (shown in red) were replaced by lysine residues (shown in green) in $E2F1^{R4K}$ mutant. (B) HCT116 p53-/- cells were transfected with 3 µg HA-$E2F1^{WT}$ or HA-$E2F1^{R4K}$ and treated with 10 µM GSK484 or DMSO. Isolated RNA samples were analysed by RT-qPCR, shown in the form of the I/E ratio of GSK484-treated cells relative to that of DMSO-treated cells within each experimental condition. (± S.D. $p > 0.05$, * $p < 0.05$, *** $p < 0.0005$, **** $p < 0.0001$). (ii) Immunoblots displaying the expression of HA-$E2F1^{WT}$ and HA-$E2F1^{R4K}$. $N = 2$.

4.6 Chapter Summary

We previously revealed that the PAD4 inhibition and E2F1 knockdown have a significant impact on the alternative splicing profile of HCT116 cells (Figure 3-9), and in this chapter, we tried to validate this result by RT-qPCR. Since the magnitude of splicing changes observed with RT-qPCR in HCT116 cells was not significant enough (Figure 4-3), we switched the cell line to HCT116 p53-/- cell lines which were potentially less sensitive to an AS disturbance-induced apoptosis as per our previous work (Barczak et al., 2020). We successfully validated several AS events so that they were regulated by PAD4 and E2F1 in HCT116 p53-/- cells (Figure 4-4), in which we observed the same trend of splicing changes as in HCT116 p53+/+ but with a greater magnitude. Given that the tumour suppressor p53 protein is frequently mutated or inactivated in cancer cells (Hafner et al., 2019; Olivier et al., 2002), this result may emphasise the relevance of this E2F1/PAD4-mediated splicing to cancer biology. In addition, it indicates that p53 may directly influence the interplay between E2F1 and PAD4. This will be further discussed in Chapter 8.

Since all of the four validated AS genes were identified as E2F1-targets through the E2F1 ChIP-seq analysis (Figure 4-6), we next examined E2F1 binding to the promoter of those genes by E2F1 ChIP experiments. Interestingly, the majority of genes demonstrated E2F1 binding to their promoter regions, though not all did, suggesting the potential influence by factors like tissue type, mitogenic signals, and post-translational modifications (Figure 4-7). Indeed, we subsequently demonstrated that an increased expression of PAD4 can enhance such E2F1-chromatin interactions at the target promoter regions. The PAD4-driven increased binding of E2F1 was suppressed when cells were subsequently treated with GSK484 (Figure 4-8), which further supports the view that PAD4-dependent

citrullination undertakes a critical role in regulating the DNA binding affinity of E2F1. Interestingly, this impact of the GSK484 treatment was more apparent on the AS target promoters, and the E2F1 enrichment at its transcription-based targets appears to be relatively maintained (Figure 4-8 and Figure 4-9). Even though the view of 'stronger association at the transcription targets and weaker association at the AS targets' looks a bit over-simplified, such difference would be an interesting avenue of future research and potentially provide a good mechanistic insight underlying the AS perturbations regulated by the E2F1/PAD4 axis. Direct influence of E2F1 citrullination was also examined by using the citrullination-defective E2F1 mutant ($E2F1^{R4K}$). The result using the EXOC4 SE event as an example demonstrated a direct role that PAD4-mediated citrullination of E2F1 undertakes in regulating such AS event, whereby the previously observed effect of PAD4 inhibition to increase the inclusion of exon 7 in this gene was effectively abolished in the cells expressing the $E2F1^{R4K}$ mutant (Figure 4-10).

In summary, using validated AS events from the RNA-seq, we have provided some initial insight into the mechanism underlying the E2F1/PAD4-driven AS perturbations, whereby PAD4 seems to undertake a pivotal role in promoting the E2F1-chromatin association. The results also highlighted the intricate nature of the E2F1/PAD4 axis and suggested more than a single pathway appears to be influenced by the inhibition of PAD4 and E2F1 knockdown. It would therefore be of importance to characterise the precise functions of both E2F1 and PAD4, as well as identify other potential components taking part in this AS regulation. Among the validated genes, the regulation of EXOC4 SE7 represented an insubstantial involvement of both E2F1 and PAD4 and exhibited a clear dependence on the citrullination of E2F1. These findings suggest that EXOC4 SE7 could be a promising

candidate as the alternative splicing event to investigate the molecular mechanism further in future experiments.

Chapter 5

PAD4 and E2F1 regulate Alternative Splicing in combination with the Splicing Factor SRSF3

5.1 Introduction

In Chapter 4, we successfully validated the RNA-seq results by demonstrating some alternative splicing (AS) events were regulated by E2F1 and PAD4 using RT-qPCR, especially in HCT116 p53-/- cells that displayed the same tread of splicing changes with greater magnitudes as compared to p53+/+ cells (Figure 4-3 and Figure 4-4). The subsequent ChIP experiment illustrated that E2F1 bound to the promoter regions of some of these validated genes, and interestingly this binding was affected by PAD4 activity (Figure 4-8). Furthermore, we elucidated that a validated splicing event was directly dependent on the citrullination of E2F1 by utilising citrullination-defective mutants of E2F1 previously made in the lab (Figure 4-10) (Ghari et al., 2016). Taken together, the results shed light on the mechanistic insight of E2F1/PAD4-regulated splicing regulation. However, they also highlight the need to examine downstream regulation, especially to identify splicing factors which work in tandem with E2F1/PAD4 and have a more direct role in the regulation of splicing.

To further uncover this, we decided to investigate protein-protein interactions between E2F1/PAD4 and other splicing-related factors. As a starting point, the association between E2F1 and p100/TSN protein was examined, and the influence of PAD4-mediated citrullination on the interaction was also assessed. The p100/TSN (also known as SND1) protein has a highly conserved structure among eukaryotes, consisting of a tandem repeat of four staphylococcal nuclease (SN)-like domains (SN domains) at the N terminus and a fusion of a Tudor domain with a partial SN domain at the C terminus (Gutierrez-Beltran et al., 2016). This single protein has been implicated in multiple facets of gene expression; from transcription to RNA splicing, interference, stability, and editing (Ochoa et al., 2018). p100/TSN in particular has a pivotal role in the regulation of pre-mRNA splicing,

during which this protein directly interacts with the spliceosome components including a range of U snRNPs and Sm proteins in order to facilitate the sequential recruitment of this protein machinery and form the complex A then B (Gutierrez-Beltran et al., 2016). In addition, this splicing regulator also acts as a reader protein for PRMT5-driven methylation marks of E2F1 at R111 and R113, which allows E2F1 to associate with the splicing machinery and regulate AS (Roworth et al., 2019; Zheng et al., 2013). Given that PAD4 and PRMT1 competitively target the R109 residue of E2F1 (Ghari et al., 2016; Zheng et al., 2013), and such PRMT1-mediated methylation can impede PRMT5-dependent methylation at R111 and R113 (Cho et al., 2012; Zheng et al., 2013), it would be plausible to hypothesise that PAD4-regulated citrullination may affect the activity of PRMT5 over E2F1 and downstream interaction between E2F1 and p100/TSN.

Furthermore, we were also interested in the Serine/Arginine (SR)-rich splicing factors, that are characterised by the possession of a domain rich in arginine and serine residues (the RS domain) (Jeong, 2017). This family of proteins has a key role in defining alternative splice sites by recognising the *cis*-acting splicing regulatory elements (SREs) of mRNA. This is usually required for the spliceosome to essentially assemble around these ‘weak’ alternative splice sites that are less similar to the consensus sequences. Emerging research indicates that SR protein members are among the substrates of PAD2 and PAD4 (Lewallen et al., 2015; Tanikawa et al., 2018), and we reasoned to investigate its potential relevance in the E2F1/PAD4 axis.

5.2 PAD4 enhances the association between E2F1 and p100/TSN.

Our previous study established that PRMT5-driven methylation of E2F1 can augment the interaction between E2F1 and p100/TSN, which further facilitates the association between E2F1 and the splicing machinery at the target mRNA (Figure 5-1A) (Roworth et al., 2019). This symmetrical methylation takes place on R111 and R113 arginine residues and is known to be antagonised by PRMT1-mediated asymmetrical methylation at R109 (Zheng et al., 2013). Given that PAD4-mediated citrullination occurs primarily at R109 (Ghari et al., 2016), we reasoned that this citrullination may effectively inhibit PRMT1-mediated methylation of E2F1 on the same arginine residue and thus allow PRMT5 to predominantly methylate E2F1 on R111 and R113 as well as encourage p100/TSN recruitment as a potential regulatory mechanism of AS.

To test this hypothesis, we carried out the immunoprecipitation of E2F1 in cells to check the intracellular interaction between E2F1 and p100/TSN. HCT116 p53-/- cells were transfected with HA-PAD4-expressing plasmid or an empty plasmid vector. Both cell lines were lysed and immunoprecipitated with an anti-E2F1 antibody (or IgG for a negative control). The subsequent immunoblot successfully demonstrated an interaction of E2F1 with p100/TSN in cells with ectopically expressed PAD4, whilst this interaction was not observed in the control group (Figure 5-1B). The result strongly suggests that the increased level of PAD4 expression can enhance the binding of p100/TSN to E2F1, and therefore supports the hypothesis that PAD4-mediated citrullination at R109 of E2F1 may promote efficient methylation of E2F1 by PRMT5 at R111 and R113. We successfully revealed that PAD4 has a significant role in modulating the interaction between E2F1 and p100/TSN, and this protein represents a good candidate splicing factor that works downstream of the E2F1/PAD4 axis.

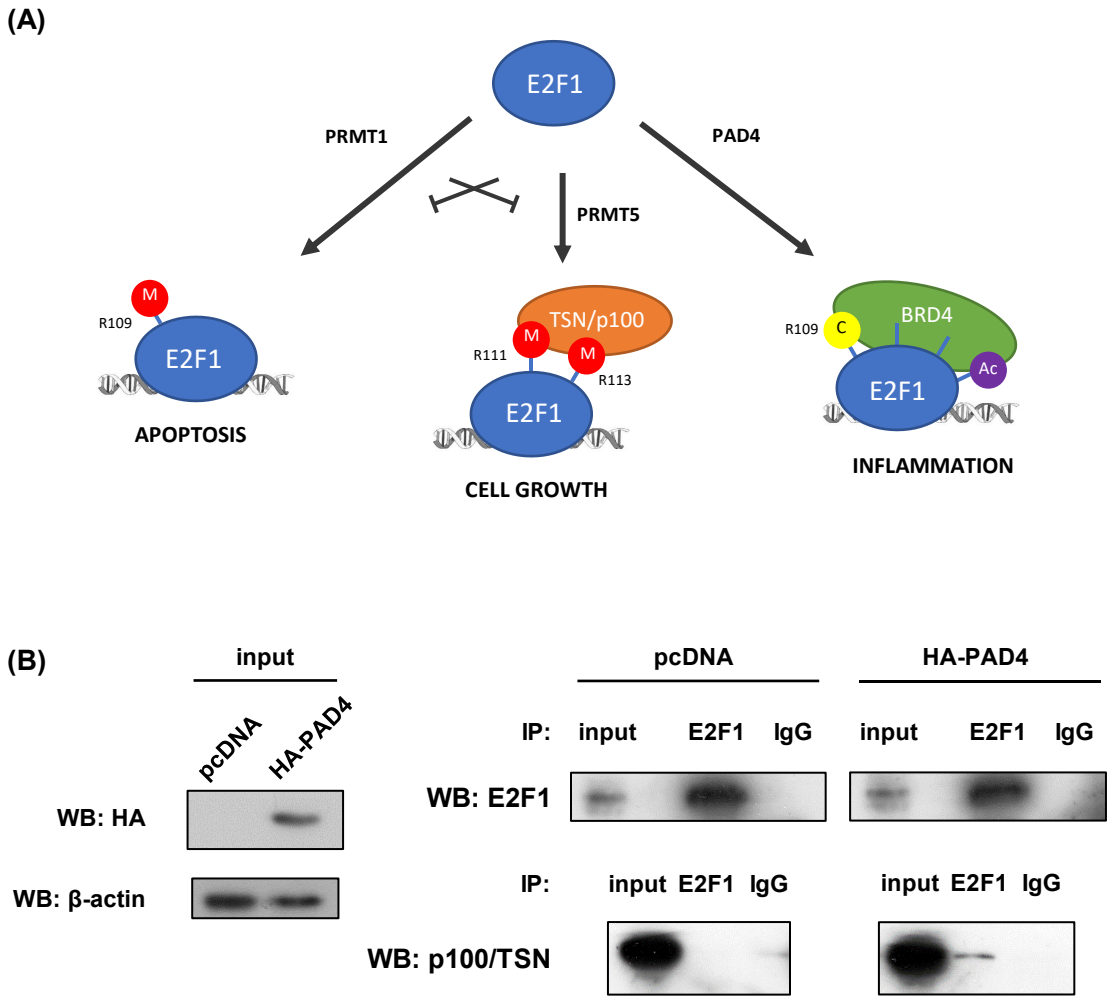


Figure 5-1 PAD4 overexpression enhances the interaction of E2F1 with p100/TSN.

(A) Schematic representation of how E2F1 activity is modulated by its post-translation modification: methylation-citrullination interplay as an example. E2F1 can be methylated at R109 by PRMT1 and at R111/R113 by PRMT5. The former promotes the E2F1 activity in pro-apoptotic pathways, whilst the latter channels into proliferation pathways. The activities of these two PRMTs are mutually exclusive. PRMT5-driven methylation marks are known to be read by the p100/TSN protein. R109 (and R127) can also be citrullinated by PAD4, which enhances the E2F1-chromatin association at the pro-inflammatory promoters. Citrullination can assist acetylation marks to recruit BRD4. (Zheng, et al. 2013, Roworth et al. 2019, Ghari et al. 2018). (B) HCT116 p53-/- cells were transfected with 2 μg of HA-PAD4, and the lysates were immunoprecipitated using an E2F1 antibody (or IgG for a negative control). Immunoblots showed a successful ectopic expression of PAD4 as well as the interaction of p100/TSN with E2F1 when PAD4 was over-expressed. *N* = 2.

5.3 SRSF3 is citrullinated by PAD4 and interacts with E2F1.

We have identified the p100/TSN protein as a candidate splicing-related factor playing a mechanistic role downstream of the E2F1/PAD4 axis. This protein is believed to regulate the sequential assembly of the spliceosome complex at the splice sites by directly interacting with its components such as U snRNPs and Sm proteins (Gutierrez-Beltran et al., 2016; Ochoa et al., 2018). In general, alternative splice sites are defined by relatively 'weak' signals with a low degree of similarity to the consensus sequences. As a result, these splice sites are usually not sufficient to facilitate the spliceosome assembly and instead require some help from extra auxiliary *trans*-acting factors that bind to *cis*-acting splicing regulatory elements (SREs) in proximity to these alternative splice sites (Clark & Thanaraj, 2002; X. Song et al., 2019). There are two major families of such *trans*-acting splicing regulators: Serine- and Arginine-rich proteins (SR proteins) and heterogeneous nuclear ribonucleoproteins (hnRNPs). SR proteins in particular caught our attention. Their characteristic RS domains, which play a pivotal role in protein-protein or protein-RNA interactions, are rich in arginine residues (Shepard & Hertel, 2009). Not surprisingly, a growing body of evidence suggested that these arginine residues of SR proteins are targets for PRMT-directed methylation and PAD-mediated citrullination (Rajyaguru & Parker, 2012; Tanikawa et al., 2018; Thandapani et al., 2013). We therefore wondered whether some members of the SR proteins family may work together with p100/TSN downstream of the E2F1/PAD4 pathway, and selected SR splicing factor 3 (SRSF3) as a candidate. This smallest member of SR proteins appears to undertake a critical role in regulating AS and is abnormally expressed in various types of tumours. Recently, proteomic analysis has revealed this protein is a substrate of PAD2, but the biological outcomes of this citrullination are yet to be elucidated (Lewallen et al., 2015;

Z. Zhou et al., 2020). To begin with, we tried to confirm whether SRSF3 can also be citrullinated by PAD4 in cells. U2OS cells with a Tet-On inducible expression system for PAD4, which was previously generated in the lab (Ghari et al., 2016), were utilised to augment the intracellular expression of PAD4. By immunoprecipitating endogenous SRSF3, we successfully demonstrated that this SR protein was citrullinated in PAD4-overexpressing cells (Figure 5-2A), which also suggests a functional redundancy between PAD2 and PAD4 for this splicing factor. This experiment was performed in collaboration with Amit Shrestha in our research group.

This interplay between SRSF3 and PAD4 stimulated us to further investigate the potential role of SRSF3 in E2F1/PAD4-mediated AS regulation. First, we examined whether SRSF3 can interact with E2F1, PAD4 or p100/TSN in our experimental setting. HCT116 cells were transfected with either or both HA-PAD4 and Flag-E2F1 plasmids, and the lysates were immunoprecipitated with an anti-SRSF3 antibody. Of great interest is that the immunoblot showed a clear interaction of SRSF3 with PAD4, E2F1 and p100/TSN when intracellular levels of E2F1 and PAD4 were increased by ectopic protein expression (Figure 5-2B). This strongly suggests that SRSF3 can form a complex with PAD4, E2F1, and p100/TSN, and made SRSF3 a strong candidate as a splicing factor working downstream of the E2F1/PAD4-axis for splicing regulation. Interestingly, a single transfection of PAD4 or E2F1 plasmids alone was not sufficient to recapitulate the interaction between SRSF3 and p100/TSN. This result emphasised the importance of both PAD4 and E2F1 for this interaction, thus further supporting the relevance of SRSF3 in the E2F1/PAD4 axis.

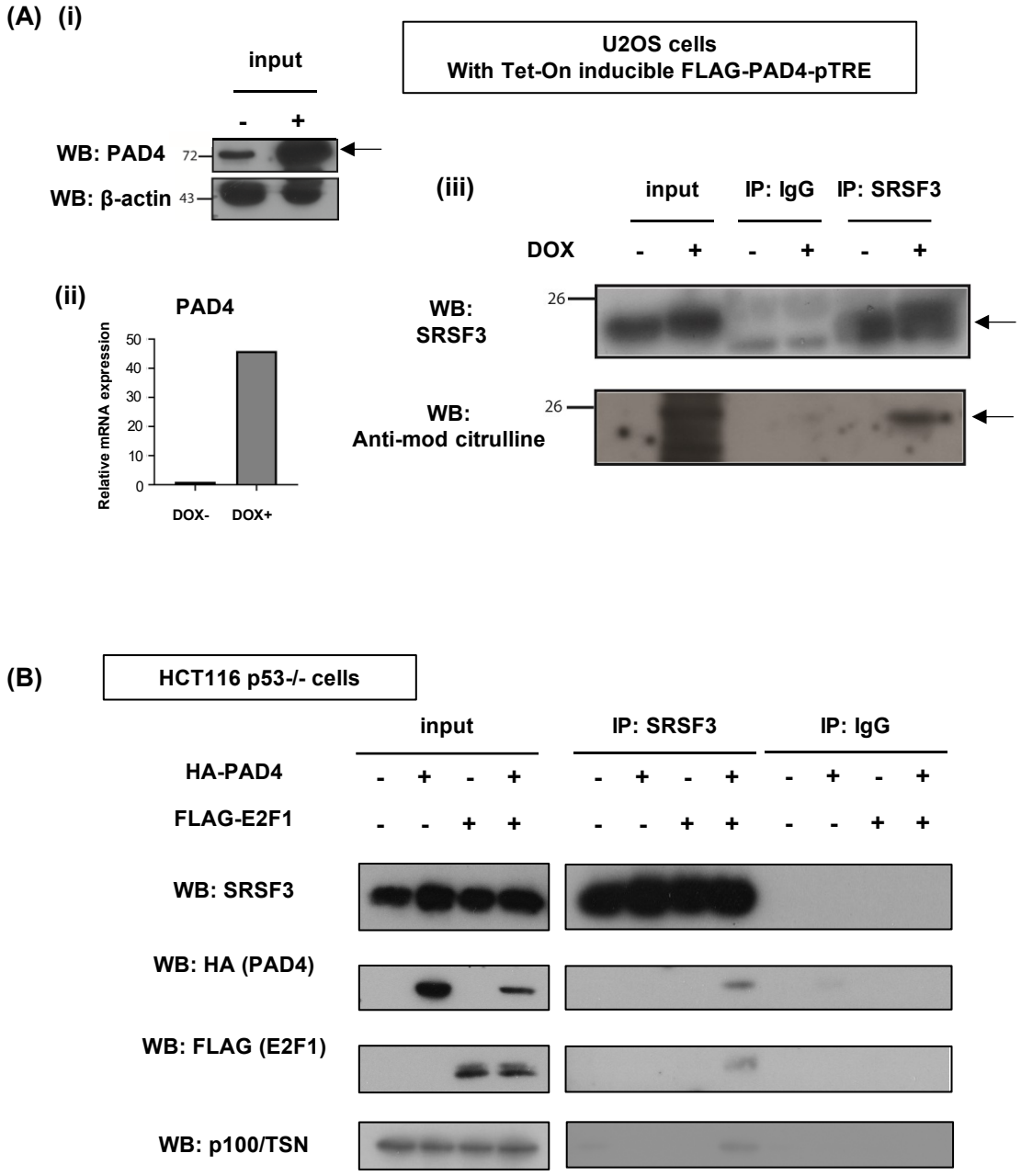


Figure 5-2 SRSF3 is a candidate splicing factor working in the E2F1/PAD4 axis.

(A) U2OS cell line with Tet-On inducible FLAG-PAD4-pTRE was previously generated in our research group (Ghari et al. 2018). PAD4 expression was induced with 1 μg/ml doxycycline treatment for 24 hours as shown in (i) immunoblot and (ii) RT-qPCR using primers specific to PAD4. The cell lysates were immunoprecipitated using an anti-SRSF3 antibody. (iii) Citrullinated peptides were immunoblotted using the anti-citrulline (modified) detection kit. Arrows indicate the band for SRSF3 (19 kDa). *N* = 2. This experiment was performed in collaboration with Amit Shrestha in our research group. (B) HCT116 p53-/- cells were transfected with 2 μg HA-PAD4 and/or 2 μg FLAG-E2F1 and immunoprecipitated using an anti-SRSF3 antibody and subsequently immunoblotted. *N* = 2.

5.4 SRSF3 regulates AS targeted by the E2F1/PAD4 axis.

The interaction between E2F1, PAD4, and SRSF3 sparked our interest in a potential role that SRSF3 may play as a splicing regulator downstream of the E2F1/PAD4 axis to exert the AS perturbations. To confirm this hypothesis, we decided to deplete the intracellular expression of SRSF3 as well as inhibit the PAD4 activity, and examine the impact on the AS events validated by RT-qPCR in Chapter 4 (Figure 4-4). WT and E2F1cr HCT116 p53-/- cells were treated with 2 μg SRSF3-specific siRNA (siSRSF3) or negative control GFP-specific siRNA (siGFP), as well as 10 μM GSK484 or DMSO, for 72 hours (Figure 5-3B). The isolated RNA samples were analysed by RT-qPCR using the primers previously used for the validation experiment against genes including EXOC4, RBM25, SNAP23 and SPIN1.

For the EXOC4 gene, single treatments of both GSK484 and siSRSF3 were demonstrated to promote the inclusion of exon 7 (Figure 5-3A). This finding for GSK484 was consistent with what we discovered in the RNA-seq and RT-qPCR validation (Figure 4-4), suggesting the reproducibility of this influence throughout different experiments. Interestingly, the co-treatment of GSK484 and siSRSF3 was also showed to increase the I/E ratio of this target exon, but not in a synergistic manner; the I/E ratio of cells with the combination treatment was not significantly higher than those treated with either GSK484 or siSRSF3 alone. This non-synergistic combination effect of the PAD4 inhibitor and siSRSF3 supports the view that these biochemical interventions may target the same molecular pathway, meaning that PAD4 and SRSF3 are likely to work together in the same axis in balancing this exon inclusion/exclusion in cells.

Interestingly, a similar trend was found in HCT116 p53-/- E2F1cr cells for the same AS event, EXOC4 SE7; GSK484 and siSRSF3 single treatments respectively increased the

I/E ratio of this exon compared to the untreated cells, whereas the combination treatment did not exhibit a synergistic effect to further enhance the exon inclusion. This further supports the hypothesis that PAD4 and SRSF3 can work together, but also suggests they may do so independently of E2F1. This result slightly contradicts previous findings in this study; E2F1 was shown to bind to the EXOC4 promoter, and this chromatin association was enhanced by the increased level of PAD4 expression (Figure 4-7 and Figure 4-8). Furthermore, cells expressing the citrullination-defective E2F1 mutant ($E2F1^{R4K}$) failed to respond to the GSK484 treatment in terms of the I/E ratio of this EXOC4 SE7 (Figure 4-10), which therefore appeared to be regulated in the mechanism directly involving the citrullination of E2F1. On the other hand, the initial RT-qPCR validation analysis indicated a potential synergistic effect of PAD4 inhibition and E2F1 knockdown, suggesting PAD4 and E2F1 regulate this AS event together but might be through an independent mechanism to one another (Figure 4-4). These seemingly contradicting discoveries might reflect the multifaceted landscape of the E2F1/PAD4 interplay in the AS regulation, and probably indicate that more than one biological pathway is involved; at least one dependent on E2F1 which could be linked with the citrullination of E2F1, and another independent on E2F1 potentially involving the citrullination of SRSF3 or other RNA-binding splicing factors.

On the contrary, the combination treatment of GSK484 and siSRSF3 showed a synergistic impact in increasing the I/E ratio of the SPIN1 SE3 significantly greater than those of GSK484 and siSRSF3 single treatments (Figure 5-3A). PAD4 and SRSF3 are therefore likely to regulate this splicing event by acting in parallel pathways in cells. The results for RBM25 SE2 and SNAP23 RI3 exhibited a large degree of variations between biological repeats which makes it difficult to draw a conclusion. However, a few trends

are indicating a potential mechanistic insight. For example, the I/E ratio of RBM25 SE2 appears to be only influenced by the GSK484 treatment and not by siSRSF3, suggesting that this AS event is regulated by PAD4 but not SRSF3. Future research needs to confirm this by improving deviations or using alternative tools to modulate the expression of target enzymes.

In summary, SRSF3 may work downstream of the E2F1/PAD4 axis to induce the AS disturbances but only be involved in a subset of the AS target events. Additionally, it was indicated that at least two pathways (one E2F1-dependent and one E2F1-independent) exist in which PAD4 and SRSF3 together control RNA splicing, highlighting the intricate relationship E2F1, PAD4, and SRSF3. The result also emphasises the need for identifying other splicing factors involved in this process.

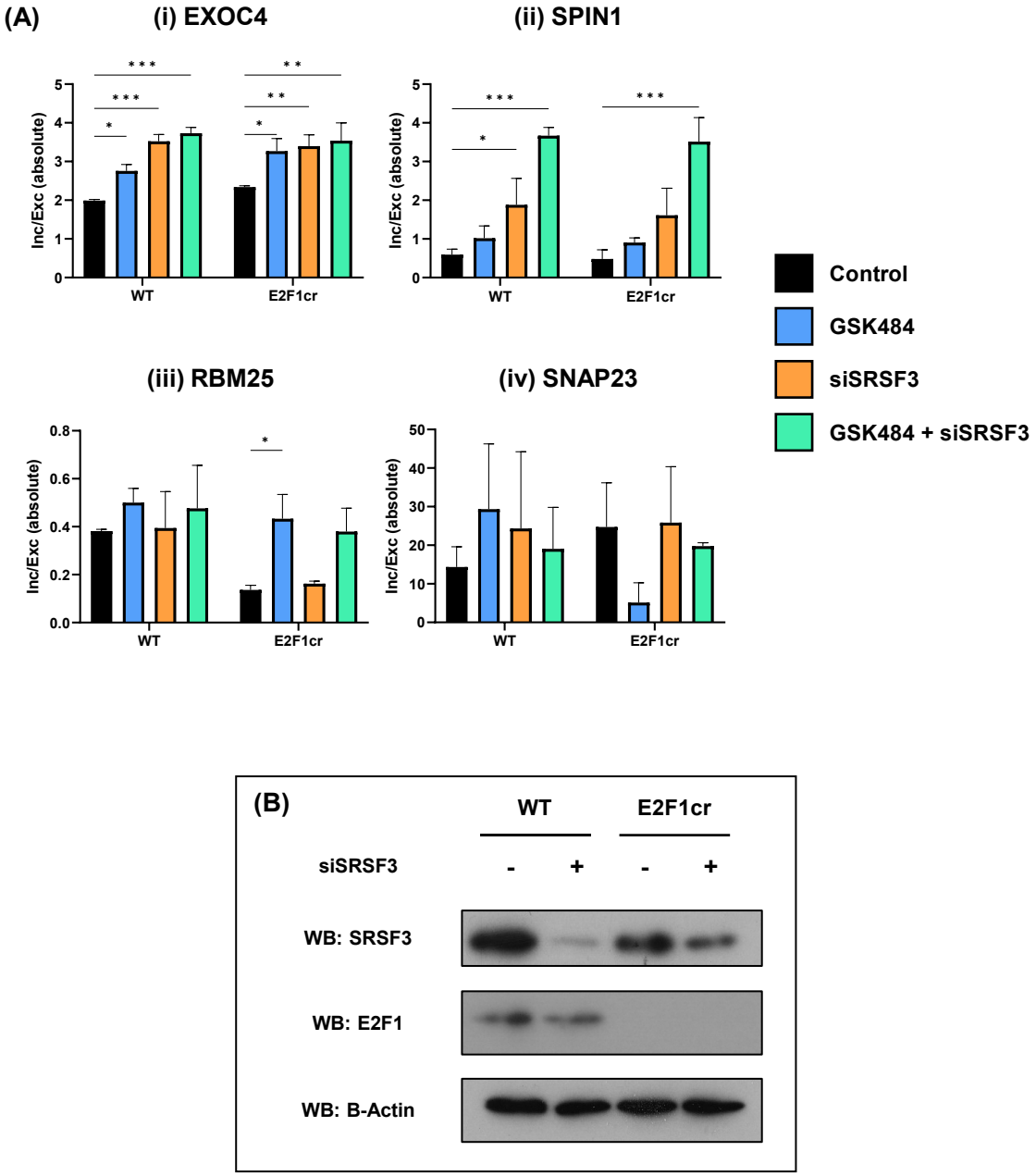


Figure 5-3 SRSF3 regulates AS events targeted by the E2F1/PAD4 axis.

WT and E2F1cr HCT116 p53-/- cells were treated with 2 µg SRSF3 siSRSF3 and/or 10 µM GSK484. (A) Isolated RNA samples were analysed by RT-qPCR using primers previously designed for the RNA-seq validation (Figure 4-4). (i) EXOC4 SE, (ii) SPIN1 SE, (iii) RBM25 SE, and (iv) SNAP23 RI. *N* = 3. Difference between each treatment group and the control group of the respective cell line was calculated by *t*-test. (± S.D. * $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$) (B) The lysates were also immunoblotted to confirm the silencing of SRSF3. *N* = 3.

5.5 Chapter Summary

In this chapter, we investigated the molecular mechanisms underlying the AS regulation by the E2F1/PAD4 axis, especially trying to reveal RNA-binding splicing factors working downstream of this pathway. To begin with, the interaction between E2F1 and p100/TSN was investigated. This multifunctional protein has previously been shown to associate with E2F1 in a PRMT5-dependent fashion and allows it to regulate gene expression at the level of AS (Roworth et al., 2019). In this study, we demonstrated that the increased expression of PAD4 can also augment this interaction of p100/TSN with E2F1 (Figure 5-1). One of the PAD4-target sites of E2F1 (R109 residue) is also targeted by PRMT1, and this PRMT1-dependent methylation is known to repress the PRMT5 activity to modify R111 and R113 residues of E2F1 (Ghari et al., 2016; Zheng et al., 2013). The described effect of PAD4 augmenting the E2F1-p100/TSN interaction might therefore be attributed to the citrullination at R109, which may competitively inhibit methylation mark at the same residue and thus relieve the suppression of PRMT5-dependent methylation at R111/R113, leading to subsequent recruitment of p100/TSN onto these modified arginine residues.

The p100/TSN protein has been described to facilitate the assembly of the spliceosome complex at the splice sites of pre-mRNA (Gutierrez-Beltran et al., 2016). However, the alternative splice sites are often composed of ‘weak’ signals with a lower degree of similarity to the consensus sequence, making it difficult to be recognised and defined by the spliceosome components themselves. This explains why alternative splicing would usually require some help from auxiliary splicing factors to efficiently occur. These RNA-binding proteins assist defining the alternative splice sites and facilitate the assembly of the spliceosome complex by recognising adjacent regulatory elements. Given that the

E2F1/PAD4 axis has a profound role in the regulation of AS, we reasoned there must exist some splicing factors working directly downstream of this pathway. SRSF3 was identified as a potential candidate, as this SR protein has been well-characterised in the AS regulation both physiologically and pathologically (Z. Zhou et al., 2020) and also is known as a substrate of PAD2 (Lewallen et al., 2015). We first demonstrated that SRSF3 can be citrullinated by PAD4, and subsequently showed the interaction of SRSF3 with E2F1, PAD4, and p100/TSN under ectopic expression of PAD4 and E2F1 (Figure 5-2). In addition, by RT-qPCR using previously validated AS events (Figure 4-4), we observed a significant impact of the SRSF3 siRNA (siSRSF3) treatment to cause AS perturbations (Figure 5-3). For the EXOC4 gene, the combination treatment of GSK484 and siSRSF3 was demonstrated to have no synergistic effect on the I/E ratio of exon 7 compared to their single treatments, suggesting PAD4 and SRSF3 working in the same molecular pathway to regulate this AS event. On the contrary, a synergistic effect of the co-treatment was discovered for genes like SPIN1. This indicates that PAD4 and SRSF3 can also work in parallel pathways depending on the target genes.

Overall, the findings presented in this chapter have elucidated that p100/TSN and SRSF3 may act as splicing factors or co-factors downstream of the E2F1/PAD4 axis. However, a complex and interdependent relationship between E2F1, PAD4, and SRSF3 in splicing regulation was also highlighted. Taken together, in addition to studying the protein-protein interactions, it will be of immense importance to investigate this biological interplay at other levels, such as protein-RNA interactions.

Chapter 6

Characterisation of E2F1/PAD4/SRSF3 Interplay in Alternative Splicing Regulation

6.1 Introduction

In Chapter 5, we examined the molecular mechanism of E2F1/PAD4 interplay in alternative splicing (AS) regulation. First, we uncovered that increased PAD4 expression augmented the interaction of E2F1 with p100/TSN (Figure 5-1), which plays a role in facilitating the spliceosome complex at the splice sites of pre-mRNA. Furthermore, the splicing factor SRSF3 was found to be citrullinated by PAD4 and interact with E2F1, PAD4, and p100/TSN (Figure 5-2). We further confirmed the involvement of SRSF3 in E2F1/PAD4-mediated splicing regulation by siRNA experiments, whereby the silencing of SRSF3 was demonstrated to affect a subset of the validated AS events in a similar manner as E2F1 knockdown and PAD4 inhibition (Figure 5-3). For AS events like the skipping of EXOC4 exon 7, combination treatment with siSRSF3 and PAD4 inhibitor exhibited no additional effect as compared to either single treatment, suggesting that these two factors may regulate specific splicing events in the same pathway. However, the results from other splicing events, such as SPIN1, indicated that PAD4 and SRSF3 could also act in parallel pathways and independently of E2F1, highlighting the need for further characterisation of this axis.

To further elucidate the complex nature of AS perturbations regulated by PAD4, E2F1, and SRSF3, we have decided to investigate the molecular interactions of this axis at another level, RNA-protein interaction. We aimed to confirm the involvement of SRSF3 in splicing, as well as to uncover further mechanistic details of E2F1/PAD4 crosstalk which potentially involves multiple pathways. Here, RNA immunoprecipitation was selected as a tool to examine the interaction between E2F1/SRSF3 and mRNA transcripts and was performed under different treatment conditions.

6.2 RNA Immunoprecipitation

To examine the RNA-protein interactions in cells, we designed the protocol of RNA immunoprecipitation (RIP) as illustrated in Figure 6-1. Before harvesting cells, protein-RNA complexes may or may not be cross-linked. The technique in which these complexes are cross-linked is known as cross-linking immunoprecipitation (CLIP), which involves using UV irradiation to generate irreversible covalent bonds between RNA and its binding protein (Baldini & Labialle, 2021). CLIP is therefore a useful technique to stabilise RNA-protein interactions and may allow more stringent purification of the RNA content. However, this might not be suitable for proteins whose association with RNA is mediated through another direct binding partner protein in the same protein complex. Given no RNA-binding domains were identified throughout its structure, E2F1 is believed not to directly interact with RNA (Roworth et al., 2015, 2019). We therefore decided to perform RIP in a native condition for E2F1, and CLIP for SRSF3 which is a known RNA-binding protein.

Following the cross-linking, cells were lysed and sonicated to be completely broken open. In the native RIP protocol, harvested cells were directly lysed and sonicated without any cross-linking. The cell lysates were then immunoprecipitated using an antibody of interest (or negative control IgG) and the magnetic beads which demonstrated better RNA yields and reduced background signals compared to protein A/G beads in some optimisation experiments (data not shown). GlycoBlue was utilised during the isolation of RNA, whereby this co-precipitant improves precipitation and visibility of low RNA yields as it acts as a carrier. RNA samples were extracted using TRIzol and analysed by RT-qPCR.

To confirm that this optimised RIP protocol performs as anticipated, a native RIP was carried out using an anti-Sm protein antibody as a positive technical control. Sm proteins

are best known as components of the spliceosome complex together with U-rich small nuclear RNPs (U snRNPs), and their interactions with several U snRNAs have long been described (Lerner & Steitz, 1979). In this study, the RNA isolated from an Sm RIP was analysed by RT-qPCR using primer sets specific to U1, U4, and U5 snRNAs. As illustrated in Figure 6-2A, RIP with this protocol demonstrated that Sm proteins were associated with U1, U4, and U5. This is consistent with previous reports (Hafner et al., 2019; Lerner & Steitz, 1979; Lu et al., 2014), and therefore indicates the present RIP protocol can pull down proteins of interest with their associated RNA.

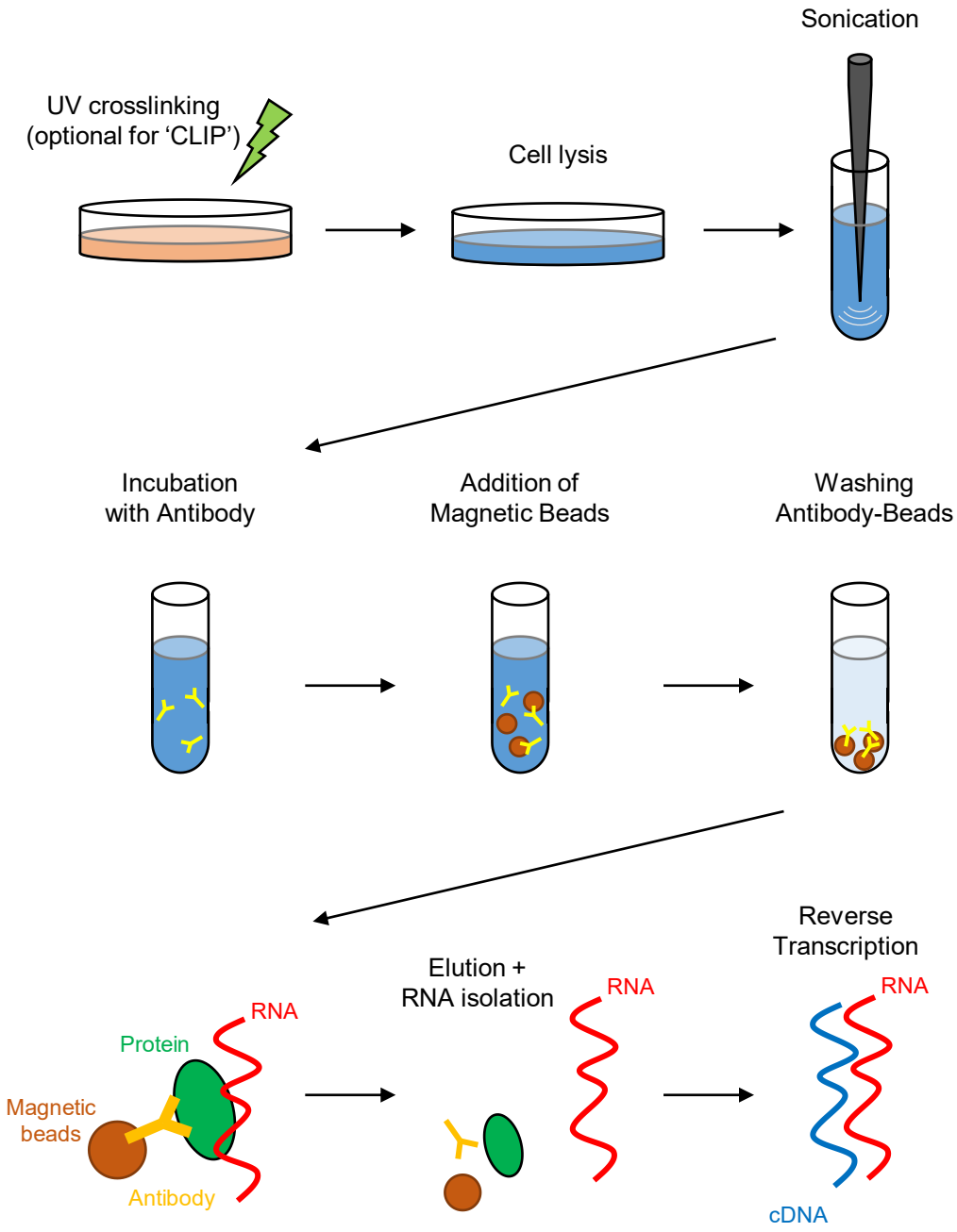


Figure 6-1 RNA immunoprecipitation (RIP) protocol.

Diagram showing the RNA IP experimental setup. In CLIP (Crosslinking RNA IP), cells were subject to UV crosslinking prior to harvesting. The lysates were sonicated to break open cells, followed by incubation with antibody and magnetic beads for immunoprecipitation. RNA was extracted using TRIzol and cDNA was subsequently synthesised for qPCR analysis.

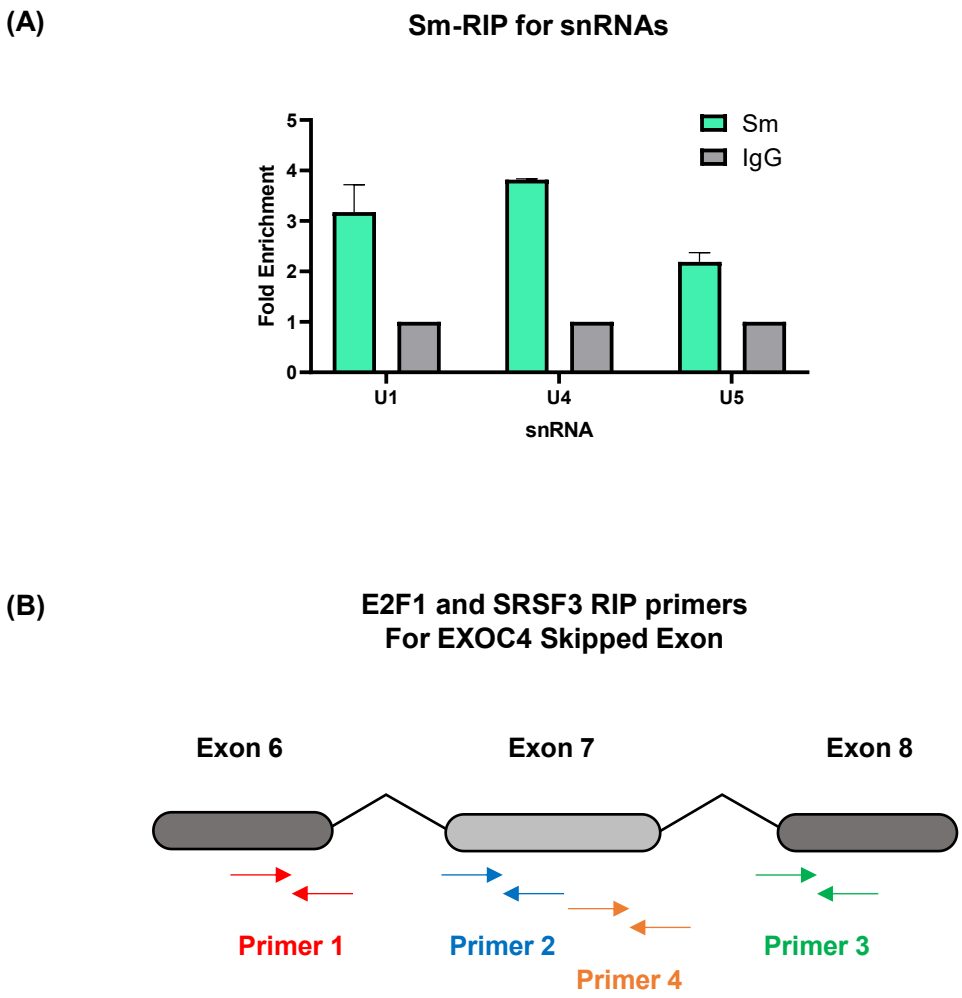


Figure 6-2 Optimisation of RIP protocol and primer design.

(A) RNA samples from Sm RIP in HCT116 p53-/- cells was analysed by RT-qPCR using primers specific to U snRNAs (U1, U4, and U5). *N* = 3 (± S.D.). Enrichment of U1, U4 and U5 was demonstrated as consistent with the previous literature (Urlaub et al., 2001). (B) Schematic representation showing the primer design for E2F1 and SRSF3 RIP to examine its binding to EXOC4 RNA transcripts. Primer sets 1-3 were initially designed for E2F1 RIP to target exon 6, 7, and 8 respectively, and an additional primer set 4 was designed before performing FLAG-SRSF3 RIP. Exon 7 was a target for skipping by E2F1/PAD4-dependent mechanism.

6.3 E2F1 interacts with RNA targeted by the E2F1/PAD4 axis.

The result from Sm protein RIP successfully demonstrated that a protein of interest can be enriched with its associated RNA using the current protocol in HCT116 p53-/- cells. Next, we tried to examine whether E2F1 interacts with the RNA of its splicing target genes by performing an E2F1 RIP. EXOC4 skipped exon 7 (SE7) was selected as an example of an AS event validated to be regulated by E2F1 and PAD4, and three primer sets were designed around exon 7 and the flanking exon as illustrated in Figure 6-2B. Firstly, HCT116 p53-/- WT and E2F1cr cells were lysed and sonicated, and each cell line was divided into three groups to be incubated with E2F1, IgG, and Sm (control) antibodies. We took a small portion of beads from each sample to check the pull-down of E2F1 by immunoblotting (Figure 6-3A). RT-qPCR analysis of Sm RIP was also performed for U1 snRNA as another control (Figure 6-3B(i)) to confirm the technical success of this RIP experiment. The RNA from E2F1 RIP was analysed by RT- qPCR using primer sets specific to exon 6, 7, and 8 of the EXOC4 mRNA (primer 1, 2, and 3 respectively as illustrated in Figure 6-2B) and displayed in the form of fold enrichment (Figure 6-3B(ii)-(iv)). For all three primer sets, EXOC4 RNA was shown to be significantly enriched in E2F1 pulldown over IgG in WT cells, while no enrichment was observed in negative control E2F1cr cells. The result demonstrated that E2F1, or the complex associated with E2F1, was bound to alternatively spliced RNA transcripts genes, and therefore further supports our hypothesis that E2F1 can directly regulate the expression of its target genes through alternative splicing.

Previously, we demonstrated that E2F1 can also associate with these AS target genes including EXOC4 at the protein-DNA level, whereby the increased level of PAD4 expression was shown to enhance this association of E2F1 with the AS target promoter

and the GSK484 treatment impeded such effect (Figure 4-8). In addition, previous work from our research group provided evidence that the interaction of E2F1 with RNA might be dependent on its DNA binding domain (DBD), as the DBD mutant construct (L(leucine)132 and R(arginine)166 residues were replaced by E (glutamic acid) and H (histidine) respectively) was unable to show redundancy for the WT protein binding to U6 snRNA (Roworth et al., 2019). Together this suggests that E2F1 may interact with AS target RNAs while binding to the chromatin, meaning the E2F1/PAD4 regulation of AS may likely undertake co-transcriptionally.

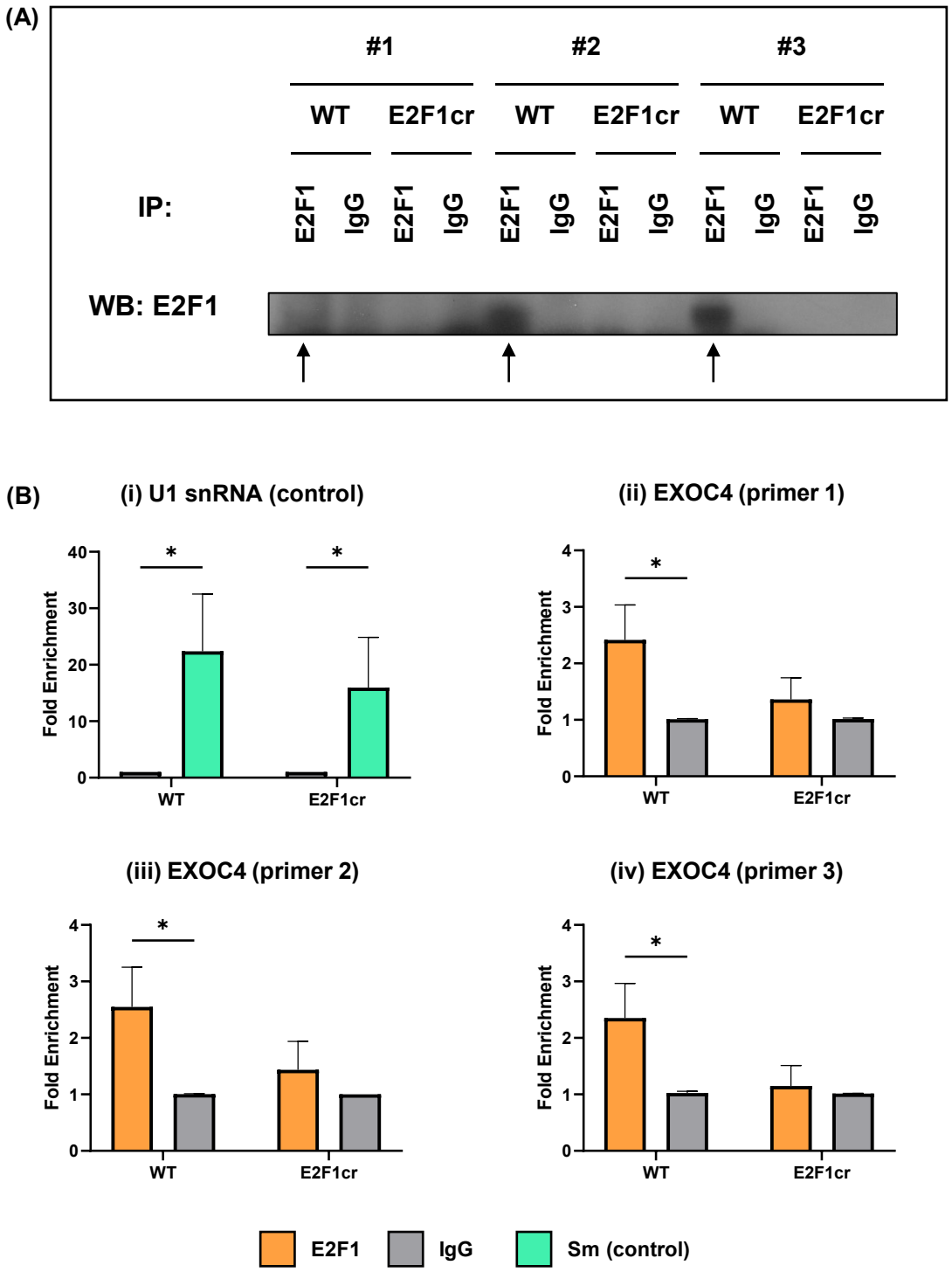


Figure 6-3 E2F1 interacts with EXOC4 RNA transcripts.

RNA immunoprecipitation was carried out in untreated HCT116 p53-/- cells using an anti-E2F1 antibody (and anti-Sm antibody for technical control). (A) An equal volume of beads (10 %) used for the RIP from each condition was analysed by immunoprecipitation to confirm the pull-down of E2F1. (B) RIP sample was analysed by RT-qPCR using (i) U1 snRNA as a technical control for Sm RIP, and (ii)-(iv) EXOC4 primer sets 1-3. *N* = 3. (± S.D. * $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$).

6.4 SRSF3 interacts with RNA targeted by the E2F1/PAD4 axis.

An E2F1 RIP experiment underlined an association between E2F1 and its target RNA for splicing regulation. Previously, Roworth et al also indicated that the E2F1-RNA association is dependent on the binding of p100/TSN protein to PRMT5-driven methylation marks of E2F1, whilst no RNA-binding motifs have been identified in E2F1 (Roworth et al., 2015, 2019). In this study, as well as p100/TSN, SRSF3 was identified to be a binding partner of E2F1 with an elevated expression of PAD4 (Figure 5-2B) and silencing of this SR protein was demonstrated to affect some AS events in a similar manner to the GSK484 treatment and E2F1 knockout (Figure 5-3). We therefore hypothesised that SRSF3 acts as an RNA-binding factor in the same protein complex with E2F1 to mediate AS, and decided to investigate the direct association of SRSF3 with splicing target RNA.

HCT116 p53-/- WT cells were transfected with FLAG-SRSF3 (or empty plasmid vector for a negative control), and CLIP was performed with UV crosslinking using an anti-FLAG antibody (alongside IgG negative control and anti-Sm technical control). First, we confirmed the technical success of the experiment by checking the U1 snRNA enrichment with Sm antibodies (Figure 6-4B(i)). In addition, the immunoblotting affirmed that FLAG-SRSF3 protein was successfully immunoprecipitated whilst no unspecific binding was found with negative control IgG (Figure 6-4A). The RNA content associated with FLAG-SRSF3 was analysed by RT-qPCR using primer sets specific to exon 7 of the EXOC4 transcript (indicated as primers 3 and 4 in Figure 6-2B). For a potential positive control, we used a primer set specific to PKM, to which FLAG-SRSF3 was previously demonstrated to bind in the RIP experiment (Kuranaga et al., 2018). The result showed a successful enrichment of FLAG-SRSF3 with the EXOC4 RNA as well as the positive

control PKM (Figure 6-4(ii)-(iv)), whilst such enrichment was not observed in the cells expressing an empty plasmid vector. This suggests that SRSF3 can bind to the target RNA transcripts of E2F1/PAD4- mediated AS regulation, and further supports the view that this splicing factor acts a direct role in inducing an AS disturbance in this pathway.

The sequential assembly of the spliceosome complex at the splice site underlies the initiation of the pre-mRNA splicing process. Alternative splice sites, which often consist of weak signals with a low level of similarity to the consensus sequence, are normally not sufficient to facilitate this recruitment of the spliceosome by themselves. The definition of such splice sites is therefore largely dependent on the *trans*-acting auxiliary splicing factors, which can recognise the adjacent *cis*-acting splicing regulatory elements to promote or repress the assembly of the spliceosome components. SR proteins including SRSF3 constitute one major family of such auxiliary splicing factors and can reportedly act as activators or repressors of splicing events, by typically binding to the enhancer elements on pre-mRNA. Nonetheless, the molecular mechanisms underlying how SRSF3 contributes to the spliceosome assembly remain to be elucidated. We therefore decided to examine the interaction of SRSF3 with different RNA variants of the target gene EXOC4, in the hope of discovering further information about the binding preferences of SRSF3 as well as in which stages of the sequential spliceosome assembly this splicing factor plays a role.

Here, ‘inclusion’ and ‘exclusion’ primer pairs targeting the EXOC4 exon 7 from our previous RT-qPCR validation (Figure 4-4) were utilised; the inclusion primer pair recognises the junction between exons 6 and 7, whilst the PCR product of the exclusion pair spans the junction between exons 6 and 8 (Figure 6-5B). The FLAG-SRSF3 CLIP samples (Figure 6-4) were analysed by RT-qPCR using these inclusion/exclusion primers

for EXOC4 to examine to which junctions SRSF3 binds with greater affinity. Figure 6-5B illustrates the % input of EXOC4 RNA signals in the cells expressing FLAG-SRSF3 and empty plasmid vector. It was demonstrated that FLAG-SRSF3 was enriched with both variants of EXOC4 transcripts over IgG levels, and significantly greater enrichment was found for the inclusion form. This difference in enrichment might simply reflect the difference in abundance between two RNA variants, but of interest was that SRSF3 seems to bind to both variants of EXOC4 RNA. Since the SRSF3 siRNA treatment was shown to promote the inclusion of this exon (Figure 5-3), this splicing factor is expected to act as an enhancer for the exon 7 skipping of EXOC4. This result probably may indicate that the *cis*-acting regulatory element, most likely enhancer, is located around the 3' end of the exon 6, as this region was contained in both EXOC4 transcript variants.

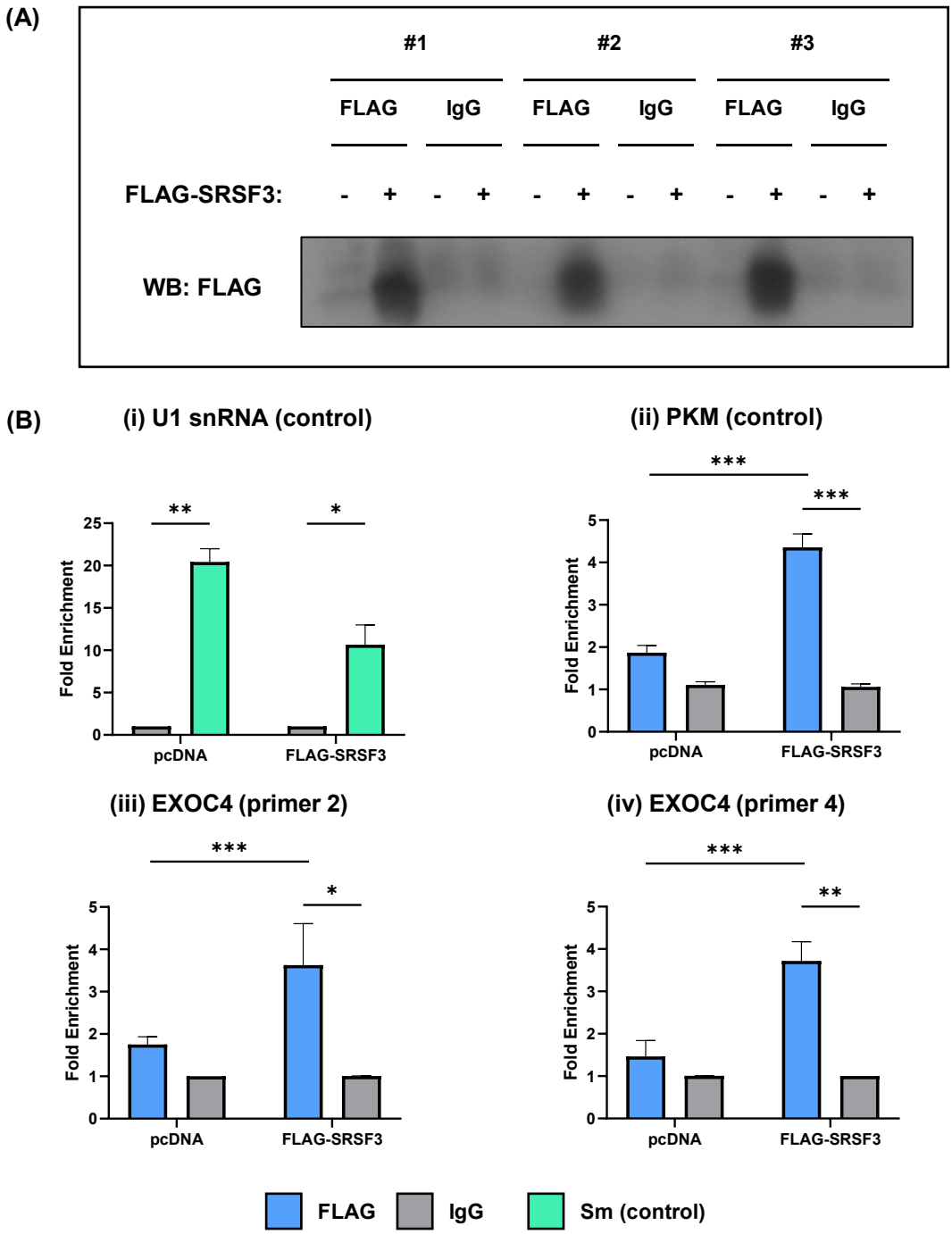


Figure 6-4 FLAG-SRSF3 interacts with EXOC4 RNA transcripts.

RNA immunoprecipitation was carried out in HCT116 p53-/- cells expressing FLAG-SRSF3 with an anti-FLAG antibody. (A) An equal volume of beads (10 %) used for the RIP from each condition was analysed by immunoprecipitation to confirm the pull-down of SRSF3. (B) RIP sample was analysed by RT-qPCR using primers specific to (i) U1 snRNA as a technical control for Sm RIP, (ii) PKM as a positive control for FLAG-SRSF3 RIP, and (iii)-(iv) EXOC4 (primer 2/4 (Figure 6-2)). Cells transfected with an empty vector and IgG pull-down served as negative controls. *N* = 3. (± S.D. * $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$).

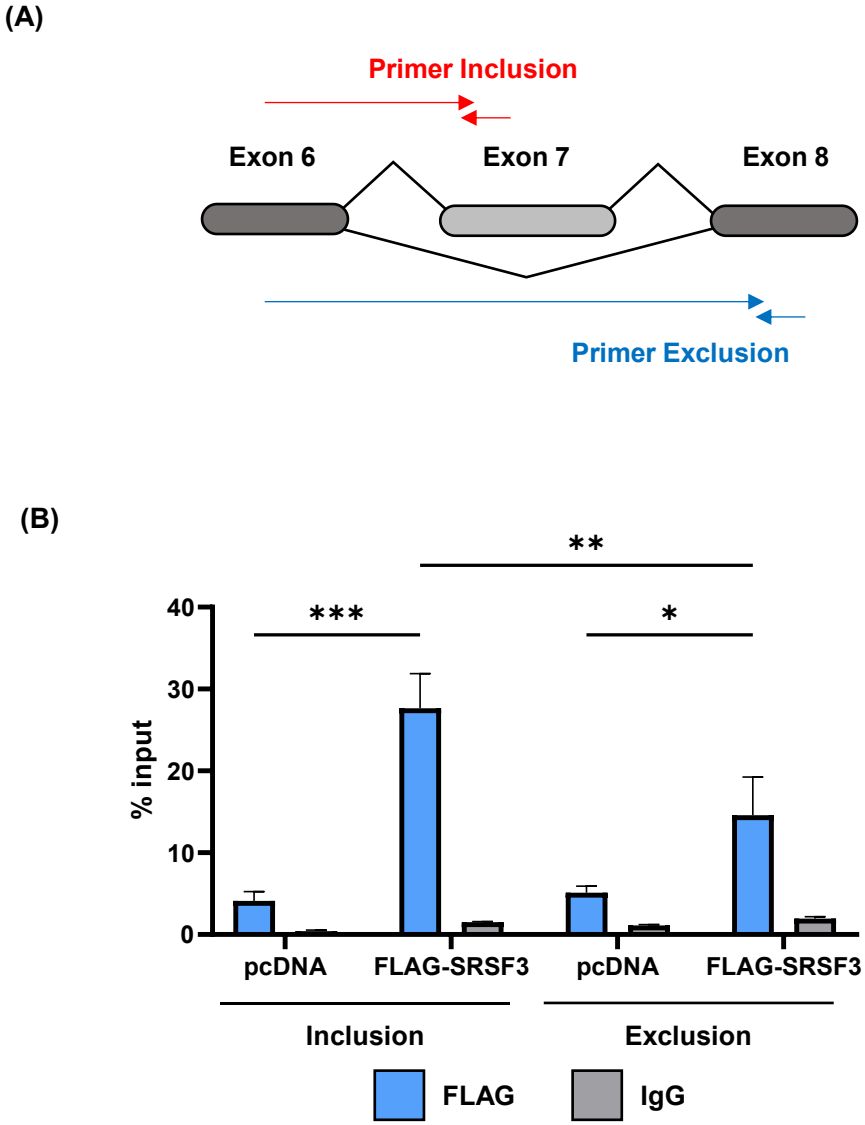


Figure 6-5 SRSF3 interacts with different variants of EXOC4 RNA transcripts.

The RNA samples from the previous RIP (Figure 6-5) were analysed by RT-qPCR using the two different primer pairs specific to 'inclusion' and 'exclusion' variants of EXOC4 RNA (the same primers used for the RNA-seq validation (Figure 4-3)). (i) Diagram illustrates the primer designs to detect the SRSF3 binding to the exon 7 inclusion and exclusion RNA species. (ii) RT-qPCR analysis showed FLAG-SRSF3 can interact with both EXOC4 variants. *N* = 3.

6.5 SRSF3-RNA interaction and the E2F1/PAD4 axis.

Having established that SRSF3 is associated with EXOC4 mRNA, we next questioned whether E2F1 and PAD4 activities have any influences on this association as a part of their alternative splicing regulation mechanism. First, a FLAG-SRSF3 RIP experiment was conducted in HCT116 p53-/- WT and E2F1cr cells to examine how the E2F1 status changes the interaction of SRSF3 with EXOC4 mRNA. Both cell lines were transfected with FLAG-SRSF3 (or empty plasmid vector) and immunoprecipitated using an anti-FLAG antibody. Associated RNA was then analysed by RT-qPCR.

The result demonstrated that SRSF3 can bind to exon 7 of the EXOC4 transcript in both WT and E2F1cr cell lines (Figure 6-6). The difference between WT and E2F1cr in the FLAG-SRSF3 enrichment was quite small, even though the statistical *t*-test indicated there is a slightly greater amount of RNA pulled down in E2F1cr cells compared to WT cells. This probably means that the loss of E2F1 may have a limited impact on the SRSF3-RNA association, which is consistent with our siSRSF3 experiment showing that GSK484 and siSRSF3 exhibited a similar influence on the inclusion of EXOC4 exon 7 regardless of the E2F1 knockdown (Figure 5-3). At the same time, another experiment in this study also demonstrated that this AS event is directly dependent on the citrullination of E2F1 using the citrullination-defective E2F1 mutant (Figure 4-10).

Given the seemingly complicated role that E2F1 plays to influence PAD4/SRSF3-regulated AS events, it would be of foremost importance to carefully examine the effect of PAD4 activity on the SRSF3-RNA association with and without E2F1. Another round of FLAG-SRSF3 RIP in WT and E2Fcr cells was therefore carried out with either enhanced or depleted PAD4 expression. Four experimental conditions were prepared for each cell line; (1) Just expression FLAG-SRSF3, (2) Expression FLAG-SRSF3 and HA-

PAD4, (3) Expressing FLAG-SRSF3 and HA-PAD4 then treated with 10 μM GSK484, and (4) negative control just expressing an empty plasmid vector. Cells were lysed and immunoprecipitated using an anti-FLAG antibody. The result was analysed by RT-qPCR using a primer set specific to EXOC4 exon 7 and presented in the form of % input to precisely compare two cell lines.

The result shows that the EXOC4 mRNA enrichment with FLAG antibody in the three treatment conditions (i.e. expressing FLAG-SRSF3) was significantly greater than that in the control condition (expressing empty plasmid) in both WT and E2F1cr cells (Figure 6-7). This suggests that a certain proportion of SRSF3 interacts with this target RNA regardless of the activities of PAD4 and E2F1. The result also indicated that increased PAD4 activity by HA-PAD4 over-expression did not have a significant impact on the SRSF3-RNA association in both cell lines. The increased expression of PAD4 was found to have a profound impact on the interaction of SRSF3 with other proteins such as E2F1, PAD4, and p100/TSN (Figure 5-2), but its effect on the RNA binding affinity seems to be quite minimal.

Interestingly, despite relatively large error bars, the GSK484 treatment in HA-PAD4 expressing WT cells was demonstrated to significantly enhance the association between SRSF3 and EXOC4 mRNA, but such increase in the association was not found in E2F1cr cells (Figure 6-7). This evidence potentially supports the conclusion that the interaction of SRSF3 with its target mRNA is regulated by PAD4 in an E2F1-dependent fashion, although further experiment will need to improve the error bars or confirm this hypothesis in a parallel experiment. Previously, we discovered that SRSF3 interacts with E2F1 and p100/TSN when PAD4 is over-expressed (Figure 5-2). In addition, PAD4 appears to play a role in regulating the localisation of E2F1 onto the splicing target promoter (Figure 4-8).

Taken together, it can be hypothesised that E2F1 may potentially act as a splicing co-factor by sequestrating SRSF3 away from the alternatively spliced exon 7 of EXOC4 in a citrullination-driven mechanism, and under the PAD4 inhibitor treatment, SRSF3 may be released from the E2F1/PAD4 complex and can closely associate with the spliced target exon. This would slightly conflict with the predicted role of SRSF3 as a splicing enhancer for the described skipping exon event, as the siSRSF3 treatment also increased the inclusion of exon 7 (Figure 5-3). The model of the molecular interaction and potential ways to explain these contradictory findings will be discussed in Chapter 8.

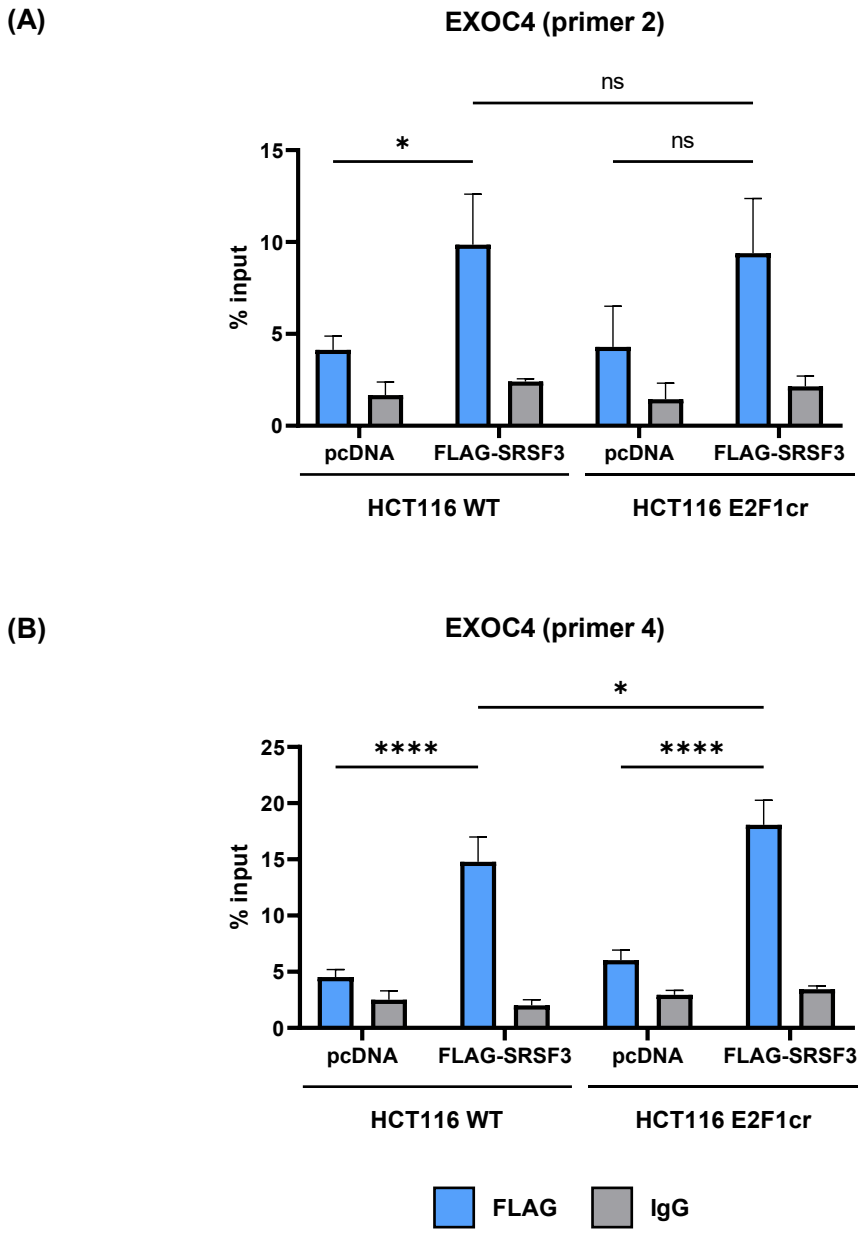


Figure 6-6 SRSF3 interacts with EXOC4 mRNA in both WT and E2F1cr cells.

RNA immunoprecipitation was carried out in HCT116 p53-/- WT and E2F1cr cells expressing FLAG-SRSF3 using an anti-FLAG antibody. The isolated RNA samples were analysed by RT-qPCR using primers specific to EXOC4 exon 7 ((A) primer 2 and (B) primer 4). Cells transfected with an empty vector and IgG pull-down served as negative controls. *N* = 3. (± S.D. * $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$, **** $p < 0.0001$).

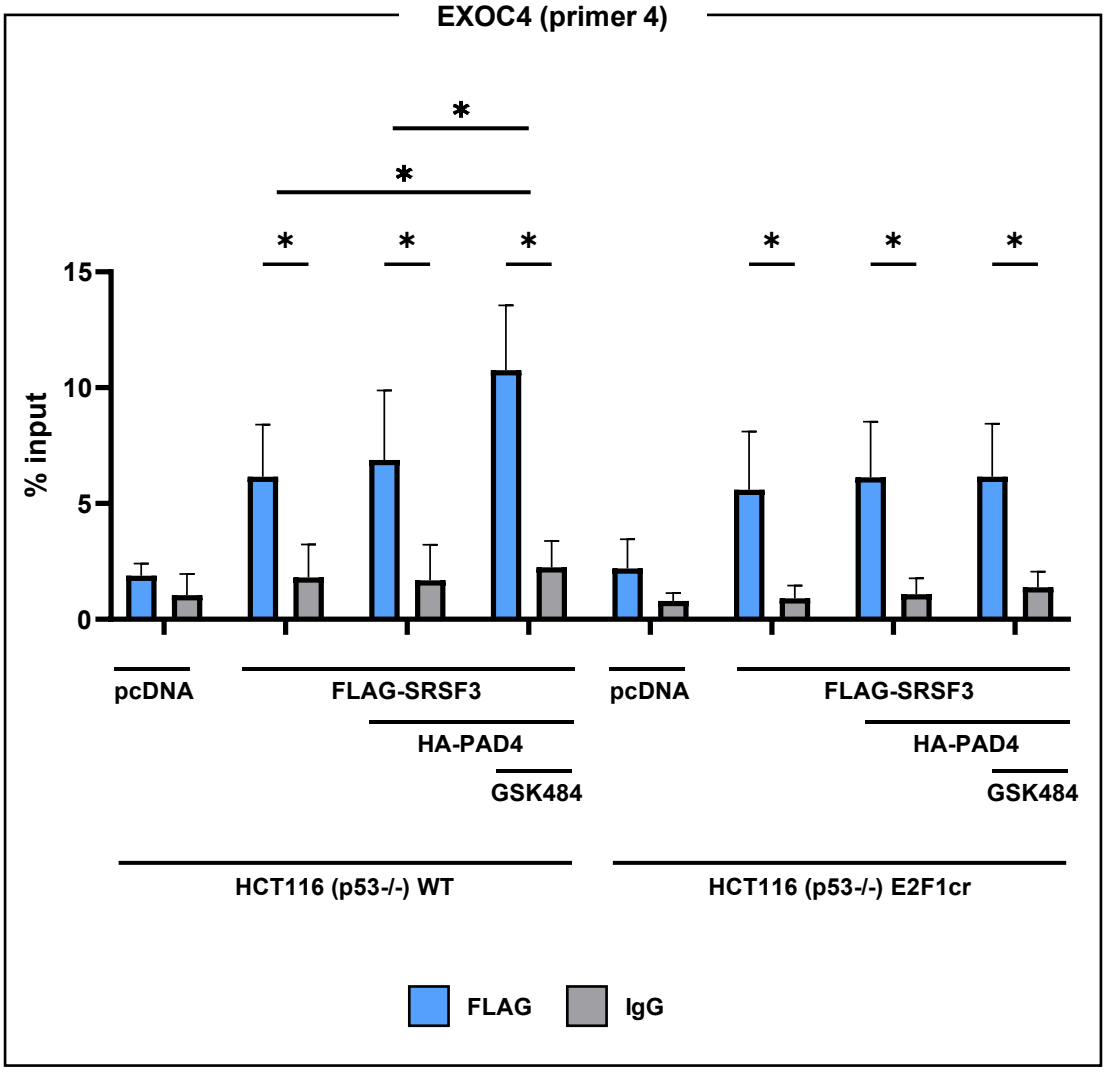


Figure 6-7 SRSF3-mRNA interaction under modulated activities of E2F1 and PAD4.

RNA immunoprecipitation was carried out in HCT116 p53-/- WT and E2F1cr cells expressing FLAG-SRSF3 with PAD4 ectopic expression and GSK484 treatment, using an anti-FLAG antibody. The isolated RNA samples were analysed by RT-qPCR using primers specific to EXOC4 exon 7 (primer 4). (A) WT and (B) E2F1cr cells. Cells transfected with an empty vector and IgG pull-down served as negative controls. *N* = 3. (± S.D. * $p < 0.05$)

6.6 Chapter Summary

Having characterised various protein-protein interactions in and downstream of the E2F1/PAD4 axis in splicing regulation, we decided to investigate the protein-RNA interaction to further examine the underlying molecular mechanisms involved. We first optimised the experimental conditions and protocol of RIP (Figure 6-2), followed by an E2F1 RIP successfully demonstrating E2F1 was associated with EXOC4 mRNA which was an AS target regulated by E2F1 and PAD4 (Figure 6-3). Since it was previously demonstrated that E2F1 bound to the promoter of this gene, this result suggests that the AS regulation by the E2F1/PAD4 axis may occur co-transcriptionally.

Since E2F1 is likely to interact with RNA through another RNA-binding protein, we next examined the interaction between target RNA and SRSF3, which was demonstrated to interact with E2F1 in a PAD4-dependent fashion in the previous chapter (Figure 5-2). Unsurprisingly, a FLAG-SRSF3 RIP experiment showed a clear interaction of SRSF3 with EXOC4 mRNA (Figure 6-4). Subsequently, we performed another SRSF3 RIP under conditions of manipulated PAD4 and E2F1 activities and revealed that the SRSF3-RNA association can be significantly enhanced in the PAD4 inhibitor GSK484-treated cells in an E2F1-dependent manner (Figure 6-7). Together these results provide evidence that E2F1 and PAD4 regulate AS of their target genes by modulating the RNA binding activity of SRSF3. Since PAD4 over-expression or inhibition did not demonstrate a significant impact on the SRSF3-RNA interaction in the absence of E2F1, this suggests that PAD4 does not directly affect the association of SRSF3 with RNA, but likely impacts splicing by modulating protein-protein interactions between E2F1 and SRSF3.

Chapter 7

Investigating the E2F1 Interactome and the Effect of PAD4 Inhibition

7.1 Introduction

A series of experiments in Chapters 5 and 6 have provided mechanistic insight into E2F1/PAD4-driven splicing regulation in SRSF3-dependent pathways, where PAD4 appears to have a pivotal role in promoting the interaction between E2F1 and SRSF3, as well as potentially involved in regulating the RNA binding affinity of this splicing factor. However, our previous siRNA experiments indicated that this involvement of SRSF3 might be limited only to a subset of their target genes (Figure 5-3). It would therefore be important to uncover the precise protein network interacting with the E2F1/PAD4 axis, and identify novel splicing-related factors in addition to SRSF3 that are involved in this AS regulation.

Therefore, we decided to perform proteomics involving immunoprecipitation-mass spectrometry (IP-MS) for E2F1 from cells treated with the PAD4 inhibitor. WT HCT116 p53-/- cells were treated with 10 μM GSK484 or DMSO for 72 hours and E2F1cr cells were treated with DMSO. The lysates were immunoprecipitated using an anti-E2F1 antibody (Figure S-6). Biological triplicates of the samples were used to identify proteins interacting with E2F1 in treatment and control conditions by liquid chromatography-mass spectrometry. Protein enrichment in the treatment condition (WT cells treated with GSK484: WT-GSK484) was analysed with reference to the background condition (E2F1cr cells treated with DMSO: E2F1cr) as well as the control condition (WT cells treated with DMSO: WT-DMSO). This experiment was performed in collaboration with Iolanda Vendrell (Target Discovery Institute, University of Oxford) and Simon Carr in our research group.

7.2 Investigation of E2F1 Interactome by Mass Spectrometry.

The lists of proteins from which more than a single peptide was enriched with an anti-E2F1 antibody in the WT-DMSO and WT-GSK484 conditions over the negative control E2F1cr condition were shown in Table 7-1. Protein enrichment was filtered for the Student's T-test p values < 0.05 (-log p > 1.301). There are 25 and 29 proteins (including E2F1 itself) enriched in the WT-DMSO and WT-GSK484 conditions respectively, and only 13 and 16 proteins from respective conditions made an additional 2-fold change cut-off (Table 7-1). Known binding partners of E2F1 such as pRB (RB1) and DPs (TFDP1/2) were enriched in both conditions as expected, which indicates a successful IP of E2F1 with its associated proteins. In Chapter 5, we observed the binding of SRSF3 and p100/TSN to E2F1 when E2F1 and PAD4 were overexpressed (Figure 5-2B), but these proteins were not included in either of the lists shown in Table 7-1. This further supports the view that the interaction of SRSF3 and p100/TSN with E2F1 may require high expression of PAD4 and therefore presumably its elevated enzymatic activity.

In addition, proteins enriched in the E2F1 interactome in the WT-GSK484 condition over the WT-DMSO condition were analysed (Table 7-2). The analysis identified 23 proteins (excluding E2F1 itself and known binding partners highlighted in green in Table 7-2) that were more enriched, and 25 proteins less enriched, in the E2F1 interactome of WT-GSK484 compared to the WT-DMSO condition. This suggests that PAD4 activity can modulate the interaction of E2F1 with other proteins both in a positive and negative fashion. This list of proteins also did not include p100/TSN or SRSF3. It was noticeable that not all proteins enriched in WT-GSK484 over WT-DMSO comparison made the significance cut-off when each treatment was compared individually against the negative control E2F1cr condition (Table 7-1). This is likely due to the fact that GSK484 treatment

itself reduced the amount of E2F1 immunoprecipitated in this experiment, as indicated by the observed negative fold change value found for E2F1 (-0.931, Table 7-2), probably for some technical reasons.

Table 7-1 E2F1 interactome: Proteins enriched with E2F1.

(A) Enrichment in WT-DMSO over E2F1cr

Protein	No. of Unique Peptides	*P* value (-log)	Log2 Fold Change
E2F1	15	4.109	4.035
RB1	36	5.772	4.764
TFDP1	8	5.689	4.924
TFDP2	3	3.761	8.369
CDK1	3	0.594	0.582
DLAT	3	1.594	3.036
UBE2N	2	1.680	2.039
SSBP1	8	2.017	1.477
GRPEL1	4	1.307	1.207
EIF5A	2	1.666	1.206
CLIC1	3	2.008	1.147
PYCR2	2	1.369	1.101
PLEC	47	1.903	1.035
LCN2	3	1.601	0.973
CAV1	3	1.494	0.970
SLC1A5	4	1.710	0.879
PGAM1	5	1.339	0.846
NOP56	6	1.449	0.808
SRP14	2	1.470	0.796
RSL1D1	5	1.674	0.721
RPS24	3	1.507	0.569
CAD	10	1.723	0.530
HSPA5	49	1.801	0.520
RPS23	3	1.507	0.449
ILF2	5	2.041	0.410

(B) Enrichment in WT-484 over E2F1cr

Protein	No. of Unique Peptides	*P* value (-log)	Log2 Fold Change
E2F1	15	3.505	3.104
RB1	36	4.082	3.986
TFDP1	8	3.920	3.915
TFDP2	3	3.700	7.753
CDK1	3	1.041	-1.008
DLAT	3	2.722	4.135
PYCR1	4	2.617	1.724
CAV1	3	1.973	1.675
PYCR2	2	1.767	1.588
UQCRFS1P1	2	2.028	1.584
SSBP1	8	2.332	1.544
ETFA	2	2.061	1.269
RRBP1	3	2.562	1.261
PLEC	47	1.945	1.195
SCD	2	1.585	1.092
SLC1A5	4	1.903	1.040
S100A9	10	2.114	0.821
SEC61A1	8	1.337	0.810
HSPB1	11	1.310	0.760
PGAM5	6	1.772	0.741
CAD	10	1.572	0.668
USP5	2	1.424	0.655
AHNAK	18	1.684	0.582
DHCR7	4	1.660	0.571
GARS1	14	1.445	0.527
VDAC2	5	3.010	0.526
ACSL3	2	1.363	0.521
HP1BP3	9	1.391	0.430
ENO1	15	1.771	0.339

Known E2F1 binding partner	RNA-binding	$P < 0.05$ (-log $P > 1.301$)	Log2 > 1

HCT116 p53-/- cells were treated with 10 μM GSK484 or DMSO for 72 hours. WT-DMSO, WT-GSK484, and untreated E2F1cr cells were immunoprecipitated using anti-E2F1 antibody and analysed by Liquid-Chromatography Mass-Spectrometry. Proteins from which more than one peptide was enriched with E2F1 in (A) the WT-DMSO and (B) WT-GSK484 over the E2F1cr condition with Student's T-test p values < 0.05 (-log p > 1.301, highlighted in yellow) were shown. Known E2F1 binding partners (RB, TFDP1/2, and CDK1) are highlighted in green. Log2 fold change of more than 1 is highlighted in red, and proteins over-representing RNA-binding (GO:0003723) were shown in red in (B).

Table 7-2 E2F1 interactome: Proteins enriched in WT-GSK484 over WT-DMSO.

Enrichment in WT-GSK484 over WT-DMSO

Protein	No. of Unique Peptides	*P* value (-log)	Log2 Fold Change
E2F1	15	2.447	-0.931
RB1	36	1.511	-0.777
TFDP1	8	1.616	-1.008
TFDP2	3	0.893	-0.616
CDK1	3	2.424	-1.590
KRT87P	7	1.754	3.976
IMPDH2	6	1.499	2.281
NDUFA4	2	1.389	1.675
SCD	2	2.864	1.492
DCD	3	1.714	1.408
ARF3	3	1.505	1.254
PYCR1	4	1.467	1.210
CA2	7	1.855	0.943
SFN	8	1.605	0.899
POF1B	9	1.429	0.844
ATP5F1A	17	1.990	0.824
ATP5F1B	23	1.337	0.821
DDX39A	5	1.547	0.790
S100A7A	3	3.006	0.755
SEC61A1	8	1.485	0.741
VAPA	2	1.633	0.597
PGAM5	6	1.632	0.582
PPIB	8	2.084	0.499
VDAC2	5	2.939	0.499
AHNAK	18	1.310	0.475
ALB	31	1.975	0.437
ENO1	15	1.360	0.386
TUFM	15	1.413	0.358
RSL1D1	5	2.055	-0.393
LRPPRC	12	1.449	-0.404
OAT	4	1.720	-0.424
GLUL	3	1.320	-0.497
TUBB	4	2.491	-0.564
TUBA1C	6	1.959	-0.623
TUBA4A	5	1.539	-0.673
RPS27	2	2.206	-0.690
HMGB1	6	1.675	-0.725
PCNA	6	1.403	-0.726
RPS24	3	1.772	-0.740
IARS1	9	2.090	-0.771
RPS15A	3	1.466	-0.840
TFRC	7	1.810	-0.850
CSE1L	5	1.571	-0.895
S100A7	7	2.370	-0.923
NUP205	3	1.302	-1.005
PNP	4	1.536	-1.107
OGT	15	3.069	-1.225
KIF5B	23	2.602	-1.349
TRAK1	20	2.491	-1.400
CRABP2	2	1.858	-1.796
CTPS1	2	2.150	-1.917
ACTR3	2	1.578	-2.573
RCBTB1	4	2.506	-2.698

Known E2F1 binding partner	RNA-binding *Enriched over E2F1cr too*	*P* < 0.05 (-log *P* > 1.301)	Log2 > 1 Log2 < -1

Proteins from which more than one peptide's enrichment was significantly changed in the WT-DMSO compared to the WT-GSK484 with Student's T-test p values < 0.05 (-log p > 1.301, highlighted in yellow) were shown. Known E2F1 binding partners (RB, TFDP1/2, and CDK1) are highlighted in green. Log2 fold change of more than 1 and less than -1 are highlighted in red and blue respectively. Proteins over-representing RNA-binding (GO:0003723) were shown in red. Proteins which were also enriched over E2F1cr were shown in italic.

7.3 GO Analysis for the E2F1 Interactome with PAD4 Inhibition

Next, using the ShinyGO bioinformatic tool, the proteins enriched in WT-GSK484 over E2F1cr or WT-DMSO were analysed for Gene Ontology (GO) molecular function terms and KEGG pathway terms (Figure 7-1). Interestingly, the molecular functional term "RNA-binding" (GO:0003723) was over-represented in the WT-GSK484 E2F1 interactome when compared to WT-DMSO (Figure 7-1B (i)), and 16 potential RNA-binding proteins (RBPs) were identified (shown in red in Table 7-2). Additionally, some of these RBPs were also enriched in WT-GSK484 when compared to E2F1cr (shown in red in Table 7-1B and italic in Table 7-2), though the GO term itself was not significant in this comparison. This is probably due to the small number of protein inputs in this analysis. Out of these 16 potential RBPs, the interaction of E2F1 with 8 of them was demonstrated to be enhanced with GSK484 treatment, whereas PAD4 inhibition appears to reduce the association of the other 8 RBPs with E2F1. Although the relevance of these proteins to E2F1/PAD4-driven AS regulation is currently unknown, the result suggests that PAD4 may have a profound effect on regulating the E2F1-RBP interactome.

To further investigate these possible RNA-binding proteins identified from the GO analysis, we mined the RefSeq database at the National Centre for Biotechnology Information (NCBI) website (https://www.ncbi.nlm.nih.gov/refseq/) (O'Leary et al., 2016) for functional annotations (Table 7-3). Despite our expectations, the list of potential RBPs did not appear to contain many splicing-related factors. DDX39A is one of the few proteins with known functions in RNA processing. The physiological role of this RNA helicase has been connected with RNA export, but it is also implicated in pre-mRNA splicing (Jarmoskaite & Russell, 2011; Linder & Jankowsky, 2011; M.-J. Luo et al., 2001; Nakata et al., 2017; Yamazaki et al., 2010). For example, DDX39A was demonstrated to

regulate the AS of the Androgen Receptor (AR) protein to induce the splice variant (AR-V7) which is often linked to poor cancer prognosis and resistance to androgen deprivation therapy (Nakata et al., 2017). Another potential splicing factor in the GSK484-treated E2F1 interactome is AHNAK, and this large (700 kDa) scaffolding protein was demonstrated to regulate the splicing of its own transcript (de Morrée et al., 2012).

It is also noteworthy that a few components of the 40S ribosome subunit (RPS27, RPS24, and RPS15A) were among the proteins whose interactions were significantly impacted by GSK484 treatment. This is also reflected in the KEGG analysis, which identified the term ribosome (M189) to be over-represented in the WT-GSK484 over WT-DMSO comparison (Figure 7-1B (ii)). The relevance of E2F1 in ribosomal biology has been described by several studies. For example, E2F1 is known to regulate the expression of ribosomal component rRNAs (Ayrault et al., 2006), whereas some recent reports suggested that ribosomal proteins were in turn involved in the regulation of E2F1 expression and activity (Ma et al., 2022; Pecoraro et al., 2019). These studies, however, have never reported a direct interaction between E2F1 and ribosomal proteins. The IP-MS analysis presented here thus indicates a novel interaction of E2F1 with ribosomal components (40S subunit), and such an interaction seems to be negatively regulated by GSK484 treatment (Table 7-2). Further supporting this, ribosomal proteins including RPS24 and RPS23 were enriched in the WT-DMSO E2F1 interactome when compared to the E2F1cr condition. This link between E2F1 and the ribosome is an interesting topic for future research, which might suggest a potential role for PAD4-dependent citrullination in extending E2F1 activity from a classical transcription factor to a translation co-regulator.

Finally, in the hope of confirming novel splicing factors that act downstream of the E2F1/PAD4 axis, we attempted to validate the interaction of E2F1 with the RBPs identified in the IP-MS in a parallel biochemical experiment. DDX39A and AHNAK were selected as potential candidates with their possible roles in RNA processing as indicated by the functional annotation analysis. However, due to the limited availability of commercial antibodies, only DDX39A was finally selected for downstream IP experiments. HCT116 p53-/- cells were transfected with wildtype or citrullination-defective mutant (Figure 4-10) E2F1 (HA-E2F1^{WT} or HA-E2F1^{R4K}) and treated with 10 μM GSK484 (or DMSO) for 72 hours before being immunoprecipitated with an anti-HA antibody. Unfortunately, it was not possible to detect an interaction between E2F1 and DDX39A under these conditions (Figure S-7), despite multiple repeat experiments. Nevertheless, mass spectrometry is considered a more sensitive technique than IP for detecting protein-protein interactions. Therefore, our inability to confirm the DDX39A interaction with E2F1 may simply reflect a technical limitation of the antibodies used in the IP. In addition, since that the protein-protein interactions between E2F1 and SRSF3 or p100/TSN were significantly enhanced by the ectopic expression of PAD4, it would be interesting to repeat the IP-MS experiment in cells over-expressing PAD4 in future studies.

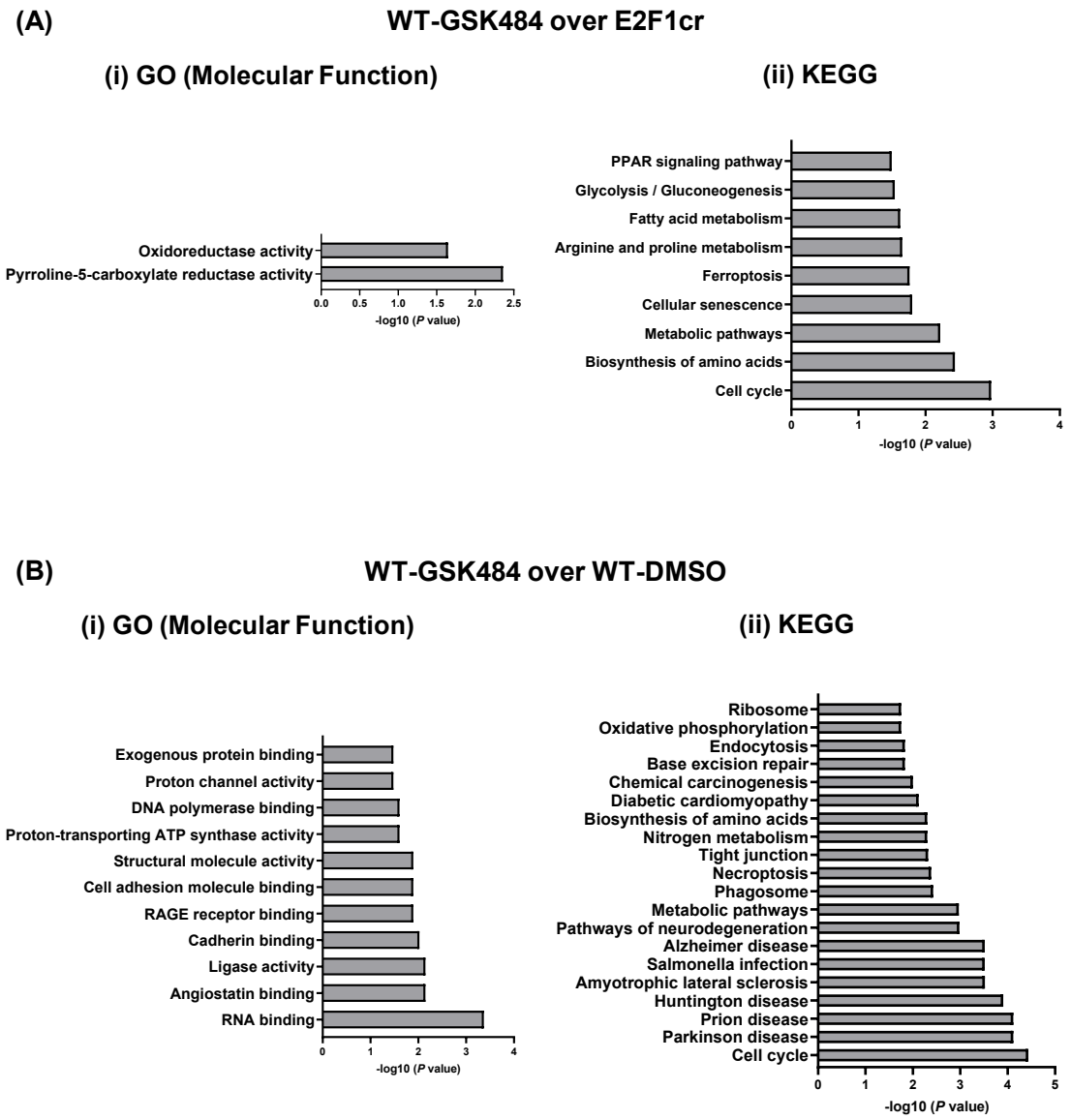


Figure 7-1 GO Analysis of E2F1 interactome with GSK484 treatment.

Gene Ontology (GO) enrichment for Molecular Functions and KEGG pathway analyses were performed for the proteins enriched with E2F1 in the WT-GSK484 condition over (A) the E2F1cr condition and (B) WT-DMSO condition, using the ShinyGO online tool (http://bioinformatics.sdstate.edu/go/) (S. X. Ge et al., 2020). Proteins with more than one peptide identified with significant Student's T-test p values (<0.05) were used in the analysis. (i) GO molecular function and (ii) KEGG. Over-represented pathways (-log10(P) > 1.301) were shown in -log10(P value).

Table 7-3 RNA-binding proteins in GSK484-treated E2F1 interactome.

RBP candidate	Description	Functions
IMPDH2	Inosine Monophosphate Dehydrogenase 2	Involved in the de novo guanine nucleotide biosynthesis.
DCD	Dermcidin	Promotes neural cell survival under conditions of severe oxidative stress.
ATP5F1A	ATP Synthase F1 Subunit Alpha	Subunit of mitochondrial ATP synthase.
DDX39A	DExD-Box Helicase 39A	RNA helicases, implicated in alteration of RNA secondary structure, translation initiation, nuclear and mitochondrial splicing, and ribosome and spliceosome assembly.
PPIB	Peptidylprolyl Isomerase B	Implicated in the secretory pathway. Involved in the regulation of cyclosporine A-mediated immunosuppression.
AHNAK	AHNAK Nucleoprotein (/ Desmoyokim)	Large (700 kDa) structural scaffold protein, implicated in blood - brain barrier formation, cell structure and migration, cardiac calcium channel regulation, and tumour metastasis.
ENO1	Enolase 1	Involved in glycolytic pathways. AS of this gene results in the shorter isoform, which is known as a tumour suppressor by regulating c-MYC expression.
TUFM	Tu Translation Elongation Factor, Mitochondrial	Involved in the protein translation in mitochondria.
RSL1D1	Ribosomal L1 Domain Containing 1	Regulates 5'-UTR and 3'-UTR binding activities of mRNA. Involved in regulation of apoptotic process and regulation of cellular senescence.
LRPPRC	Leucine Rich Pentatricopeptide Repeat Containing	Potentially involved in cytoskeletal organization, vesicular transport, or in transcriptional regulation of both nuclear and mitochondrial genes.
RPS27	40S Ribosomal Protein S27	Component of the 40S subunit, containing a zinc finger domain which may allow it to bind to nucleic acid. Mutation of this gene is found in numerous melanoma patients.
HMGB1	High Mobility Group Box 1	Regulates transcription through DNA organisation. Might be involved in inflammation, cell differentiation and tumour cell migration.
RPS24	40S Ribosomal Protein S24	Component of the 40S subunit.
IARS1	Isoleucyl-TRNA Synthetase 1	Mediates the aminoacylation reaction of tRNA.
RPS15A	40S Ribosomal Protein S15a	Component of the 40S subunit.
TFRC	Transferrin Receptor	Cell surface receptor regulating the iron influx through the process of receptor-mediated endocytosis.

List of proteins over-representing the 'RNA-binding' (GO:0003723) term by the GO Molecular Function Analysis for whose enrichment with an anti-E2F1 antibody in the WT-GSK484 condition was significantly altered compared to the WT-DMSO condition (as shown in Table 7-2). Functional annotations were provided by the RefSeq database from the National Centre for Biotechnology Information (NCBI) website (https://www.ncbi.nlm.nih.gov/refseq/) (O'Leary et al., 2016).

7.4 Chapter Summary

Even though we were lucky enough to identify SRSF3 as a splicing factor working downstream of the E2F1/PAD4 axis, this AS regulation and the involved pathways seem to be complex and likely involve multiple other splicing factors in addition to SRSF3. We therefore carried out the immunoprecipitation-mass spectrometry using an anti-E2F1 antibody to examine the E2F1 interactome under the influence of the GSK484 treatment (Table 7-1 and Table 7-2). Interestingly, the subsequent GO analysis demonstrated that the molecular function term 'RNA-binding' was enriched in GSK484-treated WT cells over untreated cells (Figure 7-1). The splicing-related RNA helicase DDX39A was among the proteins over-representing this GO term (Table 7-3), although its interaction with E2F1 was not confirmed by immunoprecipitation (Figure S-7). Additionally, this list of RBPs contained several ribosome proteins, indicating a potential link between E2F1 and translation. Overall, the IP-mass spec results indicated a potential general role of PAD4 in regulating the interaction of E2F1 with RNA-binding proteins, although each interaction may need to be confirmed in a parallel experiment *in vitro* and *in vivo*.

In future research, investigation of the E2F1 interactome with an elevated level of PAD4 expression would be of great interest, as PAD4 activity may have a greater impact on the protein-protein interactions in such conditions, as previously highlighted for the interactions of p100/TSN and SRSF3 with E2F1 (Figure 5-1 and Figure 5-2). Further characterisation of RBPs identified in our IP-MS experiment, especially DDX39A and AHNAK, would also contribute to expanding our understanding of the E2F1/PAD4 axis in the context of RNA processing regulation.

Chapter 8

Discussion

8.1 Novel Function of the E2F1/PAD4 Axis in Splicing Regulation.

8.1.1 Current understanding of the E2F1/PAD4 axis and AS.

Pre-mRNA splicing is a key component of eukaryotic gene expression. It has been suggested that up to 95 % of the human genome requires this process to remove introns from transcripts and produce functional mRNAs as templates for protein synthesis (Bonnal et al., 2020; Gallego-Paez et al., 2017). Alternative splicing (AS), by generating multiple variants of mRNA transcripts from a single gene, can substantially expand protein diversity. Mounting evidence indicates the deregulation of this mechanism in the progression of various diseases, and cancer is among such pathological conditions affected by these splicing disturbances (Baralle & Giudice, 2017; Bonnal et al., 2020; Gallego-Paez et al., 2017).

In the last decade, the incorporation of state-of-the-art mass spectrometry and analytical methodology has revolutionised our understanding of the cellular 'citrullinome', and the evidence connecting PADs with RNA processing is becoming more compelling (Lewallen et al., 2015; Tanikawa et al., 2018). For example, one study, which identified around 50 intracellular proteins citrullinated by PAD2, discovered more than 20 of them were RNA processing or splicing factors (Lewallen et al., 2015). Another proteomic analysis found more than 150 substrates of PAD4 in cells, and interestingly about 20 % of them were shown to possess the RG/RGG motif (Tanikawa et al., 2018). This motif is a known molecular signature of RNA-binding proteins and implicated in the regulation of a variety of intracellular molecular interactions by serving as binding sites for RNA or other proteins (Rajyaguru & Parker, 2012; Thandapani et al., 2013). Since citrullination results in the loss of positive charge from an arginine residue (Figure 1-5), one can predict

that PADs here may act to modulate the interaction of these RGG-containing RNA-binding proteins (RBPs) with RNA which is negatively charged or with other proteins.

Indeed, it was demonstrated that PAD4-driven citrullination of a splicing factor called SFPQ (Splicing factor proline- and glutamine-rich, also known as PSF) results in the downregulation of its association with target mRNA (Snijders et al., 2015). Interestingly, the target arginine residues in the N-terminal of SFPQ for PAD4 can also be methylated by PRMT1, and thereby this modification in turn enhances the interaction of SFPQ with mRNA. Experiments using recombinant SFPQ and SFPQ-derived peptides illustrated that prior citrullination by PAD4 could block subsequent methylation by PRMT1, and *vice versa*, suggesting that PAD4 may play a key role in regulating SFPQ-mediated splicing in orchestration with PRMT1. Curiously, SFPQ is not the sole example of RBPs subjected to such crosstalk between citrullination and methylation. For instance, the human ribosomal protein S2 (RPS2), whose expression is reportedly dysregulated in multiple types of tumours (Q. Guo et al., 2011), contains an N-terminal RGG motif targeted by PAD4 for citrullination (Q. Guo et al., 2011) and PRMT3 for methylation (Bachand & Silver, 2004; Swiercz et al., 2005, 2007). Furthermore, methylation of the arginine residue in the RGG motifs can be prevalently found, whereby PRMTs are thought to play a key regulatory role in modulating protein-RNA interactions (C. Chen et al., 2011; Fong et al., 2019; C.-Y. Lee et al., 2018; Lewallen et al., 2015; W.-J. Li et al., 2021; Lim, Lee, et al., 2020; Lim, Park, et al., 2020; Musiani et al., 2019; Radzisheuskaya et al., 2019; Tanikawa et al., 2018). Together, these studies suggest that PADs may undertake a key role in post-translationally regulating RNA splicing, and the crosstalk between citrullination and methylation may serve as a general regulatory mechanism for RNA-binding factors. Nevertheless, no direct evidence has yet been proposed to

demonstrate the involvement of PADs in RNA processing, and underlying mechanisms remain to be elucidated.

E2F1 is another example of a substrate for the citrullination-methylation interplay and its potential involvement in RNA splicing (Harada et al., 2023). This transcription factor is a known master regulator of the cell cycle progression (Figure 1-3), as well as a potent mediator of apoptosis in response to DNA damage (Roworth et al., 2015). The activity of E2F1 in the cell cycle is primarily regulated by the pocket proteins including the retinoblastoma protein (pRB), whose interaction with E2F1 at the designated binding site within the transactivation domain suppresses its transcriptional activity (Figure 1-2). The E2F1-pRB pathway is integral to the regulation of the cell cycle progression, and unsurprisingly often deregulated in numerous proliferative diseases including cancer (Di Fiore et al., 2013; Roworth et al., 2015). In addition, the E2F1 transcription factor is subject to various post-translational modifications, that govern the different biological outcomes of E2F1 activity in an orchestrated manner (Figure 1-4) (Carr et al., 2015).

Previous work in our lab has demonstrated that the arginine residues in an RGRGR sequence of E2F1 are targets for methylation by PRMT5 and PRMT1 (Figure 1-4), and the balance of activity between these two methyltransferases acts as a molecular switch to channel E2F1 activity into distinct biological pathways (Figure 8-1) (Cho et al., 2012; Zheng et al., 2013). PRMT5, which symmetrically methylates R111 and R113 residues, allows E2F1 to activate cell proliferation pathways, whilst PRMT1-dependent asymmetric methylation of R109 drives E2F1-dependent apoptosis (Cho et al., 2012; Zheng et al., 2013). The activities of PRMT5 and PRMT1 are considered mutually exclusive. Methylation at R109 mediated by PRMT1 in DNA-damaged cells can antagonise the subsequent PRMT5-dependent methylation at R111 and R113 (Figure

8-1A). Conversely, this PRMT1-driven methylation is hindered by the binding of cyclin A to E2F1 in proliferating cells, whereas it instead augments the activity of PRMT5 methylating R111 and R113 residues as well as ensuring that E2F1 is committed to the cell cycle progression (Figure 8-1B) (Cho et al., 2012; Zheng et al., 2013). PRMT1 and PRMT5 allow E2F1 to regulate the expression of different subsets of genes, which partly contributes to the distinct biological outcomes. Furthermore, PRMT5-driven methylation at R111 and R113 is a known mark to be recognised by a reader protein, p100/TSN, which acts as a transcription co-factor of E2F1 in promoting pro-growth gene expression and downregulating apoptotic pathways (Zheng et al., 2013).

Interestingly, the R109 residue of E2F1, which can be methylated by PRMT1, was also identified as a primary target residue of PAD4 for citrullination along the R127 residue (Ghari et al., 2016). In inflammatory cells, PAD4-dependent citrullination of E2F1 was shown to augment the binding and transactivating activity of E2F1 at the pro-inflammatory promoters, resulting in the enhanced expression of cytokine genes involved in the immune response and inflammation (Ghari et al., 2016). Given that citrullination and methylation are antagonistic to one another for the same arginine residue in a competitive fashion, these findings together suggest the interplay between PAD4 and PRMTs may play an essential role in regulating the decision-making process of E2F1 activity to influence different subsets of its target genes, and greatly expand its genome-wide impact.

More recently, PRMT5-dependent arginine methylation of E2F1 was demonstrated to extend E2F1 activity to regulate a distinct group of genes at the level of alternative splicing, whereby the reader protein p100/TSN plays a key role in recruiting a large number of RNA and the spliceosome components to E2F1 (Figure 8-1B) (Roworth et al.,

2019). Given the interplay between PRMT1 and PRMT5, as well as the potential competitive crosstalk between PRMT1 and PAD4, it would be plausible that PAD4-mediated citrullination of E2F1 may also have an influence over the E2F1-target genome outside of its classical transcriptional regulation.

8.1.2 RNA-seq revealed the impact of PAD4 inhibition on splicing of genes.

This study set out to examine the genome-wide impact of PAD4-dependent citrullination of E2F1, and an RNA-seq experiment was performed in wildtype (WT) and E2F1 CRISPR-knockout (E2F1cr) HCT116 colorectal cancer cell lines treated with the PAD4 inhibitor GSK484 (Figure 3-3C). We indeed discovered a global impact of PAD4 inhibition and E2F1 knockdown on both transcription and splicing of numerous genes (Figure 3-5 and Figure 3-9). This result is consistent with previous studies indicating a potential link between citrullination and RNA processing (Lewallen et al., 2015; Tanikawa et al., 2018), and clearly provides novel, direct evidence that PAD4 regulates RNA splicing of genes in cancer cells.

Of great interest was the observation that GSK484 treatment in WT cells (WT-GSK484) only showed a significant impact at the level of AS, and our analysis did not detect any significant DEGs in this treatment condition (Figure 3-5 and Figure 3-9). This finding appears to conflict with the previously established function of PAD4 as a transcriptional co-factor, with various reports suggesting that transcription can be significantly influenced by the citrullination of histones (Cuthbert et al., 2004; Y. Wang et al., 2004) and other proteins such as transcription factors like Elk1 or E2F1 (Ghari et al., 2016; X. Zhang et al., 2011). That being said, it is important to note that most of these studies focused on the role that PAD4 plays under the conditions where its expression was significantly augmented; for example, by an ectopic expression of PAD4 (Cuthbert et al.,

2004; Ghari et al., 2016; Y. Wang et al., 2004; X. Zhang et al., 2011). In addition, such experiments were often coupled with the ionophore treatment, to further augment PAD4 activity in cells by inducing high intracellular calcium ion concentrations (Cuthbert et al., 2004; Ghari et al., 2016; Y. Wang et al., 2004). The investigation of the physiological roles played by PAD4 is, to a certain extent, limited to cell lines that express high levels of PAD4, such as differentiated or stimulated HL60 cells (Ghari et al., 2016; Y. Wang et al., 2004). On the contrary, the RNA-seq in this study was performed in HCT116 cells, which are known to express low to moderate levels of PAD4 (X. Luo et al., 2022), and the activity of PAD4 was further depleted by the PAD4-specific inhibitor GSK484. Therefore, our finding in this study likely indicates that PAD4 may principally act as a splicing regulator rather than as a modulator of transcription when its expression is not over-expressed.

Our understanding of the roles of PAD4 and citrullination in normal physiological conditions is currently limited. PADs, including PAD4, require a high concentration of calcium ions (10^3-10^4 nM) for enzymatic activation. The conformational change of the C-terminal active site, induced by Ca^{2+} binding, is believed to increase enzymatic activity by over 10,000-fold (Arita et al., 2004; Slade et al., 2015). Since the intracellular concentration of calcium ions is typically maintained at around 100 nM (Clapham, 2007), it was commonly believed that PAD4 would be catalytically inactive in a normal physiological condition. However, several studies have reported potential substrates of PAD4, including histone and non-histone proteins, in cells under physiological conditions without ectopic PAD4 expression (Hagiwara et al., 2002; Tanikawa et al., 2018). This indicates the existence of a yet-unknown mechanism through which PAD4 evades the Ca^{2+} dependency.

One obvious candidate mechanism is post-translational modifications (PTMs), which widely regulate the activity and structure of substrate proteins in a variety of biological processes (G. Duan & Walther, 2015; Ramazi & Zahiri, 2021). There exist only a few studies investigating the PTMs on PAD4 to date. Probably the most studied example is citrullination itself, whereby the auto-citrullination of this PAD member has been previously reported (Andrade et al., 2010; Slack et al., 2011). However, a recent study demonstrated that auto-citrullinated PAD4 exhibits a similar enzymatic activity with unmodified PAD4 *in vitro*; despite the auto-citrullination arginine residues (R372, R374, and R639) located in proximity to the substrate binding pocket of PAD4, such modification does not influence the conformation of the protein (X. Liu et al., 2021). The authors nevertheless did not exclude the possibility of auto-citrullination affecting protein-protein interactions with as-yet unknown binding partners or co-factors of PAD4 (X. Liu et al., 2021), which may compensate for or lower the Ca^{2+} requirement for its efficient enzymatic activity. Continuous investigation of PAD4 auto-citrullination, as well as further identification and characterisation of other PTMs on PAD4, would be a promising direction for future investigations and should contribute to expanding our knowledge about intracellular functions of PAD4 under normal physiology. Alternatively, it is plausible that supraphysiological concentrations of Ca^{2+} may exist in specific subcellular locations or within certain organelles, which could enable intracellular activation of PAD4 through a conventional calcium-dependent mechanism. Future studies could therefore explore the dynamics of Ca^{2+} concentration in cells and their effects on the molecular pathways related to PAD4.

In addition to its physiological roles, it will also be fascinating to investigate the genome-wide impact of augmented PAD4 activity, for example, by performing an RNA-seq

experiment in cells with ectopic expression of PAD4 and treated with calcium ionophore. Given the abnormal expression and activity of PAD4 found in numerous diseases including cancer (Christophorou, 2022; Yuzhalin, 2019), this type of experiments may provide a more precise indication of the pathological relevance of PAD4-mediated regulation of RNA splicing. In addition, considering the transcriptional function that PAD4 plays through citrullinating histones and non-histone substrates when over-expressed (Cuthbert et al., 2004; Ghari et al., 2016; Y. Wang et al., 2004; X. Zhang et al., 2011), an RNA-seq experiment with ectopic PAD4 expression would be able to more accurately profile the transcriptome regulated by PAD4-dependent citrullination at the transcriptional level. By comparing this RNA-seq with the one in this study, it could potentially confirm the hypothesis that PAD4 may act as a transcriptional co-factor or splicing regulator depending on its expression level.

8.1.3 Both E2F1 and PAD4 regulate alternative splicing.

On the contrary to the WT-GSK484 condition, we discovered large numbers of DEGs in the E2F1cr-DMSO and E2F1cr-GSK484 conditions (Figure 3-5). Only less than 50 % of them were predicted E2F1-targets through E2F1 ChIP-seq analysis using ENCODE datasets (Figure 3-7), suggesting that the E2F1/PAD4 axis can indirectly regulate the expression of many genes, probably through the gene regulation of other transcription factors or co-factors. In addition, most of these DEGs were shared between both conditions. This probably means their differential expression was as a result of E2F1 knockdown and not hugely affected by PAD4 inhibition, highlighting the significant role that E2F1 plays as a master transcription regulator. Nonetheless, it is also worth noting that subsets of genes were differentially expressed only in either E2F1cr-DMSO or E2F1cr-GSK484 (Figure 3-5), suggesting that PAD4 inhibition had an impact on

transcription of such genes in E2F1cr cells, unlike in the WT-GSK484 condition. We later revealed that E2F1 and PAD4 can interact with each other in the protein complex with p100/TSN and SRSF3 when their expressions are augmented (Figure 5-2). Taken together, E2F1, by physically associating with PAD4 and other splicing-related factors, may have a role to mask the transcriptional activity of PAD4 and instead engage it to splicing regulation.

Our RNA-seq also identified a large number of genes alternatively spliced in the E2F1cr-DMSO and E2F1cr-GSK484 conditions as well as in WT-GSK484 (Figure 3-9). The E2F1 ChIP-seq analysis identified the majority of the spliced genes were predicted E2F1 targets (Figure 3-10), indicating that the DNA binding ability of E2F1 might still be required for its function to regulate splicing of genes. Interestingly, however, we found only minimal overlap between DEGs and spliced genes in these two conditions of E2F1cr cells (Figure 3-11). This means that E2F1 may regulate distinct subsets of genes, with some representing poor transcription targets that are regulated instead at the AS level. A similar phenomenon was observed in our previous study, whereby E2F1 methylated by PRMT5 was shown to regulate the expression of distinct populations of genes by transcription or AS (Roworth et al., 2019). Our result further supports this previous finding and suggests citrullination, as well as methylation, can greatly extend the genomic landscape under E2F1 control.

8.1.4 Potential involvement of p53 in the E2F1/PAD4 axis in AS regulation.

After uncovering a novel function for the E2F1/PAD4 axis in AS regulation through our RNA-seq analysis, we confirmed that some of these AS changes are regulated in an E2F1 and PAD4-dependent manner in HCT116 p53-/- cells by RT-qPCR (Figure 4-4). As discussed in Chapter 4, our RT-qPCR validation experiments in HCT116 p53+/+ cells,

the original cell line in which the RNA-seq was performed, displayed a trend of AS changes consistent with the RNA-seq results, though did not meet statistical significance (Figure 4-3).

Previously, our research group has conducted another RNA-seq experiment in WT and E2F1cr HCT116 p53-/- cell lines to investigate the impact of the PRMT5 inhibitor T1-44, whereby PRMT5-driven methylation of E2F1 was demonstrated to have an important function in regulating cell migration and invasion through transcriptional regulation of several motility-related genes (Barczak et al., 2020). The rMATS analysis of this RNA-seq dataset in p53-/- cells discovered a much larger number of AS events as compared to the RNA-seq in p53+/+ cells in this study; for example, approximately 3 times as many AS events were found in the E2F1cr-DMSO condition with respect to the control WT-DMSO condition in HCT116 p53-/- cells (Figure S-3). The p53 protein is an important tumour suppressor, and plays a critical role during the response to various cellular stresses such as ionizing radiation, hypoxia, carcinogens, and oxidative stress, typically by promoting cell-cycle arrest and apoptosis (Pflaum et al., 2014). Reflecting its importance in cancer progression, p53 marks one of the most frequently mutated genes in the human cancer genome (Hafner et al., 2019; Olivier et al., 2002). Considering the forementioned observation in HCT116 p53-/- cell lines, as well as the fact that p53 is an important mediator of apoptosis, we hypothesised that the p53 knockout might make HCT116 cells less sensitive to apoptosis induced by the AS disturbances and allow us to examine cells bearing greater levels of changes in their splicing profile. To support this hypothesis, we discovered the same trends of AS changes with greater magnitude and statistical significance in HCT116 p53-/- cells compared to p53-/- cells (Figure 4-4).

Programmed cell death or Apoptosis is known as one of the most common biological processes affected by alternative splicing, and several apoptotic genes have been identified to encode multiple splicing variants with opposite effects on cell survival (Paronetto et al., 2016). A well-known example is a member of the Bcl-2 family called Bcl-x (Dou et al., 2021). An AS event targeting its exon 2 lead to two isoforms of Bcl-x transcripts with antagonistic outcomes; Bcl-xL (long isoform) acts against apoptosis, whilst Bcl-xS (short isoform) has a pro-apoptotic effect on cells (Dou et al., 2021). P53 has been reported to directly interact with and antagonise Bcl-xL, resulting in apoptosis through the mitochondrial pathway (Mihara et al., 2003). Even though our RNA-seq did not detect significant splice changes for Bcl-x gene, it is still possible that p53 induces apoptosis in response to splicing events of other apoptotic genes in a similar manner in HCT116 p53+/+ cells. This would explain why we observed a larger impact at the level of RNA processing in the absence of p53, but further characterisation and investigation would be required to confirm this.

In addition, p53 itself is a known target for alternative splicing. Splice variants contributes to a production of around 12 isoforms of this tumour suppressor protein along with other mechanisms such as alternative initiation of translation and usage of alternative promoters (Surget et al., 2013). Aberrant splicing of p53 has been reported to affect its activity; for example, a spliced variant called Δ133-p53 was shown not to bind p53 response elements , as well inhibiting full-length p53 from doing so (Marcel et al., 2010). Again, p53 was not in the list of our spliced genes affected by the E2F1/PAD4 axis in this study, but the future study may explore the possibility of PAD4-mediated splicing regulation of this protein.

Finally, one also has to bear in mind that there is substantial evidence showing a direct link between p53 and both E2F1/PAD4. For example, E2F1 can activate the p53-

dependent apoptotic pathway through the expression of $p14^{ARF}$ in response to DNA damage (Laurie et al., 2006; Pierce et al., 1999; Pierce, Gimenez-Conti, et al., 1998). In the absence of p53, E2F1 instead mediates the expression of a p53-related protein called p73 to induce apoptosis (Roworth et al., 2015). On the other hand, multiple reports also highlighted the crosstalk between PAD4 and p53 in multiple cellular contexts. PAD4 was demonstrated to interact with p53 in cells and inhibit the expression of p53-target genes related to cell cycle arrest and apoptosis (Yao et al., 2008). In addition, another study suggested that PAD4 can indirectly reduce the p53 transcription ability; Guo and Fast showed that the citrullination of a tumour suppressor protein called inhibitor of growth 4 (ING4) resulted in the reduced association of this protein with p53 (Q. Guo & Fast, 2011). ING4 is a known co-factor of p53, whereby it recruits the histone acetyltransferase p300 and promotes the acetylation status of p53 at the K382 residue (Gu & Roeder, 1997). Citrullinated ING4, which dissociates from p53, can no longer activate p53 through this p300 pathway, leading to the repression of p53-target genes including p21 (Q. Guo & Fast, 2011). Collectively, these studies provide evidence of intricate molecular interactions among E2F1, PAD4, and p53 in diverse cellular pathways, suggesting that the knockdown of p53 may have affected the activities of E2F1 and PAD4 in the present study too. Future studies are required to carefully examine this interplay in the context of RNA processing, especially the relevance of p53 to the PAD4/E2F1-mediated regulation of splicing.

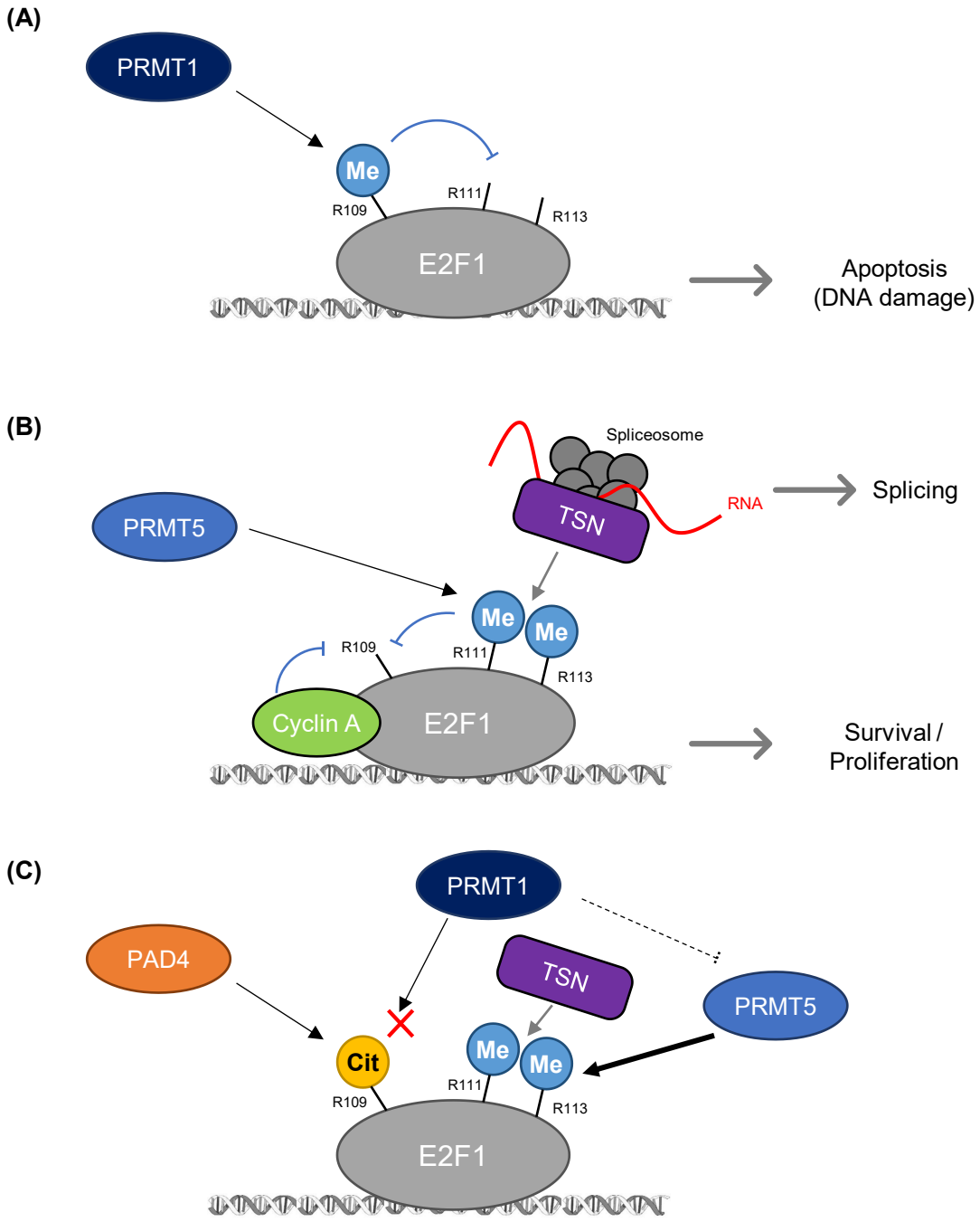


Figure 8-1 Potential Interplay between PAD4 and PRMTs on E2F1.

Model diagram explaining the potential crosstalk between citrullination and methylation on E2F1. (A) In DNA-damaged cells, PRMT1 methylates R109 of E2F1, which antagonises PRMT5-driven arginine methylation at R111 and R113. This methylation event promotes E2F1-dependent apoptosis. (Zheng et al., 2013) (B) In cycling cells, cyclin A binding to E2F1 hinders PRMT1-driven methylation, resulting in PRMT5-dependent methylation events at R111 and R113. This instead promotes proliferation (Zheng et al., 2013). This methylation marks are read by p100/TSN, which recruits RNA and spliceosome components to E2F1 (Roworth et al., 2019). (C) PAD4-mediated citrullination also occurs at R109 on E2F1 (Ghari et al., 2016), which should effectively block the subsequent arginine methylation by PRMT1. This may de-repress the activity of PRMT5, resulting in the p100/TSN interaction as seen in this study.

8.2 Molecular Mechanisms of E2F1/PAD4-regulated AS.

8.2.1 PAD4 activity influences the interaction of E2F1 with its target promoters.

Using the validated AS target genes from our RNA-seq, we performed a series of biochemical experiments in HCT116 p53-/- cells to elucidate the molecular mechanisms underlying the E2F1/PAD4-dependent regulation of splicing. EXOC4 was one of the validated examples whereby skipping of its exon (SE) 7 was regulated by the interplay of E2F1 and PAD4 (Figure 4-4). This gene was predicted to be an E2F1 target after analysing the E2F1 ChIP-seq data deposited in the ENCODE project (Figure 4-6), and we subsequently confirmed that E2F1 can bind to the promoter of EXOC4 by E2F1 ChIP (Figure 4-7). We also observed that the increased level of PAD4 expression can enhance this E2F1-DNA association (Figure 4-8). Interestingly, subsequent chemical inhibition of PAD4 had an effect to suppress this enhanced binding, but on the spliced genes, whilst the binding of E2F1 to most of its transcription target promoters remained relatively unaffected (Figure 4-8 and Figure 4-9). This might be related to an essential but uncharted mechanism by which E2F1 distinguishes its target genes for transcriptional regulation or regulation at the level of RNA processing, but the number of tested genes was too limited to make a conclusion and a large-scale confirmation experiment would still be required.

In addition, we subsequently revealed that splicing events including EXOC4 SE7 are directly dependent on the citrullination of E2F1 (Figure 4-10), by using a citrullination-defected mutant of E2F1 that was used in our previous study (Ghari et al., 2016). Together with the aforementioned E2F1 ChIP experiments (Figure 4-7 and Figure 4-8), this finding suggests that PAD4-mediated citrullination of E2F1 plays a crucial role in the regulation of AS for target genes, potentially by strengthening the E2F1 binding to the promoters of such genes. This is consistent with our previous finding that the association of methylate

E2F1 with spliceosome snRNA requires the E2F1-chromatin association, whereby the mutation in the DNA-binding domain of E2F1 disrupted the binding of U6 snRNA to methylated E2F1 (Roworth et al., 2019). However, since E2F1 binding was not observed at the promoters of some AS target genes such as SPIN1 (Figure 4-7), this mechanism may not apply to all of the splicing targets, indicating the possibility of the participation of E2F1 and PAD4 in various pathways to regulate RNA splicing.

One limitation of this citrullination-defective mutant assay, however, is that we were unable to detect the direct citrullination of E2F1 by immunoblot. Citrullination is technically challenging to detect since the quality and efficacy of commercially available antibodies are still poor. We therefore used the anti-citrulline (modified) kit to attempt to detect citrullination on immunoprecipitated E2F1, but unfortunately it was not possible to achieve a clear result (data not shown). Nevertheless, both endogenous and ectopically-expressed E2F1 have been found to be citrullinated in a previous study (Ghari et al., 2016) and intracellular PAD4 activity and inhibition by GSK484 were confirmed in this project (Figure 3-1). We therefore conclude that our findings are still sufficient evidence to suggest that citrullination of E2F1 is central to the regulatory mechanism of AS for genes including EXOC4 SE7, though continued development of citrullination detection tools would permit us to investigate this question more thoroughly.

Future research would further characterise this citrullination-defective E2F1 mutant; for example, a ChIP experiment with this mutant would measure the impact that E2F1 citrullination has on its DNA binding affinity. Nonetheless, it is noteworthy that site-specific mutagenesis of E2F1 arginine residues R109, R111, R113, and R127 can not only inhibit citrullination but also efficiently block arginine methylation. Future experiments therefore need to carefully distinguish between the effects of citrullination

and methylation of these residues by the use of careful control experiments and the specific PAD/PRMT inhibitors.

8.2.2 PAD4 activity influences the interaction of E2F1 with p100/TSN.

Considering no RNA-binding motifs were identified either in E2F1 and PAD4, it would be more plausible to predict that these two proteins regulate RNA splicing by interacting with and modulating the activity of other RNA-binding proteins. Indeed, as well as the protein-DNA interaction, we demonstrated that PAD4 has key functions to regulate two additional levels of molecular interactions in cells: protein-protein and protein-RNA interactions.

First, we showed that ectopic PAD4 can enhance the interaction of E2F1 with p100/Tudor staphylococcal nuclease (TSN) (Figure 5-1). In addition to being described as a transcriptional co-activator, p100/TSN reportedly plays a central role in RNA splicing (Gutierrez-Beltran et al., 2016). This protein directly interacts with components of the spliceosome including U1, U2, U4, U5 and U6 snRNPs and a few Sm proteins, and thereby it facilitates the sequential assembly of this complex at the splice sites (Figure 1-9) and enhances the rate of splicing *in vitro* (Gao et al., 2012; J. Yang et al., 2007). Previous work in our lab demonstrated that E2F1 is also associated with p100/TSN in a PRMT5-dependent manner, where this binding of p100/TSN was only observed for E2F1 symmetrically methylated by PRMT5 at R111 and R113 (Figure 8-1B), and not for E2F1 asymmetrically methylated by PRMT1 at R109 or unmethylated E2F1 (Roworth et al., 2019; Zheng et al., 2013). Methylation events of E2F1 catalysed by PRMT5 and PRMT1 are considered mutually exclusive, and the balance in activity between these two enzymes has an essential role in channelling E2F1 into different pathways with distinct biological outcomes (Cho et al., 2012; Zheng et al., 2013). In response to DNA damage, PRMT1

principally acts to methylate E2F1 at R109 and hinder methylation at R111 and R113 by PRMT5, resulting in the induction of E2F1-dependent apoptosis (Figure 8-1A). In cycling cells, the interaction of cyclin A with E2F1 instead hinders PRMT1-driven methylation and allows methylation at R111 and R113 to be catalysed by PRMT5, which engages E2F1 into a proliferative mode (Figure 8-1B) (Zheng et al., 2013). The p100/TSN protein was initially thought to act as a transcription co-factor of methylated E2F1 in promoting gene expression within growth pathways and downregulating pro-apoptotic genes; p100/TSN ChIP experiments demonstrated the localisation of this protein at the promoters of E2F-target pro-growth genes such as Cyclin E and CDC6 in a PRMT5-dependent fashion, whereas such binding was not observed at pro-apoptotic E2F1-target genes including APAF-1 and p73 (Zheng et al., 2013). However, more recently, p100/TSN was also shown to play an important role beyond transcriptional regulation; this multifunctional protein recruits a large body of pre-mRNA and components of the spliceosome complex to E2F1 in a PRMT5-dependent fashion, and allows it to regulate alternative splicing of a subset of its target genes (Roworth et al., 2019).

In this study, increased PAD4 expression was found to enhance the interaction between p100/TSN and E2F1 (Figure 5-1). Citrullination of peptidyl arginine residue effectively blocks the recognition and methylation by PRMTs (Raijmakers et al., 2007), and PAD4 primarily citrullinates R109 residue which is also a target of PRMT1-mediated methylation (Ghari et al., 2016; Zheng et al., 2013). Taken together, this enhanced interaction of E2F1 with p100/TSN in PAD4 over-expressing cells might be a result of the crosstalk between PAD4 and PRMTs; increased PAD4 expression may competitively inhibit PRMT1-dependent methylation at R109 via targeting this arginine residue for citrullination, which, given that this methylation event by PRMT1 is known to hinder the

activity of PRMT5 methylating R111/R113, may result in the relieved suppression of PRMT5-driven methylation events and subsequently enhance the recruitment of p100/TSN reading these marks (Figure 8-1C). Future research needs to confirm this interplay biochemically, for example, by using the PRMT5-mediated methylation-defective E2F1 mutant whose R111/R113 residues are site-specifically mutated, as used in previous studies (Cho et al., 2012; Roworth et al., 2019; Zheng et al., 2013).

8.2.3 PAD4 activity influences the interaction of E2F1 with SRSF3.

In addition to p100/TSN, we also revealed that the splicing factor SRSF3 was bound to E2F1 in a PAD4-dependent manner (Figure 5-2). Serine/arginine-rich (SR) proteins are one of the major families of AS regulators in eukaryotes (Bradley et al., 2015; Jeong, 2017; Shepard & Hertel, 2009). By recognising splicing regulatory elements (SRE) in the pre-mRNA, particularly splicing enhancer sequences, they can either promote or suppress constitutive and alternative splicing events through multiple mechanisms; such as helping the recruitment of spliceosome components, physically blocking the splice site, or 'looping out' the target exon to help the splicing machinery remove it (Bradley et al., 2015; Dvinge et al., 2016; Stamm et al., 2012). A growing body of evidence suggests that splicing factors including SR proteins are often deregulated in disease conditions including cancer, and SRSF3 is no exception. Aberrant expression of SRSF3 can be found in multiple types of tumours, and increasing efforts are being devoted to exploring the possibility of targeting SRSF3 therapeutically (Z. Zhou et al., 2020).

We started to investigate SRSF3 because this SR protein member was identified as a substrate of PAD2 (Lewallen et al., 2015). In this study, we discovered that PAD4 can also citrullinate SRSF3 (Figure 5-2A), as well as promote its association with E2F1, PAD4, and p100/TSN (Figure 5-2B). In addition, siRNA treatment for SRSF3 had a

similar impact on AS changes as knockout of E2F1 or PAD4 inhibitor treatment (Figure 5-3). Notably, for the EXOC4 gene, the treatment of cells with GSK484 or SRSF3 siRNA increased the inclusion of exon 7, whilst the co-treatment with both did not significantly increase the degree of inclusion (Figure 5-3). This indicates that PAD4 and SRSF3 are likely to regulate this splicing event through the same pathway.

However, many questions remain unanswered regarding the functional consequence of PAD4-dependent citrullination of SRSF3; for example, whether this modification is necessary for its interaction with E2F1 and p100/TSN, or has any impacts on its activity in regulating splicing. Therefore, future study should concentrate on precisely identifying the site(s) of citrullination on SRSF3 by mass spectrometry, and subsequent site-directed mutagenesis studies may provide further insight towards the functional relevance of PAD4-mediated SRSF3 citrullination.

8.2.4 PAD4 activity potentially influences the protein-RNA interactions of SRSF3.

Another level of molecular interactions regulated by PAD4 in cells is the protein-RNA interaction. We identified that both E2F1 and SRSF3 are associated with EXOC4 mRNA by RNA immunoprecipitation (RIP) (Figure 6-3 and Figure 6-4). Considering that they interact with each other in cells in a PAD4-dependent manner, and no RNA-binding sequence has yet been identified in E2F1, one can predict that E2F1 is associated with mRNA via the splicing machinery including SRSF3 under the effect of PAD4 activity. Of our interest is that citrullination of E2F1 appears to be a cornerstone of this RNA-protein association, whereby GSK484 treatment significantly increased the association between SRSF3 and EXOC4 mRNA in an E2F1-dependent fashion despite some relatively large error bars (Figure 6-7). This observation was only made in the WT cells and was not observed in the absence of E2F1. This provides a potential insight into the

molecular mechanism of E2F1/PAD4-regulated AS; the citrullination of E2F1 may control the RNA binding affinity of the associated splicing factor and regulates the splicing events of target genes.

Citrullination results in the loss of positive charge from an arginine residue (Figure 1-5). Given that nucleic acids are negatively charged, one may plausibly expect PADs to negatively regulate protein-RNA interactions of their target substrates. To support this view, it has previously been demonstrated that the citrullination of splicing factor SFPQ results in its reduced affinity for mRNA transcripts (Snijders et al., 2015). Our finding supports this observation, since PAD4 inhibitor treatment caused an increased binding of SRSF3 to target mRNA exon (Figure 6-7). However, ectopic expression of HA-PAD4 in cells did not decrease RNA enrichment by SRSF3 RIP, and the described effect of GSK484 was only found in the presence of E2F1 (Figure 6-7). This E2F1 dependency suggests that the observed change in the RNA binding affinity of SRSF3 might not simply be due to PAD4-mediated citrullination of its RNA binding site, but likely also involves the regulation of protein-protein interaction between SRSF3 and E2F1.

8.2.5 PAD4 activity influences the interaction of E2F1 with RNA-binding proteins.

Furthermore, in hope of identifying the novel interaction of E2F1 with splicing-related factors that were affected by the PAD4 inhibition, we performed the E2F1 immuno-precipitation-mass spectrometry (IP-MS) experiment in cells treated with GSK484 (Table 7-1 and Table 7-2). Interestingly, the Gene Ontology (GO) term ‘RNA-binding’ was over-represented by 16 potential RNA-binding proteins (RBPs) whose interaction with E2F1 was significantly influenced by the GSK484 treatment (Figure 7-1), suggesting that PAD4 may have a significant effect on regulating the E2F1-RBP interactome. Since many RBPs have been identified as targets for citrullination (Lewallen et al., 2015; Tanikawa

et al., 2018), it is plausible to speculate that PAD4 could play a wider role as a master regulator of RBPs by modulating their interactions with other proteins. The RNA helicase DDX39A and the scaffolding protein AHNAK were potential splicing-related factors among these RBPs (Table 7-3). Given that our SRSF3 siRNA experiments indicated that the E2F1/PAD4-dependent AS regulation was likely involved with multiple pathways (Figure 5-3), the relevance of these splicing factors to our validated AS events warrants further investigation.

8.3 Model of AS Regulation by the E2F1/PAD4/SRSF3 Axis.

Taking into account previous literature, this study highlights a possible mechanism whereby PAD4-dependent citrullination extends E2F1 activity to splicing regulation via its crosstalk with PRMT-mediated methylation, as illustrated in Figure 8-2. In this model, PAD4-directed citrullination allows E2F1 to associate with the promoter regions of its splicing target genes. In addition, this citrullination at the R109 residue competitively inhibits the activity of PRMT1 which targets the same arginine residue of E2F1 for methylation. This results in the de-repression of antagonistic PRMT5 activity, and thus promotes methylation events at the R111 and R113. PRMT5-driven methylation marks are recognised by p100/TSN, which may recruit target pre-mRNA and splicing machinery to E2F1 as proposed by Roworth et al. (Roworth et al., 2019). Furthermore, either by this p100/TSN protein or via direct citrullination of itself or E2F1, SRSF3 is also recruited to chromatin-bound E2F1 in a PAD4-dependent manner. This recruitment engages the E2F1/SRSF3 complex to mRNA, and thereby, for example, positively regulates exon 7 skipping of EXOC4 co-transcriptionally. In the absence of PAD4 activity, the interaction of E2F1 with the splicing machinery containing SRSF3 is disrupted, leading to the change in alternative splicing profile as visualised in our RNA-seq.

8.3.1 Molecular mechanism of SRSF3 regulating AS in the E2F1/PAD4 axis.

However, there remain a few uncertainties to be clarified in this model, including the precise role that SRSF3 undertakes in regulating AS in the E2F1/PAD4 axis. In this research, we discovered that GSK484 treatment can promote the inclusion of EXOC4 exon 7 by the RNA-seq and RT-qPCR validation (Figure 4-4). In addition, we later revealed that the same PAD4 inhibitor treatment enhanced the interaction of SRSF3 with this target exon of EXOC4 RNA in an E2F1-dependent fashion (Figure 6-7). However,

our result demonstrated that SRSF3 siRNA treatment also promoted the inclusion of the same exon of EXOC4 (Figure 5-3). Together, these results indicate that the seemingly contradicting states of SRSF3, one increasingly associated with its target exon and the other silenced by siRNA, may nonetheless produce the same biological outcome in terms of EXOC4 exon 7 usage.

One potential explanation is that SRSF3 might regulate this AS event in a dual manner, exhibiting both positive and negative effects contingent upon its specific binding location (Figure 8-3). In alternative splicing, SR proteins, including SRSF3, are generally implicated as a splicing activator; they bind to *cis*-acting exonic or intronic splicing enhancers (E/ISEs) on pre-mRNA via their RNA-recognition motifs (RRM), and recruit the splicing machinery via RS domain (Graveley, 2000; Lam & Hertel, 2002; Matlin et al., 2005). This means that our SRSF3 siRNA experiment, which demonstrated suppression of this AS event (skipping of EXOC4 exon 7) when SRSF3 was silenced (Figure 5-3), may likely reflect this classical SRSF3 activity as an AS activator. Nevertheless, mounting evidence suggests that the notion of SR proteins simply serving as splicing activators is overly simplified, with several studies reporting their opposite roles as splicing suppressors (Bradley et al., 2015; Pandit et al., 2013; M. Shen & Mattox, 2012). Their functional properties appear to be dependent on several factors, such as the sequence and location of the *cis*-acting element to which SR proteins bind with regards to the splice site in the pre-mRNA, and whether the adjacent exons are constitutive or not (Bradley et al., 2015; Pandit et al., 2013; M. Shen & Mattox, 2012). Accordingly, with the assumption that EXOC4 exon 7 contains the *cis*-acting exonic splicing silencer (ESS) rather than ESE, SRSF3 could possibly act as a suppressor for the skipping of this exon when binding to this exon itself. This would explain why the GSK484 treatment, which

enhanced the interaction of SRSF3 with EXOC4 mRNA at the target exon 7, resulted in the increased inclusion of this exon (Figure 8-3). Based on the fact that the described effect of GSK484 on the RNA binding affinity of SRSF3 was only observed in WT cells and not in E2F1cr cells (Figure 6-7), it can be hypothesised that the citrullination of E2F1 plays a significant role in regulating the target selection of SRSF3 and isolating this splicing factor from the silencer element. PAD4-dependent regulation of the protein-protein interaction between E2F1 and SRSF3 (Figure 5-2), as well as E2F1-chromatin association in a similar manner (Figure 4-8), might therefore be a key mechanism underlying the E2F1/PAD4 axis facilitating the SRSF3 binding into distinct *cis*-acting elements.

As discussed here, this study highlights a complicated nature of SRSF3 biology in AS regulation. Its RNA target specificity, particularly in response to PAD4 and E2F1 activity, would therefore be a fruitful area for further work. Experiments including SRSF3 RIP-seq performed in E2F1cr cells, or cells treated with PAD4 inhibitor, may enable effective tracking of SRSF3 splicing targets regulated by the E2F1/PAD4 axis. Additionally, a technique called Targets of RNA-binding proteins Identified By Editing (TRIBE), which employs the partial fusion of the RNA-editing enzyme ADAR with a target RNA-binding protein (RBP) and detects the RNA-editing activity at the target RNAs by RNA-seq, has recently emerged as a tool to study the target specificity of RBPs (Burjoski & Reddy, 2021; Jin et al., 2021; McMahon et al., 2016). One recent study indeed exploited this technique for SRSF3 and identified several consensus motifs (Jin et al., 2021), although we were unable to locate such motifs around the respective spliced region of EXOC4 (exon 6, 7, and 8) through the motif scanning analysis using the online tool FIMO (Data not shown) (https://meme-suite.org/meme/tools/fimo) (Grant et al., 2011). Nevertheless,

with the use of this novel technique in our experimental setting, we might be able to profile SRSF3 binding sites around our target genes more precisely, which should help us further elucidate the molecular mechanisms underpinning the interplay between E2F1, PAD4 and SRSF3.

8.3.2 Involvement of p100/TSN and DNA-binding domain of E2F1

Previous work in our lab has demonstrated that PRMT5-driven arginine methylation at R111/R113 allows E2F1 to regulate a subset of its target genes at the level of alternative splicing (Roworth et al., 2019). The authors proposed that this activity of methylated E2F1 requires the reader protein p100/TSN, which functions to recruit a large body of RNAs and the splicing machinery components to E2F1 (Roworth et al., 2019). Indeed, both p100/TSN siRNA and PRMT5 inhibitor treatments were shown to have a similar impact to deplete the association between E2F1 and RNA in cells (Roworth et al., 2019).

In this study, we demonstrated that the interaction of E2F1 with p100/TSN was also enhanced by the ectopic expression of PAD4 (Figure 5-1). It was hypothesised that this might be mediated through the de-repression of PRMT5 activity, by PAD4 competitively inhibiting antagonistic PRMT1-mediated methylation (Figure 8-1C). However, our subsequent experiments rather focused on another splicing factor SRSF3, and the question of whether the E2F1/PAD4-regulated AS events are dependent on p100/TSN remained unanswered. Furthermore, we have demonstrated that SRSF3 can be associated with p100/TSN, E2F1, and PAD4 under ectopic expression of E2F1 and PAD4 (Figure 5-2). Given that p100/TSN can recruit components of splicing machinery to methylated E2F1 (Roworth et al., 2019), one could hypothesise that p100/TSN might also facilitate the recruitment of SRSF3 to citrullinated E2F1, and thereby regulate splicing in the E2F1/PAD4 axis. Conversely, the recruitment of this SR protein could be entirely

independent of p100/TSN; given that both E2F1 and SRSF3 are substrates of PAD4 (Figure 5-1) (Ghari et al., 2016), this protein-protein interaction might simply be mediated by PAD4-driven citrullination of E2F1 and/or SRSF3. Future research is therefore needed to delineate the role of the p100/TSN protein in the E2F1/PAD4 axis thoroughly, especially regarding its potential function in recruiting splicing machinery components to citrullinated E2F1, and whether this included SRSF3 or not. Since the interaction between E2F1 and p100/TSN should be primarily dependent on PRMT5-driven methylation (Roworth et al., 2019; Zheng et al., 2013), this investigation might be able to provide an experimental confirmation for the potential interplay between PAD4 and PRMTs on E2F1, as indicated throughout this study (Figure 8-1).

In addition to the p100/TSN dependency, the study by Roworth et al. demonstrated that the DNA binding activity of E2F1 is essential for this transcription factor to regulate splicing of its target genes (Roworth et al., 2019). They utilised the E2F1 mutant with compromised chromatin-binding activity, in which Leucine L132 and arginine R166 in the DNA-binding domain (DBD) were replaced by glutamine acid and histidine respectively, and showed that the interaction of E2F1 with U6 snRNA was disrupted in this DBD mutant (Roworth et al., 2019). In this study, we demonstrated that the association of E2F1 with its AS target promoters can be enhanced by ectopic PAD4 expression and reduced by subsequent GSK484 treatment (Figure 4-8). This result indicates that E2F1/PAD4-regulated alternative splicing is also likely to occur at chromatin-associated E2F1, and in part regulated by PAD4 activity which modulates the interaction between E2F1 and DNA. Future studies should therefore examine this potential co-transcriptional mechanism of the E2F1/PAD4-regulated splicing, possibly

with the use of the aforementioned DBD mutant of E2F1 in combination with PAD4 over-expression or inhibition.

8.3.3 SRSF3-independent mechanism

The model proposed in Figure 8-2 may also fail to account for certain AS events that are regulated by E2F1 and/or PAD4 but independently of SRSF3. Our SRSF3 siRNA experiment demonstrated a clear involvement of this SR protein in the exon skipping of genes like EXOC4 and SPIN1 (Figure 5-3). However, the influence of siSRSF3 appeared to be limited on other validated splicing events such as RBM25 SE2 and SNAP23 RI3 (Figure 5-3). It can therefore be hypothesised that other downstream splicing factors, apart from SRSF3, might also take part in the E2F1/PAD4 axis. Our IP-MS experiment for the E2F1 interactome and GO analysis demonstrated that the E2F1 interactome affected by GSK484 treatment contained many potential RBPs (Figure 7-1), and subsequent functional annotation analysis identified two candidate splicing-related factors, DDX39A and ANHAK (Table 7-3). Further characterisation of these RBPs in the E2F1/PAD4 axis might prove their involvement in the E2F1/PAD4-regulated pathways.

In addition, a series of immunoprecipitation experiments performed in this study indicated that over-expression of PAD4 had a greater impact on the interaction of E2F1 with other proteins, such as SRSF3 and p100/TSN (Figure 5-1 and Figure 5-2), as compared to PAD4 inhibition (Figure S-7). Accordingly, an IP-MS experiment for E2F1 from cells expressing ectopic PAD4 and/or treated with calcium ionophore, instead of the PAD4 inhibition in this study, could potentially detect a wider spectrum of the E2F1 interactome influenced by PAD4 activity. Given the abnormal expression of PAD4 found in many diseases including cancer, this kind of experiments would also provide a better indication of the pathological or cancer relevance of citrullinated E2F1 interactome.

8.3.4 E2F1-independent mechanism

Finally, it will be fascinating to investigate other facets of PAD4-regulated splicing mechanisms, particularly via the citrullination of RBPs, for example at their RGG motifs, and independently of E2F1. This sequence element is often found in RBPs and functions to mediate protein-protein or protein-RNA interactions (Rajyaguru & Parker, 2012; Thandapani et al., 2013). Given that many RGG-containing RBPs were targeted by PAD4 and PAD2 for citrullination (Lewallen et al., 2015; Tanikawa et al., 2018), it is plausible to hypothesise a physiological or pathological role that PADs undertake in altering the splicing profile of cells by modulating the interaction of RBPs with their target RNAs or other partner proteins. The present study mainly focused on the E2F1/PAD4 axis and the citrullination of RBPs was slightly out of the scope. However, we discovered that SRSF3 can be citrullinated by PAD4 (Figure 5-2), as well as demonstrating that some AS events like SPIN1 SE3 were regulated by PAD4 and SRSF3 but independent of E2F1 through RT-qPCR experiments using SRSF3 siRNA (Figure 5-3). Together, these splicing events are likely regulated by the citrullination of SRSF3, rather than via an E2F1-dependent pathway proposed in the aforementioned model (Figure 8-2). Future research can test this E2F1-independent mechanism by investigating the RNA interactions of PAD-targeted RBPs including SRSF3, for example by RIP-seq in cells with modulated activity of PADs. In addition, an IP-MS experiment targeting any spliceosome complex component(s) from cells over-expressing PAD4 could produce interesting findings that account more for the splicing-related protein interactome and the influence of citrullination.

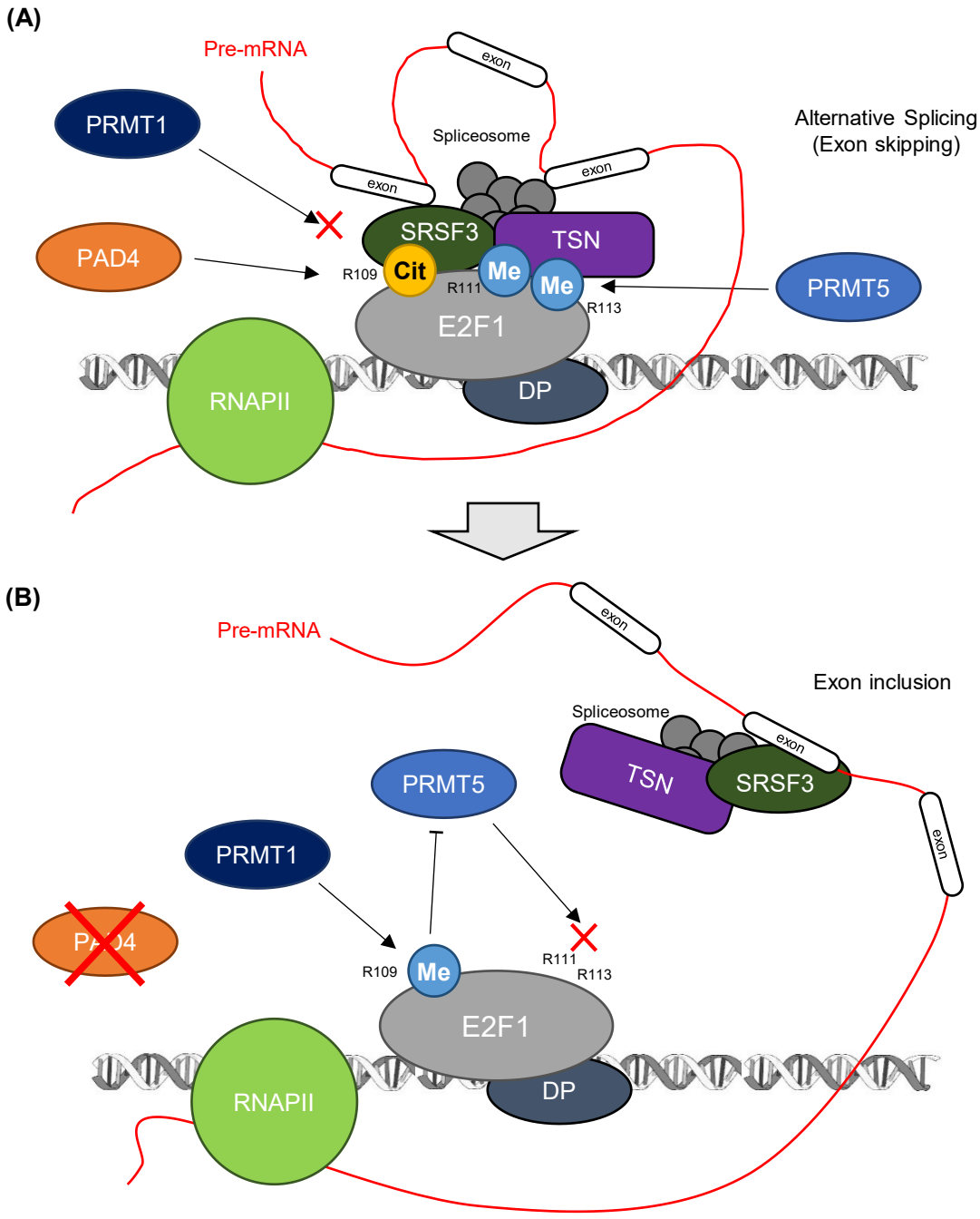


Figure 8-2 Potential mechanism of PAD4 influencing PRMT activities on E2F1.

Model diagram explaining the potential crosstalk between citrullination and methylation on E2F1. (A) In DNA-damaged cells, PRMT1 methylates R109 of E2F1, which antagonises PRMT5-driven arginine methylation at R111 and R113. This methylation event promotes E2F1-dependent apoptosis (Zheng et al., 2013). (B) In cycling cells, cyclin A binding to E2F1 hinders PRMT1-driven methylation, resulting in PRMT5-dependent methylation events at R111 and R113. This instead promotes proliferation (Zheng et al., 2013). This methylation marks are read by p100/TSN, which recruits RNA and spliceosome components to E2F1 (Roworth et al., 2019). (C) PAD4-mediated citrullination also occurs at R109 on E2F1 (Ghari et al., 2016), which should effectively block the subsequent arginine methylation by PRMT1. This may relieve the suppression of the PRMT5 activity, resulting in the p100/TSN interaction as seen in this study.

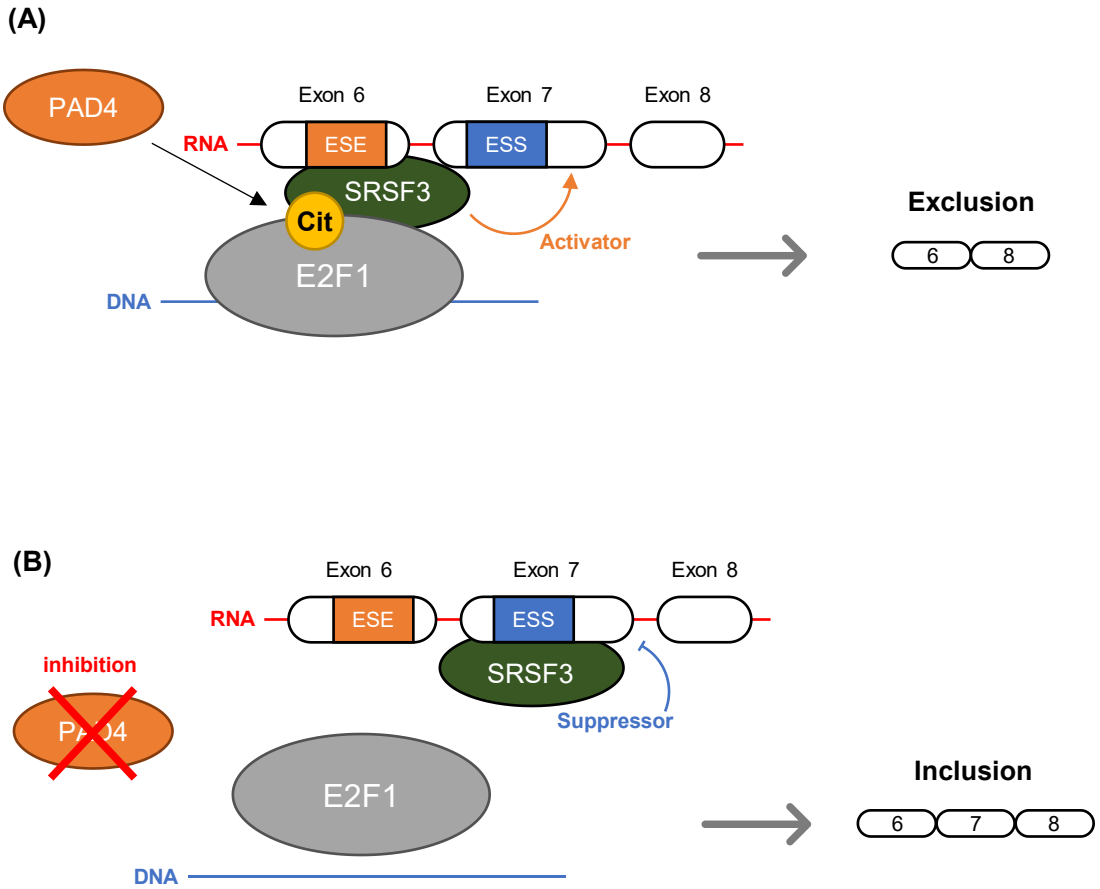


Figure 8-3 Potential dual roles of SRSF3 in the E2F1/PAD4 axis.

Model diagram illustrating potential dual roles that SRSF3 plays as a splicing activator and suppressor depending on its binding site. (A) Over-expression of PAD4 enhances the interaction between E2F1 and SRSF3 as well as E2F1-chromatin association at the target gene, by citrullinating E2F1 and/or SRSF3. These interactions would direct SRSF3 binding to the exonic or intronic splicing enhancer away from exon 7 (E/ISE; ESE in exon 6 as an example here), promoting the exclusion/skipping of EXOC4 exon 7. This might be a primary activity of SRSF3, as SRSF3 siRNA resulted in the inclusion of exon 7. (B) With GSK484 treatment, the E2F1 interactions with SRSF3 and the target promoter are no longer intact, releasing SRSF3. SRSF3 now binds to the exonic splicing silencer (ESS) within exon 7 and promotes the inclusion of this exon.

8.4 Biological and Clinical Significance of the E2F1/PAD4 Interplay.

8.4.1 PAD4 inhibitor in cancer therapy.

In recent years, we have seen profound developments in the pharmacology of citrullination, and several potent pan-PAD and isozyme-specific inhibitors have been used to investigate the functions of these enzymes both *in vitro* and *in vivo* (Lewis et al., 2015; Lewis & Nacht, 2016; Y. Wang et al., 2012). With a growing body of evidence suggesting that the deregulation of PADs is integral to many diseases including cancer, the therapeutic use of such small molecule PAD inhibitors is increasingly becoming attractive in the field of oncology. Indeed, some PAD4-specific inhibitors have already been reported to have promising anti-tumour effects in pre-clinical models; the most recent examples include GSK484 which was also used in this study (Wei et al., 2021), and another small molecule PAD4-specific inhibitor called JBI-589 (Deng et al., 2022). The former was recently shown to sensitise triple-negative breast cancer cells to radiotherapy in a mouse xenograft model (Wei et al., 2021), whilst the latter reportedly suppresses primary tumour growth in mouse models via the downregulation of a chemokine receptor, CXCR2, in neutrophils (Deng et al., 2022). Further research is required to better understand the molecular interactions of PAD4 with cancer-related pathways and signalling cascades, and continued development of potent inhibitors could improve cancer treatment options and outcomes.

Similarly, the protein arginine methyltransferase (PRMT) family of proteins have attracted scientific interest during the past decade, underlined by their deregulated expression in multiple types of cancer, important roles in various pathways contributing to tumour progression, and thus their potential as druggable therapeutic targets in cancer (Y. Yang & Bedford, 2013). The anti-cancer therapeutic use of PRMT5-specific

inhibitors has been widely explored, some of which have already shown encouraging preliminary clinical efficacy (Feustel & Falchook, 2022). Recent work from our lab demonstrated that a PRMT5-specific inhibitor, T1-44, in combination with the TGF-β1 signalling inhibitor Vactosertib, can significantly reduce tumour size and improve the long-term survival of mouse models (Hong et al., 2023). In addition, a few other PRMT5 inhibitors including AMG193 and GSK3326595 have already entered clinical trials for use in cancer therapy (Amgen, 2023; Siu et al., 2019).

In this study, we highlighted a potential crosstalk between PAD4 and PRMTs on E2F1-dependent activities (Figure 8-2), whereby we observed the enhanced interaction of E2F1 with p100/TSN in cells over-expressing PAD4 (Figure 5-1). Previously, the p100/TSN protein was identified as a reader of the PRMT5-driven arginine methylation marks on E2F1 at R111 and R113 (Zheng et al., 2013). This methylation event was known to be hindered by the activity of another methyltransferase, PRMT1, which targets the R109 residue of E2F1 in DNA-damaged cells and promotes E2F1-dependent apoptosis (Zheng et al., 2013). In proliferative cells, the binding of cyclin A to E2F1 impedes PRMT1-mediated arginine methylation and allows PRMT5 primarily acts to methylate E2F1 to channel its activity into pro-growth pathways (Cho et al., 2012; Zheng et al., 2013). Citrullination and arginine methylation competitively inhibit one another by targeting the same arginine residue (Cuthbert et al., 2004; Hagiwara et al., 2005), and PAD4 and PRMT1 share the target arginine residue (R109) of E2F1 (Ghari et al., 2016; Roworth et al., 2019). This probably means that our observation of the increased interaction between E2F1 and p100/TSN with ectopic PAD4 expression might be a result of the citrullination-methylation interplay; PAD4 may inhibit PRMT1-mediated methylation by citrullinating R109 residue, which leads to the de-repression of PRMT5 activity catalysing arginine

methylation at R111/R113, leading to recruitment of the reader protein p100/TSN (Figure 8-1C).

Given that PRMT5-driven methylation favours cell growth and proliferation (Cho et al., 2012; Zheng et al., 2013), and PRMT1 promotes E2F1-dependent apoptosis (Zheng et al., 2013), it can be hypothesised that PAD4 might contribute to cancer progression via PRMT5-regulated pathways and overcome PRMT1-mediated apoptosis by competitively citrullinating R109 residue of E2F1. This study therefore proposes PAD4 and its crosstalk with PRMTs as a potential attractive target pathway for cancer therapy. For example, it might be of great interest to explore the potential use of PAD4 and PRMT5 inhibitors in combination, as this co-treatment would synergistically relieve the suppression of PRMT1 activity and prompt E2F1-dependent apoptosis in theory. However, it is noteworthy that previous work in our lab also suggested that, although the predominant citrullination sites of E2F1 were R109 and R127, arginine residues R111 and R113 (the residues targeted by PRMT5) could also be targeted for PAD4-directed citrullination (Ghari et al., 2016). This indicates that an intricate, context-dependent relationship must exist between PAD4 and PRMTs over E2F1 activity. The substrate specificities and preferences of PADs, including PAD4, have not been clearly defined to date (Christophorou, 2022; Darrah et al., 2012; Knuckley et al., 2010), and further investigation in this area, such as the identification of consensus target motifs, would be essential for characterisation and potential therapeutic use of the co-inhibition of PAD4 and PRMT5.

8.4.2 Therapeutic relevance of alternative splicing.

An ever-increasing number of publications suggests that AS disturbances are intricately connected with cancer occurrence and progression, which highlights the potential for splicing-based therapeutics to be used in the treatment of cancer (Bonnal et al., 2020; Gallego-Paez et al., 2017). Multiple methods, including the use of small molecule inhibitors and splice-switching antisense oligonucleotides (SSOs), are currently implemented to modulate splicing activity in cancer cells (Bonnal et al., 2020; Gallego-Paez et al., 2017).

In this study, it was demonstrated that SRSF3 plays a significant role in the E2F1/PAD4-regulated splicing process. SRSF3 is a splicing factor that has been found to be over-expressed in various types of tumours, including breast cancer, ovarian cancer, retinoblastoma, gastric cancer, colorectal cancer, and hepatocellular carcinoma (Iborra et al., 2013; Ke et al., 2018; Park & Jeong, 2016; Villegas et al., 2014; H. Wang et al., 2019). This SR protein is therefore amongst the candidate splicing factors implicated as a potential therapeutic target (Z. Zhou et al., 2020). Interestingly, SRSF3 has been linked with the anti-tumour effects of some drugs whose repositioned use in cancer therapy are explored; such as the anti-arrhythmic drug Amiodarone (Y.-L. Chang et al., 2018) and a potassium-sparing diuretic Amiloride (J.-G. Chang et al., 2011). Direct targeting of SRSF3 has also been examined; for example, a recent study demonstrated that the SRSF3-specific inhibitor SFI003 can exhibit anti-tumour effects in suppressing growth, inducing apoptosis and changing metastatic profiles of colorectal cancer cells (Y. Zhang et al., 2022). The authors discovered such impacts of the SRSF3 inhibitor treatment were caused by the suppression of a cholesterol biosynthesis-related gene called DHCR24, whose expression was regulated by SRSF3-dependent splicing (Y. Zhang et al., 2022).

Downregulation of DHCR24 led to a significant increase in the level of reactive oxygen species (ROS), which resulted in the activation of apoptotic pathways in cells (Y. Zhang et al., 2022).

In addition, there is a growing body of evidence to suggest that cancer-related splicing abnormalities may generally sensitise tumour cells to therapeutic interventions of splicing (S. C.-W. Lee et al., 2016; Obeng et al., 2016; Seiler et al., 2018; Shirai et al., 2017). For example, leukaemia cells with SRSF2 mutation were demonstrated to display greater sensitivity to inhibitors that target splicing network components, such as the SF3B complex, in comparison to wild-type cells (Fong et al., 2019; S. C.-W. Lee et al., 2016). Interestingly, it was observed that PRMT inhibitors to block arginine methylation pathways were particularly effective in killing leukaemia cells with SRSF2 mutations, both *in vitro* and *in vivo*, and such response is partially explained by the fact that PRMTs primarily target RNA-binding proteins that are known to be involved in regulating splicing (Fong et al., 2019). Furthermore, the combination treatment of PRMT inhibitors (like the PRMT-specific inhibitor GSK591 and Type I PRMT inhibitor MS023) with other drugs targeting the spliceosome, such as the SF3B inhibitor E7107, was shown to reduce tumour growth and improved survival in murine models in a synergistic manner (Fong et al., 2019). This result signals an immense potential for co-treatment with compounds targeting distinct components of the splicing network as means of therapeutic strategy for cancer with AS disturbances (Harada et al., 2023). Therefore, given the fact that many RNA-binding proteins are substrates of methylation and citrullination (C. Chen et al., 2011; Fong et al., 2019; C.-Y. Lee et al., 2018; Lewallen et al., 2015; W.-J. Li et al., 2021; Lim, Lee, et al., 2020; Lim, Park, et al., 2020; Musiani et al., 2019; Radzisheuskaya et al., 2019; Tanikawa et al., 2018), it would be reasonable to anticipate

that the combination treatment of PAD inhibitor with PRMT inhibitor or other compounds targeting splicing machinery may have a synergistic anti-tumour effect in cells with AS abnormalities. Successful therapeutic use of a co-treatment strategy involving both PADs and PRMTs would require significant advancements in the field. This would entail further understanding of the molecular interplay between PADs and PRMTs, precise characterisation of the role of PADs in regulating the activity of splicing factors, and continued efforts towards the development of potent pan-PAD or PAD4-specific inhibitors that are safe and effective for use in patients.

8.4.3 Therapeutic implications of E2F1/PAD4-driven AS regulation.

Future studies also need to elucidate the biological consequences of the described E2F1/PAD4-regulated AS events. The spliced genes identified in our RNA-seq contained many implicated in cancer occurrence and progression. For example, EXOC4 (also known as Sec8), one of the validated AS targets from our RNA-seq (Figure 4-4), is encoding a component of the Exocyst complex (Martin-Urdiroz et al., 2016; Tanaka et al., 2017). This evolutionary-conserved protein machinery is known to regulate exocytosis, the process of transporting intracellular contents such as hormones, neural transmitters, RNAs, and proteins into the extracellular environment, by mediating the tethering of secretory vesicles to the plasma membrane (Martin-Urdiroz et al., 2016). A wide spectrum of cellular events, including cell polarity, migration, cell cycle progression, and apoptosis, as well as various diseases including cancer, are reportedly influenced by Exocyst-dependent exocytosis (Ashktorab et al., 2010; J. Liu et al., 2009; Sakurai-Yageta et al., 2008; Tanaka et al., 2012, 2016, 2017). Even though the underlying mechanisms remain unclear, EXOC4 appears to have a pivotal role in the functionality and regulation of this protein complex, whereby the knockdown of EXOC4 has a significant impact on

apoptosis, cell cycle arrest, and DNA repair (Tanaka et al., 2016; Tanaka & Iino, 2014; Torres et al., 2015). More recently, a study proposed that EXOC4 mediates the repair of the plasma membrane of cancer cells through exocytosis, resulting in resistance to antibody-dependent cellular cytotoxicity (ADCC) by neutrophils (van Rees et al., 2022). Indeed, an enhanced level of neutrophil-dependent killing of cancer cells was observed with EXOC4 knockdown, suggesting the potential of EXOC4 as a therapeutic target or clinical biomarker for antibody therapy in the context of neutrophil trogocytosis (van Rees et al., 2022). Interestingly, a considerable link between PAD4 and exocytosis has also been reported (Kholia et al., 2015; Kosgodage et al., 2018). For example, a treatment of prostate cancer cells with the PAD4 inhibitor Cl-amidine was found to significantly reduce the release of macrovesicles (Kholia et al., 2015). Later, the same inhibitor affected the biogenesis and release of extracellular vesicles in glioblastoma multiforme cells (Kosgodage et al., 2018). Taken together, it would be valuable for future research to examine the biochemical and functional differences of EXOC4 protein isoforms that result from exon 7 skipping or inclusion as highlighted in our RNA-seq experiment. Particularly, investigating the exon 7 inclusion isoform, which is induced by the GSK484 treatment in cancer cells, might prove important in establishing a therapeutic strategy based on this PAD4-specific inhibitor via the EXOC4 pathway. This is currently an active topic of research in our lab.

Other validated spliced genes were also shown to have important roles in many biological processes and pathologies. For example, RBM25, whose exon 2 skipping is regulated by E2F1 and PAD4, was demonstrated to have a tumour-suppressive effect in acute myeloid leukaemia (AML) cells through splicing regulation of genes such as the apoptotic regulator BCL-X and the MYC inhibitor BIN1 (Y. Ge et al., 2019). Since PAD4 is usually

found expressed at high levels in immune cells and in AML, it will also be interesting to examine the function and activity of different E2F1/PAD4-induced isoforms of RBM25 (G. Song et al., 2015).

8.5 Conclusion

In conclusion, this study has demonstrated that the interplay between E2F1 and PAD4 has a significant genome-wide impact at the level of alternative splicing in colorectal cancer cells. This regulation is involved with PAD4 activity enhancing the interaction of E2F1 with splicing machinery components including p100/TSN and SRSF3, and results in AS changes of genes, the majority of which represent poor E2F1 transcriptional targets. Furthermore, we revealed the pivotal role that PAD4 undertakes in regulating the RNA binding profile of SRSF3 in an E2F1-dependent manner, which seems to underpin a key regulatory mechanism in the E2F1/PAD4/SRSF3 pathway.

Future work in this area should focus on further elucidating the molecular functions of SRSF3 in splicing regulation within and outside of the E2F1/PAD4 axis. Techniques such as RIP-seq and TRIBE could explore its target RNA preferences and potential dual role as an activator and suppressor of alternative splicing, as well as examining how the activity of E2F1 and PAD4 influence these aspects of SRSF3 biology. Additionally, selecting cell lines or systems that exhibit high expression of PAD4, and treating them with a calcium ionophore, may enable a more pronounced visualisation of the impact of citrullination on the cancer transcriptome at both splicing and transcriptional levels, and provide important implications for its pathological or cancer relevance.

Furthermore, it is crucial to develop a deeper understanding of the E2F1/PAD4 crosstalk, particularly with respect to the biological outcomes that arise from splicing perturbations. Comparative functional analysis of distinct splice variants resulting from E2F1/PAD4-dependent splice events, as well as further characterisation of citrullination-methylation crosstalk on E2F1 and RBPs, will provide additional insights into the potential of targeting PAD4 for cancer treatment. A growing body of research suggests that targeting

splicing in cancer cells with alternative splicing disturbances can have therapeutic benefits, and the relevance of PADs and PRMTs in regulating RNA-binding proteins is increasingly becoming evident. Therefore, continued investigation into the therapeutic use of PAD4 inhibitors is a highly topical area in oncological clinical research, with the perspectives of being combined with splice-targeted therapy or PRMT inhibitor treatment.

Supplementary Information

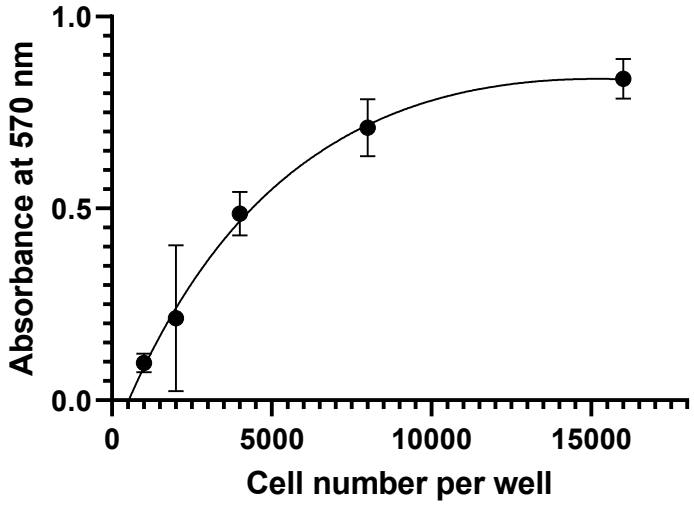
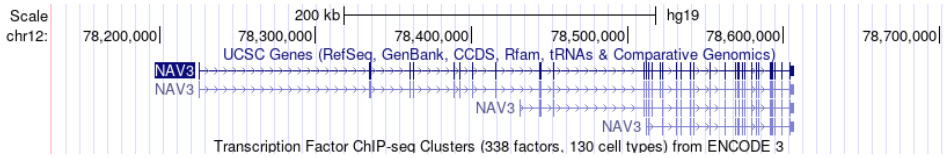


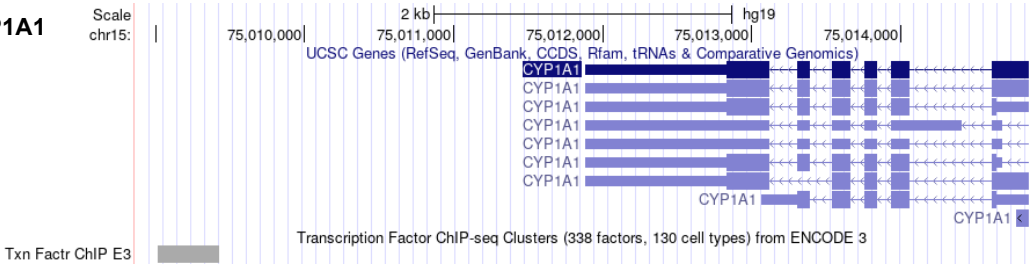
Figure S-1 MTT assay to determine cell numbers to seed.

Between 1000 and 16000 HCT116 WT cells were seeded in each well of a 96-well plate and incubated for 72 hours. The result indicated that 3000 cells per well are optimum for the cell line where it did not reach the maximum confluency in the tested incubation periods.

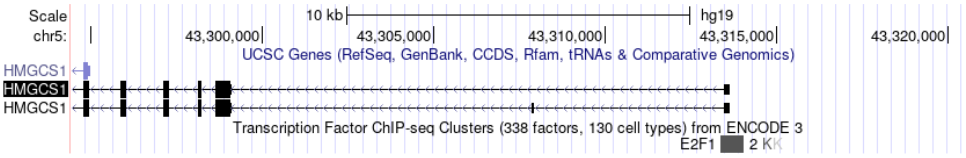
NAV3



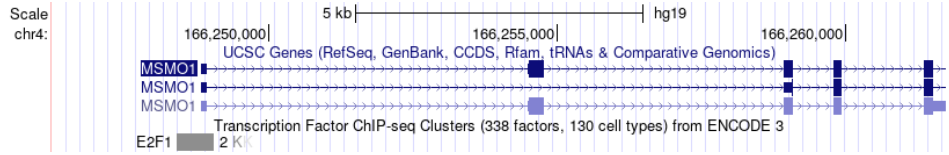
CYP1A1



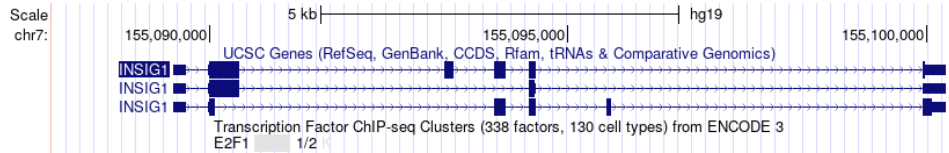
HMGCS1



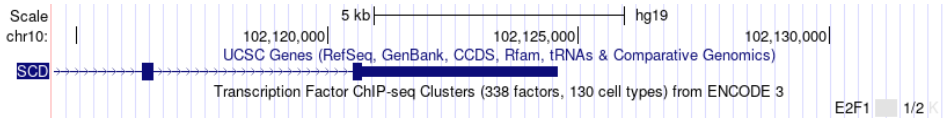
MSMO1



INSIG1



SCD



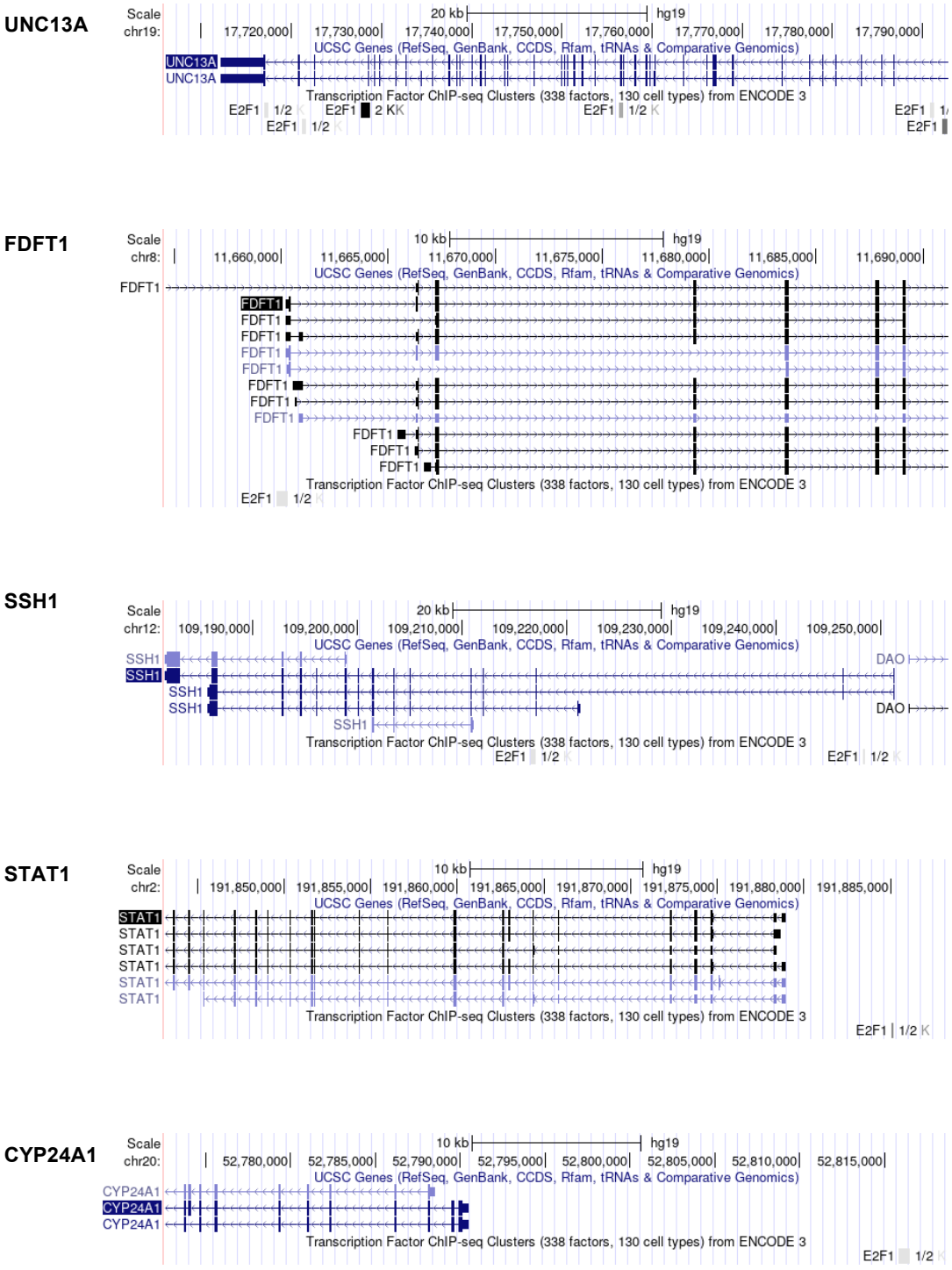


Figure S-2 DEGs found in WT-GSK484 are mostly E2F1 targets.

Screenshots from UCSC Genome Browser (http://genome.ucsc.edu, GRCh37/h19 assembly) (Kent et al., 2002) illustrating the gene structures of DEGs found in WT-GSK484 with statistically significant adjusted *P* (P_{adj}) values (< 0.01) (None of them made 2-fold change cut-off). With indicated E2F1 ChIP-seq peaks (grey boxes) from the data deposited on the ENCODE project (http://genome.ucsc.e du/ENCODE/).

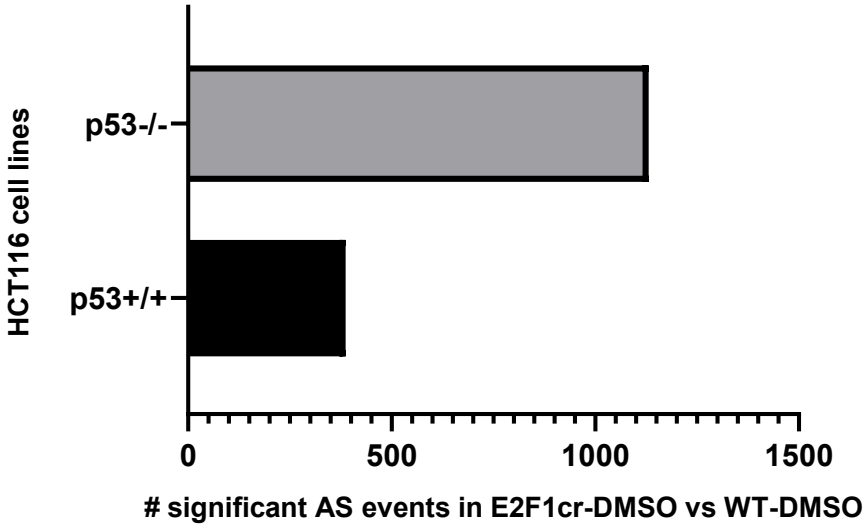


Figure S-3 HCT116 p53-/- cells might be more sensitive to splicing perturbations.

Numbers of significant AS events (FDR < 0.01) in the E2F1cr-DMSO condition with respect to the control WT-DMSO condition in HCT116 p53+/+ and p53-/- cell lines.

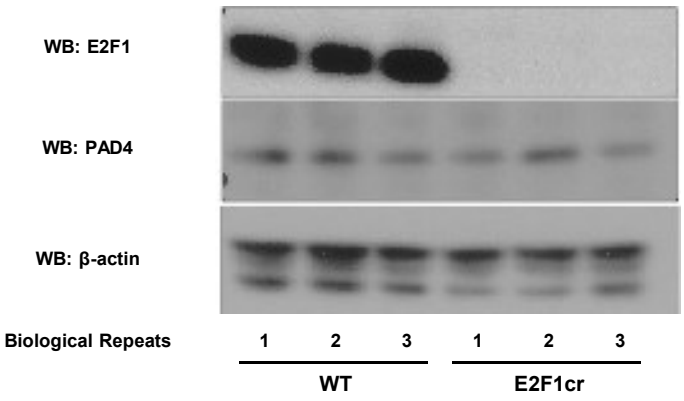


Figure S-4 E2F1 and PAD4 expression in HCT116 p53-/- cell lines

Immunoblot showing the expression of E2F1, PAD4 and β-actin in HCT116 p53-/- WT and E2F1cr cell lines. *N = 3*. Performed in collaboration with Dr Amit Shrestha.

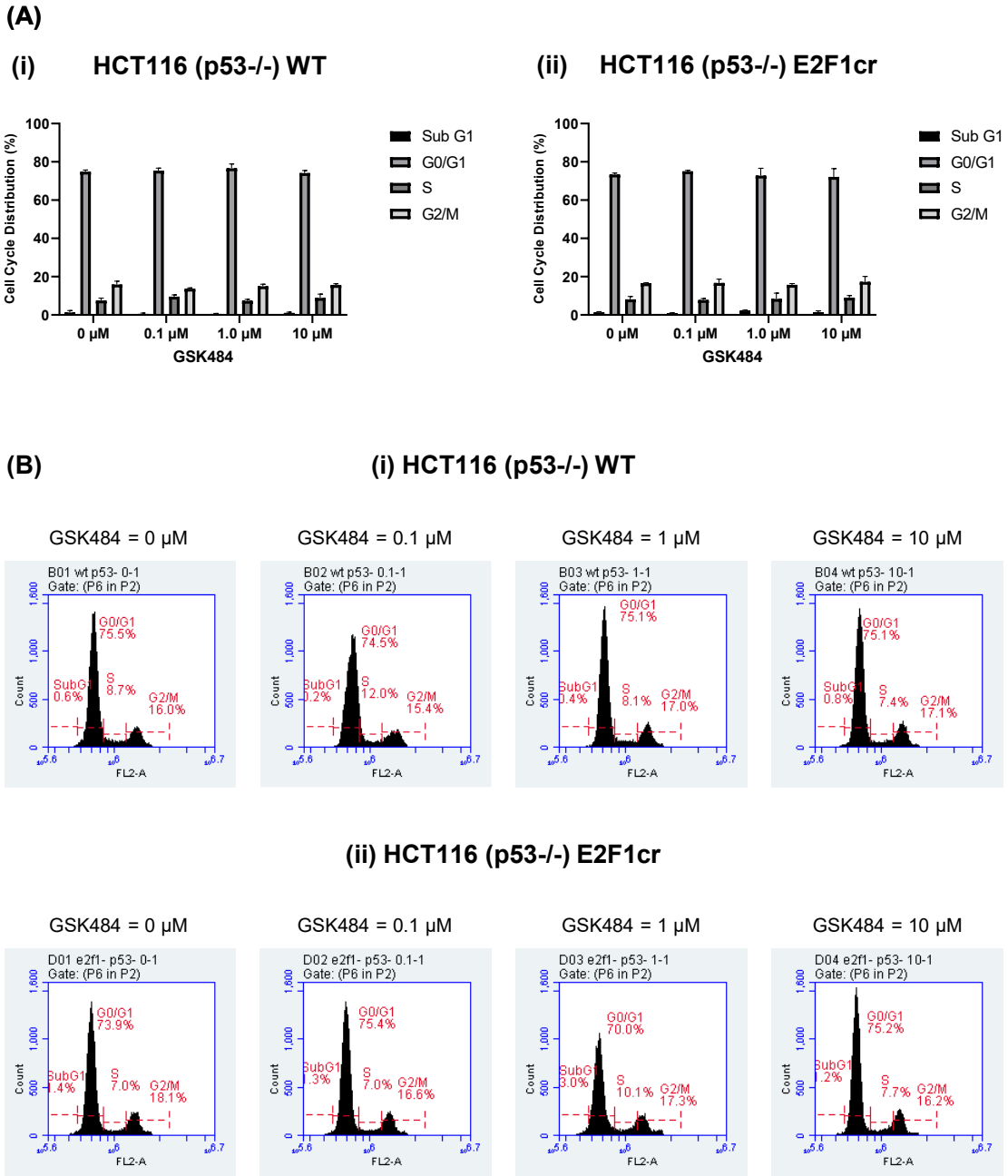


Figure S-5 FACS cell cycle profile analysis for HCT116 p53-/- cell lines.

Flow cytometric analysis using propidium iodide to check the cell cycle states of HCT116 p53-/- cell lines treated with various concentrations of the PAD4 inhibitor GSK484. (A) Cell cycle distributions for (i) WT and (ii) E2F1cr cells. *N* = 3. (± S.D.) (B) FACS cell cycle profiles for HCT116 p53-/- (i) WT and (ii) E2F1cr cells. A representative example from *N* = 3.

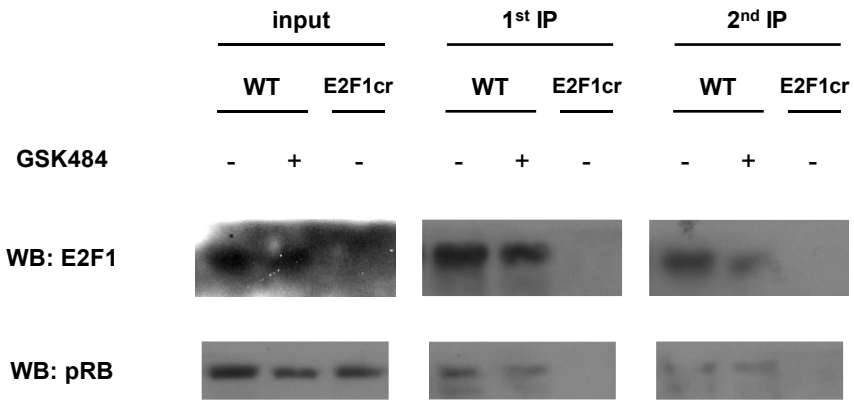


Figure S-6 IP validation for IP-MS experiment for the E2F1 interactome

Immunoblot showing the successful immunoprecipitation of E2F1 for the IP-MS experiment for the E2F1 interactome. *N* = 3. Samples were IP-ed twice to make sure E2F1 proteins were thoroughly pulled-down. The blot of pRB as a positive control.

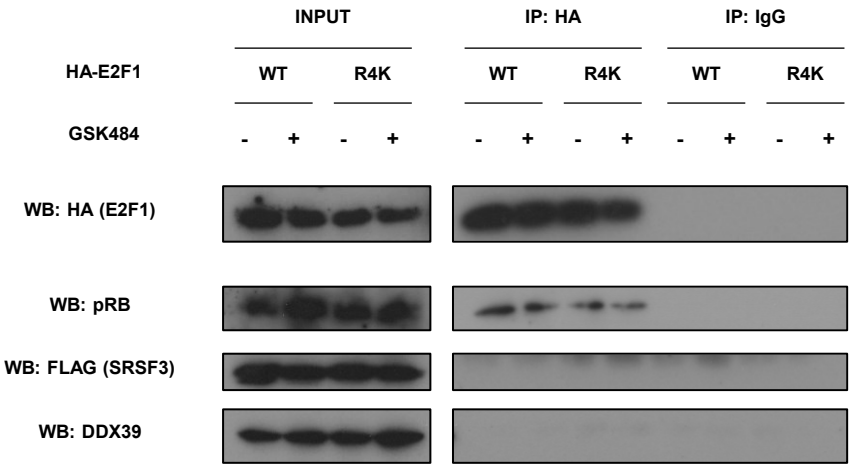


Figure S-7 Immunoblot to examine the binding of DDX39A to E2F1.

HCT116 p53-/- cells were transfected with 2 µg FLAG-SRSF3 and HA-E2F1^{WT} or HA-E2F1^{R4K} and treated with 10 µM GSK484 (or DMSO for a negative control) and immunoprecipitated using an anti-HA antibody. *N* = 2. The successful pull-down of HA-E2F1$^{WT/R4K}$ and a known binding partner, pRB, was confirmed, but no binding of DDX39 or SRSF3 was observed.

Table S-1 List of DEGs found in WT-GSK484.

Symbol	Ensembl ID	log2FoldChange	P_{adj} value	E2F1 target
NAV3	ENSG00000067798	0.946	0.00027	No
CYP1A1	ENSG00000140465	0.934	0.00242	Yes
HMGCS1	ENSG00000112972	0.812	0.00182	Yes
MSMO1	ENSG00000052802	0.797	0.00356	Yes
INSIG1	ENSG00000186480	0.764	0.00107	Yes
SCD	ENSG00000099194	0.630	0.00044	Yes
UNC13A	ENSG00000130477	0.509	0.00027	Yes
FDFT1	ENSG00000079459	0.489	0.00027	Yes
SSH1	ENSG00000084112	0.432	0.00110	Yes
STAT1	ENSG00000115415	-0.489	0.00356	Yes
CYP24A1	ENSG00000019186	-0.553	0.00397	Yes

DEGs found in WT-GSK484 with statistically significant adjusted *P* (P_{adj}) values (< 0.01). None of them made 2-fold change cut-off. 10 out of 11 genes were identified as known E2F1 targets through E2F1 ChIP-seq analysis using dataset deposited on the ENCODE project (http://genome.ucsc.edu/ENCODE/).

References

Adams, J. M., & Cory, S. (2007). The Bcl-2 apoptotic switch in cancer development and therapy. *Oncogene*, *26*(9), 1324–1337. https://doi.org/10.1038/sj.onc.1210220

Amgen. (2023). AMG 193, Methylthioadenosine (MTA) Cooperative Protein Arginine Methyltransferase 5 (PRMT5) Inhibitor, Alone and in Combination With Docetaxel in Advanced Methylthioadenosine Phosphorylase (MTAP)-Null Solid Tumors (MTAP). *Clinicaltrials.Gov*. https://clinicaltrials.gov/ct2/show/NCT05094336

Andrade, F., Darrah, E., Gucek, M., Cole, R. N., Rosen, A., & Zhu, X. (2010). Autocitrullination of human peptidyl arginine deiminase type 4 regulates protein citrullination during cell activation. *Arthritis & Rheumatism*, *62*(6), 1630–1640. https://doi.org/10.1002/art.27439

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*(48), 7632–7641. https://doi.org/10.1038/sj.onc.1206840

Arita, K., Hashimoto, H., Shimizu, T., Nakashima, K., Yamada, M., & Sato, M. (2004). Structural basis for Ca-induced activation of human PAD4. *Nature Structural & Molecular Biology*, *11*(8), Article 8. https://doi.org/10.1038/nsmb799

Asaga, H., Yamada, M., & Senshu, T. (1998). Selective Deimination of Vimentin in Calcium Ionophore-Induced Apoptosis of Mouse Peritoneal Macrophages. *Biochem. Biophys. Res. Commun.*, *243*, 641–646. https://doi.org/10.1006/bbrc.1998.8148

Ashktorab, H., Schäffer, A. A., Daremipouran, M., Smoot, D. T., Lee, E., & Brim, H. (2010). Distinct Genetic Alterations in Colorectal Cancer. *PLoS ONE*, *5*(1). https://doi.org/10.1371/journal.pone.0008879

Ayrault, O., Andrique, L., & Séité, P. (2006). Involvement of the transcriptional factor E2F1 in the regulation of the rRNA promoter. *Experimental Cell Research*, *312*(7), 1185–1193. https://doi.org/10.1016/j.yexcr.2006.01.027

Bachand, F., & Silver, P. A. (2004). PRMT3 is a ribosomal protein methyltransferase that affects the cellular levels of ribosomal subunits. *The EMBO Journal*, *23*(13), 2641–2650. https://doi.org/10.1038/sj.emboj.7600265

Baldini, L., & Labialle, S. (2021). Using Native RIP, UV-CLIP or fCLIP to Address Protein-RNA Interactions In Vivo. *Methods in Molecular Biology (Clifton, N.J.)*, *2300*, 89–98. https://doi.org/10.1007/978-1-0716-1386-3_9

Bandara, L., Lam, E., Sørensen, T., Zamanian, M., Girling, R., & La Thangue, N. (1994). DP-1 a cell cycle-regulated and phosphorylated component of transcription factor DRTF1-E2F which is functionally important for recognition by pRb and the adenovirus E4 orf 6-7 protein. *The EMBO Journal*, *13*(13), 3104–3114. https://doi.org/10.1002/j.1460-2075.1994.tb06609.x

Bandara, L. R., Buck, V. M., Zamanian, M., Johnston, L. H., & La Thangue, N. B. (1993). Functional synergy between DP-1 and E2F-1 in the cell cycle-regulating transcription factor DRTF1-E2F. *The EMBO Journal*, *12*(11), 4317–4324. https://doi.org/10.1002/j.1460-2075.1993.tb06116.x

Baralle, F., & Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nature Reviews. Molecular Cell Biology*, *18*(7). https://doi.org/10.1038/nrm.2017.27

Barczak, W., Carr, S. M., Liu, G., Munro, S., Nicastri, A., Lee, L. N., Hutchings, C., Ternette, N., Klenerman, P., Kanapin, A., Samsonova, A., & La Thangue, N. B. (2023). Long non-coding RNA-derived peptides are immunogenic and drive a potent anti-tumour response. *Nature Communications*, *14*(1), Article 1. https://doi.org/10.1038/s41467-023-36826-0

Barczak, W., Jin, L., Carr, S. M., Munro, S., Ward, S., Kanapin, A., Samsonova, A., & La Thangue, N. B. (2020). PRMT5 promotes cancer cell migration and invasion through the E2F pathway. *Cell Death & Disease*, *11*(7), 572. https://doi.org/10.1038/s41419-020-02771-9

Barker, N., Ridgway, R. A., van Es, J. H., van de Wetering, M., Begthel, H., van den Born, M., Danenberg, E., Clarke, A. R., Sansom, O. J., & Clevers, H. (2009). Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature*, *457*(7229), Article 7229. https://doi.org/10.1038/nature07602

Bates, S., Phillips, A. C., Clark, P. A., Stott, F., Peters, G., Ludwig, R. L., & Vousden, K. H. (1998). p14ARF links the tumour suppressors RB and p53. *Nature*, *395*(6698), 124–125. https://doi.org/10.1038/25867

Bedard, K. M., Daijogo, S., & Semler, B. L. (2007). A nucleo-cytoplasmic SR protein functions in viral IRES-mediated translation initiation. *The EMBO Journal*, *26*(2), 459–467. https://doi.org/10.1038/sj.emboj.7601494

Berget, S. M., Moore, C., & Sharp, P. A. (1977). Spliced segments at the 5 terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences*, *74*(8), 3171–3175. https://doi.org/10.1073/pnas.74.8.3171

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*(5), 755–764. https://doi.org/10.1038/cdd.2013.6

Bertoli, C., Skotheim, J. M., & Bruin, R. A. M. de. (2013). Control of cell cycle transcription during G1 and S phases. *Nature Reviews. Molecular Cell Biology*, *14*(8), 518. https://doi.org/10.1038/nrm3629

Biamonti, G., Infantino, L., Gaglio, D., & Amato, A. (2020). An Intricate Connection between Alternative Splicing and Phenotypic Plasticity in Development and Cancer. *Cells*, *9*(1), Article 1. https://doi.org/10.3390/cells9010034

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*(5), 595–605. https://doi.org/10.1101/gr.4887606

Birch, J., & Gil, J. (2020). Senescence and the SASP: many therapeutic avenues. *Genes & Development*, *34*(23–24), 1565–1576. https://doi.org/10.1101/gad.343129.120

Biron, B. M., Chung, C.-S., O'Brien, X. M., Chen, Y., Reichner, J. S., & Ayala, A. (2017). Cl-Amidine Prevents Histone 3 Citrullination and Neutrophil Extracellular Trap Formation, and Improves Survival in a Murine Sepsis Model. *Journal of Innate Immunity*, *9*(1), 22–32. https://doi.org/10.1159/000448808

Biswas, A. K., & Johnson, D. G. (2012). Transcriptional and nontranscriptional functions of E2F1 in response to DNA damage. *Cancer Research*, *72*(1), 13–17. https://doi.org/10.1158/0008-5472.CAN-11-2196

Blanc, R. S., & Richard, S. (2017). Arginine Methylation: The Coming of Age. *Molecular Cell*, *65*(1), 8–24. https://doi.org/10.1016/j.molcel.2016.11.003

Blanchet, E., Annicotte, J.-S., Lagarrigue, S., Aguilar, V., Clapé, C., Chavey, C., Fritz, V., Casas, F., Apparailly, F., Auwerx, J., & Fajas, L. (2011). E2F transcription factor-1 regulates oxidative metabolism. *Nature Cell Biology*, *13*(9), 1146–1152. https://doi.org/10.1038/ncb2309

Blasco, M. A. (2005). Telomeres and human disease: Ageing, cancer and beyond. *Nature Reviews Genetics*, *6*(8), 611–622. https://doi.org/10.1038/nrg1656

Bonnal, S. C., López-Oreja, I., & Valcárcel, J. (2020). Roles and mechanisms of alternative splicing in cancer - Implications for care. *Nature Reviews Clinical Oncology*, *17*(8), Article 8. https://doi.org/10.1038/s41571-020-0350-x

Brabletz, T., Kalluri, R., Nieto, M. A., & Weinberg, R. A. (2018). EMT in cancer. *Nature Reviews Cancer*, *18*(2), Article 2. https://doi.org/10.1038/nrc.2017.118

Bradley, T., Cook, M. E., & Blanchette, M. (2015). SR proteins control a complex network of RNA-processing events. *Rna*, *21*(1), 75–92. https://doi.org/10.1261/rna.043893.113

Brown, R. L., Reinke, L. M., Damerow, M. S., Perez, D., Chodosh, L. A., Yang, J., & Cheng, C. (2011). CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. *The Journal of Clinical Investigation*, *121*(3), 1064–1074. https://doi.org/10.1172/JCI44540

Budhavarapu, V., White, E., Mahanic, C., Chen, L., Lin, F.-T., & Lin, W.-C. (2012). Regulation of E2F1 by APC-CCdh1 via K11 linkage-specific ubiquitin chain formation. *Cell Cycle*, *11*(10), 2030–2038. https://doi.org/10.4161/cc.20643

Buratti, E., Stuani, C., De Prato, G., & Baralle, F. E. (2007). SR protein-mediated inhibition of CFTR exon 9 inclusion: Molecular characterization of the intronic splicing silencer. *Nucleic Acids Research*, *35*(13), 4359–4368. https://doi.org/10.1093/nar/gkm444

Burjoski, V., & Reddy, A. S. N. (2021). The Landscape of RNA-Protein Interactions in Plants: Approaches and Current Status. *International Journal of Molecular Sciences*, *22*(6), Article 6. https://doi.org/10.3390/ijms22062845

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*(1), 245–255. https://doi.org/10.1016/j.jmb.2013.09.031

Cáceres, J. F., Stamm, S., Helfman, D. M., & Krainer, A. R. (1994). Regulation of Alternative Splicing in Vivo by Overexpression of Antagonistic Splicing Factors. *Science*, *265*(5179), 1706–1709. https://doi.org/10.1126/science.8085156

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*(14), 11985. https://doi.org/10.1074/jbc.M110.217158

Carlson, S. M., Soulette, C. M., Yang, Z., Elias, J. E., Brooks, A. N., & Gozani, O. (2017). RBM25 is a global splicing factor promoting inclusion of alternatively spliced exons and is itself regulated by lysine mono-methylation. *Journal of Biological Chemistry*, *292*(32), 13381–13390. https://doi.org/10.1074/jbc.M117.784371

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*(5), 900–912. https://doi.org/10.1128/MCB.06286-11

Carr, S. M., Poppy Roworth, A., Chan, C., & La Thangue, N. B. (2015). Post-translational control of transcription factors: Methylation ranks highly. *The FEBS Journal*, *282*(23), 4450–4465. https://doi.org/10.1111/febs.13524

Chang, J.-G., Yang, D.-M., Chang, W.-H., Chow, L.-P., Chan, W.-L., Lin, H.-H., Huang, H.-D., Chang, Y.-S., Hung, C.-H., & Yang, W.-K. (2011). Small molecule amiloride modulates oncogenic RNA alternative splicing to devitalize human cancer cells. *PloS One*, *6*(6), e18643. https://doi.org/10.1371/journal.pone.0018643

Chang, X., & Han, J. (2006). Expression of Peptidylarginine Deiminase Type 4 (PAD4) in various Tumours. *Molecular Carcinogenesis*, *45*, 183–196. https://doi.org/10.1002/mc.20169

Chang, X., Han, J., Pang, L., Zhao, Y., Yang, Y., & Shen, Z. (2009). Increased PADI4 expression in blood and tissues of patients with malignant tumors. *BMC Cancer*, *9*, 40. https://doi.org/10.1186/1471-2407-9-40

Chang, Y.-L., Liu, S.-T., Wang, Y.-W., Lin, W.-S., & Huang, S.-M. (2018). Amiodarone promotes cancer cell death through elevated truncated SRSF3 and downregulation of miR-224. *Oncotarget*, *9*(17), 13390–13406. https://doi.org/10.18632/oncotarget.24385

Charlet-B, N., Logan, P., Singh, G., & Cooper, T. A. (2002). Dynamic antagonism between ETR-3 and PTB regulates cell type-specific alternative splicing. *Molecular Cell*, *9*(3), 649–658. https://doi.org/10.1016/s1097-2765(02)00479-3

Chasin, L. A. (2007). Searching for splicing motifs. *Advances in Experimental Medicine and Biology*, *623*, 85–106. https://doi.org/10.1007/978-0-387-77374-2_6

Chen, B., Wen, P., Hu, G., Gao, Y., Qi, X., Zhu, K., Chen, S., Wu, L., Xu, A., & Zhao, G. (2020). Antagonizing CDK8 Sensitizes Colorectal Cancer to Radiation Through Potentiating the Transcription of e2f1 Target Gene apaf1. *Frontiers in Cell and Developmental Biology*, *8*. https://doi.org/10.3389/fcell.2020.00408

Chen, C., Nott, T. J., Jin, J., & Pawson, T. (2011). Deciphering arginine methylation: Tudor tells the tale. *Nature Reviews Molecular Cell Biology*, *12*(10), Article 10. https://doi.org/10.1038/nrm3185

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. *The EMBO Journal*, *31*(7), 1785–1797. https://doi.org/10.1038/emboj.2012.17

Chow, L. T., Gelinas, R. E., Broker, T. R., & Roberts, R. J. (1977). An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell*, *12*(1), 1–8. https://doi.org/10.1016/0092-8674(77)90180-5

Christofk, H. R., Vander Heiden, M. G., Harris, M. H., Ramanathan, A., Gerszten, R. E., Wei, R., Fleming, M. D., Schreiber, S. L., & Cantley, L. C. (2008). The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature*, *452*(7184), Article 7184. https://doi.org/10.1038/nature06734

Christophorou, M. A. (2022). The virtues and vices of protein citrullination. *Royal Society Open Science*, *9*(6), 220125. https://doi.org/10.1098/rsos.220125

Christophorou, M. A., Castelo-Branco, G., Halley-Stott, R. P., Oliveira, C. S., Loos, R., Radzisheuskaya, A., Mowen, K. A., Bertone, P., Silva, J. C. R., Zernicka-Goetz, M., Nielsen, M. L., Gurdon, J. B., & Kouzarides, T. (2014). Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature*, *507*(7490), Article 7490. https://doi.org/10.1038/nature12942

Chumanevich, A. A., Causey, C. P., Knuckley, B. A., Jones, J. E., Poudyal, D., Chumanevich, A. P., Davis, T., Matesic, L. E., Thompson, P. R., & Hofseth, L. J. (2011). Suppression of colitis in mice by Cl-amidine: A novel peptidylarginine deiminase inhibitor. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, *300*(6), G929–G938. https://doi.org/10.1152/ajpgi.00435.2010

Ciesielski, O., Biesiekierska, M., Panthu, B., Soszyński, M., Pirola, L., & Balcerczyk, A. (2022). Citrullination in the pathology of inflammatory and autoimmune disorders: Recent advances and future perspectives. *Cellular and Molecular Life Sciences: CMLS*, *79*(2), 94. https://doi.org/10.1007/s00018-022-04126-3

Clapham, D. E. (2007). Calcium Signaling. *Cell*, *131*(6), 1047–1058. https://doi.org/10.1016/j.cell.2007.11.028

Clark, F., & Thanaraj, T. A. (2002). Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human. *Human Molecular Genetics*, *11*(4), 451–464. https://doi.org/10.1093/hmg/11.4.451

Coenye, T. (2021). Do results obtained with RNA-sequencing require independent verification? *Biofilm*, *3*, 100043. https://doi.org/10.1016/j.bioflm.2021.100043

Curto, M., Cole, B. K., Lallemand, D., Liu, C.-H., & McClatchey, A. I. (2007). Contact-dependent inhibition of EGFR signaling by Nf2-Merlin. *The Journal of Cell Biology*, *177*(5), 893–903. https://doi.org/10.1083/jcb.200703010

Cuthbert, G. L., Daujat, S., Snowden, A. W., Erdjument-Bromage, H., Hagiwara, T., Yamada, M., Schneider, R., Gregory, P. D., Tempst, P., Bannister, A. J., & Kouzarides, T. (2004). Histone deimination antagonizes arginine methylation. *Cell*, *118*(5), 545–553. https://doi.org/10.1016/j.cell.2004.08.020

Darrah, E., Rosen, A., Giles, J. T., & Andrade, F. (2012). Peptidylarginine deiminase 2, 3 and 4 have distinct specificities against cellular substrates: Novel insights into autoantigen selection in rheumatoid arthritis. *Annals of the Rheumatic Diseases*, *71*(1), 92–98. https://doi.org/10.1136/ard.2011.151712

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*(7279), Article 7279. https://doi.org/10.1038/nature08697

Davies, M., & Samuels, Y. (2010). Analysis of the genome to personalize therapy for melanoma. *Oncogene*, *29*(41), 5545–5555. https://doi.org/10.1038/onc.2010.323

de la Luna, S., Burden, M. J., Lee, C.-W., & La Thangue, N. B. (1996). Nuclear accumulation of the E2F heterodimer regulated by subunit composition and alternative splicing of a nuclear localization signal. *Journal of Cell Science*, *109*(10), 2443–2452. https://doi.org/10.1242/jcs.109.10.2443

de la Mata, M., Alonso, C. R., Kadener, S., Fededa, J. P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., & Kornblihtt, A. R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. *Molecular Cell*, *12*(2), 525–532. https://doi.org/10.1016/j.molcel.2003.08.001

de la Mata, M., & Kornblihtt, A. R. (2006). RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nature Structural & Molecular Biology*, *13*(11), Article 11. https://doi.org/10.1038/nsmb1155

de Morrée, A., Droog, M., Moursel, L. G., Bisschop, I. J. M., Impagliazzo, A., Frants, R. R., Klooster, R., & Maarel, S. M. van der. (2012). Self-regulated alternative splicing at the AHNAK locus. *The FASEB Journal*, *26*(1), 93–103. https://doi.org/10.1096/fj.11-187971

de Thé, H. (2018). Differentiation therapy revisited. *Nature Reviews Cancer*, *18*(2), 117–127. https://doi.org/10.1038/nrc.2017.103

DeGregori, J., & Johnson, D. G. (2006). Distinct and Overlapping Roles for E2F Family Members in Transcription, Proliferation and Apoptosis. *Current Molecular Medicine*, *6*(7), 739–748. https://doi.org/10.2174/1566524010606070739

Deng, H., Lin, C., Garcia-Gerique, L., Fu, S., Cruz, Z., Bonner, E. E., Rosenwasser, M., Rajagopal, S., Sadhu, M. N., Gajendran, C., Zainuddin, M., Gosu, R., Sivanandhan, D., Shelef, M. A., Nam, B., Vogl, D. T., Gabrilovich, D. I., & Nefedova, Y. (2022). A Novel Selective Inhibitor JBI-589 Targets PAD4-Mediated Neutrophil Migration to Suppress Tumor Progression. *Cancer Research*, *82*(19), 3561–3572. https://doi.org/10.1158/0008-5472.CAN-21-4045

Dewaele, M., Tabaglio, T., Willekens, K., Bezzi, M., Teo, S. X., Low, D. H. P., Koh, C. M., Rambow, F., Fiers, M., Rogiers, A., Radaelli, E., Al-Haddawi, M., Tan, S. Y., Hermans, E., Amant, F., Yan, H., Lakshmanan, M., Koumar, R. C., Lim, S. T., … Guccione, E. (2016). Antisense

oligonucleotide-mediated MDM4 exon 6 skipping impairs tumor growth. *The Journal of Clinical Investigation*, *126*(1), 68–84. https://doi.org/10.1172/JCI82534

Di Fiore, R., D'Anneo, A., Tesoriere, G., & Vento, R. (2013). RB1 in cancer: Different mechanisms of RB1 inactivation and alterations of pRb pathway in tumorigenesis. *Journal of Cellular Physiology*, *228*(8), 1676–1687. https://doi.org/10.1002/jcp.24329

Dou, Z., Zhao, D., Chen, X., Xu, C., Jin, X., Zhang, X., Wang, Y., Xie, X., Li, Q., Di, C., & Zhang, H. (2021). Aberrant Bcl-x splicing in cancer: From molecular mechanism to therapeutic modulation. *Journal of Experimental & Clinical Cancer Research*, *40*(1), 194. https://doi.org/10.1186/s13046-021-02001-w

Dreyfuss, G., Matunis, M. J., Piñol-Roma, S., & Burd, C. G. (1993). hnRNP proteins and the biogenesis of mRNA. *Annual Review of Biochemistry*, *62*, 289–321. https://doi.org/10.1146/annurev.bi.62.070193.001445

Du, L., Fakih, M. G., Rosen, S. T., & Chen, Y. (2020). SUMOylation of E2F1 Regulates Expression of EZH2. *Cancer Research*, *80*(19), 4212–4223. https://doi.org/10.1158/0008-5472.CAN-20-1259

Duan, G., & Walther, D. (2015). The Roles of Post-translational Modifications in the Context of Protein Interaction Networks. *PLoS Computational Biology*, *11*(2), e1004049. https://doi.org/10.1371/journal.pcbi.1004049

Duan, Q., Pang, C., Chang, N., Zhang, J., & Liu, W. (2016). Overexpression of PAD4 suppresses drug resistance of NSCLC cell lines to gefitinib through inhibiting Elk1-mediated epithelial-mesenchymal transition. *Oncology Reports*, *36*(1), 551–558. https://doi.org/10.3892/or.2016.4780

Dubrez, L. (2017). Regulation of E2F1 Transcription Factor by Ubiquitin Conjugation. *International Journal of Molecular Sciences*, *18*(10), 2188. https://doi.org/10.3390/ijms18102188

Dvinge, H., Kim, E., Abdel-Wahab, O., & Bradley, R. K. (2016). RNA splicing factors as oncoproteins and tumour suppressors. *Nature Reviews Cancer*, *16*(7), Article 7. https://doi.org/10.1038/nrc.2016.51

Dzutsev, A., Badger, J. H., Perez-Chanona, E., Roy, S., Salcedo, R., Smith, C. K., & Trinchieri, G. (2017). Microbes and cancer. *Annual Review of Immunology*, *35*, 199–228. https://doi.org/10.1146/annurev-immunol-051116-052133

Esposito, G., Vitale, A., Leijten, F., Strik, A., Koonen-Reemst, A., Yurttas, P., Robben, T., Coonrod, S., & Gossen, J. (2007). Peptidylarginine deiminase (PAD) 6 is essential for oocyte cytoskeletal sheet formation and female fertility. *Molecular and Cellular Endocrinology*, *273*(1–2), 25–31. https://doi.org/10.1016/j.mce.2007.05.005

Fang, Z., Lin, M., Li, C., Liu, H., & Gong, C. (2020). A comprehensive review of the roles of E2F1 in colon cancer. *American Journal of Cancer Research*, *10*(3), 757–768.

Fares, J., Fares, M. Y., Khachfe, H. H., Salhab, H. A., & Fares, Y. (2020). Molecular principles of metastasis: A hallmark of cancer revisited. *Signal Transduction and Targeted Therapy*, *5*(1), 28. https://doi.org/10.1038/s41392-020-0134-x

Ferlay, J., Ervik, M., Lam, F., Colombet, M., Mery, L., Piñeros, M., Znaor, A., Soerjomataram, I., & Bray, F. (2020). *Global Cancer Observatory: Cancer Today*. https://gco.iarc.fr/today

Feustel, K., & Falchook, G. S. (2022). Protein Arginine Methyltransferase 5 (PRMT5) Inhibitors in Oncology Clinical Trials: A review. *Journal of Immunotherapy and Precision Oncology*, *5*(3), 58–67. https://doi.org/10.36401/JIPO-22-1

Folco, E. J., Mawson, T. L., Vromman, A., Bernardes-Souza, B., Franck, G., Persson, O., Nakamura, M., Newton, G., Luscinskas, F. W., & Libby, P. (2018). Neutrophil Extracellular Traps Induce Endothelial Cell Activation and Tissue Factor Production Through Interleukin-1α and Cathepsin G. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *38*(8), 1901–1912. https://doi.org/10.1161/ATVBAHA.118.311150

Fong, J. Y., Pignata, L., Goy, P.-A., Kawabata, K. C., Lee, S. C.-W., Koh, C. M., Musiani, D., Massignani, E., Kotini, A. G., Penson, A., Wun, C. M., Shen, Y., Schwarz, M., Low, D. HP., Rialdi, A., Ki, M., Wollmann, H., Mzoughi, S., Gay, F., … Guccione, E. (2019). Therapeutic Targeting of RNA Splicing Catalysis through Inhibition of Protein Arginine Methylation. *Cancer Cell*, *36*(2), 194-209.e9. https://doi.org/10.1016/j.ccell.2019.07.003

Förch, P., Puig, O., Martínez, C., Séraphin, B., & Valcárcel, J. (2002). The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5′ splice sites. *The EMBO Journal*, *21*(24), 6882–6892. https://doi.org/10.1093/emboj/cdf668

Fuhrmann, J., Clancy, K., & Thompson, P. (2015). Chemical Biology of Protein Arginine Modifications in Epigenetic Regulation. *Chemical Reviews*, *115*, 5413–5461. https://doi.org/10.1021/acs.chemrev.5b00003

Fuhrmann, J., & Thompson, P. R. (2016). Protein Arginine Methylation and Citrullination in Epigenetic Regulation. *ACS Chemical Biology*, *11*(3), 654–668. https://doi.org/10.1021/acschembio.5b00942

Furukawa, Y., Nishimura, N., Furukawa, Y., Satoh, M., Endo, H., Iwase, S., Yamada, H., Matsuda, M., Kano, Y., & Nakamura, M. (2002). Apaf-1 Is a Mediator of E2F-1-induced Apoptosis. *Journal of Biological Chemistry*, *277*(42), 39760–39768. https://doi.org/10.1074/jbc.M200805200

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*(7), 930–939. https://doi.org/10.4161/cc.4.7.1784

Gallego-Paez, L., Bordone, M., Leote, A., Saraiva-Agostinho, N., Ascensão-Ferreira, M., & Barbosa-Morais, N. (2017). Alternative splicing: The pledge, the turn, and the prestige: The key role of alternative splicing in human biological systems. *Human Genetics*, *136*(9). https://doi.org/10.1007/s00439-017-1790-y

Gao, X., Zhao, X., Zhu, Y., He, J., Shao, J., Su, C., Zhang, Y., Zhang, W., Saarikettu, J., Silvennoinen, O., & others. (2012). Tudor staphylococcal nuclease (Tudor-SN) participates in small ribonucleoprotein (snRNP) assembly via interacting with symmetrically dimethylated Sm proteins. *Journal of Biological Chemistry*, *287*(22), 18130–18141. https://doi.org/10.1074/jbc.M111.311852

Ge, S. X., Jung, D., & Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics (Oxford, England)*, *36*(8), 2628–2629. https://doi.org/10.1093/bioinformatics/btz931

Ge, Y., Schuster, M. B., Pundhir, S., Rapin, N., Bagger, F. O., Sidiropoulos, N., Hashem, N., & Porse, B. T. (2019). The splicing factor RBM25 controls MYC activity in acute myeloid leukemia. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-018-08076-y

Gehring, N. H., & Roignant, J.-Y. (2021). Anything but Ordinary - Emerging Splicing Mechanisms in Eukaryotic Gene Regulation. *Trends in Genetics*, *37*(4), 355–372. https://doi.org/10.1016/j.tig.2020.10.008

Geuens, T., Bouhy, D., & Timmerman, V. (2016). The hnRNP family: Insights into their role in health and disease. *Human Genetics*, *135*(8), 851–867. https://doi.org/10.1007/s00439-016-1683-5

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. *Science Advances*, *2*(2), e1501257. https://doi.org/10.1126/sciadv.1501257

Girling, R., Partridge, J. F., Bandara, L. R., Burden, N., Totty, N. F., Hsuan, J. J., & La Thangue, N. B. (1993). A new component of the transcription factor DRTF1-E2F. *Nature*, *362*(6415), 83–87. https://doi.org/10.1038/362083a0

Glorian, V., Allègre, 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 & Disease*, *8*(5), Article 5. https://doi.org/10.1038/cddis.2017.222

Gorgoulis, V., Adams, P. D., Alimonti, A., Bennett, D. C., Bischof, O., Bishop, C., Campisi, J., Collado, M., Evangelou, K., Ferbeyre, G., & others. (2019). Cellular senescence: Defining a path forward. *Cell*, *179*(4), 813–827. https://doi.org/10.1016/j.cell.2019.10.005Sorting out the complexity of SR protein functions

Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: Scanning for occurrences of a given motif. *Bioinformatics*, *27*(7), 1017–1018. https://doi.org/10.1093/bioinformatics/btr064

Graveley, B. R. (2000). Sorting out the complexity of SR protein functions. *Rna*, *6*(9), 1197–1211. https://doi.org/10.1017/s1355838200000960

Graves, J. D., Lee, Y.-J., Liu, K., Li, G., Lin, F.-T., & Lin, W.-C. (2020). E2F1 sumoylation as a protective cellular mechanism in oxidative stress response. *Proceedings of the National Academy of Sciences*, *117*(26), 14958–14969. https://doi.org/10.1073/pnas.1921554117

Gromak, N., Matlin, A. J., Cooper, T. A., & Smith, C. W. J. (2003). Antagonistic regulation of α-actinin alternative splicing by CELF proteins and polypyrimidine tract binding protein. *RNA*, *9*(4), 443–456. https://doi.org/10.1261/rna.2191903

Gu, W., & Roeder, R. G. (1997). Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell*, *90*(4), 595–606. https://doi.org/10.1016/s0092-8674(00)80521-8

Guo, J., Jia, J., & Jia, R. (2015). PTBP1 and PTBP2 impaired autoregulation of SRSF3 in cancer cells. *Scientific Reports*, *5*(1), Article 1. https://doi.org/10.1038/srep14548

Guo, Q., Bedford, M. T., & Fast, W. (2011). Discovery of Peptidylarginine Deiminase-4 Substrates by Protein Array: Antagonistic Citrullination and Methylation of Human Ribosomal Protein S2. *Molecular bioSystems*, *7*(7), 2286–2295. https://doi.org/10.1039/c1mb05089c

Guo, Q., & Fast, W. (2011). Citrullination of inhibitor of growth 4 (ING4) by peptidylarginine deiminase 4 (PAD4) disrupts the interaction between ING4 and p53. *J. Biol. Chem.*, *19*, 17096–17078. https://doi.org/10.1074/jbc.M111.230961

Gutierrez-Beltran, E., Denisenko, T. V., Zhivotovsky, B., & Bozhkov, P. V. (2016). Tudor staphylococcal nuclease: Biochemistry and functions. *Cell Death & Differentiation*, *23*(11), Article 11. https://doi.org/10.1038/cdd.2016.93

György, B., Tóth, E., Tarcsa, E., Falus, A., & Buzás, E. I. (2006). Citrullination: A posttranslational modification in health and disease. *Int. J. Biochem. Cell Biol.*, *38*, 1662–1677. https://doi.org/10.1016/j.biocel.2006.03.008

Hafner, A., Bulyk, M. L., Jambhekar, A., & Lahav, G. (2019). The multiple mechanisms that regulate p53 activity and cell fate. *Nature Reviews Molecular Cell Biology*, *20*(4), Article 4. https://doi.org/10.1038/s41580-019-0110-x

Hagiwara, T., Hidaka, Y., & Yamada, M. (2005). Deimination of histone H2A and H4 at arginine 3 in HL-60 granulocytes. *Biochemistry*, *44*, 5827–5834. https://doi.org/10.1021/bi047505c

Hagiwara, T., Nakashima, K., Hirano, H., Senshu, T., & Yamada, M. (2002). Deimination of Arginine Residues in Nucleophosmin-B23 and Histones in HL-60 Granulocytes. *Biochemical and Biophysical Research Communications*, *290*(3), 979–983. https://doi.org/10.1006/bbrc.2001.6303

Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discovery*, *12*(1), 31–46. https://doi.org/10.1158/2159-8290.CD-21-1059

Hanahan, D., & Weinberg, R. A. (2000). The Hallmarks of Cancer. *Cell*, *100*(1), 57–70. https://doi.org/10.1016/S0092-8674(00)81683-9

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, *144*(5), 646–674. https://doi.org/10.1016/j.cell.2011.02.013

Harada, K., Carr, S. M., Shrestha, A., & La Thangue, N. B. (2023). Citrullination and the Protein Code: Crosstalk between Post-translational Modifications in Cancer. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2022.0243

He, L.-Z., Merghoub, T., & Pandolfi, P. P. (1999). In vivo analysis of the molecular pathogenesis of acute promyelocytic leukemia in the mouse and its therapeutic implications. *Oncogene*, *18*(38), 5278–5292. https://doi.org/10.1038/sj.onc.1203088microbiome

Helin, K., Lees, J. A., Vidal, M., Dyson, N., Harlow, E., & Fattaey, A. (1992). A cDNA encoding a pRB-binding protein with properties of the transcription factor E2F. *Cell*, *70*(2), 337–350. https://doi.org/10.1016/0092-8674(92)90107-n

Helmink, B. A., Khan, M. W., Hermann, A., Gopalakrishnan, V., & Wargo, J. A. (2019). The microbiome, cancer, and cancer therapy. *Nature Medicine*, *25*(3), 377–388. https://doi.org/10.1038/s41591-019-0377-7

Hershko, T., Chaussepied, M., Oren, M., & Ginsberg, D. (2005). Novel link between E2F and p53: Proapoptotic cofactors of p53 are transcriptionally upregulated by E2F. *Cell Death & Differentiation*, *12*(4), Article 4. https://doi.org/10.1038/sj.cdd.4401575

Hershko, T., & Ginsberg, D. (2004). Up-regulation of Bcl-2 homology 3 (BH3)-only proteins by E2F1 mediates apoptosis. *The Journal of Biological Chemistry*, *279*(10), 8627–8634. https://doi.org/10.1074/jbc.m312866200

Hidaka, Y., Hagiwara, T., & Yamada, M. (2005). Methylation of the guanidino group of arginine residues prevents citrullination by peptidylarginine deiminase IV. *FEBS Lett.*, *579*, 4088–4092. https://doi.org/10.1016/j.febslet.2005.06.035

Hollander, D., Naftelberg, S., Lev-Maor, G., Kornblihtt, A. R., & Ast, G. (2016). How Are Short Exons Flanked by Long Introns Defined and Committed to Splicing? *Trends in Genetics: TIG*, *32*(10), 596–606. https://doi.org/10.1016/j.tig.2016.07.003

Holmberg, C., Helin, K., Sehested, M., & Karlström, O. (1998). E2F-1-induced p53-independent apoptosis in transgenic mice. *Oncogene*, *17*(2), 143–155. https://doi.org/10.1038/sj.onc.1201915

Hong, E., Barczak, W., Park, S., Heo, J. S., Ooshima, A., Munro, S., Hong, C. P., Park, J., An, H., Park, J. O., Park, S. H., La Thangue, N. B., & Kim, S.-J. (2023). Combination treatment of T1-44, a PRMT5 inhibitor with Vactosertib, an inhibitor of TGF-β signaling, inhibits invasion and prolongs survival in a mouse model of pancreatic tumors. *Cell Death & Disease*, *14*(2), Article 2. https://doi.org/10.1038/s41419-023-05630-5

Houghton, A. M., Rzymkiewicz, D. M., Ji, H., Gregory, A. D., Egea, E. E., Metz, H. E., Stolz, D. B., Land, S. R., Marconcini, L. A., Kliment, C. R., Jenkins, K. M., Beaulieu, K. A., Mouded, M., Frank, S. J., Wong, K. K., & Shapiro, S. D. (2010). Neutrophil elastase-mediated degradation of IRS-1 accelerates lung tumor growth. *Nature Medicine*, *16*(2), Article 2. https://doi.org/10.1038/nm.2084

Iaquinta, P. J., & Lees, J. A. (2007). Life and death decisions by the E2F transcription factors. *Current Opinion in Cell Biology*, *19*(6), 649–657. https://doi.org/10.1016/j.ceb.2007.10.006

Iborra, S., Hirschfeld, M., Jaeger, M., Zur Hausen, A., Braicu, I., Sehouli, J., Gitsch, G., & Stickeler, E. (2013). Alterations in expression pattern of splicing factors in epithelial ovarian cancer and its clinical impact. *International Journal of Gynecologic Cancer*, *23*(6). https://doi.org/10.1097/IGC.0b013e31829783e3

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 and Differentiation*, *18*(1), 122–132. https://doi.org/10.1038/cdd.2010.70

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 Jr, W. G. (2000). Role for the p53 homologue p73 in E2F-1-induced apoptosis. *Nature*, *407*(6804), Article 6804. https://doi.org/10.1038/35036614

Jackson, S. P., & Bartek, J. (2009). The DNA-damage response in human biology and disease. *Nature*, *461*(7267), 1071–1078. https://doi.org/10.1038/nature08467

Jang, H. N., Lee, M., Loh, T. J., Choi, S.-W., Oh, H. K., Moon, H., Cho, S., Hong, S.-E., Kim, D. H., Sheng, Z., Green, M. R., Park, D., Zheng, X., & Shen, H. (2014). Exon 9 skipping of apoptotic caspase-2 pre-mRNA is promoted by SRSF3 through interaction with exon 8. *Biochimica et Biophysica Acta*, *1839*(1), 25–32. https://doi.org/10.1016/j.bbagrm.2013.11.006

Jarmoskaite, I., & Russell, R. (2011). DEAD-box proteins as RNA helicases and chaperones. *Wiley Interdisciplinary Reviews: RNA*, *2*(1), 135–152. https://doi.org/10.1002/wrna.50

Jeong, S. (2017). SR Proteins: Binders, Regulators, and Connectors of RNA. *Molecules and Cells*, *40*(1), 1–9. https://doi.org/10.14348/molcells.2017.2319

Jin, S., Xue, Z., Zhang, J., Wang, Z., Zhang, J., Chen, D., Liu, W., & Lin, J. (2021). Identification of SRSF3 target mRNAs using inducible TRIBE. *Biochemical and Biophysical Research Communications*, *578*, 21–27. https://doi.org/10.1016/j.bbrc.2021.09.019

Kaelin, W. G., Krek, W., Sellers, W. R., DeCaprio, J. A., Ajchenbaum, F., Fuchs, C. S., Chittenden, T., Li, Y., Farnham, P. J., & Blanar, M. A. (1992). Expression cloning of a cDNA encoding a retinoblastoma-binding protein with E2F-like properties. *Cell*, *70*(2), 351–364. https://doi.org/10.1016/0092-8674(92)90108-o

Kalsotra, A., & Cooper, T. A. (2011). Functional consequences of developmentally regulated alternative splicing. *Nature Reviews Genetics*, *12*(10), Article 10. https://doi.org/10.1038/nrg3052

Kanopka, A., Mühlemann, O., & Akusjärvi, G. (1996). Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature*, *381*(6582), Article 6582. https://doi.org/10.1038/381535a0

Kastan, M. B. (2008). DNA damage responses: Mechanisms and roles in human disease: 2007 GHA Clowes Memorial Award Lecture. *Molecular Cancer Research*, *6*(4), 517–524. https://doi.org/10.1158/1541-7786.MCR-08-0020

Kawalkowska, J., Quirke, A.-M., Ghari, F., Davis, S., Subramanian, V., Thompson, P. R., Williams, R. O., Fischer, R., La Thangue, N. B., & Venables, P. J. (2016). Abrogation of collagen-induced arthritis by a peptidyl arginine deiminase inhibitor is associated with modulation of T cell-mediated immune responses. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep26430

Kazerounian, S., Yee, K., & Lawler, J. (2008). Thrombospondins: From structure to therapeutics: Thrombospondins in cancer. *Cellular and Molecular Life Sciences*, *65*, 700–712. https://doi.org/10.1007/s00018-007-7486-z

Ke, H., Zhao, L., Zhang, H., Feng, X., Xu, H., Hao, J., Wang, S., Yang, Q., Zou, L., Su, X., & others. (2018). Loss of TDP43 inhibits progression of triple-negative breast cancer in coordination with

SRSF3. *Proceedings of the National Academy of Sciences*, *115*(15), E3426–E3435. https://doi.org/10.1073/pnas.1714573115

Kearney, P. L., Bhatia, M., Jones, N. G., Luo, Y., Glascock, M. C., Catchings, K. L., Yamada, M., & Thompson, P. R. (2005). Kinetic characterization of protein arginine deiminase 4: A transcriptional corepressor implicated in the onset and progression of rheumatoid arthritis. *Biochemistry*, *44*, 10570–10582. https://doi.org/10.1021/bi050292m

Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M., & Stamm, S. (2013). Function of alternative splicing. *Gene*, *514*(1), 1–30. https://doi.org/10.1016/j.gene.2012.07.083

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The Human Genome Browser at UCSC. *Genome Research*, *12*(6), 996–1006. https://doi.org/10.1101/gr.229102

Kessler, O., Jiang, Y., & Chasin, L. A. (1993). Order of intron removal during splicing of endogenous adenine phosphoribosyltransferase and dihydrofolate reductase pre-mRNA. *Molecular and Cellular Biology*, *13*(10), 6211–6222. https://doi.org/10.1128/mcb.13.10.6211-6222.1993

Kholia, S., Jorfi, S., Thompson, P. R., Causey, C. P., Nicholas, A. P., Inal, J. M., & Lange, S. (2015). A novel role for peptidylarginine deiminases in microvesicle release reveals therapeutic potential of PAD inhibition in sensitizing prostate cancer cells to chemotherapy. *Journal of Extracellular Vesicles*, *4*(1), 26192. https://doi.org/10.3402/jev.v4.26192

Kim, R., Emi, M., & Tanabe, K. (2007). Cancer immunoediting from immune surveillance to immune escape. *Immunology*, *121*(1), 1–14. https://doi.org/10.1111/j.1365-2567.2007.02587.x

Kim, S. W., Taggart, A. J., Heintzelman, C., Cygan, K. J., Hull, C. G., Wang, J., Shrestha, B., & Fairbrother, W. G. (2017). Widespread intra-dependencies in the removal of introns from human transcripts. *Nucleic Acids Research*, *45*(16), 9503–9513. https://doi.org/10.1093/nar/gkx661

Knight, J. S., Subramanian, V., O'Dell, A. A., Yalavarthi, S., Zhao, W., Smith, C. K., Hodgin, J. B., Thompson, P. R., & Kaplan, M. J. (2015). Peptidylarginine deiminase inhibition disrupts NET formation and protects against kidney, skin and vascular disease in lupus-prone MRL-lpr mice. *Annals of the Rheumatic Diseases*, *74*(12), 2199–2206. https://doi.org/10.1136/annrheumdis-2014-205365

Knuckley, B., Causey, C. P., Jones, J. E., Bhatia, M., Dreyton, C. J., Osborne, T. C., Takahara, H., & Thompson, P. R. (2010). Substrate Specificity and Kinetic Studies of PADs 1, 3, and 4 Identify Potent and Selective Inhibitors of Protein Arginine Deiminase 3. *Biochemistry*, *49*(23), 4852–4863. https://doi.org/10.1021/bi100363t

Komori, H., Goto, Y., Kurayoshi, K., Ozono, E., Iwanaga, R., Bradford, A. P., Araki, K., & Ohtani, K. (2018). Differential requirement for dimerization partner DP between E2F-dependent activation of tumor suppressor and growth-related genes. *Scientific Reports*, *8*, 8438. https://doi.org/10.1038/s41598-018-26860-0

Kontaki, H., & Talianidis, I. (2010). Lysine Methylation Regulates E2F1-Induced Cell Death. *Molecular Cell*, *39*(1), 152–160. https://doi.org/10.1016/j.molcel.2010.06.006

Kosgodage, U. S., Uysal-Onganer, P., MacLatchy, A., Kraev, I., Chatterton, N. P., Nicholas, A. P., Inal, J. M., & Lange, S. (2018). Peptidylarginine Deiminases Post-Translationally Deiminate Prohibitin and Modulate Extracellular Vesicle Release and MicroRNAs in Glioblastoma Multiforme. *International Journal of Molecular Sciences*, *20*(1), 103. https://doi.org/10.3390/ijms20010103

Kowald, A., Passos, J. F., & Kirkwood, T. B. (2020). On the evolution of cellular senescence. *Aging Cell*, *19*(12), e13270. https://doi.org/10.1111/acel.13270

Kuan, C. Y., Roth, K. A., Flavell, R. A., & Rakic, P. (2000). Mechanisms of programmed cell death in the developing brain. *Trends in Neurosciences*, *23*(7), 291–297. https://doi.org/10.1016/s0166-2236(00)01581-2

Kuranaga, Y., Sugito, N., Shinohara, H., Tsujino, T., Taniguchi, K., Komura, K., Ito, Y., Soga, T., & Akao, Y. (2018). SRSF3, a Splicer of the PKM Gene, Regulates Cell Growth and Maintenance of Cancer-Specific Energy Metabolism in Colon Cancer Cells. *International Journal of Molecular Sciences*, *19*(10). https://doi.org/10.3390/ijms19103012

Kurowska, W., Kuca-Warnawin, E. H., Radzikowska, A., & Maśliński, W. (2017). The role of anti-citrullinated protein antibodies (ACPA) in the pathogenesis of rheumatoid arthritis. *Central-European Journal of Immunology*, *42*(4), 390–398. https://doi.org/10.5114/ceji.2017.72807

Labi, V., Erlacher, M., Kiessling, S., & Villunger, A. (2006). BH3-only proteins in cell death initiation, malignant disease and anticancer therapy. *Cell Death & Differentiation*, *13*(8), Article 8. https://doi.org/10.1038/sj.cdd.4401940

Ladomery, M. (2013). Aberrant Alternative Splicing Is Another Hallmark of Cancer. *International Journal of Cell Biology*, *2013*, 463786. https://doi.org/10.1155/2013/463786

Lam, B. J., & Hertel, K. J. (2002). A general role for splicing enhancers in exon definition. *Rna*, *8*(10), 1233–1241. https://doi.org/10.1017/s1355838202028030

Laurie, N. A., Donovan, S. L., Shih, C.-S., Zhang, J., Mills, N., Fuller, C., Teunisse, A., Lam, S., Ramos, Y., Mohan, A., Johnson, D., Wilson, M., Rodriguez-Galindo, C., Quarto, M., Francoz, S., Mendrysa, S. M., Guy, R. K., Marine, J.-C., Jochemsen, A. G., & Dyer, M. A. (2006). Inactivation of the p53 pathway in retinoblastoma. *Nature*, *444*(7115), 61–66. https://doi.org/10.1038/nature05194

Ledet, M. M., Anderson, R., Harman, R., Muth, A., Thompson, P. R., Coonrod, S. A., & Van de Walle, G. R. (2018). BB-Cl-Amidine as a novel therapeutic for canine and feline mammary cancer via activation of the endoplasmic reticulum stress pathway. *BMC Cancer*, *18*(1), 412. https://doi.org/10.1186/s12885-018-4323-8

Lee, C.-Y., Lin, C.-C., Liu, Y.-L., Liu, G.-Y., Liu, J.-H., & Hung, H.-C. (2017). Molecular Interplay between the Dimer Interface and the Substrate-Binding Site of Human Peptidylarginine Deiminase 4. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/srep42662

Lee, C.-Y., Wang, D., Wilhelm, M., Zolg, D. P., Schmidt, T., Schnatbaum, K., Reimer, U., Pontén, F., Uhlén, M., Hahne, H., & Kuster, B. (2018). Mining the Human Tissue Proteome for Protein

Citrullination. *Molecular & Cellular Proteomics*, *17*(7), 1378–1391. https://doi.org/10.1074/mcp.RA118.000696

Lee, S. C.-W., Dvinge, H., Kim, E., Cho, H., Micol, J.-B., Chung, Y. R., Durham, B. H., Yoshimi, A., Kim, Y. J., Thomas, M., Lobry, C., Chen, C.-W., Pastore, A., Taylor, J., Wang, X., Krivtsov, A., Armstrong, S. A., Palacino, J., Buonamici, S., … Abdel-Wahab, O. (2016). Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins. *Nature Medicine*, *22*(6), 672–678. https://doi.org/10.1038/nm.4097

Lee, S., & Schmitt, C. A. (2019). The dynamic nature of senescence in cancer. *Nature Cell Biology*, *21*(1), Article 1. https://doi.org/10.1038/s41556-018-0249-2

Lee, Y.-H., Coonrod, S. A., Kraus, W. L., Jelinek, M. A., & Stallcup, M. R. (2005). Regulation of coactivator complex assembly and function by protein arginine methylation and demethylimination. *Proc. Natl. Acad. Sci.*, *102*, 3611–3616. https://doi.org/10.1073/pnas.0407159102

Lerner, M. R., & Steitz, J. A. (1979). Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proceedings of the National Academy of Sciences of the United States of America*, *76*(11), 5495–5499. https://doi.org/10.1073/pnas.76.11.5495

Lewallen, D. M., Bicker, K. L., Subramanian, V., Clancy, K. W., Slade, D. J., Martell, J., Dreyton, C. J., Sokolove, J., Weerapana, E., & Thompson, P. R. (2015). Chemical Proteomic Platform To Identify Citrullinated Proteins. *ACS Chemical Biology*, *10*(11), 2520–2528. https://doi.org/10.1021/acschembio.5b00438

Lewis, H. D., Liddle, J., Coote, J. E., Atkinson, S. J., Barker, M. D., Bax, B. D., Bicker, K. L., Bingham, R. P., Campbell, M., Chen, Y. H., Chung, C.-W., Craggs, P. D., Davis, R. P., Eberhard, D., Joberty, G., Lind, K. E., Locke, K., Maller, C., Martinod, K., … Wilson, D. M. (2015). Inhibition of PAD4 activity is sufficient to disrupt mouse and human NET formation. *Nature Chemical Biology*, *11*(3), 189–191. https://doi.org/10.1038/nchembio.1735

Lewis, H. D., & Nacht, M. (2016). iPAD or PADi - Tablets with therapeutic disease potential? *Current Opinion in Chemical Biology*, *33*, 169–178. https://doi.org/10.1016/j.cbpa.2016.06.020

Li, P., Li, M., Lindberg, M. R., Kennett, M. J., Xiong, N., & Wang, Y. (2010). PAD4 is essential for antibacterial innate immunity mediated by neutrophil extracellular traps. *Journal of Experimental Medicine*, *207*(9), 1853–1862. https://doi.org/10.1084/jem.20100239

Li, P., Wang, D., Yao, H., Doret, P., Hao, G., Shen, Q., Qiu, H., Zhang, X., Wang, Y., Chen, G., & Wang, Y. (2010). Coordination of PAD4 and HDAC2 in the regulation of p53-target gene expression. *Oncogene*, *29*(21), Article 21. https://doi.org/10.1038/onc.2010.51

Li, P., Yao, H., Zhang, Z., Li, M., Luo, Y., Thompson, P. R., Gilmour, D. S., & Wang, Y. (2008). Regulation of p53 target gene expression by peptidylarginine deiminase 4. *Mol. Cell Biol.*, *283*, 20060–20068. https://doi.org/10.1128/MCB.01747-07

Li, W.-J., He, Y.-H., Yang, J.-J., Hu, G.-S., Lin, Y.-A., Ran, T., Peng, B.-L., Xie, B.-L., Huang, M.-F., Gao, X., Huang, H.-H., Zhu, H. H., Ye, F., & Liu, W. (2021). Profiling PRMT methylome

reveals roles of hnRNPA1 arginine methylation in RNA splicing and cell growth. *Nature Communications*, *12*(1), 1946. https://doi.org/10.1038/s41467-021-21963-1

Liban, T. J., Medina, E. M., Tripathi, S., Sengupta, S., Henry, R. W., Buchler, N. E., & Rubin, S. M. (2017). Conservation and divergence of C-terminal domain structure in the retinoblastoma protein family. *Proceedings of the National Academy of Sciences*, *114*(19), 4942–4947. https://doi.org/10.1073/pnas.1619170114

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*(20), 3960–3971. https://doi.org/10.1016/j.jmb.2016.08.017

Lim, Y., Lee, J. Y., Ha, S. J., Yu, S., Shin, J. K., & Kim, H. C. (2020). Proteome-wide identification of arginine methylation in colorectal cancer tissues from patients. *Proteome Science*, *18*, 6. https://doi.org/10.1186/s12953-020-00162-8

Lim, Y., Park, Y. E., Ha, S., Lee, J. E., & Kim, H. C. (2020). A Comprehensive Analysis of Symmetric Arginine Dimethylation in Colorectal Cancer Tissues Using Immunoaffinity Enrichment and Mass Spectrometry. *Proteomics*, e1900367. https://doi.org/10.1002/pmic.201900367

Lin, C. H., & Patton, J. G. (1995). Regulation of alternative 3 splice site selection by constitutive splicing factors. *RNA (New York, N.Y.)*, *1*(3), 234–245.

Linder, P., & Jankowsky, E. (2011). From unwinding to clamping - The DEAD box RNA helicase family. *Nature Reviews Molecular Cell Biology*, *12*(8), 505–516. https://doi.org/10.1038/nrm3154

Liu, G.-Y., Liao, Y.-F., Chang, W.-H., Liu, C.-C., Hsieh, M.-C., Hsu, P.-C., Tsay, G. J., & Hung, H.-C. (2006). Overexpression of peptidylarginine deiminase IV features in apoptosis of haematopoietic cells. *Apoptosis*, *11*(2), 183–196. https://doi.org/10.1007/s10495-006-3715-4

Liu, J., Yue, P., Artym, V., Mueller, S., & Guo, W. (2009). The role of the exocyst in matrix metalloproteinase secretion and actin dynamics during tumor cell invadopodia formation. *Molecular Biology of the Cell*, *20*(16). https://doi.org/10.1091/mbc.e08-09-0967

Liu, X., Wichapong, K., Lamers, S., Reutelingsperger, C. P. M., & Nicolaes, G. A. F. (2021). Autocitrullination of PAD4 does not alter its enzymatic activity: In vitro and in silico studies. *The International Journal of Biochemistry & Cell Biology*, *134*, 105938. https://doi.org/10.1016/j.biocel.2021.105938

Liu, Y.-L., Lee, C.-Y., Huang, Y.-N., Chen, H.-Y., Liu, G.-Y., & Hung, H.-C. (2017). Probing the Roles of Calcium-Binding Sites during the Folding of Human Peptidylarginine Deiminase 4. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-02677-1

Logan, N., Graham, A., Zhao, X., Fisher, R., Maiti, B., Leone, G., & Thangue, N. B. L. (2005). E2F-8: An E2F family member with a similar organization of DNA-binding domains to E2F-7. *Oncogene*, *24*(31), Article 31. https://doi.org/10.1038/sj.onc.1208703

Lou, H., Neugebauer, K. M., Gagel, R. F., & Berget, S. M. (1998). Regulation of alternative polyadenylation by U1 snRNPs and SRp20. *Molecular and Cellular Biology*, *18*(9), 4977–4985. https://doi.org/10.1128/MCB.18.9.4977

Lu, Z., Guan, X., Schmidt, C. A., & Matera, A. G. (2014). RIP-seq analysis of eukaryotic Sm proteins identifies three major categories of Sm-containing ribonucleoproteins. *Genome Biology*, *15*(1), R7. https://doi.org/10.1186/gb-2014-15-1-r7

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

Luo, X., Chang, S., Xiao, S., Peng, Y., Gao, Y., Hu, F., Liang, J., Xu, Y., Du, K., Chen, Y., Qin, J., Meltzer, S. J., Deng, S., Feng, X., Fan, X., Hou, G., Jin, Z., & Zhang, X. (2022). PAD4-dependent citrullination of nuclear translocation of GSK3β promotes colorectal cancer progression via the degradation of nuclear CDKN1A. *Neoplasia (New York, N.Y.)*, *33*, 100835. https://doi.org/10.1016/j.neo.2022.100835

Ma, X., Li, Y., & Zhao, B. (2022). Ribosomal protein L5 (RPL5)- E2F transcription factor 1 (E2F1) signaling suppresses breast cancer progression via regulating endoplasmic reticulum stress and autophagy. *Bioengineered*, *13*(4), 8076–8086. https://doi.org/10.1080/21655979.2022.2052672

Malumbres, M., & Barbacid, M. (2001). To cycle or not to cycle: A critical decision in cancer. *Nature Reviews Cancer*, *1*(3), 222–231. https://doi.org/10.1038/35106065

Manickavinayaham, S., Vélez-Cruz, R., Biswas, A. K., Bedford, E., Klein, B. J., Kutateladze, T. G., Liu, B., Bedford, M. T., & Johnson, D. G. (2019). E2F1 acetylation directs p300-CBP-mediated histone acetylation at DNA double-strand breaks to facilitate repair. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-12861-8

Marasco, L. E., & Kornblihtt, A. R. (2022). The physiology of alternative splicing. *Nature Reviews Molecular Cell Biology*, 1–13. https://doi.org/10.1038/s41580-022-00545-z

Marcel, V., Vijayakumar, V., Fernández-Cuesta, L., Hafsi, H., Sagne, C., Hautefeuille, A., Olivier, M., & Hainaut, P. (2010). P53 regulates the transcription of its Δ133p53 isoform through specific response elements contained within the TP53 P2 internal promoter. *Oncogene*, *29*(18), Article 18. https://doi.org/10.1038/onc.2010.26

Marguerat, S., & Bähler, J. (2010). RNA-seq: From technology to biology. *Cellular and Molecular Life Sciences: CMLS*, *67*(4), 569–579. https://doi.org/10.1007/s00018-009-0180-6

Marti, A., Wirbelauer, C., Scheffner, M., & Krek, W. (1999). Interaction between ubiquitin-protein ligase SCFSKP2 and E2F-1 underlies the regulation of E2F-1 degradation. *Nature Cell Biology*, *1*(1), Article 1. https://doi.org/10.1038/8984

Martins-Cardoso, K., Almeida, V. H., Bagri, K. M., Rossi, M. I. D., Mermelstein, C. S., König, S., & Monteiro, R. Q. (2020). Neutrophil Extracellular Traps (NETs) Promote Pro-Metastatic Phenotype in Human Breast Cancer Cells through Epithelial-Mesenchymal Transition. *Cancers*, *12*(6), Article 6. https://doi.org/10.3390/cancers12061542

Martin-Urdiroz, M., Deeks, M. J., Horton, C. G., Dawe, H. R., & Jourdain, I. (2016). The Exocyst Complex in Health and Disease. *Frontiers in Cell and Developmental Biology*, *4*. https://doi.org/10.3389/fcell.2016.00024

Mastronardi, F. G., Wood, D. D., Mei, J., Raijmakers, R., Tseveleki, V., Dosch, H.-M., Probert, L., Casaccia-Bonnefil, P., & Moscarello, M. A. (2006). Increased citrullination of histone H3 in multiple sclerosis brain and animal models of demyelination: A role for tumor necrosis factor-induced peptidylarginine deiminase 4 translocation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *26*(44), 11387–11396. https://doi.org/10.1523/JNEUROSCI.3349-06.2006

Masucci, M. T., Minopoli, M., Del Vecchio, S., & Carriero, M. V. (2020). The Emerging Role of Neutrophil Extracellular Traps (NETs) in Tumor Progression and Metastasis. *Frontiers in Immunology*, *11*, 1749. https://doi.org/10.3389/fimmu.2020.01749

Mateyak, M. K., Obaya, A. J., & Sedivy, J. M. (1999). C-Myc regulates cyclin D-Cdk4 and-Cdk6 activity but affects cell cycle progression at multiple independent points. *Molecular and Cellular Biology*, *19*(7), 4672–4683. https://doi.org/10.1128/MCB.19.7.4672

Matlin, A. J., Clark, F., & Smith, C. W. (2005). Understanding alternative splicing: Towards a cellular code. *Nature Reviews Molecular Cell Biology*, *6*(5), 386–398. https://doi.org/10.1038/nrm1645

McElwee, J. L., Mohanan, S., Griffith, O. L., Breuer, H. C., Amguish, L. J., Cherrington, B. D., Palmer, A. M., Howe, L. R., Subramanian, V., Causey, C. P., Thompson, P. R., Gray, J. W., & Coonrod, S. A. (2012). Identification of PADI2 as a potential breast cancer biomarker and therapeutic target. *BMC Cancer*, *12*, 500. https://doi.org/10.1186/1471-2407-12-500

McMahon, A. C., Rahman, R., Jin, H., Shen, J. L., Fieldsend, A., Luo, W., & Rosbash, M. (2016). TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins. *Cell*, *165*(3), 742–753. https://doi.org/10.1016/j.cell.2016.03.007

Mihara, M., Erster, S., Zaika, A., Petrenko, O., Chittenden, T., Pancoska, P., & Moll, U. M. (2003). P53 Has a Direct Apoptogenic Role at the Mitochondria. *Molecular Cell*, *11*(3), 577–590. https://doi.org/10.1016/S1097-2765(03)00050-9

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*(22), 3212–3218. https://doi.org/10.1038/sj.onc.1209343

Mondal, S., & Thompson, P. R. (2019). Protein Arginine Deiminases (PADs): Biochemistry and Chemical Biology of Protein Citrullination. *Accounts of Chemical Research*, *52*(3), 818–832. https://doi.org/10.1021/acs.accounts.9b00024

Morgunova, E., Yin, Y., Jolma, A., Dave, K., Schmierer, B., Popov, A., Eremina, N., Nilsson, L., & Taipale, J. (2015). Structural insights into the DNA-binding specificity of E2F family transcription factors. *Nature Communications*, *6*(1), Article 1. https://doi.org/10.1038/ncomms10050

Moroni, M. C., Hickman, E. S., Denchi, E. L., 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*(6), Article 6. https://doi.org/10.1038/35078527

Morris, E. J., Ji, J.-Y., Yang, F., Di Stefano, L., Herr, A., Moon, N.-S., Kwon, E.-J., Haigis, K. M., Näär, A. M., & Dyson, N. J. (2008). E2F1 represses β-catenin transcription and is antagonized by both pRB and CDK8. *Nature*, *455*(7212), 552–556. https://doi.org/10.1038/nature07310

Müller-McNicoll, M., Botti, V., de Jesus Domingues, A. M., Brandl, H., Schwich, O. D., Steiner, M. C., Curk, T., Poser, I., Zarnack, K., & Neugebauer, K. M. (2016). SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes & Development*, *30*(5), 553–566. https://doi.org/10.1101/gad.276477.115

Musiani, D., Bok, J., Massignani, E., Wu, L., Tabaglio, T., Ippolito, M. R., Cuomo, A., Ozbek, U., Zorgati, H., Ghoshdastider, U., Robinson, R. C., Guccione, E., & Bonaldi, T. (2019). Proteomics profiling of arginine methylation defines PRMT5 substrate specificity. *Science Signaling*, *12*(575), eaat8388. https://doi.org/10.1126/scisignal.aat8388

Mutua, V., & Gershwin, L. J. (2021). A Review of Neutrophil Extracellular Traps (NETs) in Disease: Potential Anti-NETs Therapeutics. *Clinical Reviews in Allergy & Immunology*, *61*(2), 194. https://doi.org/10.1007/s12016-020-08804-7

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*(37), 5341–5357. https://doi.org/10.1038/sj.onc.1210604

Nahle, Z., Polakoff, J., Davuluri, R. V., McCurrach, M. E., Jacobson, M. D., Narita, M., Zhang, M. Q., Lazebnik, Y., Bar-Sagi, D., & Lowe, S. W. (2002). Direct coupling of the cell cycle and cell death machinery by E2F. *Nature Cell Biology*, *4*(11), Article 11. https://doi.org/10.1038/ncb868

Nakashima, K., Hagiwara, T., Ishigami, A., Nagata, S., Asaga, H., Kuramoto, M., Senshu, T., & Yamada, M. (1999). Molecular Characterization of Peptidylarginine Deiminase in HL-60 Cells Induced by Retinoic Acid and 1α,25-Dihydroxyvitamin D3. *Journal of Biological Chemistry*, *274*(39), 27786–27792. https://doi.org/10.1074/jbc.274.39.27786

Nakashima, K., Hagiwara, T., & Yamada, M. (2002). Nuclear Localization of Peptidylarginine Deiminase V and Histone Deimination in Granulocytes. *J Biol Chem.*, *277*, 49562–49568. https://doi.org/10.1074/jbc.M208795200

Nakata, D., Nakao, S., Nakayama, K., Araki, S., Nakayama, Y., Aparicio, S., Hara, T., & Nakanishi, A. (2017). The RNA helicase DDX39B and its paralog DDX39A regulate androgen receptor splice variant AR-V7 generation. *Biochemical and Biophysical Research Communications*, *483*(1), 271–276. https://doi.org/10.1016/j.bbrc.2016.12.153

Neeli, I., Khan, S. N., & Radic, M. (2008). Histone Deimination As a Response to Inflammatory Stimuli in Neutrophils1. *The Journal of Immunology*, *180*(3), 1895–1902. https://doi.org/10.4049/jimmunol.180.3.1895

Obeng, E. A., Chappell, R. J., Seiler, M., Chen, M. C., Campagna, D. R., Schmidt, P. J., Schneider, R. K., Lord, A. M., Wang, L., Gambe, R. G., McConkey, M. E., Ali, A. M., Raza, A., Yu, L., Buonamici, S., Smith, P. G., Mullally, A., Wu, C. J., Fleming, M. D., & Ebert, B. L. (2016). Physiologic expression of SF3B1K700E causes impaired erythropoiesis, aberrant splicing, and

sensitivity to pharmacologic spliceosome modulation. *Cancer Cell*, *30*(3), 404–417. https://doi.org/10.1016/j.ccell.2016.08.006

Ochoa, B., Chico, Y., & Martínez, M. J. (2018). Insights Into SND1 Oncogene Promoter Regulation. *Frontiers in Oncology*, *8*. https://doi.org/10.3389/fonc.2018.00606

Okumura, S., Konishi, Y., Narukawa, M., Sugiura, Y., Yoshimoto, S., Arai, Y., Sato, S., Yoshida, Y., Tsuji, S., Uemura, K., & others. (2021). Gut bacteria identified in colorectal cancer patients promote tumourigenesis via butyrate secretion. *Nature Communications*, *12*(1), 5674. https://doi.org/10.1038/s41467-021-25965-x

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., … Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, *44*(D1), D733–D745. https://doi.org/10.1093/nar/gkv1189

Olivier, M., Eeles, R., Hollstein, M., Khan, M. A., Harris, C. C., & Hainaut, P. (2002). The IARC TP53 database: New online mutation analysis and recommendations to users. *Human Mutation*, *19*(6), 607–614. https://doi.org/10.1002/humu.10081

Ormondroyd, E., De la Luna, S., & La Thangue, N. B. (1995). A new member of the DP family, DP-3, with distinct protein products suggests a regulatory role for alternative splicing in the cell cycle transcription factor DRTF1-E2F. *Oncogene*, *11*(8), 1437–1446.

Pahlich, S., Zakaryan, R. P., & Gehring, H. (2006). Protein arginine methylation: Cellular functions and methods of analysis. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1764*(12), 1890–1903. https://doi.org/10.1016/j.bbapap.2006.08.008

Pandit, S., Zhou, Y., Shiue, L., Coutinho-Mansfield, G., Li, H., Qiu, J., Huang, J., Yeo, G. W., Ares Jr, M., & Fu, X.-D. (2013). Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Molecular Cell*, *50*(2), 223–235. https://doi.org/10.1016/j.molcel.2013.03.001

Papayannopoulos, V., Metzler, K. D., Hakkim, A., & Zychlinsky, A. (2010). Neutrophil elastase and myeloperoxidase regulate the formation of neutrophil extracellular traps. *Journal of Cell Biology*, *191*(3), 677–691. https://doi.org/10.1083/jcb.201006052

Park, S. K., & Jeong, S. (2016). SRSF3 represses the expression of PDCD4 protein by coordinated regulation of alternative splicing, export and translation. *Biochemical and Biophysical Research Communications*, *470*(2), 431–438. https://doi.org/10.1016/j.bbrc.2016.01.019

Paronetto, M. P., Passacantilli, I., & Sette, C. (2016). Alternative splicing and cell survival: From tissue homeostasis to disease. *Cell Death & Differentiation*, *23*(12), Article 12. https://doi.org/10.1038/cdd.2016.91

Pecoraro, A., Carotenuto, P., Russo, G., & Russo, A. (2019). Ribosomal protein uL3 targets E2F1 and Cyclin D1 in cancer cell response to nucleolar stress. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-51723-7

Perekatt, A. O., Shah, P. P., Cheung, S., Jariwala, N., Wu, A., Gandhi, V., Kumar, N., Feng, Q., Patel, N., Chen, L., Joshi, S., Zhou, A., Taketo, M. M., Xing, J., White, E., Gao, N., Gatza, M. L., & Verzi, M. P. (2018). SMAD4 Suppresses WNT-Driven Dedifferentiation and Oncogenesis in the Differentiated Gut Epithelium. *Cancer Research*, *78*(17), 4878–4890. https://doi.org/10.1158/0008-5472.CAN-18-0043

Pierce, A. M., Fisher, S. M., Conti, C. J., & Johnson, D. G. (1998). Deregulated expression of E2F1 induces hyperplasia and cooperates with ras in skin tumor development. *Oncogene*, *16*(10), 1267–1276. https://doi.org/10.1038/sj.onc.1201666

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*(15), 8858–8863. https://doi.org/10.1073/pnas.95.15.8858

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*(9), 6408–6414. https://doi.org/10.1128/MCB.19.9.6408

Pleguezuelos-Manzano, C., Puschhof, J., Rosendahl Huber, A., van Hoeck, A., Wood, H. M., Nomburg, J., Gurjao, C., Manders, F., Dalmasso, G., Stege, P. B., & others. (2020). Mutational signature in colorectal cancer caused by genotoxic pks E. coli. *Nature*, *580*(7802), 269–273. https://doi.org/10.1038/s41586-020-2080-8

Powers, J. T., Hong, S., Mayhew, C. N., Rogers, P. M., Knudsen, E. S., & Johnson, D. G. (2004). E2F1 Uses the ATM Signaling Pathway to Induce p53 and Chk2 Phosphorylation and Apoptosis1. *Molecular Cancer Research*, *2*(4), 203–214. https://doi.org/10.1158/1541-7786.203.2.4

Pruijn, G. J., Wiik, A., & van Venrooij, W. J. (2010). The use of citrullinated peptides and proteins for the diagnosis of rheumatoid arthritis. *Arthritis Research & Therapy*, *12*(1), 203. https://doi.org/10.1186/ar2903

Qin, H., Liu, X., Li, F., Miao, L., Li, T., Xu, B., An, X., Muth, A., Thompson, P. R., Coonrod, S. A., & Zhang, X. (2017). PAD1 promotes epithelial-mesenchymal transition and metastasis in triple-negative breast cancer cells by regulating MEK1-ERK1-2-MMP2 signaling. *Cancer Letters*, *409*, 30–41. https://doi.org/10.1016/j.canlet.2017.08.019

Qin, X. Q., Livingston, D. M., Kaelin, W. G. J., & 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 of the United States of America*, *91*(23), 10918–10922. https://doi.org/10.1073/pnas.91.23.10918

Radzisheuskaya, A., Shliaha, P. V., Grinev, V., Lorenzini, E., Kovalchuk, S., Shlyueva, D., Gorshkov, V., Hendrickson, R. C., Jensen, O. N., & Helin, K. (2019). PRMT5 methylome profiling uncovers a direct link to splicing regulation in acute myeloid leukemia. *Nature Structural & Molecular Biology*, *26*(11), 999–1012. https://doi.org/10.1038/s41594-019-0313-z

Raijmakers, R., Zendman, A. J. W., Egberts, W. V., Vossenaar, E. R., Raats, J., Soede-Huijbregts, C., Rutjes, F. P. J. T., van Veelen, P. A., Drijfhout, J. W., & Pruijn, G. J. M. (2007). Methylation of

arginine residues interferes with citrullination by peptidylarginine deiminases in vitro. *J. Mol. Biol.*, *367*, 1118–1129. https://doi.org/10.1016/j.jmb.2007.01.054

Rajyaguru, P., & Parker, R. (2012). RGG motif proteins: Modulators of mRNA functional states. *Cell Cycle*, *11*(14), 2594–2599. https://doi.org/10.4161/cc.20716

Ramazi, S., & Zahiri, J. (2021). Post-translational modifications in proteins: Resources, tools and prediction methods. *Database*, *2021*. https://doi.org/10.1093/database/baab012

Rambout, X., Dequiedt, F., & Maquat, L. E. (2018). Beyond Transcription: Roles of Transcription Factors in Pre-mRNA Splicing. *Chemical Reviews*, *118*(8), 4339–4364. https://doi.org/10.1021/acs.chemrev.7b00470

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, *8*(11), Article 11. https://doi.org/10.1038/nprot.2013.143

Ren, B., Cam, H., Takahashi, Y., Volkert, T., Terragni, J., Young, R. A., & Dynlacht, B. D. (2002). E2F integrates cell cycle progression with DNA repair, replication, and G(2)-M checkpoints. *Genes & Development*, *16*(2), 245–256. https://doi.org/10.1101/gad.949802

Riley, T., Sontag, E., Chen, P., & Levine, A. (2008). Transcriptional control of human p53-regulated genes. *Nature Reviews Molecular Cell Biology*, *9*(5), Article 5. https://doi.org/10.1038/nrm2395

Rivlin, N., Brosh, R., Oren, M., & Rotter, V. (2011). Mutations in the p53 tumor suppressor gene: Important milestones at the various steps of tumorigenesis. *Genes & Cancer*, *2*(4), 466–474. https://doi.org/10.1177/1947601911408889

Robertson, K. D., & Jones, P. A. (1998). The Human ARF Cell Cycle Regulatory Gene Promoter Is a CpG Island Which Can Be Silenced by DNA Methylation and Down-Regulated by Wild-Type p53. *Molecular and Cellular Biology*, *18*(11), 6457–6473. https://doi.org/10.1128/MCB.18.11.6457

Rogoff, H. A., Pickering, M. T., Frame, F. M., Debatis, M. E., Sanchez, Y., Jones, S., & Kowalik, T. F. (2004). Apoptosis Associated with Deregulated E2F Activity Is Dependent on E2F1 and Atm-Nbs1-Chk2. *Molecular and Cellular Biology*, *24*(7), 2968–2977. https://doi.org/10.1128/MCB.24.7.2968-2977.2004

Roscigno, R. F., & Garcia-Blanco, M. A. (1995). SR proteins escort the U4-U6-U5 tri-snRNP to the spliceosome. *RNA (New York, N.Y.)*, *1*(7), 692–706.

Roworth, A. P., Carr, S. M., Liu, G., Barczak, W., Miller, R. L., Munro, S., Kanapin, A., Samsonova, A., & La Thangue, N. B. (2019). Arginine methylation expands the regulatory mechanisms and extends the genomic landscape under E2F control. *Science Advances*, *5*(6), eaaw4640. https://doi.org/10.1126/sciadv.aaw4640

Roworth, A. P., Ghari, F., & La Thangue, N. B. (2015). To live or let die - Complexity within the E2F1 pathway. *Molecular & Cellular Oncology*, *2*(1), e970480. https://doi.org/10.4161/23723548.2014.970480

Rubin, S. M. (2013). Deciphering the retinoblastoma protein phosphorylation code. *Trends in Biochemical Sciences*, *38*(1), 12–19. https://doi.org/10.1016/j.tibs.2012.10.007

Rumble, J., Fackelman, E., & Mobley, J. (2017). Histone H3 citrullination as a measure of PAD4 activity-inhibition. *Report, Cayman Chemical*.

Sakurai-Yageta, M., Recchi, C., Le Dez, G., Sibarita, J., Daviet, L., Camonis, J., D'Souza-Schorey, C., & Chavrier, P. (2008). The interaction of IQGAP1 with the exocyst complex is required for tumor cell invasion downstream of Cdc42 and RhoA. *The Journal of Cell Biology*, *181*(6). https://doi.org/10.1083/jcb.200709076

Santoni-Rugiu, E., Falck, J., Mailand, N., Bartek, J., & Lukas, J. (2000). Involvement of Myc activity in a G1-S-promoting mechanism parallel to the pRb-E2F pathway. *Molecular and Cellular Biology*, *20*(10), 3497–3509. https://doi.org/10.1128/mcb.20.10.3497-3509.2000

Schiliro, C., & Firestein, B. L. (2021). Mechanisms of metabolic reprogramming in cancer cells supporting enhanced growth and proliferation. *Cells*, *10*(5), 1056. https://doi.org/10.3390/cells10051056

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), Article 7. https://doi.org/10.1038/nmeth.2089

Schulze, A., Zerfass, K., Spitkovsky, D., Middendorp, S., Bergès, J., Helin, K., Jansen-Dürr, 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 of the United States of America*, *92*(24), 11264–11268. https://doi.org/10.1073/pnas.92.24.11264

Sears, C. L., & Garrett, W. S. (2014). Microbes, microbiota, and colon cancer. *Cell Host & Microbe*, *15*(3), 317–328. https://doi.org/10.1016/j.chom.2014.02.007

Seiler, M., Yoshimi, A., Darman, R., Chan, B., Keaney, G., Thomas, M., Agrawal, A. A., Caleb, B., Csibi, A., Sean, E., Fekkes, P., Karr, C., Klimek, V., Lai, G., Lee, L., Kumar, P., Lee, S. C.-W., Liu, X., Mackenzie, C., … Buonamici, S. (2018). H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. *Nature Medicine*, *24*(4), Article 4. https://doi.org/10.1038/nm.4493

Shammas, M. A. (2011). Telomeres, lifestyle, cancer, and aging. *Current Opinion in Clinical Nutrition and Metabolic Care*, *14*(1), 28. https://doi.org/10.1097/MCO.0b013e32834121b1

Shan, B., & Lee, W. H. (1994). Deregulated expression of E2F-1 induces S-phase entry and leads to apoptosis. *Molecular and Cellular Biology*, *14*(12), 8166–8173. https://doi.org/10.1128/mcb.14.12.8166-8173.1994

Shan, B., Zhu, X., Chen, P. L., Durfee, T., Yang, Y., Sharp, D., & Lee, W. H. (1992). Molecular cloning of cellular genes encoding retinoblastoma-associated proteins: Identification of a gene with properties of the transcription factor E2F. *Molecular and Cellular Biology*, *12*(12), 5620–5631. https://doi.org/10.1128/mcb.12.12.5620-5631.1992

Shay, J. W., & Wright, W. E. (2000). Hayflick, his limit, and cellular ageing. *Nature Reviews Molecular Cell Biology*, *1*(1), 72–76. https://doi.org/10.1038/35036093

Shen, M., & Mattox, W. (2012). Activation and repression functions of an SR splicing regulator depend on exonic versus intronic-binding position. *Nucleic Acids Research*, *40*(1), 428–437. https://doi.org/10.1093/nar/gkr713

Shen, S., Park, J. W., Lu, Z., Lin, L., Henry, M. D., Wu, Y. N., Zhou, Q., & Xing, Y. (2014). rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(51), E5593-5601. https://doi.org/10.1073/pnas.1419161111

Shepard, P. J., & Hertel, K. J. (2009). The SR protein family. *Genome Biology*, *10*(10), 242. https://doi.org/10.1186/gb-2009-10-10-242

Shih, I.-M., Wang, T.-L., Traverso, G., Romans, K., Hamilton, S. R., Ben-Sasson, S., Kinzler, K. W., & Vogelstein, B. (2001). Top-down morphogenesis of colorectal tumors. *Proceedings of the National Academy of Sciences*, *98*(5), 2640–2645. https://doi.org/10.1073/pnas.051629398

Shin, C., & Manley, J. L. (2002). The SR Protein SRp38 Represses Splicing in M Phase Cells. *Cell*, *111*(3), 407–417. https://doi.org/10.1016/S0092-8674(02)01038-3

Shirai, C. L., White, B. S., Tripathi, M., Tapia, R., Ley, J. N., Ndonwi, M., Kim, S., Shao, J., Carver, A., Saez, B., Fulton, R. S., Fronick, C., O'Laughlin, M., Lagisetti, C., Webb, T. R., Graubert, T. A., & Walter, M. J. (2017). Mutant U2AF1-expressing cells are sensitive to pharmacological modulation of the spliceosome. *Nature Communications*, *8*(1), Article 1. https://doi.org/10.1038/ncomms14060

Shiseki, M., Nagashima, M., Pedeux, R. M., Kitahama-Shiseki, M., Miura, K., Okamura, S., Onogi, H., Higashimoto, Y., Appella, E., Yokota, J., & Harris, C. C. (2003). p29ING4 and p28ING5 bind to p53 and p300, and enhance p53 activity. *Cancer Res.*, *63*, 2373–2378.

Siegel, R. L., Miller, K. D., Wagle, N. S., & Jemal, A. (2023). Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, *73*(1), 17–48. https://doi.org/10.3322/caac.21763

Sigal, A., & Rotter, V. (2000). Oncogenic mutations of the p53 tumor suppressor: The demons of the guardian of the genome. *Cancer Research*, *60*(24), 6788–6793.

Singh, R., Valcárcel, J., & Green, M. R. (1995). Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science (New York, N.Y.)*, *268*(5214), 1173–1176. https://doi.org/10.1126/science.7761834

Siu, L. L., Rasco, D. W., Vinay, S. P., Romano, P. M., Menis, J., Opdam, F. L., Heinhuis, K. M., Egger, J. L., Gorman, S. A., Parasrampuria, R., Wang, K., Kremer, B. E., & Gounder, M. M. (2019). METEOR-1: A phase I study of GSK3326595, a first-in-class protein arginine methyltransferase 5 (PRMT5) inhibitor, in advanced solid tumours. *Annals of Oncology*, *30*, v159. https://doi.org/10.1093/annonc/mdz244

Slack, J. L., Causey, C. P., Luo, Y., & Thompson, P. R. (2011). Development and Use of Clickable Activity Based Protein Profiling Agents for Protein Arginine Deiminase 4. *ACS Chemical Biology*, *6*(5), 466–476. https://doi.org/10.1021/cb1003515

Slade, D. J., Fang, P., Dreyton, C. J., Zhang, Y., Fuhrmann, J., Rempel, D., Bax, B. D., Coonrod, S. A., Lewis, H. D., Guo, M., Gross, M. L., & Thompson, P. R. (2015). Protein Arginine Deiminase 2 Binds Calcium in an Ordered Fashion: Implications for Inhibitor Design. *ACS Chemical Biology*, *10*(4), 1043–1053. https://doi.org/10.1021/cb500933j

Slade, D. J., Subramanian, V., Fuhrmann, J., & Thompson, P. R. (2014). Chemical and biological methods to detect post-translational modifications of arginine. *Biopolymers*, *101*(2), 133–143. https://doi.org/10.1002/bip.22256

Snijders, A. P., Hautbergue, G. M., Bloom, A., Williamson, J. C., Minshull, T. C., Phillips, H. L., Mihaylov, S. R., Gjerde, D. T., Hornby, D. P., Wilson, S. A., Hurd, P. J., & Dickman, M. J. (2015). Arginine methylation and citrullination of splicing factor proline- and glutamine-rich (SFPQ-PSF) regulates its association with mRNA. *RNA*, *21*(3), 347–359. https://doi.org/10.1261/rna.045138.114

Song, G., Shi, L., Guo, Y., Yu, L., Wang, L., Zhang, X., Li, L., Han, Y., Ren, X., Guo, Q., Bi, K., & Jiang, G. (2015). A novel PAD4-SOX4-PU.1 signaling pathway is involved in the committed differentiation of acute promyelocytic leukemia cells into granulocytic cells. *Oncotarget*, *7*(3), 3144–3157. https://doi.org/10.18632/oncotarget.6551

Song, X., Wan, X., Huang, T., Zeng, C., Sastry, N., Wu, B., James, C., Horbinski, C., Nakano, I., Zhang, W., Hu, B., & Cheng, S. (2019). SRSF3-Regulated RNA Alternative Splicing Promotes Glioblastoma Tumorigenicity by Affecting Multiple Cellular Processes. *Cancer Research*, *79*(20). https://doi.org/10.1158/0008-5472.CAN-19-1504

Sousa-Luís, R., & Carmo-Fonseca, M. (2022). Pseudouridylation: A new player in co-transcriptional splicing regulation. *Molecular Cell*, *82*(3), 495–496. https://doi.org/10.1016/j.molcel.2022.01.010

Sousa-Luís, R., Dujardin, G., Zukher, I., Kimura, H., Weldon, C., Carmo-Fonseca, M., Proudfoot, N. J., & Nojima, T. (2021). POINT technology illuminates the processing of polymerase-associated intact nascent transcripts. *Molecular Cell*, *81*(9), 1935-1950.e6. https://doi.org/10.1016/j.molcel.2021.02.034

Stadler, S. C., Vincent, C. T., Fedorov, V. D., Patsialou, A., Cherrington, B. D., Wakshlag, J. J., Mohanan, S., Zee, B. M., Zhang, X., Garcia, B. A., Condeelis, J. S., Brown, A. M. C., Coonrod, S. A., & Allis, C. D. (2013). Dysregulation of PAD4-mediated citrullination of nuclear GSK3β activates TGF-β signaling and induces epithelial-to-mesenchymal transition in breast cancer cells. *Proceedings of the National Academy of Sciences*, *110*(29), 11851–11856. https://doi.org/10.1073/pnas.1308362110

Staknis, D., & Reed, R. (1994). SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. *Molecular and Cellular Biology*, *14*(11), 7670–7682. https://doi.org/10.1128/mcb.14.11.7670-7682.1994

Stamm, S., Smith, C., & Lührmann, R. (2012). *Alternative Pre-MRNA Splicing: Theory and Protocols*. John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=837577

Stevens, C., Smith, L., & La Thangue, N. B. (2003). Chk2 activates E2F-1 in response to DNA damage. *Nature Cell Biology*, *5*(5), 401–409. https://doi.org/10.1038/ncb974

Stiewe, T., & Pützer, B. M. (2000). Role of the p53-homologue p73 in E2F1-induced apoptosis. *Nature Genetics*, *26*(4), Article 4. https://doi.org/10.1038/82617

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. *Journal of Biological Chemistry*, *290*(11), 7208–7220. https://doi.org/10.1074/jbc.M114.625046

Sun, B., Dwivedi, N., Bechtel, T. J., Paulsen, J. L., Muth, A., Bawadekar, M., Li, G., Thompson, P. R., Shelef, M. A., Schiffer, C. A., Weerapana, E., & Ho, I.-C. (2017). Citrullination of NF-κB p65 promotes its nuclear localization and TLR-induced expression of IL-1β and TNFα. *Science Immunology*, *2*(12), eaal3062. https://doi.org/10.1126/sciimmunol.aal3062

Sun, C., Mezzadra, R., & Schumacher, T. N. (2018). Regulation and function of the PD-L1 checkpoint. *Immunity*, *48*(3), 434–452. https://doi.org/10.1016/j.immuni.2018.03.014

Sun, H., & Chasin, L. A. (2000). Multiple Splicing Defects in an Intronic False Exon. *Molecular and Cellular Biology*, *20*(17), 6414–6425. https://doi.org/10.1128/MCB.20.17.6414-6425.2000

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Surget, S., Khoury, M. P., & Bourdon, J.-C. (2013). Uncovering the role of p53 splice variants in human malignancy: A clinical perspective. *OncoTargets and Therapy*, *7*, 57–68. https://doi.org/10.2147/OTT.S53876

Swiercz, R., Cheng, D., Kim, D., & Bedford, M. T. (2007). Ribosomal protein rpS2 is hypomethylated in PRMT3-deficient mice. *The Journal of Biological Chemistry*, *282*(23), 16917–16923. https://doi.org/10.1074/jbc.M609778200

Swiercz, R., Person, M. D., & Bedford, M. T. (2005). Ribosomal protein S2 is a substrate for mammalian PRMT3 (protein arginine methyltransferase 3). *The Biochemical Journal*, *386*(Pt 1), 85–91. https://doi.org/10.1042/BJ20041466

Tanaka, T., Goto, K., & Iino, M. (2017). Diverse Functions and Signal Transduction of the Exocyst Complex in Tumor Cells. *Journal of Cellular Physiology*, *232*(5), 939–957. https://doi.org/10.1002/jcp.25619

Tanaka, T., & Iino, M. (2014). Knockdown of Sec8 promotes cell-cycle arrest at G1-S phase by inducing p21 via control of FOXO proteins. *The FEBS Journal*, *281*(4), 1068–1084. https://doi.org/10.1111/febs.12669

Tanaka, T., Iino, M., & Goto, K. (2012). Knockdown of Sec6 improves cell-cell adhesion by increasing α-E-catenin in oral cancer cells. *FEBS Letters*, *586*(6). https://doi.org/10.1016/j.febslet.2012.02.026

Tanaka, T., Kikuchi, N., Goto, K., & Iino, M. (2016). Sec6-8 regulates Bcl-2 and Mcl-1, but not Bcl-xl, in malignant peripheral nerve sheath tumor cells. *Apoptosis: An International Journal on Programmed Cell Death*, *21*(5). https://doi.org/10.1007/s10495-016-1230-9

Tanikawa, C., Espinosa, M., Suzuki, A., Masuda, K., Yamamoto, K., Tsuchiya, E., Ueda, K., Daigo, Y., Nakamura, Y., & Matsuda, K. (2012). Regulation of histone modification and chromatin structure by the p53-PADI4 pathway. *Nature Communications*, *3*(1), Article 1. https://doi.org/10.1038/ncomms1676

Tanikawa, C., Ueda, K., Suzuki, A., Iida, A., Nakamura, R., Atsuta, N., Tohnai, G., Sobue, G., Saichi, N., Momozawa, Y., Kamatani, Y., Kubo, M., Yamamoto, K., Nakamura, Y., & Matsuda, K. (2018). Citrullination of RGG Motifs in FET Proteins by PAD4 Regulates Protein Aggregation and ALS Susceptibility. *Cell Reports*, *22*(6), 1473–1483. https://doi.org/10.1016/j.celrep.2018.01.031

Tao, Y., Kassatly, R., Cress, W., & Horowitz, J. (1997). Subunit composition determines E2F DNA-binding site specificity. *Molecular and Cellular Biology*, *17*(12). https://doi.org/10.1128/MCB.17.12.6994

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*(10), 4546–4556. https://doi.org/10.1128/MCB.24.10.4546-4556.2004

Teng, M. W., Swann, J. B., Koebel, C. M., Schreiber, R. D., & Smyth, M. J. (2008). Immune-mediated dormancy: An equilibrium with cancer. *Journal of Leukocyte Biology*, *84*(4), 988–993. https://doi.org/10.1189/jlb.1107774

Thandapani, P., O'Connor, T. R., Bailey, T. L., & Richard, S. (2013). Defining the RGG-RG Motif. *Molecular Cell*, *50*(5), 613–623. https://doi.org/10.1016/j.molcel.2013.05.021

Thomas, S., Izard, J., Walsh, E., Batich, K., Chongsathidkiet, P., Clarke, G., Sela, D. A., Muller, A. J., Mullin, J. M., Albert, K., & others. (2017). The Host Microbiome Regulates and Maintains Human Health: A Primer and Perspective for Non-MicrobiologistsHost Microbiome in Human Health. *Cancer Research*, *77*(8), 1783–1812. https://doi.org/10.1158/0008-5472.CAN-16-2929

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 Science*, *105*(10), 1360–1368. https://doi.org/10.1111/cas.12495

Torres, M. J., Pandita, R. K., Kulak, O., Kumar, R., Formstecher, E., Horikoshi, N., Mujoo, K., Hunt, C. R., Zhao, Y., Lum, L., Zaman, A., Yeaman, C., White, M. A., & Pandita, T. K. (2015). Role of the Exocyst Complex Component Sec6-8 in Genomic Stability. *Molecular and Cellular Biology*, *35*(21), 3633–3645. https://doi.org/10.1128/MCB.00768-15

Uhlen, M., Karlsson, M. J., Zhong, W., Tebani, A., Pou, C., Mikes, J., Lakshmikanth, T., Forsström, B., Edfors, F., Odeberg, J., Mardinoglu, A., Zhang, C., von Feilitzen, K., Mulder, J., Sjöstedt, E., Hober, A., Oksvold, P., Zwahlen, M., Ponten, F., … Brodin, P. (2019). A genome-wide

transcriptomic analysis of protein-coding genes in human blood cells. *Science*, *366*(6472), eaax9198. https://doi.org/10.1126/science.aax9198

Urlaub, H., Raker, V. A., Kostka, S., & Lührmann, R. (2001). Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *The EMBO Journal*, *20*(1–2), 187–196. https://doi.org/10.1093/emboj/20.1.187

Valesini, G., Gerardi, M., Iannuccelli, C., Pacucci, V., Pendolino, M., & Shoenfeld, Y. (2016). Citrullination and Autoimmunity. *Autoimmunity Reviews*, *14*, 490–497. https://doi.org/10.1016/j.autrev.2015.01.013

van Rees, D. J., Bouti, P., Klein, B., Verkuijlen, P. J. H., van Houdt, M., Schornagel, K., Tool, A. T. J., Venet, D., Sotiriou, C., El-Abed, S., Izquierdo, M., Guillaume, S., Saura, C., Di Cosimo, S., Huober, J., Roylance, R., Kim, S.-B., Kuijpers, T. W., van Bruggen, R., … Matlung, H. L. (2022). Cancer cells resist antibody-mediated destruction by neutrophils through activation of the exocyst complex. *Journal for Immunotherapy of Cancer*, *10*(6), e004820. https://doi.org/10.1136/jitc-2022-004820

Vélez-Cruz, R., & Johnson, D. G. (2017). The retinoblastoma (RB) tumor suppressor: Pushing back against genome instability on multiple fronts. *International Journal of Molecular Sciences*, *18*(8), 1776. https://doi.org/10.3390/ijms18081776

Villegas, V. M., Gold, A. S., Wildner, A., Ehlies, F., & Murray, T. G. (2014). Genomic landscape of retinoblastoma. *Clinical & Experimental Ophthalmology*, *42*(1), 2–3. https://doi.org/10.1111/ceo.12277

Vorobjeva, N. V., & Chernyak, B. V. (2020). NETosis: Molecular Mechanisms, Role in Physiology and Pathology. *Biochemistry. Biokhimiia*, *85*(10), 1178. https://doi.org/10.1134/S0006297920100065

Vossenaar, E. R., Zendman, A. J. W., van Venrooij, W. J., & Pruijn, G. J. M. (2003). PAD, a growing family of citrullinating enzymes: Genes, features and involvement in disease. *BioEssays*, *25*(11), 1106–1118. https://doi.org/10.1002/bies.10357

Wagner, K.-D., El Maï, M., Ladomery, M., Belali, T., Leccia, N., Michiels, J.-F., & Wagner, N. (2019). Altered VEGF Splicing Isoform Balance in Tumor Endothelium Involves Activation of Splicing Factors Srpk1 and Srsf1 by the Wilms Tumor Suppressor Wt1. *Cells*, *8*(1). https://doi.org/10.3390/cells8010041

Wang, B., Kohli, J., & Demaria, M. (2020). Senescent cells in cancer therapy: Friends or foes? *Trends in Cancer*, *6*(10), 838–857. https://doi.org/10.1016/j.trecan.2020.05.004

Wang, B., Liu, K., Lin, F., & Lin, W. (2004). A role for 14-3-3 tau in E2F1 stabilization and DNA damage-induced apoptosis. *The Journal of Biological Chemistry*, *279*(52). https://doi.org/10.1074/jbc.M410493200

Wang, B., Su, X., Zhang, B., & Pan, S. (2023). GSK484, an inhibitor of peptidyl arginine deiminase 4, increases the radiosensitivity of colorectal cancer and inhibits neutrophil extracellular traps. *The Journal of Gene Medicine*, *25*(9), e3530. https://doi.org/10.1002/jgm.3530

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P., & Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, *456*(7221), Article 7221. https://doi.org/10.1038/nature07509

Wang, H., Lekbaby, B., Fares, N., Augustin, J., Attout, T., Schnuriger, A., Cassard, A.-M., Panasyuk, G., Perlemuter, G., Bieche, I., & others. (2019). Alteration of splicing factors expression during liver disease progression: Impact on hepatocellular carcinoma outcome. *Hepatology International*, *13*, 454–467. https://doi.org/10.1007/s12072-019-09950-7

Wang, J.-L., Guo, C.-R., Sun, T.-T., Su, W.-Y., Hu, Q., Guo, F.-F., Liang, L.-X., Xu, J., Xiong, H., & Fang, J.-Y. (2020). SRSF3 functions as an oncogene in colorectal cancer by regulating the expression of ArhGAP30. *Cancer Cell International*, *20*, 120. https://doi.org/10.1186/s12935-020-01201-2

Wang, Y., Li, M., Stadler, S., Correll, S., Li, P., Wang, D., Hayama, R., Leonelli, L., Han, H., Grigoryev, S. A., Allis, C. D., & Coonrod, S. A. (2009). Histone hypercitrullination mediates chromatin decondensation and neutrophil extracellular trap formation. *Journal of Cell Biology*, *184*(2), 205–213. https://doi.org/10.1083/jcb.200806072

Wang, Y., Li, P., Wang, S., Hu, J., Chen, X. A., Wu, J., Fisher, M., Oshaben, K., Zhao, N., Gu, Y., Wang, D., Chen, G., & Wang, Y. (2012). Anticancer Peptidylarginine Deiminase (PAD) Inhibitors Regulate the Autophagy Flux and the Mammalian Target of Rapamycin Complex 1 Activity. *Journal of Biological Chemistry*, *287*(31), 25941–25953. https://doi.org/10.1074/jbc.M112.375725

Wang, Y., Wysocka, J., Sayegh, J., Lee, Y.-H., Perlin, J., Leonelli, L., Sonbuchner, L., McDonald, C., Cook, R., Dou, Y., Roeder, R., Clarke, S., Stallcup, M., Allis, D., & Coonrod, S. (2004). Human PAD4 Regulates Histone Arginine Methylation Levels via Demethylimination. *Science*, *306*, 279–283. https://doi.org/10.1126/science.1101400

Wang, Z., Chatterjee, D., Jeon, H. Y., Akerman, M., Vander Heiden, M. G., Cantley, L. C., & Krainer, A. R. (2012). Exon-centric regulation of pyruvate kinase M alternative splicing via mutually exclusive exons. *Journal of Molecular Cell Biology*, *4*(2), 79–87. https://doi.org/10.1093/jmcb/mjr030

Warrell Jr, R. P., de The, H., Wang, Z.-Y., & Degos, L. (1993). Acute promyelocytic leukemia. *New England Journal of Medicine*, *329*(3), 177–189. https://doi.org/10.1056/NEJM199307153290307

Wei, L., Wang, X., Luo, M., Wang, H., Chen, H., & Huang, C. (2021). The PAD4 inhibitor GSK484 enhances the radiosensitivity of triple-negative breast cancer. *Human & Experimental Toxicology*, *40*(7), 1074–1083. https://doi.org/10.1177/0960327120979028

Weijts, B. G. M. W., Bakker, W. J., Cornelissen, P. W. A., Liang, K.-H., Schaftenaar, F. H., Westendorp, B., de Wolf, C. A. C. M. T., Paciejewska, M., Scheele, C. L. G. J., Kent, L., Leone, G., Schulte-Merker, S., & de Bruin, A. (2012). E2F7 and E2F8 promote angiogenesis through transcriptional activation of VEGFA in cooperation with HIF1. *The EMBO Journal*, *31*(19), 3871–3884. https://doi.org/10.1038/emboj.2012.231

Will, C. L., & Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harbor Perspectives in Biology*, *3*(7). https://doi.org/10.1101/cshperspect.a003707

Willis, V. C., Gizinski, A. M., Banda, N. K., Causey, C. P., Knuckley, B., Cordova, K. N., Luo, Y., Levitt, B., Glogowska, M., Chandra, P., Kulik, L., Robinson, W. H., Arend, W. P., Thompson, P. R., & Holers, V. M. (2011). N-α-benzoyl-N5-(2-chloro-1-iminoethyl)-L-ornithine amide, a protein arginine deiminase inhibitor, reduces the severity of murine collagen-induced arthritis. *Journal of Immunology (Baltimore, Md.: 1950)*, *186*(7), 4396–4404. https://doi.org/10.4049/jimmunol.1001620

Wolach, O., Sellar, R. S., Martinod, K., Cherpokova, D., McConkey, M., Chappell, R. J., Silver, A. J., Adams, D., Castellano, C. A., Schneider, R. K., Padera, R. F., DeAngelo, D. J., Wadleigh, M., Steensma, D. P., Galinsky, I., Stone, R. M., Genovese, G., McCarroll, S. A., Iliadou, B., … Ebert, B. L. (2018). Increased neutrophil extracellular trap formation promotes thrombosis in myeloproliferative neoplasms. *Science Translational Medicine*, *10*(436), eaan8292. https://doi.org/10.1126/scitranslmed.aan8292

Wong, A., Bryzek, D., Dobosz, E., Scavenius, C., Svoboda, P., Rapala-Kozik, M., Lesner, A., Frydrych, I., Enghild, J., Mydel, P., Pohl, J., Thompson, P. R., Potempa, J., & Koziel, J. (2018). A Novel Biological Role for Peptidyl-Arginine Deiminases: Citrullination of Cathelicidin LL-37 Controls the Immunostimulatory Potential of Cell-Free DNA. *The Journal of Immunology*, *200*(7), 2327–2340. https://doi.org/10.4049/jimmunol.1701391

Wu, J. Y., & Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell*, *75*(6), 1061–1070. https://doi.org/10.1016/0092-8674(93)90316-I

Wu, Q., Schapira, M., Arrowsmith, C. H., & Barsyte-Lovejoy, D. (2021). Protein arginine methylation: From enigmatic functions to therapeutic targeting. *Nature Reviews Drug Discovery*, *20*(7), Article 7. https://doi.org/10.1038/s41573-021-00159-8

Xie, D., Pei, Q., Li, J., Wan, X., & Ye, T. (2021). Emerging Role of E2F Family in Cancer Stem Cells. *Frontiers in Oncology*, *11*. https://www.frontiersin.org/articles/10.3389/fonc.2021.723137

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. *Journal of Receptor and Signal Transduction Research*, *31*, 139–146. https://doi.org/10.3109/10799893.2011.552914

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*(12), 8420–8431. https://doi.org/10.1128/mcb.14.12.8420-8431.1994

Xu, X., Bieda, M., Jin, V. X., Rabinovich, A., Oberley, M. J., Green, R., & Farnham, P. J. (2007). A comprehensive ChIP-chip analysis of E2F1, E2F4, and E2F6 in normal and tumor cells reveals interchangeable roles of E2F family members. *Genome Research*, *17*(11), 1550–1561. https://doi.org/10.1101/gr.6783507

Yamazaki, T., Fujiwara, N., Yukinaga, H., Ebisuya, M., Shiki, T., Kurihara, T., Kioka, N., Kambe, T., Nagao, M., Nishida, E., & others. (2010). The closely related RNA helicases, UAP56 and URH49, preferentially form distinct mRNA export machineries and coordinately regulate mitotic progression. *Molecular Biology of the Cell*, *21*(16), 2953–2965. https://doi.org/10.1091/mbc.E09-10-0913

Yang, J., Välineva, 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*(13), 4485–4494. https://doi.org/10.1093/nar/gkm470

Yang, L., Liu, Q., Zhang, X., Liu, X., Zhou, B., Chen, J., Huang, D., Li, J., Li, H., Chen, F., Liu, J., Xing, Y., Chen, X., Su, S., & Song, E. (2020). DNA of neutrophil extracellular traps promotes cancer metastasis via CCDC25. *Nature*, *583*(7814), Article 7814. https://doi.org/10.1038/s41586-020-2394-6

Yang, Y., & Bedford, M. T. (2013). Protein arginine methyltransferases and cancer. *Nature Reviews Cancer*, *13*(1), Article 1. https://doi.org/10.1038/nrc3409

Yao, H., Li, P., Venters, B., Zheng, S., Thompson, P., Pugh, F., & Wang, Y. (2008). Histone Arg Modifications and p53 Regulate the Expression of OKL38, a Mediator of Apoptosis. *J Biol Chem.*, *283*, 20060–20068. https://doi.org/10.1074/jbc.M802940200

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*(1), 134. https://doi.org/10.1186/1471-2105-13-134

Yilmaz, M., & Christofori, G. (2009). EMT, the cytoskeleton, and cancer cell invasion. *Cancer and Metastasis Reviews*, *28*, 15–33. https://doi.org/10.1007/s10555-008-9169-0 Full text linksCite

Yuan, S., Norgard, R. J., & Stanger, B. Z. (2019). Cellular Plasticity in Cancer - Cancer Cells Change Identity during Tumor Progression. *Cancer Discovery*, *9*(7), 837–851. https://doi.org/10.1158/2159-8290.CD-19-0015

Yuzhalin, A. E. (2019). Citrullination in Cancer. *Cancer Research*, *79*(7), 1274–1284. https://doi.org/10.1158/0008-5472.CAN-18-2797

Yuzhalin, A. E., Gordon-Weeks, A. N., Tognoli, M. L., Jones, K., Markelc, B., Konietzny, R., Fischer, R., Muth, A., O'Neill, E., Thompson, P. R., Venables, P. J., Kessler, B. M., Lim, S. Y., & Muschel, R. J. (2018). Colorectal cancer liver metastatic growth depends on PAD4-driven citrullination of the extracellular matrix. *Nature Communications*, *9*(1), Article 1. https://doi.org/10.1038/s41467-018-07306-7

Zeng, Y., Fair, B. J., Zeng, H., Krishnamohan, A., Hou, Y., Hall, J. M., Ruthenburg, A. J., Li, Y. I., & Staley, J. P. (2022). Profiling lariat intermediates reveals genetic determinants of early and late co-transcriptional splicing. *Molecular Cell*, *82*(24), 4681-4699.e8. https://doi.org/10.1016/j.molcel.2022.11.004

Zhang, C., Chen, Y., Li, F., Yang, M., Meng, F., Zhang, Y., Chen, W., & Wang, W. (2021). B7-H3 is spliced by SRSF3 in colorectal cancer. *Cancer Immunology, Immunotherapy: CII*, *70*(2), 311–321. https://doi.org/10.1007/s00262-020-02683-9

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*(1), 79–89. https://doi.org/10.1016/S0092-8674(00)80625-X

Zhang, M. Q. (1998). Statistical Features of Human Exons and Their Flanking Regions. *Human Molecular Genetics*, *7*(5), 919–932. https://doi.org/10.1093/hmg/7.5.919

Zhang, X., Gamble, M. J., Stadler, S., Cherrington, B. D., Causey, C. P., Thompson, P. R., Roberson, M. S., Kraus, W. L., & Coonrod, S. A. (2011). Genome-Wide Analysis Reveals PADI4 Cooperates with Elk-1 to Activate c-Fos Expression in Breast Cancer Cells. *PLoS Genet.*, *7*, e1002112. https://doi.org/10.1371/journal.pgen.1002112

Zhang, X. H.-F., Heller, K. A., Hefter, I., Leslie, C. S., & Chasin, L. A. (2003). Sequence information for the splicing of human pre-mRNA identified by support vector machine classification. *Genome Research*, *13*(12), 2637–2650. https://doi.org/10.1101/gr.1679003

Zhang, X. H.-F., Kangsamaksin, T., Chao, M. S. P., Banerjee, J. K., & Chasin, L. A. (2005). Exon inclusion is dependent on predictable exonic splicing enhancers. *Molecular and Cellular Biology*, *25*(16), 7323–7332. https://doi.org/10.1128/MCB.25.16.7323-7332.2005

Zhang, Y., & Chellappan, S. P. (1995). Cloning and characterization of human DP2, a novel dimerization partner of E2F. *Oncogene*, *10*(11), 2085–2093.

Zhang, Y., Qian, J., Gu, C., & Yang, Y. (2021). Alternative splicing and cancer: A systematic review. *Signal Transduction and Targeted Therapy*, *6*(1), Article 1. https://doi.org/10.1038/s41392-021-00486-7

Zhang, Y., Wang, M., Meng, F., Yang, M., Chen, Y., Guo, X., Wang, W., Zhu, Y., Guo, Y., Feng, C., Tian, S., Zhang, H., Li, H., Sun, J., & Wang, W. (2022). A novel SRSF3 inhibitor, SFI003, exerts anticancer activity against colorectal cancer by modulating the SRSF3-DHCR24-ROS axis. *Cell Death Discovery*, *8*(1), Article 1. https://doi.org/10.1038/s41420-022-01039-9

Zhao, J., Ramos, R., & Demma, M. (2013). CDK8 regulates E2F1 transcriptional activity through S375 phosphorylation. *Oncogene*, *32*(30), Article 30. https://doi.org/10.1038/onc.2012.364

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*(1), 37–51. https://doi.org/10.1016/j.molcel.2013.08.039

Zhong, X.-Y., Wang, P., Han, J., Rosenfeld, M. G., & Fu, X.-D. (2009). SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Molecular Cell*, *35*(1), 1–10. https://doi.org/10.1016/j.molcel.2009.06.016

Zhou, K. I., Shi, H., Lyu, R., Wylder, A. C., Matuszek, Ż., Pan, J. N., He, C., Parisien, M., & Pan, T. (2019). Regulation of Co-transcriptional Pre-mRNA Splicing by m(6)A through the Low-

Complexity Protein hnRNPG. *Molecular Cell*, *76*(1), 70-81.e9. https://doi.org/10.1016/j.molcel.2019.07.005

Zhou, Z., Gong, Q., Lin, Z., Wang, Y., Li, M., Wang, L., Ding, H., & Li, P. (2020). Emerging Roles of SRSF3 as a Therapeutic Target for Cancer. *Frontiers in Oncology*, *10*. https://doi.org/10.3389/fonc.2020.577636

Zhu, J., Blenis, J., & Yuan, J. (2008). Activation of PI3K-Akt and MAPK pathways regulates Myc-mediated transcription by phosphorylating and promoting the degradation of Mad1. *Proceedings of the National Academy of Sciences*, *105*(18), 6584–6589. https://doi.org/10.1073/pnas.0802785105

Zhu, S., Chen, Z., Katsha, A., Hong, J., Belkhiri, A., & El-Rifai, W. (2016). Regulation of CD44E by DARPP-32-dependent activation of SRp20 splicing factor in gastric tumorigenesis. *Oncogene*, *35*(14), 1847–1856. https://doi.org/10.1038/onc.2015.250

Zilfou, J. T., & Lowe, S. W. (2009). Tumor suppressive functions of p53. *Cold Spring Harbor Perspectives in Biology*, *1*(5). https://doi.org/10.1101/cshperspect.a001883